Biometrika (20XX), **XX**, 1, *pp*. 1–20

Printed in Great Britain

Supplementary Material for Partially Functional Linear Regression in High Dimensions

BY DEHAN KONG

Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A.

kongdehanstat@gmail.com

KAIJIE XUE AND FANG YAO

Department of Statistical Sciences, University of Toronto, Ontario M5S 3G3, Canada

kaijie@utstat.toronto.edu fyao@utstat.toronto.edu

HAO H. ZHANG

Department of Mathematics, University of Arizona, 617 North Santa Rita Avenue, Tucson, Arizona 85721, U.S.A.

hzhang@math.arizona.edu

1. ADDITIONAL SIMULATION RESULTS

We present additional simulation results based on 200 Monte Carlo runs. In particular, Design III illustrates the situation that the response does not necessarily depend on the leading principal components and the regression coefficients may decay more slowly than theoretically required. Specifically, we use $b_{j1} = 0$, $b_{j2} = 1$, $b_{j3} = 0{\cdot}5$, $b_{jk} = (k-2)^{-3}$ for $k = 4, \ldots, 50$, $j = 1, 2$, and other settings are the same as Design I. The results across s_n follow a similar pattern, while the automated ABIC captures the regression relationship adaptively and resembles the optimal estimation and prediction. The refitting step using least squares with jointly tuned s_{nj} behaves similarly as in Design I. Designs IV contains ultra-high numbers of scalar covariates $\gamma = (1_5^{\mathrm{T}}, 0_{995}^{\mathrm{T}})^{\mathrm{T}}$ with $p_n = 1000$, and other settings the same as Design II. The results exhibit similar phenomenon as those in Design II. Moreover, Table 2 includes the results obtained from the same settings as Design I–IV except for a larger sample size $n = 400$. For Table 3, the only setting difference is that the regression error ϵ_i is generated from $N(0, 2)$ instead of $N(0, 1)$. As expected, a larger sample size reduced the estimation and prediction errors, while a higher noise level increased such errors.

2. REGULARITY CONDITIONS

Without loss of generality, we assume that $\{X_j, j = 1, \ldots, d\}$, Y and Z have been centred to have mean zero. With $W_{ijl} = x_{ij}(t_{ijl}) + \epsilon_{ijl}$, for definiteness, we consider the local linear smoother for each set of subjects using bandwidths $\{h_{ij}, j = 1, \ldots, d\}$, and denote the smoothed trajectories by $\hat{x}_{ij}$. Denote the minimum and maximum eigenvalues of a symmetric matrix A by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. Recall that the first g functional predictors are significant, while the rest are not. Define the $(gs_n + q_n) \times 1$ vector $\tilde{N} =$

Design	s_n	FZ$_f$	FN$_f$	MSE$_f$	FZ$_s$	FN$_s$	MSE$_s$	PE
III	1	2·0	0	2·5 (0)	1·5	1·1	4·2 (0·15)	26·0 (0·1)
$p_n = 20$	2	0·70	0	1·7 (0·06)	0·02	1·2	0·49 (0·028)	3·9 (0·1)
	3	0	0	0·076 (0·004)	0	0·38	0·069 (0·004)	0·21 (0·009)
	4	0	0	0·069 (0·003)	0	0·38	0·065 (0·004)	0·13 (0·005)
	5	0	0	0·11 (0·005)	0	0·40	0·067 (0·004)	0·14 (0·005)
	6	0	0	0·15 (0·006)	0	0·41	0·068 (0·004)	0·15 (0·005)
	10	0	0·01	0·60 (0·02)	0	0·37	0·071 (0·004)	0·19 (0·006)
	16	0·17	0·21	3·4 (0·1)	0	0·05	0·089 (0·007)	0·65 (0·05)
	ABIC	$\hat{s}_n$=3·73 (0·048)						
		0	0	0·072 (0·004)	0	0·38	0·066 (0·004)	0·14 (0·005)
	TUNE s_{nj}	$\hat{s}_{n1}$=3·99 (0·083), $\hat{s}_{n2}$=3·98 (0·079)						
		0	0	0·056 (0·003)	0	0·32	0·063 (0·004)	0·13 (0·005)
IV		$\hat{s}_n$=3·50 (0·040)						
$p_n = 1000$	STEP 1	0	0·09	0·10 (0·003)	0	6·5	0·41 (0·02)	0·81 (0·021)
	STEP 2	0	0	0·059 (0·003)	0	0·13	0·048 (0·003)	0·13 (0·005)
		$\hat{s}_{n1}$=4·05 (0·067), $\hat{s}_{n2}$ =4·11 (0·065)						
	TUNE s_{nj}	0	0	0·042 (0·002)	0	0·09	0·045 (0·003)	0·11 (0·004)

Table 1: Simulation results with sample size $n = 200$ based on 200 Monte Carlo replicates for Designs III and IV. Shown are the Monte Carlo averages (standard errors in parentheses) for the number of false zero functional predictors (FZ$_f$), the number of the false nonzero functional predictors (FN$_f$), the functional mean squared error (MSE$_f$), the number of false zero scalar covariates (FZ$_s$), the number of false nonzero scalar covariates (FN$_s$), the scalar mean squared error (MSE$_s$), and the prediction error (PE). We first use ABIC to choose the tuning parameter λ_n and a common truncation s_n, then tune s_{nj} jointly with AIC by refitting the selected model using ordinary least squares. In Design IV, Step 1 results are based on the original sample in each Monte Carlo run, while Step 2 contains the improved results by fitting the penalized procedure to the selected model in Step 1 with an additional sample of $n = 200$.

$(\xi_{11} w_{11}^{-1/2}, \ldots, \xi_{1s_n} w_{1s_n}^{-1/2}, \ldots, \xi_{g1} w_{g1}^{-1/2}, \ldots, \xi_{gs_n} w_{gs_n}^{-1/2}, Z_1, \ldots Z_{q_n})^{\mathrm{T}}$ to combine all functional and scalar predictors.

Condition (B1) consists of regularity assumptions for functional data, for example, a Gaussian process with Hölder continuous sample paths satisfies (B1), see Hall & Hosseini-Nasab (2006). Condition (B2) is standard for local linear smoothers, (B3)–(B4) concern how the functional predictors are sampled and smoothed, while (B5) is for the moments of non-functional covariates $Z = (Z_1, \ldots, Z_{p_n})^{\mathrm{T}}$.

(B1) For $j = 1, \ldots, d$, for any $C > 0$ there exists an $\epsilon > 0$ such that

$$\sup_{t \in T} [E\{|X_j(t)|^C\}] < \infty, \qquad \sup_{s,t \in T} (E[\{|s-t|^{-\epsilon} |X_j(s) - X_j(t)|\}^C]) < \infty.$$

For each integer $r \geq 1$, $w_{jk}^{-r} E(\xi_{jk}^{2r})$ is bounded uniformly in k.

(B2) For $j = 1, \ldots, d$, X_j is twice continuously differentiable on T with probability 1, and $\int E\{X_j^{(2)}(t)\}^4 dt < \infty$, where $X_j^{(2)}(\cdot)$ denotes the second derivative of $X_j(\cdot)$.

The following condition concerns the design on which x_{ij} is observed and the local linear smoother $\hat{x}_{ij}$. When a function is said to be smooth, we mean that it is continuously differentiable to an adequate order.

(B3) For $j = 1, \ldots, d$, $\{t_{ijl}, l = 1, \ldots, m_{ij}\}$ are considered deterministic and ordered increasingly for $i = 1, \ldots, n$. There exist densities g_{ij} uniformly smooth over i, satisfying $\int_T g_{ij}(t)\, dt = 1$

Design	s_n	FZ_f	FN_f	MSE_f	FZ_s	FN_s	MSE_s	PE
	1	0·90	0	4·1 (0·04)	0·15	5·1	2·3 (0·08)	25·8 (0·2)
	2	0	0	1·3 (0·009)	0	1·4	0·34 (0·01)	4·6 (0·04)
	3	0	0	0·58 (0·008)	0	0·23	0·11 (0·004)	1·5 (0·02)
	4	0	0	0·063 (0·002)	0	0·23	0·027 (0·002)	0·13 (0·004)
	5	0	0	0·071 (0·003)	0	0·27	0·028 (0·002)	0·12 (0·004)
I	6	0	0	0·095 (0·003)	0	0·24	0·028 (0·002)	0·12 (0·004)
$p_n = 20$	10	0	0	0·32 (0·01)	0	0·31	0·029 (0·002)	0·13 (0·004)
	16	0	0	1·2 (0·04)	0	0·10	0·026 (0·002)	0·15 (0·004)
	ABIC			$\hat{s}_n$=4·22 (0·047)				
		0	0	0·067 (0·003)	0	0·23	0·027 (0·002)	0·13 (0·004)
	TUNE s_{nj}			$\hat{s}_{n1}$=4·67 (0·058), $\hat{s}_{n2}$=4·68 (0·061)				
		0	0	0·049 (0·002)	0	0·20	0·026 (0·002)	0·12 (0·004)
				$\hat{s}_n$=4·10 (0·027)				
	STEP 1	0	0	0·11 (0·004)	0	4·6	0·13(0·007)	0·51 (0·01)
II	STEP 2	0	0	0·052 (0·002)	0	0·07	0·025 (0·001)	0·11 (0·004)
$p_n = 1000$				$\hat{s}_{n1}$=4·72 (0·056), $\hat{s}_{n2}$=4·60 (0·053)				
	TUNE s_{nj}	0	0	0·041 (0·001)	0	0·06	0·025 (0·001)	0·11 (0·004)
	1	2·0	0	2·5 (0)	1·0	0·43	2·7 (0·1)	25·2 (0·09)
	2	0·10	0	0·72 (0·04)	0	0·43	0·19 (0·01)	2·8 (0·06)
	3	0	0	0·059 (0·002)	0	0·25	0·032 (0·002)	0·16 (0·006)
	4	0	0	0·033 (0·001)	0	0·23	0·026 (0·001)	0·073 (0·004)
	5	0	0	0·048 (0·002)	0	0·26	0·027 (0·002)	0·075 (0·004)
III	6	0	0	0·072 (0·003)	0	0·25	0·027 (0·002)	0·078 (0·004)
$p_n = 20$	10	0	0	0·28 (0·01)	0	0·25	0·028 (0·002)	0·099 (0·004)
	16	0	0·03	1·2 (0·04)	0	0·02	0·024 (0·001)	0·13 (0·004)
	ABIC			$\hat{s}_n$=3·91 (0·036)				
		0	0	0·034 (0·002)	0	0·23	0·026 (0·001)	0·075 (0·004)
	TUNE s_{nj}			$\hat{s}_{n1}$=4·37 (0·064), $\hat{s}_{n2}$=4·38 (0·066)				
		0	0	0·026 (0·001)	0	0·22	0·026 (·001)	0·070 (0·004)
				$\hat{s}_n$=3·81 (0·043)				
	STEP 1	0	0·01	0·047 (0·001)	0	3·8	0·17 (0·009)	0·35 (0·008)
IV	STEP 2	0	0	0·031 (0·002)	0	0·08	0·024 (0·001)	0·073 (0·004)
$p_n = 1000$				$\hat{s}_{n1}$=4·28 (0·061), $\hat{s}_{n2}$=4·37 (0·067)				
	TUNE s_{nj}	0	0	0·023 (0·001)	0	0·07	0·024 (0·001)	0·058 (0·004)

Table 2: Simulation results using the same settings as Design I & II in Section 4 of the paper and Deisgn III & IV as above, based on 200 Monte Carlo replicates, except the sample size $n = 400$.

and $0 < c_1 < \inf_i\{\inf_{t\in T} g_{ij}(t)\} < \sup_i\{\sup_{t\in T} g_{ij}(t)\} < c_2 < \infty$ that generate t_{ijl} according to be $t_{ijl} = G_{ij}^{-1}\{l/(m_{ij}+1)\}$, where G_{ij}^{-1} is the inverse of $G_{ij}(t) = \int_{-\infty}^{t} g_{ij}(s)\,ds$. For each $j = 1, \ldots, d$, there exist a common sequence of bandwidths h_j such that $0 < c_1 < \inf_i h_{ij}/h_j < \sup_i h_{ij}/h_j < c_2 < \infty$, where h_{ij} is the bandwidth for $\hat{x}_{ij}$. The kernel density function is smooth and compactly supported.

Let $T = [a_0, b_0], t_{ij0} = a_0, t_{ij,m_{ij}+1} = b_0$, let $\Delta_{ij} = \sup\{t_{ij,l+1} - t_{ij,l}, l = 0, \ldots, m_{ij}\}$ and $m_j = m_j(n) = \inf_{i=1,\ldots,n} m_{ij}$. The condition below is to let the smooth estimate $\hat{x}_{ij}$ serve as well as the true x_{ij} in the asymptotic analysis, denoting $0 < \lim a_n/b_n < \infty$ by $a_n \sim b_n$.

(B4) For $j = 1, \ldots, d$, $\sup_i \Delta_{ij} = O(m_j^{-1})$, $h_j \sim m_j^{-1/5}$, $m_j n^{-5/4} \to \infty$.

The condition for the scalar covariates is given below.

(B5) $\{Z_l, l = 1, \ldots, p_n\}$ are subGaussian random variables such that $pr(|Z_l| > t) \leq \exp(-2^{-1}Ct^2)$ for any $t \geq 0$ and some $C > 0$ that does not depend on l, $\lambda_{\max}(\Sigma) \leq c_1 < \infty$

Design	s_n	FZ$_f$	FN$_f$	MSE$_f$	FZ$_s$	FN$_s$	MSE$_s$	PE
	1	0·94	0	4·1 (0·03)	0·41	7·3	4·9 (0·2)	28·0 (0·6)
	2	0·56	0	2·8 (0·1)	0·08	3·3	1·4 (0·07)	9·9 (0·3)
	3	0·01	0	0·63 (0·02)	0	1·7	0·38 (0·03)	1·7 (0·07)
	4	0	0	0·17 (0·007)	0	0·58	0·15 (0·009)	0·31 (0·01)
	5	0	0	0·23 (0·1)	0	0·58	0·15 (0·009)	0·30 (0·01)
I	6	0	0	0·33 (0·01)	0	0·63	0·15 (0·009)	0·32 (0·01)
$p_n = 20$	10	0·01	0·01	1·3 (0·05)	0	0·49	0·15 (0·009)	0·44 (0·05)
	16	0·31	0·18	7·7 (0·3)	0·08	0·20	0·36 (0·03)	3·6 (0·3)
	ABIC	$\hat{s}_n$=4·12 (0·026)						
		0	0	0·18 (0·009)	0	0·56	0·15 (0·009)	0·30 (0·01)
	TUNE s_{nj}	$\hat{s}_{n1}$=4·59 (0·069), $\hat{s}_{n2}$=4·61 (0·060)						
		0	0	0·14 (0·006)	0	0·48	0·14 (0·008)	0·28 (0·01)
		$\hat{s}_n$=4·04 (0·021)						
	STEP 1	0	0·04	0·29 (0·008)	0	7·3	0·55 (0·03)	1·9 (0·07)
II	STEP 2	0	0	0·15 (0·008)	0	0·16	0·096 (0·005)	0·26 (0·02)
$p_n = 1000$		$\hat{s}_{n1}$=4·53 (0·056), $\hat{s}_{n2}$=4·46 (0·050)						
	TUNE s_{nj}	0	0	0·12 (0·006)	0	0·15	0·095 (0·005)	0·24 (0·008)
	1	2·0	0	2·5 (0)	1·5	1·2	4·3 (0·15)	26·1 (0·1)
	2	0·33	0	1·1 (0·06)	0·02	0·69	0·43 (0·03)	3·3 (0·1)
	3	0	0	0·10 (0·005)	0	0·39	0·13 (0·008)	0·29 (0·01)
	4	0	0	0·12 (0·006)	0	0·39	0·13 (0·008)	0·22 (0·01)
	5	0	0	0·20 (0·009)	0	0·41	0·13 (0·008)	0·24 (0·01)
III	6	0	0	0·29 (0·01)	0	0·41	0·14 (0·008)	0·26 (0·01)
$p_n = 20$	10	0·03	0·07	1·4 (0·06)	0	0·17	0·12 (0·007)	0·41 (0·03)
	16	0·02	1·3	11·8(0·4)	0·02	0·39	0·18 (0·02)	1·0 (0·03)
	ABIC	$\hat{s}_n$=3·76 (0·065)						
		0	0	0·15 (0·02)	0	0·36	0·13 (0·008)	0·24 (0·01)
	TUNE s_{nj}	$\hat{s}_{n1}$=3·76 (0·064), $\hat{s}_{n2}$=3·79 (0·066)						
		0	0	0·070 (0·004)	0	0·29	0·12 (0·007)	0·20 (0·01)
		$\hat{s}_n$=3·33 (0·040)						
	STEP 1	0	0·06	0·18 (0·006)	0	6·4	0·63 (0·03)	1·5 (0·04)
IV	STEP 2	0	0	0·091 (0·005)	0	0·12	0·099 (0·005)	0·23 (0·009)
$p_n = 1000$		$\hat{s}_{n1}$=3·96 (0·060), $\hat{s}_{n2}$=3·73 (0·048)						
	TUNE s_{nj}	0	0	0·065 (0·003)	0	0·10	0·095 (0·005)	0·18 (0·009)

Table 3: Simulation results using the same settings as Design I & II in Section 4 of the paper and Deisgn III & IV as above, based on 200 Monte Carlo replicates, except the variance of the regression error $\sigma^2 = 2$.

and $0 < c_2 \leq \lambda_{\min}(U_1) \leq \lambda_{\max}(U_1) \leq c_3 < \infty$ for all n, where $\Sigma = E(ZZ^{\mathrm{T}})$ and $U_1 = E(\tilde{N}\tilde{N}^{\mathrm{T}})$.

3. AUXILIARY LEMMAS

For each $j = 1, \ldots, d$, given the estimated covariances $\hat{K}_j(s,t) = n^{-1}\sum_{i=1}^n \hat{x}_{ij}(s)\hat{x}_{ij}(t)$, the eigenvalues/functions and functional principal component scores are estimated by

$$\int_T \hat{K}_j(s,t)\hat{\phi}_{jk}(s)ds = \hat{\lambda}_{jk}\hat{\phi}_{jk}(t), \qquad \hat{\xi}_{ijk} = \int_T \hat{x}_{ij}(t)\hat{\phi}_{jk}(t)dt, \tag{1}$$

subject to $\int_T \hat{\phi}_{jk}^2(t)dt = 1$ and $\int_T \hat{\phi}_{jk_1}(t)\hat{\phi}_{jk_2}(t)dt = 0$ for $k_1 \neq k_2$. Denote $\hat{\Delta}_j^2 = |||\hat{K}_j - K_j|||^2 = \int_T\int_T \{\hat{K}_j(s,t) - K_j(s,t)\}^2 dsdt$, $\quad \Re_j(s,t) = n^{1/2}\{\hat{K}_j(s,t) - K_j(s,t)\}$,

$(x_{ij} \otimes x_{ij})(s,t) = x_{ij}(s)x_{ij}(t)$, the second derivative of x_{ij} by $x_{ij}^{(2)}$, the minimum and maximum eigenvalues of a symmetric matrix A by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$. Let the $n \times p_n$ matrix $Z_M = (z_1, \ldots, z_n)^{\mathrm{T}} = (Z_M^{(1)}, Z_M^{(2)})$ with $Z_M^{(1)}$ containing the first q_n columns of Z_M. Define M_j as the $n \times s_n$ matrix with (i,k)th element ξ_{ijk}, and $M = (M_1, \ldots, M_d)$, $M^{(1)} = (M_1, \ldots, M_g)$ and $M^{(2)} = (M_{g+1}, \ldots, M_d)$. Similarly, we define $\tilde{M}_j$, $\tilde{M}^{(1)}$ and $\tilde{M}^{(2)}$, where ξ_{ijk} is replaced by $\xi_{ijk} w_{jk}^{-1/2}$. Moreover, we denote $\hat{M}_j$ as the $n \times s_n$ matrix with (i,k)th element $\hat{\xi}_{ijk}$, and $\hat{M} = (\hat{M}_1, \ldots, \hat{M}_d)$, $\hat{M}^{(1)} = (\hat{M}_1, \ldots, \hat{M}_g)$ and $\hat{M}^{(2)} = (\hat{M}_{g+1}, \ldots, \hat{M}_d)$. Similarly, we define $\check{M}_j$, $\check{M}^{(1)}$ and $\check{M}^{(2)}$, where $\hat{\xi}_{ijk}$ is replaced by $\hat{\xi}_{ijk} w_{jk}^{-1/2}$. Combining the functional and scalar covariates, we have $\tilde{N}_1 = (\tilde{M}^{(1)}, Z_M^{(1)}) = (\tilde{N}_1, \ldots, \tilde{N}_n)^{\mathrm{T}}$, where $\{\tilde{N}_i = (\tilde{N}_{i1}, ..., \tilde{N}_{i,gs_n+q_n})^{\mathrm{T}}, i = 1, \ldots, n\}$ are independently and identically distributed as $\tilde{N}$. We further denote $N_1 = (M^{(1)}, Z_M^{(1)}) = (N_1, \ldots, N_n)^{\mathrm{T}}$, $\check{N}^{(1)} = (\check{M}^{(1)}, Z_M^{(1)})$, $\hat{N}^{(1)} = (\hat{M}^{(1)}, Z_M^{(1)})$, and let $E(\tilde{N}_i \tilde{N}_i^{\mathrm{T}}) = U_1$. Recall that $\check{b}^{(1)}$ denotes the estimate of $\tilde{b}^{(1)}$, $\check{\eta} = (\check{b}^{(1)\mathrm{T}}, \hat{\gamma}^{\mathrm{T}})^{\mathrm{T}}$, and we further denote $\check{b}_j^{(1)}$ the estimate of $\tilde{b}_j^{(1)}$. Suppose $\alpha(\cdot)$, $\beta(\cdot)$ and $G(\cdot,\cdot)$ are square-integrable functions on T and $T \times T$. Write $||\alpha||$, $\int \alpha\beta$ (or $\langle \alpha, \beta \rangle$) and $\int G\alpha\beta$ for $\{\int_T \alpha^2(t)dt\}^{1/2}$, $\int_T \alpha(t)\beta(t)dt$ and $\int\int_{T^2} G(s,t)\alpha(s)\beta(t)dsdt$.

Lemma 1 provides results for the estimates obtained by functional principal component analysis. We first quantify the smoothing error of $\hat{x}_{ij}$ and its influence carried over to the covariance and functional principal component estimates, $\hat{\xi}_{ijk} - \xi_{ijk} = \tilde{\xi}_{ijk} - \xi_{ijk} + \hat{\xi}_{ijk} - \tilde{\xi}_{ijk} = \int x_{ij}(\hat{\phi}_{jk} - \phi_{jk}) + \int (\hat{x}_{ij} - x_{ij})\hat{\phi}_{jk}$. Then we obtain upper bounds for the differences between the estimated and true eigenfunctions in terms of covariance perturbation and eigenvalue spacings for quantifying the increased estimation error of the higher order terms. For each $j = 1, \ldots, d$, denote

$$\delta_{jk} = \min_{l=1,\ldots,k} (w_{jl} - w_{j,l+1}), \quad \mathcal{J}_{jn} = \{k = 1, \ldots, \infty : w_{jk} - w_{j,k+1} > 2\hat{\Delta}_j\},$$

where $\hat{\Delta}_j = |||\hat{K}_j - K_j|||$, that is, consider $k \in \mathcal{J}_{jn}$ for which the distance of w_{jk} to the nearest other eigenvalues does not fall below $2\hat{\Delta}_j$. It is known that ϕ_{jk} can be consistently estimated for $k \in \mathcal{J}_{jn}$ (Theorem 1, Hall & Hosseini-Nasab, 2006), $\|\hat{\phi}_{jk} - \phi_{jk}\| = O_p(\delta_{jk}\hat{\Delta}_j) = O_p(k^{a+1}n^{-1/2})$, implying $\sup \mathcal{J}_{jn} = o\{n^{1/(2a+2)}\}$. However, for our theoretical analysis, we need a sharper bound without compromising the number of eigenfunctions considered. Define the set of realizations such that, for sample size n, some C and any $\tau < 1$,

$$\mathcal{F}_{s_n,j} = \{(\hat{w}_{jk_1} - w_{jk_2})^{-2} \leq 2(w_{jk_1} - w_{jk_2})^{-2} \leq Cn^{\tau}, k_1, k_2 = 1, \ldots, s_n, k_1 \neq k_2\}.$$

LEMMA 1*(a) Under conditions (B1)–(B4), for each $i = 1, \ldots, n$ and each $j = 1, \ldots, d$, we have $E(\|\hat{x}_{ij} - x_{ij}\|^2) = o(n^{-1})$, $E\Big\{\int (\hat{x}_{ij} - x_{ij})^4\Big\} = o(n^{-2})$. For each $j = 1, \ldots, d$, we have $E(|||\hat{K}_j - K_j|||^2) = O(n^{-1})$.*

(b) Under conditions (A1), (A3), (B1)–(B4), for each $j = 1, \ldots, d$, we have $s_n \in J_{jn}$ for large n, $pr(\mathcal{F}_{s_n,j}) \to 1$ as $n \to \infty$ for any $\tau < 1$. Moreover, for each $k = 1, \ldots, s_n$, $\|\hat{\phi}_{jk} - \phi_{jk}\| = O_p(k\, n^{-1/2})$, where $O_p(\cdot)$ is uniform in $k = 1, \ldots, s_n$.

(c) Under conditions (A1), (A3), (B1)–(B4), for each $j = 1, \ldots, d$,

$$\hat{\phi}_{jk}(t) - \phi_{jk}(t) = n^{-1/2} \sum_{v: v \neq k} (w_{jk} - w_{jv})^{-1} \phi_{jv}(t) \int \Re_j \phi_{jk} \phi_{jv} + \alpha_{jk}(t),$$

where $||\alpha_{jk}|| = O_p(k^{a+2}n^{-1})$, $\Re_j = n^{1/2}(\hat{K}_j - K_j)$, *and* $O_p(\cdot)$ *is uniform in* $k = 1, \ldots, s_n$.

(d) Under conditions (A1), (B1)–(B4) and $n^{-1/2}s_n = o_p(1)$, *for any* $j = 1, \ldots, d$, *we have* $c_1\lambda_n \le \lambda_{jn} \le c_2\lambda_n$ *for some positive constants* c_1, c_2.

The next lemma quantifies the asymptotic orders of several important types of expressions that will be encountered in the proofs of our lemmas and main theorems. For convenience, define the following notations, $l, l_1, l_2 = 1, \ldots, p_n$, $k, k_1, k_2 = 1, \ldots, s_n$, $j, j_1, j_2 = 1, \ldots, d$,

$$\theta_{jk}^{(1)} = \sum_{i=1}^{n}(\hat{\xi}_{ijk} - \xi_{ijk})^2 w_{jk}^{-1}, \qquad \theta_{k_1k_2}^{j_1j_2\,(2)} = n^{-1}\sum_{i=1}^{n}(\hat{\xi}_{ij_1k_1}\hat{\xi}_{ij_2k_2} - \xi_{ij_1k_1}\xi_{ij_2k_2})(w_{j_1k_1}w_{j_2k_2})^{-1/2},$$

$$\theta_{k_1k_2}^{j_1j_2\,(3)} = n^{-1}\sum_{i=1}^{n}\{\xi_{ij_1k_1}\xi_{ij_2k_2} - E(\xi_{j_1k_1}\xi_{j_2k_2})\}(w_{j_1k_1}w_{j_2k_2})^{-1/2}, \quad \theta_{k_1k_2}^{j_1j_2\,(4)} = \theta_{k_1k_2}^{j_1j_2\,(2)} + \theta_{k_1k_2}^{j_1j_2\,(3)},$$

$$\theta_{jkl}^{(5)} = n^{-1}\sum_{i=1}^{n}(\hat{\xi}_{ijk} - \xi_{ijk})z_{il}w_{jk}^{-1/2}, \qquad \theta_{jkl}^{(6)} = n^{-1}\sum_{i=1}^{n}\{\xi_{ijk}z_{il} - E(\xi_{ijk}z_{il})\}w_{jk}^{-1/2},$$

$$\theta_{jkl}^{(7)} = \theta_{jkl}^{(5)} + \theta_{jkl}^{(6)}, \qquad \theta_{l_1l_2}^{(8)} = n^{-1}\sum_{i=1}^{n}\{z_{il_1}z_{il_2} - E(z_{il_1}z_{il_2})\},$$

$$\vartheta_{jl}^{(1)} = \sum_{i=1}^{n}\sum_{k=s_n+1}^{\infty}\xi_{ijk}b_{jk0}z_{il}, \qquad \vartheta_{jl}^{(2)} = \sum_{i=1}^{n}\sum_{k=1}^{s_n}(\hat{\xi}_{ijk} - \xi_{ijk})(\hat{b}_{jk} - b_{jk0})z_{il},$$

$$\vartheta_{jl}^{(3)} = \sum_{i=1}^{n}\sum_{k=1}^{s_n}(\hat{\xi}_{ijk} - \tilde{\xi}_{ijk})b_{jk0}z_{il}, \qquad \vartheta_{jl}^{(4)} = \sum_{i=1}^{n}\sum_{k=1}^{s_n}(\tilde{\xi}_{ijk} - \xi_{ijk})b_{jk0}z_{il},$$

$$\vartheta_{jl}^{(5)} = \sum_{k=1}^{s_n}b_{jk0}\int(\sum_{i=1}^{n}x_{ij}z_{il} - n\Xi_{jl})(\hat{\phi}_{jk} - \phi_{jk}), \quad \vartheta_{jl}^{(6)} = n\sum_{k=1}^{s_n}b_{jk0}\langle\Xi_{jl}, \alpha_{jk}\rangle,$$

$$\vartheta_{jl}^{(7)} = n^{1/2}\sum_{k=1}^{s_n}\sum_{v\neq k}b_{jk0}(w_{jk} - w_{jv})^{-1}\langle\Xi_{jl}, \phi_{jv}\rangle\int(\Re_j - \Re_j^*)\phi_{jk}\phi_{jv},$$

$$\vartheta_{jl}^{(8)} = n^{1/2}\sum_{k=1}^{s_n}\sum_{v\neq k}b_{jk0}(w_{jk} - w_{jv})^{-1}\langle\Xi_{jl}, \phi_{jv}\rangle\int\Re_j^*\phi_{jk}\phi_{jv},$$

where $\vartheta_j{}^{(m)} = (\vartheta_{j1}^{(m)}, \ldots\vartheta_{jq_n}^{(m)})^{\mathrm{T}}$ denote corresponding vectors and $\Re_j^* = n^{1/2}(\tilde{K}_j - K_j)$, where $\tilde{K}_j = n^{-1}\sum_{i=1}^{n}x_{ij}\otimes x_{ij}$.

LEMMA 2*(a) Under conditions (A1), (A3), (B1)–(B5), we have*

$$\begin{aligned}
&\theta_{jk}^{(1)} = O_p(k^{a+2}), && \theta_{k_1k_2}^{j_1j_2\,(2)} = O_p\big(k_1^{a/2+1}n^{-1/2} + k_2^{a/2+1}n^{-1/2}\big),\\
&\theta_{k_1k_2}^{j_1j_2\,(3)} = O_p(n^{-1/2}), && \theta_{k_1k_2}^{j_1j_2\,(4)} = O_p\big(k_1^{a/2+1}n^{-1/2} + k_2^{a/2+1}n^{-1/2}\big),\\
&\theta_{jkl}^{(5)} = O_p\big(k^{a/2+1}n^{-1/2}\big), && \theta_{jkl}^{(6)} = O_p(n^{-1/2}),\\
&\theta_{jkl}^{(7)} = O_p\big(k^{a/2+1}n^{-1/2}\big), && \theta_{l_1l_2}^{(8)} = O_p(n^{-1/2}),
\end{aligned}$$

where the $O_p(\cdot)$ *and* $o_p(\cdot)$ *terms are uniform for* $k, k_1, k_2 = 1, \ldots, s_n$ *and* $l, l_1, l_2 = 1, \ldots, p_n$.

(b) Under conditions (A1)–(A7), (B1)–(B5), uniformly for $l = 1, \dots, p_n$,

$$\vartheta_{jl}^{(1)} = \vartheta_{jl}^{(2)} = \vartheta_{jl}^{(3)} = \vartheta_{jl}^{(5)} = \vartheta_{jl}^{(6)} = \vartheta_{jl}^{(7)} = o_p(n^{1/2}),$$
$$\vartheta_{jl}^{(4)} = \vartheta_{jl}^{(5)} + \vartheta_{jl}^{(6)} + \vartheta_{jl}^{(7)} + \vartheta_{jl}^{(8)}.$$

Lemma 3 characterizes the eigenvalues of the design matrices $\check{N}_1$, and Lemma 4 concerns the asymptotic order of $\Lambda_1 = P_{\check{N}_1}(Y_M - \check{N}_1\tilde{\eta}_{10})$, where $P_{\check{N}_1} = \check{N}_1(\check{N}_1^{\mathrm{T}}\check{N}_1)^{-1}\check{N}_1^{\mathrm{T}}$, $Y_M = (y_1, ..., y_n)^{\mathrm{T}}$, and $\tilde{\eta}_{10}$ is the true parameter of $\tilde{\eta}_1$.

LEMMA 3. *Under conditions (A1), (A3), (A5), (B1)–(B5), we have* $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| = o_p(1)$, $|\lambda_{\max}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\max}(U_1)| = o_p(1)$.

LEMMA 4. *Under conditions (A1)–(A5), (B1)–(B5), we have* $\|\Lambda_1\|_2^2 = O_p(r_n^2)$, *where* $r_n^2 = q_n + s_n$.

4. PROOF OF LEMMAS

Proof of Lemma 1. We shall show Lemma 1 for any fixed $j = 1, \dots, d$, we suppress the subscript j in this proof for convenience. For part (a), recall that $W_{il} = x_i(t_{il}) + \varepsilon_{il}$, where the error ε_{il} are independent of x_i. Thus one can factor the probability space $\Omega = \Omega_1 \times \Omega_2$, where Ω_1 is for $\{x_i\}_{i=1,\dots,n}$ and Ω_2 for $\{\varepsilon_{il}\}_{l=1,\dots,m_i}$. Given the data for a single subject, a specific realization x_i corresponds to fixing a value $\omega_i \in \Omega_1$, that is $x_i(\cdot) = x_i(\cdot, \omega_i)$. We write E_ε for expectations with regard to the probability measure on Ω_2, and E_X with respect to Ω_1. For each fixed ω_i, the error $\{\varepsilon_{il}, i = 1, \dots, n\}$ are independent and identically distributed for different l, with $E_\varepsilon(\varepsilon_{il}) = 0$ and $E_\varepsilon(\varepsilon_{il}^2) = \sigma_x^2$. Given condition (B3) together with a local linear smoother, one has $E(\|\hat{x}_i - x_i\|^2) = E_X\big[E_\varepsilon\big\{\int(\hat{x}_i - x_i)^2\big\}\big]$. Using standard Taylor expansion argument, together with the dominant convergence theorem given $E(\|x_i^{(2)}\|^4) < \infty$, it is easy to verify $E_X\big[E_\varepsilon\big\{\int(\hat{x}_i - x_i)^2\big\}\big] \le C\int\big[E_X\big\{(x_i^{(2)})^2\big\}h^4 + (mh)^{-1}\big]$ under (B3), yielding $E(\|\hat{x}_i - x_i\|^2) = o(n^{-1})$ by (B4). Similar arguments will also lead to $E_X\big[E_\varepsilon\{\int(\hat{x}_i - x_i)^4\}\big] \le C\int\big[E_X\{(x_i^{(2)})^4\}h^8 + (mh)^{-2}\big]$, thus $E\big\{\int(\hat{x}_i - x_i)^4\big\}\big] = o(n^{-2})$. Regarding $\hat{K} = n^{-1}\sum_{i=1}^n \hat{x}_i \otimes \hat{x}_i$, notice that $\hat{K} - K = n^{-1}\sum_{i=1}^n\big\{(\hat{x}_i - x_i)\otimes x_i + (\hat{x}_i - x_i)\otimes(\hat{x}_i - x_i)\big\} + \big(n^{-1}\sum_{i=1}^n x_i \otimes x_i - K\big)$. It is easy to see that the first term on the right hand side is dominated by $n^{-1}\sum_{i=1}^n |||(\hat{x}_i - x_i)\otimes x_i|||$. Since $\{\hat{x}_i, x_i\}$ is independent of $\{\hat{x}_{i'}, x_{i'}\}$ for $i \ne i'$, we have $E\Big[\Big\{n^{-1}\sum_{i=1}^n |||(\hat{x}_i - x_i)\otimes x_i|||\Big\}^2\Big] = \big\{E|||(\hat{x}_i - x_i)\otimes x_i|||\big\}^2 + n^{-1}\mathrm{var}(|||(\hat{x}_i - x_i)\otimes x_i|||) \le E(\|\hat{x}_i - x_i\|^2)(E\|x_i\|^2) + n^{-1}\big\{E(\|(\hat{x}_i - x_i)\|^4)(E\|x_i\|^4)\big\}^{1/2} = o(n^{-1})$. The second term $E(|||\big(n^{-1}\sum_{i=1}^n x_i \otimes x_i - K|||^2) = O(n^{-1})$ by Lemma 3.3 of Hall & Hosseini-Nasab (2009). This leads to $E(|||\hat{K} - K|||^2) = O(n^{-1})$.

We next obtain the bound in (b) for $k = 1, \dots, s_n$. First notice that $\mathcal{F}_n$ can be implied by $\min_{k=1,\dots,s_n}(w_k - w_{k+1}) \ge s_n^{-a-1} \ge Cn^{-\tau/2}$ for some C and any $\tau < 1$, which is fulfilled by $s_n^{a+1}n^{-1/2} \to 0$ in (A3), leading to $pr(\mathcal{F}_n) \to 1$ as $n \to \infty$. It is easy to see that for $k = 1, \dots, s_n$, $w_k - w_{k+1} \ge s_n^{-a-1} > 2\hat{\Delta}$ as $n \to \infty$, that is $s_n \in \mathcal{J}_n$ for large n. Now we will bound $||\hat{\phi}_k - \phi_k||^2$ for $k = 1, \dots, s_n$. By (5.16) in Hall & Horowitz (2007), one has $\|\hat{\phi}_k - \phi_k\|^2 \le 2\hat{u}_k^2$, where $\hat{u}_k^2 = \sum_{v:v\ne k}(\hat{w}_k - w_v)^{-2}\Big\{\int(\hat{K} - K)\hat{\phi}_k\phi_v\Big\}^2$. Also $pr(\mathcal{F}_n) \to 1$

implies that $(\hat{w}_k - w_v)^{-2} \le Cn^{\tau}$ with probability tending to 1. Then we have

$$\begin{aligned}
\hat{u}_k^2 &\le \sum_{v:v\neq k} (\hat{w}_k - w_v)^{-2}\Big[2\Big\{\int(\hat{K}-K)\phi_j\phi_v\Big\}^2 \\
&\quad +2\Big\{\int(\hat{K}-K)(\hat{\phi}_k-\phi_k)\phi_v\Big\}^2\Big] \\
&\le 2\sum_{v:v\neq k}(\hat{w}_k - w_v)^{-2}\Big\{\int(\hat{K}-K)\phi_k\phi_v\Big\}^2 \\
&\quad +2Cn^{\tau}\sum_{v\neq k}\Big\{\int(\hat{K}-K)(\hat{\phi}_k-\phi_k)\phi_v\Big\}^2 \\
&\le 4\sum_{v:v\neq k}(w_k - w_v)^{-2}\Big\{\int(\hat{K}-K)\phi_k\phi_v\Big\}^2 + 2Cn^{\tau}\hat{\Delta}^2||\hat{\phi}_k-\phi_k||^2.
\end{aligned}$$

Plugging this into $\|\hat{\phi}_k - \phi_k\|^2 \le 2\hat{u}_k^2$, one has

$$(1-2Cn^{\tau}\hat{\Delta}^2)\|\hat{\phi}_k-\phi_k\|^2 \le 4\sum_{v:v\neq k}(w_k-w_v)^{-2}\Big\{\int(\hat{K}-K)\phi_k\phi_v\Big\}^2.$$

As $\hat{\Delta} = O_p(n^{-1/2})$ and $\tau < 1$, we have $n^{\tau}\hat{\Delta}^2 = o_p(1)$, and $\|\hat{\phi}_k - \phi_k\|^2 \le C\sum_{v\neq k}(w_k - w_v)^{-2}\{\int(\hat{K}-K)\phi_k\phi_v\}^2$. Given the results in (a), by analogy to (5.22) in Hall & Horowitz (2007), $nE[\sum_{v\neq k}(w_k-w_v)^{-2}\{\int(\hat{K}-K)\phi_k\phi_v\}^2] = O(k^2)$ still holds, the result follows by Chebyshev's inequality.

To obtain (c), using Lemma 5.1 of Hall & Horowitz (2007) with $\psi_k = \hat{\phi}_k$, $\lambda_k = \hat{w}_k$ and $L = \hat{K}$, since $\inf_{\{v:v\neq k\}}|\hat{w}_k - w_v| > 0$, we have

$$\begin{aligned}
\hat{\phi}_k - \phi_k &= \sum_{v:v\neq k}(\hat{w}_k - w_v)^{-1}\phi_v\int(\hat{K}-K)\hat{\phi}_k\phi_v + \phi_k\int(\hat{\phi}_k-\phi_k)\phi_k \\
&= \sum_{v:v\neq k}(w_k - w_v)^{-1}\phi_v\int(\hat{K}-K)\phi_k\phi_v + \phi_k\int(\hat{\phi}_k-\phi_k)\phi_k \\
&\quad + \sum_{v:v\neq k}(w_k - w_v)^{-1}\phi_v\int(\hat{K}-K)(\hat{\phi}_k-\phi_k)\phi_v \\
&\quad - \sum_{v:v\neq k}(\hat{w}_k - w_k)\{(w_k-w_v)(\hat{w}_k-w_v)\}^{-1}\phi_v\int(\hat{K}-K)\hat{\phi}_k\phi_v.
\end{aligned}$$

Denote the last three terms by $\alpha_k \equiv \alpha_{1k} + \alpha_{2k} + \alpha_{3k}$. For α_{1k}, one has

$$\|\hat{\phi}_k - \phi_k\|^2 = 2 - 2\int\hat{\phi}_k\phi_k = 2(\int\phi_k^2 - \int\hat{\phi}_k\phi_k) = -2\int(\hat{\phi}_k-\phi_k)\phi_k,$$

that is $\alpha_{1k} = -\|\hat{\phi}_k - \phi_k\|^2/2$. From part (b), one has $\|\hat{\phi}_k - \phi_k\| = O_p(kn^{-1/2})$ uniformly in $k = 1,\ldots,s_n$, then $\|\alpha_{1k}\| = O_p(k^2n^{-1})$. To bound α_{2k}, noticing $\|\sum_v c_v\phi_v\|^2 = \sum_v c_v^2$ due to orthonormal $\{\phi_k\}$,

$$\begin{aligned}\|\alpha_{2k}\| &\le \Big[\sum_{v\neq k}(w_k-w_v)^{-2}\Big\{\int(\hat{K}-K)(\hat{\phi}_k-\phi_k)\phi_v\Big\}^2\Big]^{1/2}\\ &\le \delta_k^{-1}\hat{\Delta}\|\hat{\phi}_k-\phi_k\| = O_p(k^{a+2}n^{-1}).\end{aligned}$$

For α_{3k}, noticing that $\sup_{k=1,\ldots,\infty}|\hat{w}_k-w_k| = O_p(\hat{\Delta})$, we can bound $\|\alpha_{3k}\|$ as follows,

$$\begin{aligned}&2|\hat{w}_k-w_k|\Big[\sum_{v:v\neq j}(w_k-w_v)^{-4}\Big\{\int(\hat{K}-K)\hat{\phi}_k\phi_v\Big\}^2\Big]^{1/2}\\ &\le C\hat{\Delta}\Big[\sum_{v:v\neq k}(w_k-w_v)^{-4}\Big\{\int(\hat{K}-K)\phi_k\phi_v\Big\}^2\\ &+\sum_{v:v\neq k}(w_k-w_v)^{-4}\Big\{\int(\hat{K}-K)(\hat{\phi}_k-\phi_k)\phi_v\Big\}^2\Big]^{1/2}\\ &= C\hat{\Delta}(A_1+A_2)^{1/2},\end{aligned}$$

where $A_1=\sum_{v:v\neq k}(w_k-w_v)^{-4}\big\{\int(\hat{K}-K)\phi_k\phi_v\big\}^2$ and $A_2=\sum_{v:v\neq k}(w_k-w_v)^{-4}\big\{\int(\hat{K}-K)(\hat{\phi}_k-\phi_k)\phi_v\big\}^2$. By the derivations of part (b), we have

$$\begin{aligned}A_1 &\le \delta_k^{-2}\sum_{v:v\neq k}(w_k-w_v)^{-2}\Big\{\int(\hat{K}-K)\phi_k\phi_v\Big\}^2 = O_p(k^2\delta_k^{-2}n^{-1}),\\ A_2 &\le \delta_k^{-4}\hat{\Delta}^2\|\hat{\phi}_k-\phi_k\|^2 = O_p(\delta_k^{-4}j^2\hat{\Delta}^4).\end{aligned}$$

Notice that $\delta_k^{-2}=O\{k^{2(a+1)}\}=o(n)$ by (A3), which indicates $A_2=o_p(k^2\delta_k^{-2}n^{-1})$. Thus $\|\alpha_{3k}\|=O_p(k\delta_k^{-1}n^{-1})$. It leads to $\|\alpha_k\|=O_p(k^{a+2}n^{-1})$.

For part (d), since $\sum_{k=1}^{s_n}w_k<\infty$ and $\sup_{k=1,\ldots,s_n}|\hat{w}_k-w_k|\le|||\hat{K}-K|||=O_p(n^{-1/2})$ by Theorem 1 of Hall & Hosseini-Nasab (2006) and part (a), we can see that $(\sum_{k=1}^{s_n}\hat{w}_k)^{1/2}=O_p(1)$ given $n^{-1/2}s_n=o(1)$, which completes the proof.

Proof of Lemma 2. For $\theta_{jk}^{(1)}$ and any fixed j, we have

$$\begin{aligned}\theta_{jk}^{(1)} &= \sum_{i=1}^n(\hat{\xi}_{ijk}-\xi_{ijk})^2w_{jk}^{-1} = \sum_{i=1}^n(\hat{\xi}_{ijk}-\tilde{\xi}_{ijk}+\tilde{\xi}_{ijk}-\xi_{ijk})^2w_{jk}^{-1}\\ &\le 2\sum_{i=1}^n(\hat{\xi}_{ijk}-\tilde{\xi}_{ijk})^2w_{jk}^{-1}+2\sum_{i=1}^n(\tilde{\xi}_{ijk}-\xi_{ijk})^2w_{jk}^{-1}\\ &= 2\sum_{i=1}^n\Big\{\int\hat{x}_{ij}(\hat{\phi}_{jk}-\phi_{jk})\Big\}^2w_{jk}^{-1}+2\sum_{i=1}^n\Big\{\int(\hat{x}_{ij}-x_{ij})\phi_{jk}\Big\}^2w_{jk}^{-1}\\ &\le 2\sum_{i=1}^n\big(\|\hat{x}_{ij}\|^2\|\hat{\phi}_{jk}-\phi_{jk}\|^2+\|\phi_{jk}\|^2\|\hat{x}_{ij}-x_{ij}\|^2\big)w_{jk}^{-1}.\end{aligned}$$

Given Lemma 1 and (B1), we know

$$E(\|\hat{x}_{ij}\|^2)\le 2E(\|\hat{x}_{ij}-x_{ij}\|^2)+2E(\|x_{ij}\|^2)=O(1),$$

hence, $\|\hat{x}_{ij}\|^2 = O_p(1)$. From Lemma 1, we have $\theta^{(1)}_{jk} = \sum_{i=1}^n \big\{O_p(1)O_p(k^2n^{-1}) + o_p(n^{-1})\big\}k^a = O_p(k^{a+2})$, uniformly for $k = 1, \ldots, s_n$.

For $\theta^{j_1j_2\,(2)}_{k_1k_2}$, it is obvious that

$$\begin{aligned}
|\theta^{j_1j_2\,(2)}_{k_1k_2}| &= \Big|n^{-1}\sum_{i=1}^n(\hat{\xi}_{ij_1k_1}\hat{\xi}_{ij_2k_2} - \xi_{ij_1k_1}\xi_{ij_2k_2})(w_{j_1k_1}w_{j_2k_2})^{-1/2}\Big| \\
&= n^{-1}\Big|\sum_{i=1}^n\hat{\xi}_{ij_1k_1}(\hat{\xi}_{ij_2k_2} - \xi_{ij_2k_2})(w_{j_1k_1}w_{j_2k_2})^{-1/2} + \sum_{i=1}^n\xi_{ij_2k_2}(\hat{\xi}_{ij_1k_1} - \xi_{ij_1k_1})(w_{j_1k_1}w_{j_2k_2})^{-1/2}\Big| \\
&\le n^{-1}\big(\sum_{i=1}^n\hat{\xi}^2_{ij_1k_1}w^{-1}_{j_1k_1}\big)^{1/2}\big\{\sum_{i=1}^n(\hat{\xi}_{ij_2k_2} - \xi_{ij_2k_2})^2w^{-1}_{j_2k_2}\big\}^{1/2} + \\
&\quad n^{-1}\big(\sum_{i=1}^n\xi^2_{ij_2k_2}w^{-1}_{j_2k_2}\big)^{1/2}\big\{\sum_{i=1}^n(\hat{\xi}_{ij_1k_1} - \xi_{ij_1k_1})^2w^{-1}_{j_1k_1}\big\}^{1/2}.
\end{aligned}$$

Since $E(\sum_{i=1}^n\xi^2_{ij_2k_2}w^{-1}_{j_2k_2}) = n$ for any $k_2 = 1, \ldots, s_n$, we have $\sum_{i=1}^n\xi^2_{ij_2k_2}w^{-1}_{j_2k_2} = O_p(n)$, uniformly for $k_2 = 1, \ldots, s_n$. Moreover,

$$\begin{aligned}
\sum_{i=1}^n\hat{\xi}^2_{ij_1k_1}w^{-1}_{j_1k_1} &\le 2\sum_{i=1}^n(\hat{\xi}_{ij_1k_1} - \xi_{ij_1k_1})^2w^{-1}_{j_1k_1} + 2\sum_{i=1}^n\xi^2_{ij_1k_1}w^{-1}_{j_1k_1} \\
&= O_p(k_1^{a+2} + n) = O_p(n),
\end{aligned}$$

uniformly for $k_1 = 1, \ldots, s_n$. In conclusion, we have

$$\begin{aligned}
|\theta^{j_1j_2\,(2)}_{k_1k_2}| &= n^{-1}O_p(n^{1/2})O_p(k_2^{a/2+1}) + n^{-1}O_p(n^{1/2})O_p(k_1^{a/2+1}) \\
&= O_p\Big(k_1^{a/2+1}n^{-1/2} + k_2^{a/2+1}n^{-1/2}\Big),
\end{aligned}$$

uniformly for $k_1, k_2 = 1, \ldots, s_n$.

For $\theta^{j_1j_2\,(3)}_{k_1k_2}$, $E(\theta^{j_1j_2\,(3)}_{k_1k_2})^2 \le n^{-1}\big\{E(\xi^4_{ij_1k_1}w^{-2}_{j_1k_1})E(\xi^4_{ij_2k_2}w^{-2}_{j_2k_2})\big\}^{1/2} \le c_1n^{-1}$, uniformly for $k_1, k_2 = 1, \ldots, s_n$ by (B1). It follows that $\theta^{j_1j_2\,(3)}_{k_1k_2} = O_p(n^{-1/2})$, uniformly for $k_1, k_2 = 1, \ldots, s_n$. Moreover, it is trivial that $\theta^{j_1j_2\,(4)}_{k_1k_2} = \theta^{j_1j_2\,(2)}_{k_1k_2} + \theta^{j_1j_2\,(3)}_{k_1k_2} = O_p(n^{-1/2}k_1^{a/2+1} + n^{-1/2}k_2^{a/2+1})$, uniformly for $k_1, k_2 = 1, \ldots, s_n$.

Based on (B5), we conclude that $z^4_{il} = O_p(1)$, uniformly for $l = 1, \ldots, p_n$. For $\theta^{(5)}_{jkl}$, we have

$$\begin{aligned}
|\theta^{(5)}_{jkl}| &\le n^{-1}\Big\{\sum_{i=1}^n(\hat{\xi}_{ijk} - \xi_{ijk})^2w^{-1}_{jk}\Big\}^{1/2}\big(\sum_{i=1}^n z^2_{il}\big)^{1/2} \\
&= n^{-1}O_p(k^{a/2+1})O_p(n^{1/2}) = O_p\Big(k^{a/2+1}n^{-1/2}\Big),
\end{aligned}$$

uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. Furthermore,

$$E(\theta^{(6)}_{jkl})^2 \le n^{-1}\big\{E(\xi^4_{ijk}w^{-2}_{jk})E(z^4_{il})\big\}^{1/2} = O\Big(n^{-1}\Big),$$

uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. Thus $\theta^{(6)}_{jkl} = O_p(n^{-1/2})$, uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. Then it follows that $\theta^{(7)}_{jkl} = \theta^{(5)}_{jkl} + \theta^{(6)}_{jkl} = O_p(k^{a/2+1}n^{-1/2})$, uniformly for $k = 1, \ldots, s_n$, $l = 1, \ldots, p_n$. For $\theta^{(8)}_{l_1 l_2}$, we have $E(\theta^{(8)}_{l_1 l_2})^2 \le n^{-1}\big\{E(z^4_{il_1})E(z^4_{il_2})\big\}^{1/2} = O(n^{-1})$, uniformly for $l_1 = 1, \ldots, p_n$, $l_2 = 1, \ldots, p_n$, and this entails that $\theta^{(8)}_{l_1 l_2} = O_p(n^{-1/2})$, uniformly for $l_1 = 1, \ldots, p_n$, $l_2 = 1, \ldots, p_n$.

For $\vartheta^{(1)}_{jl}$, given (A4), we have

$$\begin{aligned} E|\vartheta^{(1)}_{jl}| &\le \sum_{i=1}^{n} \sum_{k=s_n+1}^{\infty} |b_{jk0}|\, E\,|\xi_{ijk} z_{il}| \le \sum_{i=1}^{n} \sum_{k=s_n+1}^{\infty} |b_{jk0}|\, \{E(\xi^2_{ijk})E(z^2_{il})\}^{1/2} \\ &\le c_1 \sum_{i=1}^{n} \sum_{k=s_n+1}^{\infty} k^{-b} k^{-1/2} = O(n s_n^{-b+1/2}) = o(n^{1/2}). \end{aligned}$$

Hence $\vartheta^{(1)}_{jl} = o_p(n^{1/2})$, uniformly for $l = 1, \ldots, p_n$. For $\vartheta^{(2)}_{jl}$, we have

$$\begin{aligned} \vartheta^{(2)}_{jl} &= \sum_{i=1}^{n} \sum_{k=1}^{s_n} \Big\{ \int \hat{x}_{ij}(\hat{\phi}_{jk} - \phi_{jk}) + \int (\hat{x}_{ij} - x_{ij})\phi_{jk} \Big\} (\hat{b}_{jk} - b_{jk0}) z_{il} \\ &= \int (\sum_{i=1}^{n} \hat{x}_{ij} z_{il}) \{\sum_{k=1}^{s_n} (\hat{\phi}_{jk} - \phi_{jk})(\hat{b}_{jk} - b_{jk0})\} \\ &\quad + \int \{\sum_{i=1}^{n} (\hat{x}_{ij} - x_{ij}) z_{il}\} \{\sum_{k=1}^{s_n} \phi_{jk}(\hat{b}_{jk} - b_{jk0})\}. \end{aligned}$$

It follows that

$$\begin{aligned} (\vartheta^{(2)}_{jl})^2 &\le 2\|\sum_{i=1}^{n} \hat{x}_{ij} z_{il}\|^2 \|\sum_{k=1}^{s_n} (\hat{\phi}_{jk} - \phi_{jk})(\hat{b}_{jk} - b_{jk0})\|^2 \\ &\quad + 2\|\sum_{i=1}^{n} (\hat{x}_{ij} - x_{ij}) z_{il}\|^2 \|\sum_{k=1}^{s_n} \phi_{jk}(\hat{b}_{jk} - b_{jk0})\|^2. \end{aligned}$$

Since

$$E\|\sum_{i=1}^{n} \hat{x}_{ij} z_{il}\|^2 \le E(\sum_{i=1}^{n} \|\hat{x}_{ij}\||z_{il}|)^2 \le n^2\{E(\|\hat{x}_{1j}\||z_{1l}|)\}^2 + n\{E(\|\hat{x}_{1j}\|^4)E(z^4_{1l})\}^{1/2} = O(n^2),$$

we have $\|\sum_{i=1}^{n} \hat{x}_{ij} z_{il}\|^2 = O_p(n^2)$. Similarly,

$$E\|\sum_{i=1}^{n} (\hat{x}_{ij} - x_{ij}) z_{il}\|^2 \le n^2(E\|\hat{x}_{ij} - x_{ij}\||z_{il}|)^2 + n\{E(\|\hat{x}_{ij} - x_{ij}\|^4)E(z^4_{il})\}^{1/2} = o(n),$$

lead to $\|\sum_{i=1}^{n} (\hat{x}_{ij} - x_{ij}) z_{il}\|^2 = o_p(n)$. Moreover,

$$\begin{aligned} \|\sum_{k=1}^{s_n} (\hat{\phi}_{jk} - \phi_{jk})(\hat{b}_{jk} - b_{jk0})\|^2 &\le \|\check{b}^{(1)}_j - \tilde{b}^{(1)}_{j0}\|_2^2 \sum_{k=1}^{s_n} \|\hat{\phi}_{jk} - \phi_{jk}\|^2 w^{-1}_{jk} \\ &= O_p(r_n^2/n) O_p\{\sum_{k=1}^{s_n} (k^2/n) k^a\} \\ &= O_p(r_n^2 s_n^{a+3}/n^2), \end{aligned}$$

and

$$\|\sum_{k=1}^{s_n}\phi_{jk}(\hat b_{jk}-b_{jk0})\|^2 \le (\sum_{k=1}^{s_n}|\hat b_{jk}-b_{jk0}|)^2 \le \|\check b_j^{(1)}-\tilde b_{j0}^{(1)}\|_2^2\sum_{k=1}^{s_n}w_{jk}^{-1} = O_p(r_n^2 s_n^{a+1}/n),$$

by Theorem 1. In summary, we get $(\vartheta_{jl}^{(2)})^2 = o_p(r_n^2 s_n^{a+3}) = o_p(q_n s_n^{a+3}+s_n^{a+4})$, uniformly for $l=1,\ldots,p_n$. Under conditions (A3) and (A5), $\vartheta_{jl}^{(2)} = o_p(n^{1/2})$, uniformly for $l=1,\ldots,p_n$.

For $\vartheta_{jl}^{(3)}$, we have

$$\begin{aligned}(\vartheta_{jl}^{(3)})^2 &= [\int\{\sum_{i=1}^n(\hat x_{ij}-x_{ij})z_{il}\}(\sum_{k=1}^{s_n}\hat\phi_{jk}b_{jk0})]^2 \le \|\sum_{i=1}^n(\hat x_{ij}-x_{ij})z_{il}\|^2\|\sum_{k=1}^{s_n}\hat\phi_{jk}b_{jk0}\|^2\\ &\le (\sum_{k=1}^{s_n}|b_{jk0}|)^2 o_p(n) = o_p(n),\end{aligned}$$

hence $\vartheta_{jl}^{(3)} = o_p(n^{1/2})$ uniformly for $l=1,\ldots,p_n$.

Next, we show that $\vartheta_{jl}^{(4)} = \vartheta_{jl}^{(5)}+\vartheta_{jl}^{(6)}+\vartheta_{jl}^{(7)}+\vartheta_{jl}^{(8)}$. Recall that $\Xi_j = (\Xi_{j1},...,\Xi_{jq_n})^{\mathrm T}$ with $E\{x_{ij}(t)z_{il}\} = \Xi_{jl}(t)$. By lemma 1, we have the expression

$$\hat\phi_{jk}(t)-\phi_{jk}(t) = n^{-1/2}\sum_{v:v\neq k}(w_{jk}-w_{jv})^{-1}\phi_{jv}(t)\int\Re_j\phi_{jk}\phi_{jv}+\alpha_{jk}(t),$$

where $||\alpha_{jk}|| = O_p(k^{a+2}n^{-1})$, $\Re_j = n^{1/2}(\hat K_j-K_j)$, and $O_p(\cdot)$ is uniform in $k=1,\ldots,s_n$. Since

$$\vartheta_{jl}^{(4)} = \sum_{i=1}^n\sum_{k=1}^{s_n}(\tilde\xi_{ijk}-\xi_{ijk})b_{jk0}z_{il} = \int(\sum_{i=1}^n x_{ij}z_{il})\{\sum_{k=1}^{s_n}(\hat\phi_{jk}-\phi_{jk})b_{jk0}\},$$

substitute the expression for $\hat\phi_{jk}(t)-\phi_{jk}(t)$ into the above equation, we immediately get $\vartheta_{jl}^{(4)} = \vartheta_{jl}^{(5)}+\vartheta_{jl}^{(6)}+\vartheta_{jl}^{(7)}+\vartheta_{jl}^{(8)}$.

For $\vartheta_{jl}^{(5)}$, it is obvious that

$$(\vartheta_{jl}^{(5)})^2 \le \|\sum_{k=1}^{s_n}(\hat\phi_{jk}-\phi_{jk})b_{jk0}\|^2\|\sum_{i=1}^n(x_{ij}z_{il}-\Xi_{jl})\|^2,$$

where

$$\begin{aligned}E\|\sum_{i=1}^n(x_{ij}z_{il}-\Xi_{jl})\|^2 &= \int E[\{\sum_{i=1}^n(x_{ij}z_{il}-\Xi_{jl})\}^2] = n\int \mathrm{var}(x_{ij}z_{il})\\ &\le n\int\{E(x_{ij}^4)E(z_{il}^4)\}^{1/2} = O(n),\end{aligned}$$

and

$$\|\sum_{k=1}^{s_n}(\hat\phi_{jk}-\phi_{jk})b_{jk0}\| \le \sum_{k=1}^{s_n}|b_{jk0}|\|\hat\phi_{jk}-\phi_{jk}\| = O_p\big(\sum_{k=1}^{s_n}k^{-b}kn^{-1/2}\big) = O_p(n^{-1/2}),$$

thus $(\vartheta_{jl}^{(5)})^2 = O_p(1) = o_p(n)$, and it follows that $\vartheta_{jl}^{(5)} = o_p(n^{1/2})$ uniformly for $l = 1, \ldots, p_n$. For $\vartheta_{jl}^{(6)}$, we have

$$(\vartheta_{jl}^{(6)})^2 \le n^2 \int (Ex_{ij}z_{il})^2 \| \sum_{k=1}^{s_n} b_{jk0}\alpha_{jk}\|^2 \le n^2 \int \{E(x_{ij}^4)E(z_{il}^4)\}^{1/2} (\sum_{k=1}^{s_n} |b_{jk0}|\|\alpha_{jk}\|)^2$$
$$= n^2 O_p\{(\sum_{k=1}^{s_n} k^{-b}k^{a+2}n^{-1})^2\} = o_p(n),$$

hence $\vartheta_{jl}^{(6)} = o_p(n^{1/2})$ uniformly for $l = 1, \ldots, p_n$.

From the proof of lemma 1, it is easy to see that $\Re_j - \Re_j^* = n^{1/2}(\hat{K}_j - \tilde{K}_j) = o_p(1)$, which follows that $\vartheta_{jl}^{(7)} = o_p(n^{1/2})$, uniformly for $l = 1, \ldots, p_n$.

Proof of Lemma 3. First, it is obvious that $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| \le ||\check{N}_1^{\mathrm{T}}\check{N}_1/n - U_1||_1$, where $||.||_1$ is the L_1 norm for matrix. Since

$$||\check{N}_1^{\mathrm{T}}\check{N}_1/n - U_1||_1 \le O_p(\sum_{k_1=1}^{s_n} |\theta_{k_1 s_n}^{j_1 j_2\ (4)}| + \sum_{l=1}^{q_n} |\theta_{j_1 s_n l}^{(7)}| + \sum_{k_1=1}^{s_n} |\theta_{j_1 k_1 q_n}^{(7)}| + \sum_{l_1=1}^{q_n} |\theta_{l_1 q_n}^{(8)}|)$$
$$= O_p\{\sum_{k_1=1}^{s_n} (k_1^{a/2+1}n^{-1/2} + s_n^{a/2+1}n^{-1/2}) + \sum_{l=1}^{q_n} (s_n^{a/2+1}n^{-1/2})$$
$$+ \sum_{k_1=1}^{s_n} (k_1^{a/2+1}n^{-1/2}) + \sum_{l_1=1}^{q_n} n^{-1/2}\}$$
$$= O_p(s_n^{a/2+2}n^{-1/2} + q_n s_n^{a/2+1}n^{-1/2}),$$

by Lemma 2 (a), hence we have $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| = O_p(s_n^{a/2+2}n^{-1/2} + q_n s_n^{a/2+1}n^{-1/2})$. Under conditions (A3) and (A5), it is obvious that $|\lambda_{\min}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\min}(U_1)| = o_p(1)$. Similarly $|\lambda_{\max}(\check{N}_1^{\mathrm{T}}\check{N}_1/n) - \lambda_{\max}(U_1)| \le ||\check{N}_1^{\mathrm{T}}\check{N}_1/n - U_1||_1 = o_p(1)$, and we skip the details here.

Proof of Lemma 4. By Lemma 3 and (B5), we know that $\check{N}_1^{\mathrm{T}}\check{N}_1$ is invertible, hence $P_{\check{N}_1}$ exists. For Λ_1, we have

$$\Lambda_1 = P_{\check{N}_1}(Y_M - \check{N}_1\tilde{\eta}_{10}) = P_{\check{N}_1}\{Y_M - \tilde{N}_1\tilde{\eta}_{10} - \nu + \nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\}$$
$$= P_{\check{N}_1}\{\epsilon + \nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\},$$

where $\epsilon = Y_M - \tilde{N}_1\tilde{\eta}_{10} - \nu = (\epsilon_1, ..., \epsilon_n)^{\mathrm{T}}$, $\nu = (\nu_1, ..., \nu_n)^{\mathrm{T}}$ with $\nu_i = \sum_{j=1}^{g}\sum_{k=s_n+1}^{\infty} \xi_{ijk}b_{jk0}$.

For $P_{\check{N}_1}\epsilon$, we have

$$E\|P_{\check{N}_1}\epsilon\|^2 = E(\epsilon^{\mathrm{T}}P_{\check{N}_1}\epsilon) = E\{E(\epsilon^{\mathrm{T}}P_{\check{N}_1}\epsilon \mid \epsilon)\} = E[\mathrm{tr}\{P_{\check{N}_1}E(\epsilon\epsilon^{\mathrm{T}})\}]$$
$$= \sigma^2\mathrm{tr}(P_{\check{N}_1}) = \sigma^2(q_n + gs_n) = O(q_n + s_n),$$

hence $\|P_{\check{N}_1}\epsilon\|^2 = O_p(q_n + s_n)$.

For $P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}$, we have

$$\begin{aligned}
\|P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 &\leq \|(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 \leq O\big[\sum_{i=1}^{n}\{\sum_{j=1}^{g}\sum_{k=1}^{s_n}(\hat{\xi}_{ijk} - \xi_{ijk})b_{jk0}\}^2\big] \\
&\leq O\big[2g\sum_{i=1}^{n}\sum_{j=1}^{g}\{\sum_{k=1}^{s_n}(\tilde{\xi}_{ijk} - \xi_{ijk})b_{jk0}\}^2 + 2g\sum_{i=1}^{n}\sum_{j=1}^{g}\{\sum_{k=1}^{s_n}(\hat{\xi}_{ijk} - \tilde{\xi}_{ijk})b_{jk0}\}^2\big] \\
&\leq O\big[\sum_{j=1}^{g}\sum_{i=1}^{n}\|x_{ij}\|^2 O_p\big((\sum_{k=1}^{s_n}k^{-b}kn^{-1/2})^2\big]\big) + O_p\big[\sum_{j=1}^{g}\sum_{i=1}^{n}\{\|\hat{x}_{ij} - x_{ij}\|^2(\sum_{k=1}^{s_n}k^{-b})^2\}\big] \\
&= O_p(1),
\end{aligned}$$

since $b > 2$ implies that $\sum_{j=1}^{g}\sum_{i=1}^{n}||x_{ij}||^2 O_p\big\{(\sum_{k=1}^{s_n}k^{-b}kn^{-1/2})^2\big\} = O_p(1)$, and Lemma 1 entails that $\sum_{j=1}^{g}\sum_{i=1}^{n}\{\|\hat{x}_{ij} - x_{ij}\|^2(\sum_{k=1}^{s_n}k^{-b})^2\} = O_p(1)$. It then follows that $\|P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 = O_p(1)$.

For $P_{\check{N}_1}\nu$, it is obvious that $\|P_{\check{N}_1}\nu\|^2 \leq \|\nu\|^2$. For $i = 1, \ldots, n$, we have

$$\begin{aligned}
E(\nu_i^2) &= O\Big\{\sum_{j=1}^{g}\text{var}(\sum_{k=s_n+1}^{\infty}\xi_{ijk}b_{jk0})\Big\} = O\Big(\sum_{j=1}^{g}\sum_{k=s_n+1}^{\infty}b_{jk0}^2 w_{jk}\Big) \\
&= O\Big(\sum_{j=1}^{g}\sum_{k=s_n+1}^{\infty}k^{-2b}k^{-1}\Big) = O(s_n^{-2b}).
\end{aligned}$$

It follows that $\|P_{\check{N}_1}\nu\|^2 = O_p(ns_n^{-2b})$. In summary,

$$\begin{aligned}
\|\Lambda_1\|_2^2 &\leq O(\|P_{\check{N}_1}\epsilon\|^2 + \|P_{\check{N}_1}(\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10}\|^2 + \|P_{\check{N}_1}\nu\|^2) \\
&= O_p(q_n + s_n + 1 + ns_n^{-2b}) = O_p(q_n + s_n) = O_p(r_n^2),
\end{aligned}$$

where $r_n^2 = q_n + s_n$. This completes the proof.

5. PROOFS OF MAIN THEOREMS

Proof of Theorem 1. First, we constrain $Q_n(\tilde{\eta})$ on the subspace, where the true zero parameters are set as 0, that is $\{\tilde{\eta} \in R^{ds_n+p_n} : \tilde{b}_j^{(1)} = 0, k = g+1, \ldots, d, \gamma^{(2)} = 0\}$, and prove consistency in the $(gs_n + q_n)$-dimensional space. Define the constrained penalized function

$$\begin{aligned}
\bar{Q}_n(\tilde{\eta}_1) = \sum_{i=1}^{n}\{y_i - \sum_{j=1}^{g}\sum_{k=1}^{s_n}(\hat{\xi}_{ijk}w_{jk}^{-1/2})\tilde{b}_{jk} - {z_i}^{(1)\mathrm{T}}\gamma^{(1)}\}^2 + 2n\sum_{l=1}^{q_n}J_{\lambda_n}(|\gamma_l|) \\
+2n\sum_{j=1}^{g}J_{\lambda_{jn}}\big(\|b_j^{(1)}\|\big),
\end{aligned}$$

where $\tilde{\eta}_1 = (\tilde{b}_1^{(1)\mathrm{T}}, ..., \tilde{b}_g^{(1)\mathrm{T}}, \gamma^{(1)\mathrm{T}})^{\mathrm{T}}$ and $z_i^{(1)} = (z_{i1}, ..., z_{iq_n})^{\mathrm{T}}$. We now show there exists a local minimizer $\check{\eta}_1$ such that $\|\check{\eta}_1 - \tilde{\eta}_{10}\| = O_p(r_n n^{-1/2})$ with $r_n = (q_n + r_n)^{1/2}$. Let $\alpha_n = r_n n^{-1/2}$, we aim to show that for any $\epsilon > 0$, there exists a large constant C such that $pr\{\inf_{||u||=C}\bar{Q}_n(\tilde{\eta}_{10} + \alpha_n u) > \bar{Q}_n(\tilde{\eta}_{10})\} \geq 1 - \epsilon$ for large n, which implies that there exists a local minimizer $\check{\eta}_1$ of $\bar{Q}_n(\tilde{\eta}_1)$ such that $\|\check{\eta}_1 - \tilde{\eta}_{10}\| = O_p(\alpha_n)$. Here, $u = (u^{(1)\mathrm{T}}, u^{\gamma\mathrm{T}})^{\mathrm{T}}$ with

$u^{(1)} = (u_1^{(1)\mathrm{T}}, \dots, u_g^{(1)\mathrm{T}})^{\mathrm{T}}$ and $u^{\gamma} = (u_1^{\gamma}, \dots, u_{q_n}^{\gamma})^{\mathrm{T}}$. We have

$$\begin{aligned}
&\bar{Q}_n(\tilde{\eta}_{10} + \alpha_n u) - \bar{Q}_n(\tilde{\eta}_{10}) \\
&\geq \|\check{N}_1 \alpha_n u\|^2 - 2\Lambda_1^{\mathrm{T}} \check{N}_1 \alpha_n u + 2n[\sum_{l=1}^{q_n} \{J_{\lambda_n}(|\gamma_{l0} + \alpha_n u_l^{\gamma}|) - J_{\lambda_n}(|\gamma_{l0}|)\} \\
&\quad + \sum_{j=1}^{g} \{J_{\lambda_{jn}}(\|b_{j0}^{(1)} + \alpha_n A_j^{-1} u_j^{(1)}\|) - J_{\lambda_{jn}}(\|b_{j0}^{(1)}\|)\}] \\
&\geq n\lambda_{\min}(\check{N}_1^{\mathrm{T}} \check{N}_1/n)\alpha_n^2 \|u\|^2 - 2n^{1/2}\|\Lambda_1\|_2 \lambda_{\max}^{1/2}(\check{N}_1^{\mathrm{T}} \check{N}_1/n)\alpha_n \|u\| \\
&\quad + 2n[\sum_{l=1}^{q_n} \{J'_{\lambda_n}(|\gamma_{l0}|)\operatorname{sgn}(\gamma_{l0})\alpha_n u_l^{\gamma} + J''_{\lambda_n}(|\gamma_{l0}|)\alpha_n^2 (u_l^{\gamma})^2 (1 + o(1))\} \\
&\quad + \sum_{j=1}^{g} \{J'_{\lambda_{jn}}(\|b_{j0}^{(1)}\|)\alpha_n \|A_j^{-1} u_j^{(1)}\| + J''_{\lambda_{jn}}(\|b_{j0}^{(1)}\|)\alpha_n^2 \|A_j^{-1} u_j^{(1)}\|^2 (1 + o(1))\}].
\end{aligned}$$

The Taylor expansion in the second inequality holds because $\alpha_n w_{js_n}^{-1/2} \leq r_n s_n^{a/2} n^{-1/2} = o(1)$ by condition (A5). As for the smoothly clipped absolute deviation penalty, we have $J'_{\lambda_n}(|\gamma_{l0}|) = J''_{\lambda_n}(|\gamma_{l0}|) = 0$ for all $l = 1, \dots, q_n$ under condition (A7), and $J'_{\lambda_{jn}}(\|b_{j0}^{(1)}\|) = J''_{\lambda_{jn}}(\|b_{j0}^{(1)}\|) = 0$ for all $j = 1, \dots, g$ since $\|b_{j0}^{(1)}\| \geq C_1$ for some constant C_1. Thus, we can see that

$$\begin{aligned}
&\bar{Q}_n(\tilde{\eta}_{10} + \alpha_n u) - \bar{Q}_n(\tilde{\eta}_{10}) \\
&\geq n\lambda_{\min}(\check{N}_1^{\mathrm{T}} \check{N}_1/n)\alpha_n^2 \|u\|_2^2 - 2n^{1/2}\|\Lambda_1\|_2 \lambda_{\max}^{1/2}(\check{N}_1^{\mathrm{T}} \check{N}_1/n)\alpha_n \|u\|_2 \\
&\geq c_3 n\alpha_n^2 \|u\|^2 - c_4 n^{1/2} r_n \alpha_n \|u\| = c_3 r_n^2 \|u\|^2 - c_4 r_n^2 \|u\|,
\end{aligned}$$

where c_3, c_4 are some positive constants, and the second inequality is by Lemma 3 and (B5). When C is large enough, we have $\bar{Q}_n(\tilde{\eta}_{10} + \alpha_n u) - \bar{Q}_n(\tilde{\eta}_{10}) > 0$, which implies that there exists a local minimizer $\check{\eta}_1$ of $\bar{Q}_n(\tilde{\eta}_1)$ such that $\|\check{\eta}_1 - \tilde{\eta}_{10}\| = O_p(\alpha_n)$.

Next, we denote $\check{\eta} = (\check{b}_1^{(1)\mathrm{T}}, ..., \check{b}_g^{(1)\mathrm{T}}, 0^{\mathrm{T}}, ..., 0^{\mathrm{T}}, \hat{\gamma}^{(1)\mathrm{T}}, 0^{\mathrm{T}})^{\mathrm{T}}$, where $\check{\eta}_1 = (\check{b}_1^{(1)\mathrm{T}}, ..., \check{b}_g^{(1)\mathrm{T}}, \hat{\gamma}^{(1)\mathrm{T}})^{\mathrm{T}}$, and our second goal is to show that $\check{\eta}$ is a local minimizer of $Q_n(\tilde{\eta})$ over the whole space $R^{ds_n + p_n}$. Denote $S_l(\tilde{\eta}) = \partial\{(2n)^{-1}||Y_M - \hat{M}b^{(1)} - Z_M\gamma||^2\}/\partial\gamma_l$ and $S_j^*(\tilde{\eta}) = \partial\{(2n)^{-1}||Y_M - \hat{M}b^{(1)} - Z_M\gamma||^2\}/\partial b_j^{(1)}$. By Karush-Kuhn-Tucker condition, it suffices to show that $\tilde{\eta}$ satisfying the following conditions,

$$\begin{aligned}
&S_l(\tilde{\eta}) = 0, \text{ and }, |\gamma_l| \geq a\lambda_n \text{ for } l = 1, \dots, q_n, \\
&|S_l(\tilde{\eta})| \leq \lambda_n, \text{ and }, |\gamma_l| < \lambda_n \text{ for } l = q_n + 1, \dots, p_n, \\
&S_j^*(\tilde{\eta}) = 0, \text{ and }, \|b_j^{(1)}\| \geq a\lambda_{jn} \text{ for } j = 1, \dots, g, \\
&\|S_j^*(\tilde{\eta})\| \leq \lambda_{jn}, \text{ and }, \|b_j^{(1)}\| < \lambda_{jn} \text{ for } j = g + 1, \dots, d,
\end{aligned}$$

so that $\tilde{\eta}$ is a local minimizer of $Q_n(\tilde{\eta})$. When $l = 1, \dots, q_n$, since $\min_{l=1,\dots,q_n} |\hat{\gamma}_l| \geq \min_{l=1,\dots,q_n} |\gamma_{l0}| - \|\hat{\gamma}^{(1)} - \gamma_0^{(1)}\|$ and $\|\hat{\gamma}^{(1)} - \gamma_0^{(1)}\| = o_p(\lambda_n)$, hence under (A7), we have $\min_{l=1,\dots,q_n} |\hat{\gamma}_l|/\lambda_n \to \infty$ in probability. It follows that

$$pr(|\hat{\gamma}_l| \geq a\lambda_n \text{ for } l = 1, \dots, q_n) \to 1. \tag{2}$$

When $j = 1, \ldots, g$, since $\|\hat{b}_j^{(1)}\| \geq \|b_{j0}^{(1)}\| - \|\hat{b}_j^{(1)} - b_{j0}^{(1)}\|$, $\|\hat{b}_j^{(1)} - b_{j0}^{(1)}\| = o_p(s_n^{a/2}\alpha_n) = o_p(1)$ and $\min_{j=1,\ldots,g} \|b_{j0}^{(1)}\| > C_1$ for some positive constant C_1, hence, we have $\min_{j=1,\ldots,g} \|\hat{b}_j^{(1)}\|/\lambda_n \to \infty$ in probability. It follows that

$$pr(\|\hat{b}_j^{(1)}\| \geq a\lambda_{jn} \text{ for } j = 1, \ldots, g) \to 1. \tag{3}$$

When $l = 1, \ldots, q_n$ and $j = 1, \ldots, g$, $S_l(\check{\eta}) = 0$ and $S_j^*(\check{\eta}) = 0$ hold trivially since (2), (3) hold and $\check{\eta}_1$ is a local minimizer of $\bar{Q}_n(\tilde{\eta}_1)$.

Next, we show that $\check{\eta}$ satisfy

$$|S_l(\tilde{\eta})| \leq \lambda_n, \text{ and }, |\gamma_l| < \lambda_n \text{ for } l = q_n + 1, \ldots, p_n$$

and

$$||S_j^*(\tilde{\eta})|| \leq \lambda_{jn}, \text{ and }, ||b_j^{(1)}|| < \lambda_{jn} \text{ for } j = g+1, \ldots, d.$$

Since $\hat{\gamma}_l = 0$ for $l = q_n + 1, \ldots, p_n$ and $\hat{b}_j^{(1)} = 0$ for $j = g + 1, \ldots, d$, it suffices to show that

$$pr(|S_l(\check{\eta})| > \lambda_n \text{ for some } l = q_n + 1, \ldots, p_n) \to 0, \tag{4}$$

and

$$pr(||S_j^*(\check{\eta})|| > \lambda_{jn} \text{ for some } j = g + 1, \ldots, d) \to 0. \tag{5}$$

Denote $d_n = (S_{q_n+1}(\check{\eta}), ..., S_{p_n}(\check{\eta}))^{\mathrm{T}} = -n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}(Y_M - \check{N}_1\check{\eta}_1)$, where

$$Y_M - \check{N}_1\check{\eta}_1 = \epsilon + \nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1),$$

and $\nu = (\nu_1, ..., \nu_n)^{\mathrm{T}}$ with $\nu_i = \sum_{j=1}^{g}\sum_{k=s_n+1}^{\infty} \xi_{ijk} b_{jk0}$. From the proof of Lemma 4 and previous derivations, we have

$$\|\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1)\|^2 = O_p(r_n^2) = o_p(n\lambda_n^2).$$

Moreover, we have $\max_{l=q_n+1,\ldots,p_n} \sum_{i=1}^{n} z_{il}^2 = O_p(n)$ by (B5). Thus

$$\begin{aligned} &\|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}(\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1))\|_\infty \\ &\leq n^{-1}\big(\max_{l=q_n+1,\ldots,p_n} \sum_{i=1}^{n} z_{il}^2\big)^{1/2} \|\nu + (\tilde{N}_1 - \check{N}_1)\tilde{\eta}_{10} + \check{N}_1(\tilde{\eta}_{10} - \check{\eta}_1)\| = o_p(\lambda_n). \end{aligned}$$

Next, we denote $n^{-1/2}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon = (\delta_{q_n+1}, ..., \delta_{p_n})^{\mathrm{T}}$ with $\delta_l = n^{-1/2}\sum_{i=1}^{n} z_{il}\epsilon_i$ for $l = q_n + 1, \ldots, p_n$. As $\{\epsilon_i, i = 1, \ldots, n\}$ and $\{z_{il}, i = 1, \ldots, n\}$ are subGaussian random variables and independent of each other, combined with (B5), $\{\delta_l, l = 1, \ldots, p_n\}$ are subexponential random variables satisfying that, for any constant $C > 0$, there exists positive constants C_1 and C_2 that do not depend on l,

$$pr(|\delta_l| > Cn^{1/2}\lambda_n) \leq C_2 \exp(-2^{-1}C_1 n^{1/2}\lambda_n),$$

and it follows that

$$\begin{aligned} pr(\|n^{-1/2}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon\|_\infty > Cn^{1/2}\lambda_n) &\leq \sum_{l=q_n+1}^{p_n} pr(|\delta_l| > Cn^{1/2}\lambda_n) \\ &\leq O\{p_n \exp(-2^{-1}C_1 n^{1/2}\lambda_n)\} = O\{\exp(n^{\alpha} - 2^{-1}C_1 n^{1/2}\lambda_n)\} = o(1), \end{aligned}$$

provided that (A6) and (A7) hold. Hence, it is obvious that $\|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon\|_\infty = o_p(\lambda_n)$. In conclusion, we have $\|d_n\|_\infty \le \|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}\epsilon\|_\infty + \|n^{-1}\{Z_M^{(2)}\}^{\mathrm{T}}\{\nu + (\tilde N_1 - \check N_1)\tilde\eta_{10} + \check N_1(\tilde\eta_{10} - \check\eta_1)\}\|_\infty = o_p(\lambda_n)$ and this implies that (4) holds.

Next, we start to show (5). For $j = g+1, \dots, d$, we have $S_j^*(\check\eta) = -n^{-1}\hat M_j^{\mathrm{T}}(Y_M - \check N_1\check\eta_1) = (S_{j1}, ..., S_{js_n})^{\mathrm{T}}$ such that $\sum_{k=1}^{s_n} S_{jk}^2$ is bounded by

$$2\sum_{k=1}^{s_n}\big(n^{-1}\sum_{i=1}^{n}\hat\xi_{ijk}\epsilon_i\big)^2 + 2n^{-2}\sum_{k=1}^{s_n}\sum_{i=1}^{n}\hat\xi_{ijk}^2\|\nu + (\tilde N_1 - \check N_1)\tilde\eta_{10} + \check N_1(\tilde\eta_{10} - \check\eta_1)\|^2.$$

For $\sum_{k=1}^{s_n}\big(n^{-1}\sum_{i=1}^{n}\hat\xi_{ijk}\epsilon_i\big)^2$, we have

$$E\{\sum_{k=1}^{s_n}\big(n^{-1}\sum_{i=1}^{n}\hat\xi_{ijk}\epsilon_i\big)^2\} = n^{-2}\sum_{k=1}^{s_n}\sum_{i=1}^{n}\sigma^2 E(\hat\xi_{ijk}^2) \le n^{-2}\sum_{k=1}^{s_n}\sum_{i=1}^{n}\sigma^2 E(\|\hat x_{ij}\|^2) = O(s_n n^{-1}).$$

It is easy to see that $\sum_{k=1}^{s_n}\sum_{i=1}^{n}\hat\xi_{ijk}^2 = O_p(n\sum_{k=1}^{s_n} w_{jk}) = O_p(n)$. Thus

$$\sum_{k=1}^{s_n} S_{jk}^2 = O_p(s_n n^{-1}) + n^{-2}O_p(n)O_p(r_n^2) = O_p\{(q_n + s_n)n^{-1}\},$$

and it follows that

$$\|S_j^*(\check\eta)\|^2 = \sum_{k=1}^{s_n} S_{jk}^2 = O_p\{n^{-1}(q_n + s_n)\} = o_p(\lambda_n^2),$$

under (A7), which entails (5) immediately. Hence, we conclude that $\check\eta$ is a local minimizer of $Q_n(\tilde\eta)$ such that $||\hat\gamma - \gamma_0|| = O_p(r_n n^{-1/2})$, $||\check b^{(1)} - \tilde b_0^{(1)}|| = O_p(r_n n^{-1/2})$, $||\hat b^{(1)} - b_0^{(1)}|| = O_p(r_n s_n^{a/2} n^{-1/2})$, and $pr(\hat b_j^{(1)} = 0, j = g+1, \dots, d, \hat\gamma_2 = 0) \to 1$. This completes the proof.

Proof of Theorem 2. For convenience, define the following,

$$
\begin{aligned}
L_n(\tilde{\eta}) &= \sum_{i=1}^{n}\Big(y_i - \sum_{j=1}^{d}\sum_{k=1}^{\infty}\xi_{ijk}b_{jk} - z_i^{\mathrm{T}}\gamma\Big)^2,\\
T_{1n}(\tilde{\eta}) &= -\sum_{i=1}^{n}\Big(\sum_{j=1}^{d}\sum_{k=s_n+1}^{\infty}\xi_{ijk}b_{jk}\Big)^2 + 2\sum_{i=1}^{n}\Big(y_i - \sum_{j=1}^{d}\sum_{k=1}^{s_n}\xi_{ijk}b_{jk} - z_i^{\mathrm{T}}\gamma\Big)\\
&\qquad\Big(\sum_{j=1}^{d}\sum_{k=s_n+1}^{\infty}\xi_{ijk}b_{jk}\Big),\\
T_{2n}(\tilde{\eta}) &= -2\sum_{i=1}^{n}\Big(y_i - \sum_{j=1}^{d}\sum_{k=1}^{s_n}\xi_{ijk}b_{jk} - z_i^{\mathrm{T}}\gamma\Big)\Big\{\sum_{j=1}^{d}\sum_{k=1}^{s_n}(\hat{\xi}_{ijk} - \xi_{ijk})b_{jk}\Big\},\\
T_{3n}(\tilde{\eta}) &= \sum_{i=1}^{n}\Big\{\sum_{j=1}^{d}\sum_{k=1}^{s_n}(\hat{\xi}_{ijk} - \xi_{ijk})b_{jk}\Big\}^2,\\
T_{4n}(\tilde{\eta}) &= 2n\{\sum_{l=1}^{p_n}J_{\lambda_n}(|\gamma_l|) + \sum_{j=1}^{d}J_{\lambda_{jn}}\big(\|b_j^{(1)}\|\big)\},\\
P_n(\tilde{\eta}) &= L_n(\tilde{\eta}) + T_{1n}(\tilde{\eta}) = \sum_{i=1}^{n}\Big(y_i - \sum_{j=1}^{d}\sum_{k=1}^{s_n}\xi_{ijk}b_{jk} - z_i^{\mathrm{T}}\gamma\Big)^2.
\end{aligned}
$$

Then we have $Q_n(\tilde{\eta}) = L_n(\tilde{\eta}) + T_{1n}(\tilde{\eta}) + T_{2n}(\tilde{\eta}) + T_{3n}(\tilde{\eta}) + T_{4n}(\tilde{\eta})$, and denote $\nabla Q_n(\tilde{\eta}) = \partial Q_n(\tilde{\eta})/\partial\gamma^{(1)} = (\partial Q_n(\tilde{\eta})/\partial\gamma_1, ..., \partial Q_n(\tilde{\eta})/\partial\gamma_{q_n})^{\mathrm{T}}$ and $\nabla^2 Q_n(\tilde{\eta}) = \partial^2 Q_n(\tilde{\eta})/(\partial\gamma^{(1)}\partial\gamma^{(1)\mathrm{T}})$. Recall that $\breve{\eta}$ is a local minimizer of $Q_n(\tilde{\eta})$, hence, we have $\nabla Q_n(\breve{\eta}) = \nabla L_n(\breve{\eta}) + \nabla T_{1n}(\breve{\eta}) + \nabla T_{2n}(\breve{\eta}) + \nabla T_{3n}(\breve{\eta}) + \nabla T_{4n}(\breve{\eta}) = 0$.

It is obvious that $\nabla T_{3n}(\breve{\eta}) = 0$ because $T_{3n}(\tilde{\eta})$ doesn't depend on $\gamma^{(1)}$. Moreover, by properties of $\breve{\eta}$ derived before and (A7), we have that uniformly for $l = 1, \ldots, q_n$, $\partial T_{4n}(\breve{\eta})/\partial\gamma_l = 2nJ'_{\lambda_n}(|\hat{\gamma}_l|)\mathrm{sgn}(\hat{\gamma}_l) = 0$ with probability tending to one, thus $\nabla T_{4n}(\breve{\eta}) = 0$. For $\nabla P_n(\breve{\eta})$, by Taylor expansion we have,

$$
\begin{aligned}
\nabla P_n(\breve{\eta}) &= \nabla P_n(\tilde{\eta}_0) + \nabla^2 P_n(\tilde{\eta}_0)(\hat{\gamma}^{(1)} - \gamma_0^{(1)})\\
&= \nabla L_n(\tilde{\eta}_0) + \nabla T_{1n}(\tilde{\eta}_0) + \nabla^2 L_n(\tilde{\eta}_0)(\hat{\gamma}^{(1)} - \gamma_0^{(1)})\\
&= (-2\sum_{i=1}^{n}\epsilon_i z_i^{(1)}) + (-2\sum_{j=1}^{d}\vartheta_j^{(1)}) + 2\sum_{i=1}^{n}z_i^{(1)}z_i^{(1)\mathrm{T}}(\hat{\gamma}^{(1)} - \gamma_0^{(1)})\\
&= (-2\sum_{i=1}^{n}\epsilon_i z_i^{(1)}) + 2\sum_{i=1}^{n}z_i^{(1)}z_i^{(1)\mathrm{T}}(\hat{\gamma}^{(1)} - \gamma_0^{(1)}) + o_p(n^{1/2})1_{q_n},
\end{aligned}
$$

where the last equality is by Lemma 2 (b), and 1_{q_n} is the vector of ones with dimension q_n. For $\nabla T_{2n}(\check{\eta})$,

$$
\begin{aligned}
\nabla T_{2n}(\check{\eta}) &= 2\sum_{j=1}^{d}({\vartheta_j}^{(2)} + {\vartheta_j}^{(3)} + {\vartheta_j}^{(4)}) \\
&= 2\sum_{j=1}^{d}({\vartheta_j}^{(2)} + {\vartheta_j}^{(3)} + {\vartheta_j}^{(5)} + {\vartheta_j}^{(6)} + {\vartheta_j}^{(7)} + {\vartheta_j}^{(8)}) \\
&= o_p(n^{1/2})1_{q_n} + 2\sum_{j=1}^{d}{\vartheta_j}^{(8)},
\end{aligned}
$$

where the last two equalities are by Lemma 2 (b). Notice that ${\vartheta_j}^{(8)} = 0$ for $j = g+1, \ldots, d$,

$$
\begin{aligned}
\nabla Q_n(\check{\eta}) &= \nabla L_n(\check{\eta}) + \nabla T_{1n}(\check{\eta}) + \nabla T_{2n}(\check{\eta}) + \nabla T_{3n}(\check{\eta}) + \nabla T_{4n}(\check{\eta}) \\
&= o_p(n^{1/2})1_{q_n} + 2\sum_{j=1}^{g}{\vartheta_j}^{(8)} + (-2\sum_{i=1}^{n}\epsilon_i z_i^{(1)}) \\
&\quad +2\sum_{i=1}^{n} z_i^{(1)} {z_i^{(1)}}^{\mathrm{T}}(\hat{\gamma}^{(1)} - \gamma_0^{(1)}) = 0,
\end{aligned}
$$

and it follows that

$$
\begin{aligned}
n^{-1/2}\sum_{i=1}^{n} z_i^{(1)} {z_i^{(1)}}^{\mathrm{T}}(\hat{\gamma}^{(1)} - \gamma_0^{(1)}) &= n^{-1/2}\sum_{i=1}^{n}\epsilon_i z_i^{(1)} - n^{-1/2}\sum_{j=1}^{g}{\vartheta_j}^{(8)} + o_p(1)1_{q_n} \\
&= \sum_{i=1}^{n} W_{in} + o_p(1)1_{q_n},
\end{aligned}
$$

with $W_{in} = n^{-1/2}\epsilon_i z_i^{(1)} - n^{-1/2}\sum_{j=1}^{g}\sum_{k=1}^{s_n}\sum_{v\neq k} b_{jk0}(w_{jk} - w_{jv})^{-1}\langle \Xi_j, \phi_{jv}\rangle \int (x_{ij} \otimes x_{ij} - K_j)\phi_{jk}\phi_{jv}$. It is obvious that $\{W_{in}, i = 1, \ldots, n\}$ are independently and identically distributed. with zero mean and $\mathrm{cov}(W_{in}) = n^{-1}(\sigma^2\Sigma_1 + B_n)$. It satisfies to show that $V_n^{-1/2}A_n\sum_{i=1}^{n} W_{in}$ converges to $N(0_r, I_r)$ in distribution, where $V_n = A_n(\sigma^2\Sigma_1 + B_n)A_n^{\mathrm{T}} \to V^* = \sigma^2 H^* + B^*$, and V^* is positive definite. We start to show this by Linderberg Feller theorem.

First, we denote $\Pi_{in} = V_n^{-1/2}A_n W_{in}$, and it is trivial that $\{\Pi_{in}, i = 1, \ldots, n\}$ are independently and identically distributed centered random vectors with $\mathrm{cov}(\sum_{i=1}^{n}\Pi_{in}) = I_r$. Second, for any $\zeta > 0$, we have

$$
\sum_{i=1}^{n} E\left[\|\Pi_{in}\|_2^2 1_{q_n}\{\|\Pi_{in}\| > \zeta\}\right] \leq nE(\|\Pi_{1n}\|^4)^{1/2}\{pr(\|\Pi_{1n}\| > \zeta)\}^{1/2},
$$

where

$$
\begin{aligned}
\|\Pi_{1n}\|^2 = \Pi_{1n}^{\mathrm{T}}\Pi_{1n} = W_{1n}^{\mathrm{T}}A_n^{\mathrm{T}}V_n^{-1}A_n W_{1n} &\leq \lambda_{\max}(A_n^{\mathrm{T}}V_n^{-1}A_n)\|W_{1n}\|^2 \\
&\leq \lambda_{\max}(V_n^{-1})\lambda_{\max}(A_n^{\mathrm{T}}A_n)\|W_{1n}\|^2 = O(1)\|W_{1n}\|^2.
\end{aligned}
$$

It follows that $E\big(\|\Pi_{1n}\|^4\big) = O(1)E\big(\|W_{1n}\|^4\big) = O(1)q_n^2/n^2$. Moreover,

$$\begin{aligned} pr(\|\Pi_{1n}\| > \zeta) &= pr(\|V_n^{-1/2}A_nW_{1n}\| > \zeta) \le E(\|V_n^{-1/2}A_nW_{1n}\|^2)/\zeta^2 \\ &\le \lambda_{\max}(A_n^{\mathrm{T}}V_n^{-1}A_n)E(\|W_{1n}\|^2)/\zeta^2 \\ &= O(1)E(\|W_{1n}\|^2) = O(1)q_n/n. \end{aligned}$$

By previous derivations and assumptions, we know

$$\begin{aligned} \sum_{i=1}^{n} E\big[\|\Pi_{in}\|^2 1_{q_n}\{\|\Pi_{in}\| > \zeta\}\big] &= nO(q_n/n)O(q_n n^{-1/2}) \\ &= O(q_n^{3/2}n^{-1/2}) = o(1), \end{aligned}$$

since $q_n = o(n^{1/3})$. By Linderberg Feller theorem, we conclude that $V_n^{-1/2}A_n\sum_{i=1}^n W_{in}$ converges to $N(0_r, I_r)$ in distribution, which completes the proof.

REFERENCES

HALL, P. & HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Annals of Statistics* **35**, 70–91.

HALL, P. & HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B* **68**, 109–126.

HALL, P. & HOSSEINI-NASAB, M. (2009). Theory for high-order bounds in functional principal components analysis. *Mathematical Proceedings of the Cambridge Philosophical Society* **146**, 225–256.

[*Received January* $20XX$. *Revised June* $20XX$]