Supplement to "Adaptive Functional Linear Regression via Functional Principal Component Analysis and Block Thresholding"

T. Tony Cai[1], Linjun Zhang[1], and Harrison H. Zhou[2]

University of Pennsylvania and Yale University

Supplementary Material

This supplementary material includes the proof of the main results in the paper "Adaptive Functional Linear Regression via Functional Principal Component Analysis and Block Thresholding". Section S1.2 presents the proof of Theorem 1, and Section S1.3 proves the technical lemmas used in the proofs of the main results.

S1 Proofs

We shall only prove Theorem 1. The proof of Theorem 2 is similar and thus omitted. Before we present the proof of the main result, we first collect a few technical lemmas. These auxiliary lemmas will be proved in Section S1.3. We sharpen some results in Hall and Horowitz (2007) and give a risk bound for a blockwise James-Stein estimator. In this section we shall denote by C a generic constant which may vary from place to place.

S1.1 Technical lemmas

It was proposed in Hall and Horowitz (2007) to estimate b by $\sum_{j=1}^{m} \tilde{b}_j \hat{\phi}_j$ with a choice of cutoff $m = n^{\frac{1}{\alpha+2\beta}}$ to obtain minimax rate of convergence. The lemma below explains why there is no need ever to go beyond the $\hat{m}^*$-th term in defining the block thresholding procedure (16).

In Memory of Peter G. Hall.

Lemma 1. *Let γ and γ_1 be constants satisfying $\frac{1}{\alpha+2\beta} < \gamma < \frac{1}{3\alpha} < \gamma_1$ For all $D > 0$, there exists a constant C_D such that*

$$\mathbb{P}\left(n^{\gamma} \leq \hat{m}^* \leq n^{\gamma_1}\right) \geq 1 - c_D n^{-D}$$

where $\hat{m}^*$ is defined in (13).

In this section we set

$$\frac{1}{\alpha+2\beta} < \gamma < \min\left\{\frac{1+\varepsilon}{\alpha+2\beta}, \frac{1}{3\alpha}\right\}, \frac{1}{3\alpha} < \gamma_1 < \frac{1}{2(\alpha+1)} \tag{S1.1}$$

for a small $0 < \varepsilon < \min\left\{\frac{\alpha-2}{3}, \frac{2\beta-\alpha}{3\alpha+1}\right\}$. We give upper bounds to approximate eigenfunction ϕ_j by empirical eigenfunction $\hat{\phi}_j$ for $j \leq n^{\gamma_1}$.

Lemma 2. *For all $j \leq n^{\gamma_1}$, we have*

$$n\mathbb{E}\left\|\hat{\phi}_j - \phi_j\right\|^2 \leq Cj^2$$

and for any given $0 < \delta < 1$ and for all $D > 0$ there exists a constant $C_D > 0$ such that

$$\mathbb{P}\left\{n^{1-\delta}\left\|\hat{\phi}_j - \phi_j\right\|^2 \geq Cj^2\right\} \leq C_D n^{-D}.$$

Lemma 3 gives a variance bound for $\check{b}_j$, which helps us show that the variance of $\tilde{d}_j$ is approximately $\frac{\sigma^2}{n}$. This result is crucial for proposing a practical block thresholding procedure.

Lemma 3. *For $j \leq n^{\gamma_1}$ with $\gamma_1 < \frac{1}{2(\alpha+1)}$,*

$$\mathbb{E}\left(\check{b}_j - b_j\right)^2 \leq Cj^2/n.$$

In particular, this implies $\mathbb{V}ar(\check{b}_j) \leq Cj^2/n$ and $\mathbb{V}ar(\tilde{b}_j) = \sigma^2\theta_j^{-1}n^{-1}(1+o(1))$.

The following lemma gives bounds for the variance and mean squared error of $\tilde{d}_j$.

Lemma 4. *For $j \leq n^{\gamma_1}$ with $\gamma_1 < \frac{1}{2(\alpha+1)}$,*

$$\mathbb{V}ar(\tilde{d}_j) = \frac{\sigma^2}{n}(1+o(1)) \quad \text{and} \quad \mathbb{E}\left(\tilde{d}_j - \theta_j^{\frac{1}{2}}b_j\right)^2 \leq Cn^{-1}j^{2-\alpha}.$$

The following two lemmas will be used to analyze the factor ρ_j in equation (16).

Lemma 5. *Let $n^{\gamma} \leq m_1 \leq m_2 \leq n^{\gamma_1}$ and $m_2 - m_1 \geq n^{\delta}$ for some $\delta > 0$. Define $S^2 = \sum_{j=m_1}^{m_2} \tilde{d}_j^2$. For any given $\varepsilon > 0$ and all $D > 0$ there exists a constant $C_D > 0$ such that*

$$\mathbb{P}(S^2 > (1+\varepsilon)\,(m_2 - m_1)\,\frac{\sigma^2}{n}) \leq C_D n^{-D}.$$

Lemma 6. *Let $\tilde{d}_j = d'_j + \epsilon_j$ where $d'_j = E(\tilde{d}_j)$. Let $\varepsilon > 0$ be a fixed constant. If the block size $L_i = \mathrm{Card}(B_i) \geq n^{\delta}$ for some $\delta > 0$, then for any $D > 0$, there exists a constant $C_D > 0$ such that*

$$\mathbb{P}(\sum_{j \in B_i} \epsilon_j^2 > (1+\varepsilon)\, L_i \frac{\sigma^2}{n}) \leq C_D n^{-D}. \tag{S1.2}$$

And for all blocks B_i,

$$\mathbb{E} \sum_{j \in B_i} \epsilon_j^2 \leq C L_i \frac{\sigma^2}{n}. \tag{S1.3}$$

Conventional oracle inequalities were derived for Gaussian errors. In the current setting the errors are non-Gaussian. The following lemma gives an oracle inequality for a block thresholding estimator in the case of general error distributions. See Brown, Cai, Zhang, Zhao and Zhou (2010) for a proof.

Lemma 7. *Suppose $y_i = \theta_i + \epsilon_i, \quad i = 1, ..., L,$ where θ_i are constants and Z_i are random variables. Let $S^2 = \sum_{i=1}^{L} y_i^2$ and let*

$$\hat{\theta}_i = (1 - \frac{\lambda L}{S^2})_+ y_i.$$

Then

$$\mathbb{E}\|\hat{\theta} - \theta\|_2^2 \leq \min\{\|\theta\|_2^2, 4\lambda L\} + 4\mathbb{E}\|\epsilon\|_2^2 I(\|\epsilon\|_2^2 > \lambda L). \tag{S1.4}$$

S1.2 Proof of Theorem 1

We shall prove Theorem 1 for a general block thresholding estimator with the shrinkage factor $\rho_j = (1 - \frac{\lambda L_j \sigma^2}{n S_i^2})_+$ for a constant $\lambda > 1$.

Let γ and γ_1 be constants satisfying

$$\frac{1}{\alpha+2\beta} < \gamma < \min\left\{\frac{1+\varepsilon}{\alpha+2\beta}, \frac{1}{3\alpha}\right\} \le \frac{1}{3\alpha} < \gamma_1 < \frac{1}{2(\alpha+1)}$$

for a small $\varepsilon > 0$. Let $m_* = n^{\gamma}$ and write $\hat{b}$ as

$$\hat{b}(u) = \sum_{j=1}^{m_*} \rho_j \tilde{b}_j \hat{\phi}_j(u) + \sum_{j=m_*+1}^{n} \rho_j \tilde{b}_j \hat{\phi}_j(u). \tag{S1.5}$$

We shall show that $\mathbb{E}\|\hat{b}-b\|_2^2 \le Cn^{-\frac{2\beta-1}{\alpha+2\beta}}$. Note that

$$\begin{aligned}\mathbb{E}\|\hat{b}-b\|_2^2 &= \mathbb{E}\|\sum_{j=1}^{m_*} \hat{b}_j\hat{\phi}_j(u) + \sum_{j=m_*+1}^{n} \hat{b}_j\hat{\phi}_j(u) - \sum_{j=1}^{m_*} b_j\phi_j(u) - \sum_{j=m_*+1}^{\infty} b_j\phi_j(u)\|_2^2 \\ &\le 3\mathbb{E}\|\sum_{j=1}^{m_*} \hat{b}_j\hat{\phi}_j(u) - \sum_{j=1}^{m_*} b_j\phi_j(u)\|_2^2 + 3\sum_{j=m_*+1}^{n} \mathbb{E}(\hat{b}_j^2) + 3\sum_{j=m_*+1}^{\infty} b_j^2. \end{aligned} \tag{S1.6}$$

The last term (S1.6) is bounded by $Cn^{-\gamma(2\beta-1)} = o\left(n^{-(2\beta-1)/(\alpha+2\beta)}\right)$ since $\gamma > \frac{1}{\alpha+2\beta}$. We first show that the second term (S1.6) is small as well. Let $m^* = n^{\gamma_1}$ and let i_* and i^* be the corresponding block indices of the (m_*+1)-st and m^*-th term respectively. (That is, b_{m_*+1} is in the i_*-th block and b_{m^*} is in the i^*-th block.) Then it follows from Lemmas 1 and 5 that

$$\begin{aligned}\sum_{j=m_*+1}^{n} \mathbb{E}(\hat{b}_j^2) &= \left(\sum_{j=m_*+1}^{m^*} + \sum_{j=m^*+1}^{n}\right)\mathbb{E}(\rho_j^2\tilde{b}_j^2) \\ &\le \sum_{j=m_*+1}^{m^*} (\mathbb{E}\rho_j^4)^{\frac{1}{2}} (\mathbb{E}\tilde{b}_j^4)^{\frac{1}{2}} + \sum_{j=m^*+1}^{n} (\mathbb{E}\tilde{b}_j^4)^{\frac{1}{2}} \mathbb{P}^{\frac{1}{2}}\left(\hat{m}^* \ge n^{\gamma_1}+1\right) \\ &\le \sum_{i=i_*}^{i^*} \left[\mathbb{P}(S_i^2 > \lambda L\sigma^2/n)\right]^{1/2} \sum_{j\in B_i} (\mathbb{E}\tilde{b}_j^4)^{\frac{1}{2}} + \sum_{j=m^*+1}^{n} (\mathbb{E}\tilde{b}_j^4)^{\frac{1}{2}} \left[\mathbb{P}\left(\hat{m}^* \ge n^{\gamma_1}+1\right)\right]^{1/2} \\ &= o\left(n^{-\frac{2\beta-1}{\alpha+2\beta}}\right). \end{aligned}$$

We now turn to the first and dominant term in (S1.6). The Cauchy-Schwarz inequality yields

$$\begin{aligned}\mathbb{E}\|\sum_{j=1}^{m_*} \hat{b}_j\hat{\phi}_j(u) - \sum_{j=1}^{m_*} b_j\phi_j(u)\|_2^2 &\le 2\mathbb{E}\|\sum_{j=1}^{m_*}(\hat{b}_j - b_j)\hat{\phi}_j(u)\|_2^2 + 2\mathbb{E}\|\sum_{j=1}^{m_*} b_j(\hat{\phi}_j(u)-\phi_j(u))\|_2^2 \\ &\le 2\sum_{j=1}^{m_*}\mathbb{E}(\hat{b}_j-b_j)^2 + 2m_*\sum_{j=1}^{m_*} b_j^2\mathbb{E}\|\hat{\phi}_j(u)-\phi_j(u)\|_2^2. \end{aligned}$$

Lemma 2 implies the second term in the equation above is bounded by

$$C\frac{m_*}{n}\sum_{j=1}^{m_*} b_j^2 j^2 = O\left(n^{\gamma-1}\right) = o\left(n^{-\frac{2\beta-1}{\alpha+2\beta}}\right)$$

since $\sum_{j=1}^{m_*} b_j^2 j^2$ is finite and $\gamma < \frac{\alpha+1}{\alpha+2\beta}$ which implies $\gamma - 1 < -\frac{2\beta-1}{\alpha+2\beta}$. Set $d_j' = \mathbb{E}(\tilde{d}_j)$. Let κ_i be the smallest eigenvalue in the B_i-th block. Then

$$\begin{aligned}
&\sum_{j=1}^{m_*}\mathbb{E}(\hat{b}_j - b_j)^2 = \sum_{j=1}^{m_*}\mathbb{E}(\hat{\theta}_j^{-\frac{1}{2}}\hat{d}_j - \theta_j^{-\frac{1}{2}}d_j)^2 \le 2\sum_{j=1}^{m_*}\theta_j^{-1}\mathbb{E}(\hat{d}_j - d_j)^2 + 2\sum_{j=1}^{m_*}\mathbb{E}\left[\tilde{d}_j^2(\hat{\theta}_j^{-\frac{1}{2}} - \theta_j^{-\frac{1}{2}})^2\right] \\
\le\quad & 2\sum_{j=1}^{m_*}\theta_j^{-1}\mathbb{E}(\hat{d}_j - d_j)^2 + 2\sum_{j=1}^{m_*}\mathbb{E}\left[\tilde{d}_j^2(\hat{\theta}_j^{-\frac{1}{2}} - \theta_j^{-\frac{1}{2}})^2\right] \\
\le\quad & 2\sum_{i=1}^{i_*}\kappa_i^{-1}\sum_{j\in B_i}\mathbb{E}(\hat{d}_j - d_j')^2 + 2\sum_{i=1}^{i_*}\kappa_i^{-1}\sum_{j\in B_i}(d_j' - d_j)^2 + 2\sum_{j=1}^{m_*}\mathbb{E}\left[\tilde{d}_j^2(\hat{\theta}_j^{-\frac{1}{2}} - \theta_j^{-\frac{1}{2}})^2\right] \\
\equiv\quad & T_1 + T_2 + T_3.
\end{aligned}$$

From equations (S1.8) and (S1.9) and Lemma 4, it is easy to see

$$T_3 \le C\sum_{j=1}^{m_*}\mathbb{E}\{\tilde{d}_j^2\theta_j^{-3}(\hat{\theta}_j - \theta_j)^2\} = o\left(n^{-\frac{2\beta-1}{\alpha+2\beta}}\right).$$

We now turn to the dominant term $T_1 + T_2$. This term is most closely related to the block thresholding rule and we need to show that $T_1 + T_2 \le Cn^{-\frac{2\beta-1}{\alpha+2\beta}}$. To bound T_1, it is necessary to analyze the risk of the block thresholding rule for a single block B_i. It follows from Lemma 7 that

$$\sum_{j\in B_i}\mathbb{E}(\hat{d}_j - d_j')^2 \le \min\{4\lambda L_i\sigma^2/n, \sum_{j\in B_i}(d_j')^2\} + 4\mathbb{E}\{(\sum_{j\in B_i}\epsilon_j^2)\cdot I(\sum_{j\in B_i}\epsilon_j^2 > \lambda L_i\sigma^2/n)\} \quad \text{(S1.7)}$$

where $\lambda > 1$ is a constant. Lemma 4 implies

$$\left(d_j' - \theta_j^{\frac{1}{2}}b_j\right)^2 \le Cn^{-1}j^{2-\alpha}.$$

Note that for all j in B_i, we have $\theta_j^{-1} \asymp \kappa_i^{-1}$. Hence for $m_* = n^\gamma$ with $\gamma < \frac{1+\varepsilon}{\alpha+2\beta}$ we have

$$T_2 \le C\sum_{j=1}^{m_*}\theta_j^{-1}n^{-1}j^{2-\alpha} \le \frac{C_1}{n}\left(1 + m_*^3\right) = o\left(n^{-\frac{2\beta-1}{\alpha+2\beta}}\right)$$

Let $m = n^{\frac{1}{\alpha+2\beta}}$, then equation (S1.7) and Lemma 6 give

$$T_1 \leq C \sum_{j=1}^{m} \frac{j^\alpha}{n} + C \sum_{j=m+1}^{m_*} \left[\theta_j^{-1} \cdot \left(\theta_j^{1/2} b_j \right)^2 + \theta_j^{-1} n^{-1} j^{2-\alpha} \right] + C/n \leq C_1 n^{-\frac{2\beta-1}{\alpha+2\beta}}.$$

These together imply $\mathbb{E}\|\hat{b} - b\|_2^2 \leq C n^{-\frac{2\beta-1}{\alpha+2\beta}}$. ∎

S1.3 Proof of auxiliary lemmas

Let $\Delta^2 = \left\| \hat{K} - K \right\|^2 = \int\int \left(\hat{K}(u,v) - K(u,v) \right)^2 dudv$ and $\tau_j = \min_{k \leq j} (\theta_k - \theta_{k+1})$. It is known in Bhatia, Davis and McIntosh (1983) that

$$\sup_j \left| \hat{\theta}_j - \theta_j \right| \leq \Delta, \ \sup_{j \geq 1} \tau_j \left\| \hat{\phi}_j - \phi_j \right\| \leq 8^{1/2} \Delta. \tag{S1.8}$$

For $\varepsilon > 0$, it was shown in Hall and Hosseini-Nasab (2006, Lemma 3.3)

$$\mathbb{P}\left(\Delta > n^{\varepsilon - 1/2} \right) = c_D n^{-D} \tag{S1.9}$$

for each $D > 0$ under the assumption (19).

It is useful to rewrite $\tilde{b}_j$ as

$$\begin{aligned} \tilde{b}_j &= \hat{\theta}_j^{-1} \hat{g}_j = \hat{\theta}_j^{-1} \int \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y}) \{X_i(u) - \bar{X}(u)\} \hat{\phi}_j(u) \\ &= \hat{\theta}_j^{-1} \int \frac{1}{n} \sum_{i=1}^{n} \left(\langle X_i - \bar{X}, b \rangle + Z_i - \bar{Z} \right) \{X_i(u) - \bar{X}(u)\} \hat{\phi}_j(u) \\ &= \check{b}_j + \hat{\theta}_j^{-1} \frac{1}{n} \int (\underline{X} - \bar{X})' \hat{\phi}_j \cdot (\underline{Z} - \bar{Z}) = \check{b}_j + \hat{\theta}_j^{-1} \frac{1}{n} \hat{x}'_{\cdot,j} (\underline{Z} - \bar{Z}). \end{aligned}$$

Using the fact that for any two random variables X and Y, $\mathbb{V}ar(Y) = \mathbb{E}(\mathbb{V}ar(Y|X)) + \mathbb{V}ar(\mathbb{E}(Y|X))$ and the facts that $\underline{Z}$ has mean zero and is independent of $\underline{X}$, we have

$$\mathbb{V}ar(\tilde{b}_j) = \mathbb{V}ar(\check{b}_j) + \frac{\sigma^2}{n^2} \sum_{i=1}^{n} \mathbb{E}(\hat{\theta}_j^{-2} \hat{x}_{i,j}^2) = \mathbb{V}ar(\check{b}_j) + \frac{\sigma^2}{n} \mathbb{E} \hat{\theta}_j^{-1}.$$

Proof of Lemma 1

Recall that $\hat{m}^* = \arg\min \left\{ m : \hat{\theta}_m / \hat{\theta}_1 \leq n^{-1/3} \right\}$. Note that $\theta_j \geq M_0^{-1} j^{-\alpha}$. Since γ satisfies $\frac{1}{\alpha+2\beta} < \gamma < \frac{1}{3\alpha}$, then for $m \leq n^\gamma$ we have $\theta_m \geq M_0^{-1} n^{-\alpha\gamma}$. Since $\alpha\gamma < 1/3$, the equations

(S1.8) and (S1.9) imply that for any $D > 0$ there exists a constant $C_D > 0$ such that

$$\mathbb{P}\left(\cup_{m=1}^{n^\gamma}\left\{\hat{\theta}_m/\hat{\theta}_1 \leq n^{-1/3}\right\}\right) \leq c_D n^{-D}$$

and hence

$$\mathbb{P}\left(\hat{m}^* \leq n^\gamma\right) \leq c_D n^{-D}, \text{ i.e., } \mathbb{P}\left(\hat{m}^* \geq n^\gamma\right) \geq 1 - c_D n^{-D}.$$

Similarly, for $m \geq n^{\gamma_1}$ we have

$$\theta_m \leq M_0 n^{-\gamma_1 \alpha}$$

with $\alpha\gamma_1 > 1/3$, then

$$\mathbb{P}\left(\cup_{n \geq m \geq n^{\gamma_1}}\left\{\hat{\theta}_m/\hat{\theta}_1 > n^{-1/3}\right\}\right) \geq c_D n^{-D}$$

and hence

$$\mathbb{P}\left(\hat{m}^* \geq n^{\gamma_1}\right) \leq c_D n^{-D}, \text{ i.e., } \mathbb{P}\left(\hat{m}^* \leq n^{\gamma_1}\right) \geq 1 - c_D n^{-D}.$$

Thus we have

$$\mathbb{P}\left(n^{\gamma_1} \geq \hat{m}^* \geq n^\gamma\right) \geq 1 - c_D n^{-D}.$$

Proof of Lemma 2

Let $\mathcal{F}_j = \left\{\frac{1}{2}\left|\theta_j - \theta_k\right| \leq \left|\hat{\theta}_j - \theta_k\right| \leq 2\left|\theta_j - \theta_k\right|, k \neq j\right\}$, $j \leq n^{\gamma_1}$. From the assumption (18) we have $|\theta_j - \theta_k| \geq M_0^{-1} n^{-(\alpha+1)\gamma_1}$ with $(\alpha+1)\gamma_1 < \frac{1}{2}$. Then equations (S1.8) and (S1.9) imply that for any $D > 0$ there exists a constant $C_D > 0$ such that for $j \leq n^{\gamma_1}$

$$\mathbb{P}\left(\mathcal{F}_j^c\right) \leq c_D n^{-D} \tag{S1.10}$$

and consequently

$$\mathbb{P}\left(\cup_{j \leq n^{\gamma_1}, k \neq j}\left\{\frac{1}{2}\left|\theta_j - \theta_k\right| \leq \left|\hat{\theta}_j - \theta_k\right| \leq 2\left|\theta_j - \theta_k\right|\right\}\right) \geq 1 - c_D n^{-D}. \tag{S1.11}$$

Note that

$$\hat{\phi}_j - \phi_j = \sum_k \phi_k \int \left(\hat{\phi}_j - \phi_j\right)\phi_k = \sum_{k:k\neq j} \phi_k \int \hat{\phi}_j \phi_k + \phi_j \int \left(\hat{\phi}_j \phi_j - 1\right).$$

The facts $\int \hat{K}(u,v)\hat{\phi}_j(u)\,du = \hat{\theta}_j \hat{\phi}_j(v)$ and $\int K(u,v)\phi_k(v)\,dv = \theta_k \phi_k(u)$ imply

$$\int \hat{\phi}_j \phi_k = \left(\hat{\theta}_j - \theta_k\right)^{-1} \int\int \hat{K}(u,v) - K(u,v)\hat{\phi}_j(u)\,\phi_k(v)\,dudv.$$

Now it follows from the elementary inequality $1 - x \leq \sqrt{1-x} \leq 1 - x/2$ for $0 \leq x \leq 1$ (we assume that $\int \hat{\phi}_j \phi_j \geq 0$ WLOG) that

$$1 - \sum_{k\neq j} \left[\int \hat{\phi}_j \phi_k\right]^2 \leq \int \hat{\phi}_j \phi_j = \sqrt{1 - \sum_{k\neq j}\left[\int \hat{\phi}_j \phi_k\right]^2} \leq 1 - \frac{1}{2}\sum_{k\neq j}\left[\int \hat{\phi}_j \phi_k\right]^2.$$

Then we have

$$\left\|\hat{\phi}_j - \phi_j\right\|^2 \leq 2\sum_{k:k\neq j}\left[\left(\hat{\theta}_j - \theta_k\right)^{-1}\int\int\left(\hat{K}(u,v) - K(u,v)\right)\hat{\phi}_j(u)\,\phi_k(v)\,dudv\right]^2$$

which on $\mathcal{F}_j$ is further bounded by

$$\begin{aligned}
& 8\sum_{k:k\neq j}\left[(\theta_j - \theta_k)^{-1}\int\int\left(\hat{K}(u,v) - K(u,v)\right)\hat{\phi}_j(u)\,\phi_k(v)\,dudv\right]^2 \\
\leq\ & 16\sum_{k:k\neq j}(\theta_j - \theta_k)^{-2}\left\{\begin{array}{c}\left[\int\int\left(\hat{K}(u,v) - K(u,v)\right)\left(\hat{\phi}_j(u) - \phi_j(u)\right)\phi_k(v)\,dudv\right]^2 \\ +\left[\int\int\left(\hat{K}(u,v) - K(u,v)\right)\phi_j(u)\,\phi_k(v)\,dudv\right]^2\end{array}\right\} \\
\leq\ & Cn^{2\gamma_1(\alpha+1)}\Delta^2\left\|\hat{\phi}_j - \phi_j\right\|^2 + 16\sum_{k:k\neq j}(\theta_j - \theta_k)^{-2}\left[\int\int\left(\hat{K}(u,v) - K(u,v)\right)\phi_j(u)\,\phi_k(v)\,dudv\right]^2.
\end{aligned}$$

This implies for each $D > 0$

$$\mathbb{P}\left(\frac{1}{2}\left\|\hat{\phi}_j - \phi_j\right\|^2 \leq 16\sum_{k:k\neq j}(\theta_j - \theta_k)^{-2}\left[\int\int\left(\hat{K}(u,v) - K(u,v)\right)\phi_j(u)\,\phi_k(v)\,dudv\right]^2\right) \geq 1 - c_D n^{-D}.$$

Let $\eta_{i,j} = \int X_i \phi_j$ and $\overline{\eta}_j = \frac{1}{n}\sum_i \eta_{i,j}$, then

$$X_i - \bar{X} = \sum_{j=1}^{\infty}\left(\eta_{i,j} - \overline{\eta}_j\right)\phi_j.$$

Assume without loss of generality that $\mathbb{E}X = 0$ and for $k \neq j$ write

$$\int\int \left[\hat{K}(u,v) - K(u,v)\right] \phi_j(u)\,\phi_k(v)\,dudv = \frac{1}{n}\sum_{i=1}^{n}\left(\eta_{i,j} - \overline{\eta}_j\right)\left(\eta_{i,k} - \overline{\eta}_k\right) = \frac{1}{n}\sum_{i=1}^{n}\eta_{i,j}\eta_{i,k} - \overline{\eta}_k\overline{\eta}_j$$

where $\frac{1}{n}\sum_{i=1}^{n}\eta_{i,j}\eta_{i,k}$ is the dominating term. From the assumption (20) we have

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{i,j}\eta_{i,k}\right)^2 \le n^{-1}\mathbb{E}\left(\eta_{1,j}\eta_{1,k}\right)^2 \le n^{-1}\left[\mathbb{E}\eta_{1,j}^4\eta_{1,k}^4\right]^{1/2} \le C_1 n^{-1}\theta_j\theta_k.$$

Note that the spacing condition in (18) implies $\theta_m - \theta_{2m} \asymp m^{-\alpha}$, so we have

$$\begin{aligned}
\mathbb{E}\left\|\hat{\phi}_j - \phi_j\right\|^2 &\le C\sum_{k:k\neq j}\left(\theta_j - \theta_k\right)^{-2} n^{-1}\theta_j\theta_k \\
&\le Cn^{-1}\theta_j\sum_{k:k\neq j}\left\{j^{2\alpha}\sum_{k:k\ge 2j}k^{-\alpha} + \sum_{k:k\le j/2}k^{\alpha} + j^{2(\alpha+1)}\sum_{k:2j\ge k\ge j/2}\frac{k^{-\alpha}}{(1+|j-k|)^2}\right\} \\
&\le C_1 n^{-1} j^2 \qquad \text{(S1.12)}
\end{aligned}$$

and the first part of lemma is proved.

For the second part of the lemma, equation (S1.12) implies that it suffices to show that for $j \le n^{\gamma_1}$ and all $\delta > 0$

$$\mathbb{P}\left(\cup_k\left\{n^{1-\delta}k^{\alpha}j^{\alpha}\left[\int\int\left(\hat{K}(u,v) - K(u,v)\right)\phi_j(u)\,\phi_k(v)\,dudv\right]^2 \ge 1\right\}\right) \le c_D n^{-D}. \quad \text{(S1.13)}$$

For a large constant $q > 0$, we have

$$\begin{aligned}
&\mathbb{E}\sum_{k>n^q}\left(\theta_j - \theta_k\right)^{-2}\left[\int\int\left(\hat{K}(u,v) - K(u,v)\right)\phi_j(u)\,\phi_k(v)\,dudv\right]^2 \\
\le\ & C\mathbb{E}\frac{\theta_j^{-2}}{n^2}\sum_{k>n^q}\left(\sum_{i=1}^{n}\eta_{i,j}\eta_{i,k}\right)^2 \le C_1\theta_j^{-1}n^{-1}\theta_k \le C_q\theta_j^{-1}n^{-1}n^{-q\alpha},
\end{aligned}$$

which can be smaller than n^{-D} by setting q sufficiently large. It follows from the Markov inequality that

$$\mathbb{P}\left(\cup_{k>n^q}\left\{n^{1-\delta}k^{\alpha}j^{\alpha}\left[\int\int\left(\hat{K}(u,v) - K(u,v)\right)\phi_j(u)\,\phi_k(v)\,dudv\right]^2 \ge 1\right\}\right) \le c_D n^{-D}.$$

We need now only to consider $k \leq n^q$. Let w be a positive integer. Then

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\eta_{i,j}\eta_{i,k}\right)^{2w} \leq n^{-w}\mathbb{E}\left(\eta_{1,j}\eta_{1,k}\right)^{2w} \leq n^{-w}\left[\mathbb{E}\eta_{1,j}^{4w}\eta_{1,k}^{4w}\right]^{1/2} \leq C_1 n^{-w}\theta_j^w\theta_k^w$$

where the last inequality follows from (20). The Markov Inequality yields that for every integer $k > 0$

$$\mathbb{P}\left\{n^{1-\delta}k^{\alpha}j^{\alpha}\left[\int\int\left(\hat{K}(u,v)-K(u,v)\right)\phi_j(v)\,\phi_k(v)\,dudv\right]^2 \geq 1\right\} \leq C_2 n^{-w\delta}.$$

By choosing w sufficiently large, this implies

$$\mathbb{P}\left(\cup_{k\leq n^q}\left\{n^{1-\delta}k^{\alpha}j^{\alpha}\left[\int\int\left(\hat{K}(u,v)-K(u,v)\right)\phi_j(u)\,\phi_k(v)\,dudv\right]^2 \geq 1\right\}\right) \leq c_D n^{-D}.$$

The equation (S1.13) is then proved, and so is the second part of the lemma.

Proof of Lemmas 3 and 4

Since $\mathbb{V}ar(\check{b}_j) \leq \mathbb{E}(\int b\hat{\phi}_j - \int b\phi_j)^2$, we will analyze $\int b\hat{\phi}_j - \int b\phi_j = \int b\left(\hat{\phi}_j - \phi_j\right)$ instead. By the Cauchy-Schwarz inequality we have

$$\mathbb{E}\left[\int b\left(\hat{\phi}_j - \phi_j\right)\right]^2 \leq CE\left\|\hat{\phi}_j - \phi_j\right\|^2 \leq C_1 j^2/n = o\left(\frac{j^{\alpha}}{n}\right). \tag{S1.14}$$

We need to analyze $\tilde{d}_j = \hat{\theta}_j^{-\frac{1}{2}}\tilde{g}_j$. It follows from (12) that

$$\tilde{d}_j = \hat{\theta}_j^{-\frac{1}{2}}\tilde{g}_j = \hat{\theta}_j^{\frac{1}{2}}\check{b}_j + \hat{\theta}_j^{-\frac{1}{2}}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z}).$$

Hence, $\mathbb{E}(\tilde{d}_j) = \mathbb{E}(\hat{\theta}_j^{\frac{1}{2}}\check{b}_j)$. Same as before, it follows from the fact $\mathbb{V}ar(Y) = \mathbb{E}(\mathbb{V}ar(Y|X)) + \mathbb{V}ar(\mathbb{E}(Y|X))$ for any two random variables X and Y that

$$\mathbb{V}ar(\tilde{d}_j) = \mathbb{V}ar(\hat{\theta}_j^{\frac{1}{2}}\check{b}_j) + \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\mathbb{E}(\hat{\theta}_j^{-1}\hat{x}_{i,j}^2) = \mathbb{V}ar(\hat{\theta}_j^{\frac{1}{2}}\check{b}_j) + \frac{\sigma^2}{n}.$$

We need to bound $\mathbb{V}ar(\hat{\theta}_j^{\frac{1}{2}}\check{b}_j)$. Note that

$$
\begin{aligned}
\mathbb{V}ar(\hat{\theta}_j^{\frac{1}{2}}\check{b}_j) &\leq \mathbb{E}\left(\hat{\theta}_j^{\frac{1}{2}}\check{b}_j - \theta_j^{1/2}b_j\right)^2 \\
&\leq 2\mathbb{E}\left(\hat{\theta}_j^{\frac{1}{2}} - \theta_j^{1/2}\right)^2 b_j^2 + 2\theta_j\mathbb{E}\left(\check{b}_j - b_j\right)^2 \\
&\leq 2\mathbb{E}\left(\hat{\theta}_j^{\frac{1}{2}} - \theta_j^{1/2}\right)^2 b_j^2 + Cn^{-1}j^{2-\alpha} \\
&\leq 2\mathbb{E}\left(\frac{\hat{\theta}_j - \theta_j}{\theta_j^{1/2}}\right)^2 b_j^2 + Cn^{-1}j^{2-\alpha} \\
&\leq Cn^{-1}j^{-2\beta+\alpha} + Cn^{-1}j^{2-\alpha} \leq C_1 n^{-1}j^{2-\alpha}. \qquad \text{(S1.15)}
\end{aligned}
$$

Here the third inequality follows from (S1.14).

Proof of Lemma 5

Recall that

$$\tilde{d}_j = \hat{\theta}_j^{-1/2}\tilde{g}_j = \hat{\theta}_j^{1/2}\check{b}_j + \hat{\theta}_j^{-1/2}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z}).$$

The second term is dominant. We consider this term first. Since

$$\frac{1}{n}\sum_{i=1}^{n}\hat{x}_{i,j}\hat{x}_{i,k} = \hat{\theta}_j\delta_{j,k},$$

we have

$$\sum_{j=m_1}^{m_2}\left[\hat{\theta}_j^{-1/2}\frac{1}{\sqrt{n}}\hat{x}'_{\cdot,j}\underline{Z}\right]^2 \sim \frac{\sigma^2}{n}\chi^2_{m_2-m_1+1}.$$

So for any $D > 0$ there exists a constant $C_D > 0$ such that

$$\mathbb{P}\left(\sum_{j=m_1}^{m_2}\hat{\theta}_j^{-1}\left[\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z})\right]^2 > (1+\varepsilon)(m_2 - m_1)\frac{\sigma^2}{n}\right) \leq C_D n^{-D}. \qquad \text{(S1.16)}$$

Now we turn to the first term. It is easy to see

$$\sum_{j=m_1}^{m_2}\theta_j b_j^2 \leq \varepsilon\frac{m_2 - m_1}{n},$$

and for any $D > 0$

$$\mathbb{P}\left(\left|\hat{\theta}_j - \theta_j\right| \geq \varepsilon\theta_j,\ j \leq n^{\gamma_1}\right) \leq C_D n^{-D}.$$

We need only to show that for any $D > 0$

$$\mathbb{P}\left(\sum_{j=m_1}^{m_2} \theta_j \left[\int b\left(\hat{\phi}_j - \phi_j\right)\right]^2 > \varepsilon\left(m_2 - m_1\right)\frac{\sigma^2}{n}\right) \le C_D n^{-D}.$$

By the Cauchy-Schwarz inequality it suffices to show that for any $D > 0$

$$\mathbb{P}\left(\theta_j \int \left(\hat{\phi}_j - \phi_j\right)^2 > \varepsilon\frac{\sigma^2}{n}\right) \le C_D n^{-D}. \tag{S1.17}$$

This follows directly from Lemma 2.

Proof of Lemma 6

We write

$$\begin{aligned}
\sum_{j\in B_i} \epsilon_j^2 &= \sum_{j\in B_i}\left(\tilde{d}_j - d_j'\right)^2 = \sum_{j\in B_i}\left[\hat{\theta}_j^{\frac{1}{2}}\breve{b}_j - d_j' + \hat{\theta}_j^{-\frac{1}{2}}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z})\right]^2 \\
&= \sum_{j\in B_i}\left(\hat{\theta}_j^{\frac{1}{2}}\breve{b}_j - d_j'\right)^2 + 2\sum_{j\in B_i}\left(\hat{\theta}_j^{\frac{1}{2}}\breve{b}_j - d_j'\right)\hat{\theta}_j^{-\frac{1}{2}}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z}) + \sum_{j\in B_i}\left[\hat{\theta}_j^{-\frac{1}{2}}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z})\right]^2 \\
&\le \sum_{j\in B_i}\left(\hat{\theta}_j^{\frac{1}{2}}\breve{b}_j - d_j'\right)^2 + 2\left\{\sum_{j\in B_i}\left(\hat{\theta}_j^{\frac{1}{2}}\breve{b}_j - d_j'\right)^2\sum_{j\in B_i}\left[\hat{\theta}_j^{-\frac{1}{2}}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z})\right]^2\right\}^{1/2} \\
&\quad + \sum_{j\in B_i}\left[\hat{\theta}_j^{-\frac{1}{2}}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z} - \bar{Z})\right]^2
\end{aligned}$$

We first show equation (S1.2). From equation (S1.16) it suffices to prove that, when $\lambda = 1 + \varepsilon$ and $L_i \equiv |B_i| \ge n^{\delta}$ for some $\delta > 0$,

$$\mathbb{P}\left\{\sum_{j\in B_i}\left(\hat{\theta}_j^{\frac{1}{2}}\breve{b}_j - d_j'\right)^2 > \frac{\varepsilon}{3}L_i\frac{\sigma^2}{n}\right\} \le c_D n^{-D}$$

for any $D > 0$ where $C_D > 0$ is a constant. Note that, when $j \le n^{\gamma_1}$, for any $D > 0$ there exists a constant $C_D > 0$ such that

$$\mathbb{P}\left(\left|\hat{\theta}_j - \theta_j\right| \ge \varepsilon^2\theta_j\right) \le C_D n^{-D}$$

and

$$\mathbb{E}\left(\hat{\theta}_j^{\frac{1}{2}}\breve{b}_j - d_j'\right)^2 = o\left(\frac{1}{n}\right) \text{ as } j \to \infty.$$

It then suffices to show that for all $D > 0$

$$\mathbb{P}\left(\sum_{j\in B_i}\theta_j\left[\int b\left(\hat{\phi}_j-\phi_j\right)\right]^2>\varepsilon L_i\frac{\sigma^2}{n}\right)\le C_D n^{-D}.$$

This is true following similar arguments as in the proof of Lemma 5 with $L_i \ge n^\delta$ for some $\delta > 0$.

Equation (S1.3) follows easily from the fact

$$\mathbb{E}\sum_{j\in B_i}\epsilon_j^2=\mathbb{E}\sum_{j\in B_i}\left(\hat{\theta}_j^{\frac{1}{2}}\check{b}_j-d_j'\right)^2+\mathbb{E}\sum_{j\in B_i}\left[\hat{\theta}_j^{-\frac{1}{2}}\frac{1}{n}\hat{x}'_{\cdot,j}(\underline{Z}-\bar{Z})\right]^2$$

where the first term is bounded by $\frac{C}{n}L_i$ from equation (S1.15) and the second term is exactly $\frac{\sigma^2}{n}L_i$.

References

Bhatia, R., Davis, C. and McIntosh, A. (1983). Perturbation of spectral subspaces and solution of linear operator equations. *Linear Algebra Appl.* **52/53**, 45-67.

Brown, L. D., Cai, T. T., Zhang, R., Zhao, L. H. and Zhou, H. H. (2010). The Root-unroot algorithm for density estimation as implemented via wavelet block thresholding. *Probability Theory and Related Fields* **146**, 401-433.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70-91.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *J. R. Statist. Soc. B* **68**, 109-126.