Research Article

Juan Chen[1,2] and Yingchun Zhou[1,2]*

Supplementary Materials for Weighted Euclidean balancing for a matrix exposure in estimating causal effect

1. Proof of Theorem 1

The primal problem is

$$\min_{\mathbf{w}} \sum_{i=1}^{n} w_i log(w_i)$$

s.t.

$$\sum_{k=1}^{K} \{\lambda_k^2 [\sum_{i=1}^{n} w_i (m_{K,k}(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_{K,k})]^2\} \leq \delta. \tag{1}$$

Let $\| \theta \|_2 = \sqrt{\theta_1^2 + \cdots + \theta_K^2}$ be the l_2 norm for an arbitrary K-dimensional vector $\theta = (\theta_1, \ldots, \theta_K)^{'}$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_K)$, then the inequality constraint in the primal problem can be rewritten as $\| \sum_{i=1}^{n} w_i \Lambda (m_K(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_K) \|_2 \leq \sqrt{\delta}$. Let $\mathcal{A} \subseteq R^K$ be a convex set such that $\mathcal{A} = \{a \in R^K : \| a \|_2 \leq \sqrt{\delta}\}$. Define $I_{\mathcal{A}}(a) = 0$ if $a \in \mathcal{A}$ and $I_{\mathcal{A}}(a) = \infty$ otherwise. Then, the primal problem (1) is equivalent to the following optimaization problem:

$$\min_{\mathbf{w}} \sum_{i=1}^{n} w_i log(w_i) + I_{\mathcal{A}}(\sum_{i=1}^{n} w_i \Lambda (m_K(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_K)).$$

***Corresponding Author: Yingchun Zhou**[1,2]**:** [1]KLATASDS-MOE, School of Statistics, East China Normal University, 3663 North Zhongshan Road, Shanghai, 200062, P.R. China , [2]Institute of Brain and Education Innovation, East China Normal University, Shanghai, P.R. China; E-mail: yczhou@stat.ecnu.edu.cn

Let $h(w) = \sum_{i=1}^{n} w_i log(w_i)$, the conjugate function of h is

$$\begin{aligned} h^*(w) &= sup_t(\sum_{i=1}^{n} w_i t_i - \sum_{i=1}^{n} t_i log(t_i)) \\ &= sup_t \sum_{i=1}^{n}(w_i t_i - t_i log(t_i)) \\ &= \sum_{i=1}^{n} sup_{t_i}(w_i t_i - t_i log(t_i)) \\ &= \sum_{i=1}^{n} f^*(w_i), \end{aligned}$$

where $f^*(w_i) = \sup_{t_i}(w_i t_i - t_i log(t_i))$ is the conjugate function of $f(w_i) = w_i log(w_i)$. Let $g(\theta) = I_{\mathcal{A}}(\theta)$ for any $\theta \in R^K$, then the conjugate function of g is

$$\begin{aligned} g^*(\theta) &= \sup_a(\sum_{k=1}^{K} \theta_k a_k - T_{\mathcal{A}}(a)) \\ &= \sup_{||a||_2 \leq \sqrt{\delta}}(\sum_{k=1}^{K} \theta_k a_k) \\ &= \sup_{||a||_2 \leq \sqrt{\delta}}(|| \theta ||_2 || a ||_2) \\ &= \sqrt{\delta} \, || \theta ||_2 \, . \end{aligned}$$

Define the mapping $H : R^n \to R^K$ such that $Hw = \sum_{i=1}^{n} w_i \Lambda(m_K(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_K)$, then H is a bounded linear map. Let H^* be the adjoint operator of H, then for all $\theta = (\theta_1, \dots, \theta_K)' \in R^K$,

$$H^*\theta = (\sum_{k=1}^{K} \theta_k \lambda_k (m_{K,k}(\mathbf{T}_1, \mathbf{X}_1) - \bar{m}_{K,k}), \dots, \sum_{k=1}^{K} \theta_k \lambda_k (m_{K,k}(\mathbf{T}_n, \mathbf{X}_n) - \bar{m}_{K,k}))'.$$

Define $\tilde{\theta} = H\tilde{w} = \frac{1}{n^r} \sum_{i=1}^{n} \Lambda(m_K(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_K)$, where $\tilde{w} = (\frac{1}{n^r}, \dots, \frac{1}{n^r})' \in dom(F)$. Here, we choose b to be sufficiently large such that $|| \tilde{\theta} ||_2 \leq \sqrt{\delta}$, then we obtain that $g(\tilde{\theta}) = 0$ and g is continuous at $\tilde{\theta}$. Therefore, $\tilde{\theta} \in H(dom(F) \cap cont(g))$, which implies that $H(dom(F) \cap cont(g)) \neq \emptyset$. Here, $dom(F)$ and $cont(g)$ denotes the domain of F and the continuous set of g, respectively. Therefore, the strong duality condition of the Fenchel duality theorem is verified. Moreover,

$$F(H^*\theta) + g^*(-\theta) = \sum_{i=1}^{n} f^*(\sum_{k=1}^{K} \theta_k \lambda_k (m_{K,k}(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_{K,k})) + \sqrt{\delta} \, || \theta ||_2 \, .$$

According to the Fenchel duality theorem [1, Theorem B.39], we have

$$\min_w \sum_{i=1}^{n} w_i log(w_i) + I_{\mathcal{A}}(\sum_{i=1}^{n} w_i \Lambda(m_K(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_K))$$
$$= \min_\theta \sum_{i=1}^{n} f^*(\sum_{k=1}^{K} \theta_k \lambda_k(m_{K,k}(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_{K,k})) + \sqrt{\delta} \mid\mid \theta \mid\mid_2$$

Furthermore, since the strong duality condition holds, we can conclude that $H^*\hat{\theta}$ is a subgradient of F at $\hat{w}$. That is,

$$\sum_{k=1}^{K} \hat{\theta}_k \lambda_k(m_{K,k}(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_{K,k}) = log(\hat{w}_i) + 1.$$

Therefore, $\hat{w}_i = \exp(\sum_{k=1}^{K} \hat{\theta}_k \lambda_k(m_{K,k}(\mathbf{T}_i, \mathbf{X}_i) - \bar{m}_{K,k}) - 1)$. The proof of theorem 1 is completed.

2. Proof of Proposition 1

Using the law of total expectation and Assumption 1-3, we can deduce that

$$\begin{aligned}
&\mathbb{E}[w(Y - <\mathbf{B}, \mathbf{T}>)^2] \\
&= E[\frac{f(\mathbf{T})}{f(\mathbf{T} \mid \mathbf{X})}(Y - <\mathbf{B}, \mathbf{T}>)^2] \\
&= \mathbb{E}(\mathbb{E}[\frac{f(\mathbf{T})}{f(\mathbf{T} \mid \mathbf{X})}(Y - <\mathbf{B}, \mathbf{T}>)^2] \mid \mathbf{T} = \mathbf{t}, \mathbf{X} = \mathbf{x}) \quad \text{(the law of total expectation)} \\
&= \mathbb{E}(\frac{f(\mathbf{t})}{f(\mathbf{t} \mid \mathbf{x})}\mathbb{E}([(Y - <\mathbf{B}, \mathbf{T}>)^2] \mid \mathbf{T} = \mathbf{t}, \mathbf{X} = \mathbf{x})) \\
&= \int_{\mathcal{T}\times\mathcal{X}} \frac{f(\mathbf{t})}{f(\mathbf{t} \mid \mathbf{x})}\mathbb{E}[(Y(\mathbf{t}) - <\mathbf{B}, \mathbf{t}>)^2 \mid \mathbf{T} = \mathbf{t}, \mathbf{X} = \mathbf{x}]f(\mathbf{t} \mid \mathbf{x})d\mathbf{t}d\mathbf{x} \quad \text{(using Assumption 3)} \\
&= \int_{\mathcal{T}\times\mathcal{X}} \mathbb{E}[(Y(\mathbf{t}) - <\mathbf{B}, \mathbf{t}>)^2 \mid \mathbf{T} = \mathbf{t}, \mathbf{X} = \mathbf{x}]f(\mathbf{t})f(\mathbf{x})d\mathbf{t}d\mathbf{x} \\
&= \int_{\mathcal{T}\times\mathcal{X}} \mathbb{E}[(Y(\mathbf{t}) - <\mathbf{B}, \mathbf{t}>)^2 \mid \mathbf{X} = \mathbf{x}]f(\mathbf{t})f(\mathbf{x})d\mathbf{t}d\mathbf{x} \quad \text{(using Assumption 1)} \\
&= \int_{\mathcal{T}} \mathbb{E}[(Y(\mathbf{t}) - <\mathbf{B}, \mathbf{t}>)^2]f(\mathbf{t})d\mathbf{t}.
\end{aligned}$$

Hence, we complete the proof of Proposition 1.

3. Proof of Theorem 2

To prove Theorem 2, we first prove the following lemma.
Lemma 1. There exists a global minimizer $\hat{\theta}$ such that

$$|| \hat{\theta} - \theta^* ||_2 = O_p(K^{1/2}(logK)/n^{3/2} + K^{1/2-\tilde{\alpha}}). \tag{2}$$

Proof: Define $\rho(x) = e^{x-1}$, $A_i = B(T_i, X_i) = (\lambda_1 M_1(T_i, X_i), \dots, \lambda_K M_K(T_i, X_i))$, then the optimal objective is

$$G(\theta) = \frac{1}{n}\sum_{i=1}^{n} \rho(A_i^\top \theta) + \sqrt{\delta} \, || \, \theta \, ||_2, \tag{3}$$

where $G(\cdot)$ is convex in θ. To show that a minimizier $\triangle^*$ of $G(\theta^* + \triangle)$ exists in $\mathcal{C} = \{\triangle \in R^K : || \triangle ||_2 \leq C(K^{1/2}(logK)/n^{3/2} + K^{1/2-\tilde{\alpha}})\}$ for some constant C, it suffices to show that

$$E\{inf_{\triangle \in \mathcal{C}} G(\theta^* + \triangle) - G(\theta^*) > 0\} \to 1, as\ n \to \infty, \tag{4}$$

by the continuity of $G(\cdot)$.

To show (3), we use mean value theorem:

$$\begin{aligned} &G(\theta^* + \triangle) - G(\theta^*) \\ &\geq \frac{1}{n}\triangle^\top \sum_{i=1}^{n}(\rho'(A_i^\top \theta^*)A_i) + \frac{1}{2n}\triangle^\top\{\sum_{i=1}^{n}(\rho''(A_i^\top \tilde{\theta})A_i A_i^\top)\}\triangle - \sqrt{\delta} \, || \triangle ||_2 \\ &\geq - || \triangle ||_2 || \frac{1}{n}\sum_{i=1}^{n}(\rho'(A_i^\top \theta^*)A_i) ||_2 + \frac{1}{2n}\triangle^\top\{\sum_{i=1}^{n}(\rho''(A_i^\top \tilde{\theta})A_i A_i^\top)\}\triangle - \sqrt{\delta} \, || \triangle ||_2 \\ &= \frac{1}{2n}\triangle^\top\{\sum_{i=1}^{n}(\rho''(A_i^\top \tilde{\theta})A_i A_i^\top)\}\triangle - (\sqrt{\delta} + || \frac{1}{n}\sum_{i=1}^{n}(\rho'(A_i^\top \theta^*)A_i) ||_2) \, || \triangle ||_2 \end{aligned} \tag{5}$$

where $\tilde{\theta}$ lies between θ^* and $\hat{\theta}$. The first inequality is due to the triangle inequality, $|| \theta^* + \triangle ||_2 - || \theta^* ||_2 \geq - || \triangle ||_2$, the second inequality follows from Cauchy-Schwarz inequality.

Next we note that

$$
\begin{aligned}
&\| \frac{1}{n}\sum_{i=1}^{n}(\rho^{'}(A_i^{\top}\theta)A_i) \|_2 \\
&\leq \| \frac{1}{n}\sum_{i=1}^{n}(\rho^{'}(A_i^{\top}\theta)A_i - w_iA_i) \|_2 + \| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \| A_i \|_2 \, O(K^{-\tilde{\alpha}}) + \| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n} \| \beta_{K2}(\mathbf{X}_i)\otimes\alpha_{K1}(\mathbf{T}_i) - \bar{\beta}_{K2}\otimes\bar{\alpha}_{K1} \|_2 \, O(K^{-\tilde{\alpha}}) + \| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2 \\
&\leq \frac{1}{n}\sum_{i=1}^{n}(\| \beta_{K2}(\mathbf{X}_i)\otimes\alpha_{K1}(\mathbf{T}_i) \|_2 + \| \bar{\beta}_{K2}\otimes\bar{\alpha}_{K1} \|_2) O(K^{-\tilde{\alpha}}) + \| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(\| \beta_{K2}(\mathbf{X}_i) \|_2 \cdot \| \alpha_{K1}(\mathbf{T}_i) \|_2 + \| \bar{\beta}_{K2} \|_2 \cdot \| \bar{\alpha}_{K1} \|_2) O(K^{-\tilde{\alpha}}) + \| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2 \\
&\leq 2\sup \| \alpha_{K1}(\mathbf{T}_i) \|_2 \sup \| \beta_{K2}((X)_i) \|_2 \, O(K^{-\tilde{\alpha}}) + \| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2 \\
&\leq O(K^{1/2-\tilde{\alpha}}) + \| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2 \, .
\end{aligned} \tag{6}
$$

The first inequality is due to triangle inequality and the second inequality is due to Assumption 4(iv) and $\lambda_k \leq 1, k = 1, \dots, K$. Next, we use the Bernstein's inequality to bound the second term.

Recall that the Bernstain's inequality for random matrix in Tropp et al. (2015): Let $\{Z_i\}$ be a sequence of independent random matrices with dimension $d_1 \times d_2$. Assume that $EZ_i = 0$ and $\| Z_i \|_2 \leq R_n$ almost surely. Define

$$
\sigma_n^2 = \max\{\| \sum_{i=1}^{n} E(Z_iZ_i^{\top}) \|_2, \| \sum_{i=1}^{n} E(Z_i^{\top}Z_i) \|_2\}, \tag{7}
$$

then for all $\epsilon \geq 0$,

$$
P(\| \sum_{i=1}^{n} Z_i \|_2 \geq 0) \leq (d_1 + d_2)\exp(-\frac{\epsilon^2/2}{\sigma_n^2 + R_n\epsilon/3}). \tag{8}
$$

For the second term $\| \frac{1}{n}\sum_{i=1}^{n} w_iA_i \|_2$, we notice that

$$
\begin{aligned}
E(w_iA_i) &= E(w_i\beta_{K2}(\mathbf{X}_i)\otimes\alpha_{K1}(\mathbf{T}_i)) - E(w_i\bar{\beta}_{K2}\otimes\bar{\alpha}_{K1}) \\
&= E(\beta_{K2}(\mathbf{X}_i))\otimes E(\alpha_{K1}(\mathbf{T}_i)) - E(\bar{\beta}_{K2})\otimes E(\bar{\alpha}_{K1}) \\
&= 0.
\end{aligned} \tag{9}
$$

Then for $\|\frac{1}{n}\sum_{i=1}^{n} w_i A_i\|_2$, we have

$$\| \frac{1}{n} w_i A_i \|_2 \leq \frac{1}{n} \| w_i \|_2 \| A_i \|_2 \leq \eta_1 \frac{K^{1/2}}{n} \tag{10}$$

Next, for $\|\frac{1}{n}\sum_{i=1}^{n} E(w_i^2 A_i A_i^\top)\|_2$, we have

$$\begin{aligned}
&\| \frac{1}{n}\sum_{i=1}^{n} E(w_i^2 A_i A_i^\top) \|_2 \\
&=\| \frac{1}{n}\sum_{i=1}^{n} E(w_i^2(\beta_{K2}(\mathbf{X}_i)\otimes\alpha_{K1}(\mathbf{T}_i)-\bar{\beta}_{K2}\otimes\bar{\alpha}_{K1})(\beta_{K2}(\mathbf{X}_i)\otimes\alpha_{K1}(\mathbf{T}_i)-\bar{\beta}_{K2}\otimes\bar{\alpha}_{K1})^\top \|_2 \\
&=\| \frac{1}{n}\sum_{i=1}^{n} E(w_i^2\{(\beta_{K2}(\mathbf{X}_i)\beta_{K2}(\mathbf{X}_i)^\top)\otimes(\alpha_{K1}(\mathbf{T}_i)\alpha_{K1}(\mathbf{T}_i)^\top)-(\beta_{K2}(\mathbf{X}_i)\bar{\beta}_{K2}^\top)\otimes(\alpha_{K1}(\mathbf{T}_i)\bar{\alpha}_{K}^\top \\
&-(\bar{\beta}_{K2}\beta_{K2}(\mathbf{X}_i)^\top)\otimes(\bar{\alpha}_{K1}\alpha_{K1}(\mathbf{T}_i)^\top)+(\bar{\alpha}_{K1}\bar{\alpha}_{K1}^\top)\otimes(\bar{\alpha}_{K1}\bar{\alpha}_{K1}^\top)\}) \|_2 \\
&\leq\| \frac{\eta_2}{n}\sum_{i=1}^{n} E(w_i\{(\beta_{K2}(\mathbf{X}_i)\beta_{K2}(\mathbf{X}_i)^\top)\otimes(\alpha_{K1}(\mathbf{T}_i)\alpha_{K1}(\mathbf{T}_i)^\top)-(\beta_{K2}(\mathbf{X}_i)\bar{\beta}_{K2}^\top)\otimes(\alpha_{K1}(\mathbf{T}_i)\bar{\alpha}^\top \\
&-(\bar{\beta}_{K2}\beta_{K2}(\mathbf{X}_i)^\top)\otimes(\bar{\alpha}_{K1}\alpha_{K1}(\mathbf{T}_i)^\top)+(\bar{\alpha}_{K1}\bar{\alpha}_{K1}^\top)\otimes(\bar{\alpha}_{K1}\bar{\alpha}_{K1}^\top)\}) \|_2 \\
&=\| \frac{\eta_2}{n}\sum_{i=1}^{n}\{E(\beta_{K2}(\mathbf{X}_i)\beta_{K2}(\mathbf{X}_i)^\top)\otimes E(\alpha_{K1}(\mathbf{T}_i)\alpha_{K1}(\mathbf{T}_i)^\top)-E(\beta_{K2}(\mathbf{X}_i)\bar{\beta}_{K2}^\top)\otimes E(\alpha_{K1}(\mathbf{T}_i) \\
&-E(\bar{\beta}_{K2}\beta_{K2}(\mathbf{X}_i)^\top)\otimes E(\bar{\alpha}_{K1}\alpha_{K1}(\mathbf{T}_i)^\top)+E(\bar{\alpha}_{K1}\bar{\alpha}_{K1}^\top)\otimes E(\bar{\alpha}_{K1}\bar{\alpha}_{K1}^\top)\}) \|_2 \\
&\leq \eta_2\{\| E(\beta_{K2}(\mathbf{X}_i)\beta_{K2}(\mathbf{X}_i)^\top) \|_2\cdot\| E(\alpha_{K1}(\mathbf{T}_i)\alpha_{K1}(\mathbf{T}_i)^\top) \|_2 \\
&+\| E(\beta_{K2}(\mathbf{X}_i)\bar{\beta}_{K2}^\top) \|_2\cdot\| E(\alpha_{K1}(\mathbf{T}_i)\bar{\alpha}_{K1}^\top) \|_2+\| E(\bar{\beta}_{K2}\beta_{K2}(\mathbf{X}_i)^\top) \|_2\cdot\| E(\bar{\alpha}_{K1}\alpha_{K1}(\mathbf{T}_i)^\top \\
&+\| E(\bar{\beta}_{K2}\bar{\beta}_{K2}^\top) \|_2\cdot\| E(\bar{\alpha}_{K1}\bar{\alpha}_{K1}^\top)\}) \|_2\} \\
&\leq 4\eta_2 C_1 C_2 \\
&\leq C
\end{aligned} \tag{11}$$

Finally, for $|| \frac{1}{n}\sum_{i=1}^{n} E(w_i^2 A_i^\top A_i) ||_2$, we have

$$
\begin{aligned}
&|| \frac{1}{n}\sum_{i=1}^{n} E(w_i^2 A_i^\top A_i) ||_2 \\
=&|| \frac{1}{n}\sum_{i=1}^{n} E(w_i^2(\beta_{K2}(\mathbf{X}_i)\otimes\alpha_{K1}(\mathbf{T}_i)-\bar\beta_{K2}\otimes\bar\alpha_{K1})^\top(\beta_{K2}(\mathbf{X}_i)\otimes\alpha_{K1}(\mathbf{T}_i)-\bar\beta_{K2}\otimes\bar\alpha_{K1}) ||_2 \\
=&|| \frac{1}{n}\sum_{i=1}^{n} E(w_i^2\{(\beta_{K2}(\mathbf{X}_i)^\top\beta_{K2}(\mathbf{X}_i))\otimes(\alpha_{K1}(\mathbf{T}_i)^\top\alpha_{K1}(\mathbf{T}_i))-(\beta_{K2}(\mathbf{X}_i)^\top\bar\beta_{K2})\otimes(\alpha_{K1}(\mathbf{T}_i)^\top\bar\alpha_{K} \\
&-(\bar\beta_{K2}^\top\beta_{K2}(\mathbf{X}_i))\otimes(\bar\alpha_{K1}^\top\alpha_{K1}(\mathbf{T}_i)^)+(\bar\alpha_{K1}^\top\bar\alpha_{K1})\otimes(\bar\alpha_{K1}^\top\bar\alpha_{K1})\}) ||_2 \\
\le&|| \frac{\eta_2}{n}\sum_{i=1}^{n} E(w_i\{(\beta_{K2}(\mathbf{X}_i)^\top\beta_{K2}(\mathbf{X}_i))\otimes(\alpha_{K1}(\mathbf{T}_i)^\top\alpha_{K1}(\mathbf{T}_i))-(\beta_{K2}(\mathbf{X}_i)^\top\bar\beta_{K2})\otimes(\alpha_{K1}(\mathbf{T}_i)^\top\bar\alpha \\
&-(\bar\beta_{K2}^\top\beta_{K2}(\mathbf{X}_i))\otimes(\bar\alpha_{K1}^\top\alpha_{K1}(\mathbf{T}_i)^)+(\bar\alpha_{K1}^\top\bar\alpha_{K1})\otimes(\bar\alpha_{K1}^\top\bar\alpha_{K1})\}) ||_2 \\
=&|| \frac{\eta_2}{n}\sum_{i=1}^{n}\{E(\beta_{K2}(\mathbf{X}_i)^\top\beta_{K2}(\mathbf{X}_i))\otimes E(\alpha_{K1}(\mathbf{T}_i)^\top\alpha_{K1}(\mathbf{T}_i))-E(\beta_{K2}(\mathbf{X}_i)^\top\bar\beta_{K2})\otimes E(\alpha_{K1}(\mathbf{T}_i)^\top \\
&-E(\bar\beta_{K2}^\top\beta_{K2}(\mathbf{X}_i))\otimes E(\bar\alpha_{K1}^\top\alpha_{K1}(\mathbf{T}_i))+E(\bar\alpha_{K1}^\top\bar\alpha_{K1})\otimes E(\bar\alpha_{K1}^\top\bar\alpha_{K1})\}) ||_2 \\
\le& \frac{\eta_2}{n}\sum_{i=1}^{n}\{|| E(\beta_{K2}(\mathbf{X}_i)^\top\beta_{K2}(\mathbf{X}_i)) ||_2\cdot|| E(\alpha_{K1}(\mathbf{T}_i)^\top\alpha_{K1}(\mathbf{T}_i)) ||_2 \\
+&|| E(\beta_{K2}(\mathbf{X}_i)^\top\bar\beta_{K2}) ||_2\cdot|| E(\alpha_{K1}(\mathbf{T}_i)^\top\bar\alpha_{K1}) ||_2+|| E(\bar\beta_{K2}^\top\beta_{K2}(\mathbf{X}_i)) ||_2\cdot|| E(\bar\alpha_{K1}^\top\alpha_{K1}(\mathbf{T}_i)) \\
+&|| E(\bar\beta_{K2}^\top\bar\beta_{K2}) ||_2\cdot|| E(\bar\alpha_{K1}^\top\bar\alpha_{K1})\}) ||_2\} \\
\le& 4\eta_2 K^{1/2} \\
\le& C_3 K^{1/2}
\end{aligned}
\tag{12}
$$

Therefore,

$$\sigma_n^2 = C_3 K^{1/2}. \tag{13}$$

Combing (35),(36) and (37) with the Bernstain's inequality, we have

$$P(|| \sum_{i=1}^{n} w_i A_i ||_2 \ge \epsilon) \le (K+1)\exp(-\frac{\epsilon^2/2}{C_3K^{1/2}+C_1\frac{K^{1/2}}{n}\cdot\frac{\epsilon}{3}}). \tag{14}$$

The right side goes to zero as $K\to\infty$ when $\frac{\epsilon^2/2}{C_3K^{1/2}+C_1\frac{K^{1/2}}{n}\cdot\frac{\epsilon}{3}} > \log K$. It suffices when $\epsilon = O_p(K^{1/2}log(K)/n)$.

Moreover,

$$\triangle^\top\{\sum_{i=1}^{n}(\rho''(A_i^\top\tilde\theta)A_iA_i^\top)\triangle\} \ge c_1\eta_2^2\triangle^\top\{\sum_{i=1}^{n}(w_i^2A_iA_i^\top)\triangle\} \tag{15}$$

According to ? and Assumption 4(v), the smallest eigenvalue of matrix $\frac{1}{n}\sum_{i=1}^{n}(w_iA_i)(w_iA_i)^\top$ satisfies that $\lambda_{\min} = O_p(n^{-1/2})$. Then,

$$\begin{aligned}\triangle^\top\{\sum_{i=1}^{n}(\rho''(A_i^\top\theta)A_iA_i^\top)\triangle\} &\geq \eta_1 n\triangle^\top\{\frac{1}{n}\sum_{i=1}^{n}(A_iA_i^\top)\triangle\}\\ &\geq \lambda_{\min}\eta_1 n \mid\mid \triangle \mid\mid_2^2\\ &= O_p(n^{1/2}) \mid\mid \triangle \mid\mid_2^2 . \qquad (16)\end{aligned}$$

Combine (5), (13) and (15), we can obtain that

$$\begin{aligned}&G(\theta+\triangle)-G(\theta)\\ &\geq O_p(n^{1/2}) \mid\mid \triangle \mid\mid_2^2 -(\sqrt{\delta}+O_p(K^{1/2}log(K)/n+K^{1/2-\tilde{\alpha}})) \mid\mid \triangle \mid\mid_2\\ &\geq 0 \qquad (17)\end{aligned}$$

for $\triangle = C(\frac{K^{1/2}\log(K)}{n^{3/2}} + K^{1/2-\alpha})$ with large enough constant $C > 0$. Hence, Lemma 1 is proved.

Next, we will prove Theorem 2. Specifically, by Mean Value Theorem, we can deduce that

$$\begin{aligned}&\int \mid \hat{w}-w \mid^2 dF(\mathbf{t},\mathbf{x})\\ &\leq \sup_{(\mathbf{t},\mathbf{x})} \mid \exp\{\tilde{M}_K(\mathbf{t},\mathbf{x})^\top\theta_1-1\} \mid^2 \times \int \mid \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*) \mid^2 dF(\mathbf{t},\mathbf{x})\\ &\leq O(1)\cdot\int \mid \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*) \mid^2 dF(\mathbf{t},\mathbf{x}),\end{aligned}$$

where θ_1 lies between $\hat{\theta}$ and θ^*. Since

$$\begin{aligned}&\int \mid \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*) \mid^2 dF(\mathbf{t},\mathbf{x})\\ &=\int \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*)(\hat{\theta}-\theta^*)^\top\tilde{M}_K(\mathbf{t},\mathbf{x})dF(\mathbf{t},\mathbf{x})\\ &=tr\{(\hat{\theta}-\theta^*)(\hat{\theta}-\theta^*)^\top\int \tilde{M}_K(\mathbf{t},\mathbf{x})\tilde{M}_K(\mathbf{t},\mathbf{x})^\top dF(\mathbf{t},\mathbf{x})\}\\ &\leq Ctr\{(\hat{\theta}-\theta^*)(\hat{\theta}-\theta^*)^\top\}\\ &=C \mid\mid \hat{\theta}-\theta^* \mid\mid^2\\ &=O_p((\frac{K^{1/2}\log(K)}{n^{3/2}}+K^{1/2-\alpha})^2).\end{aligned}$$

Then we have

$$\int \mid \hat{w}-w \mid^2 dF(\mathbf{t},\mathbf{x}) = O_p((\frac{K^{1/2}\log(K)}{n^{3/2}}+K^{1/2-\alpha})^2)$$

Furthermore, one can show that

$$\frac{1}{n}\sum_{i=1}^{n} \mid \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*) \mid^2 - \int \mid \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*) \mid^2 dF(\mathbf{t},\mathbf{x}) = o_p(1).$$

Hence,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} \mid \hat{w}_i - w_i \mid^2 \\
&\leq \sup_{(\mathbf{t},\mathbf{x})} \ \mid \exp\{\tilde{M}_K(\mathbf{t},\mathbf{x})^\top\theta_1 - 1\} \mid^2 \cdot \frac{1}{n}\sum_{i=1}^{n} \mid \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*) \mid^2 \\
&\leq O(1)\int \mid \tilde{M}_K(\mathbf{t},\mathbf{x})^\top(\hat{\theta}-\theta^*) \mid^2 dF(\mathbf{t},\mathbf{x}) + o_p(1) \\
&= O_p((\frac{K^{1/2}\log(K)}{n^{3/2}} + K^{1/2-\alpha})^2).
\end{aligned}$$

Therefore, the proof of Theorem 2 is completed.

4. Proof of Theorem 3

We first show the conclusion of Theorem 3(i).

Since $\hat{\mathbf{B}}$ (as a estimator of $\mathbf{B}^*$) is a unique minimizer of $\frac{1}{n}\sum_{i=1}^{n}\hat{w}_i(Y_i- < \mathbf{B},\mathbf{T}_i >)^2$(regarding $\mathbb{E}[w(Y- < \mathbf{B},\mathbf{T}_i >)^2]$, according to the theory of M-estimation [3, Theorem 5.7], if

$$\sup_{\beta\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^{n}\hat{w}_i(Y_i- < \mathbf{B},\mathbf{T}_i >)^2 - \mathbb{E}[w(Y- < \mathbf{B},\mathbf{T} >)^2]) \mid \rightarrow_p 0,$$

then $\hat{\mathbf{B}} \rightarrow_p \mathbf{B}^*$. Note that

$$\begin{aligned}
&\sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^{n}\hat{w}_i(Y_i- < \mathbf{B},\mathbf{T}_i >)^2 - \mathbb{E}[w(Y- < \mathbf{B},\mathbf{T} >)^2]) \mid \\
&\leq \sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^{n}(\hat{w}_i - w_i)(Y_i- < \mathbf{B},\mathbf{T}_i >)^2 \mid \\
&+\sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^{n}w_i(Y_i- < \mathbf{B},\mathbf{T}_i >)^2 - \mathbb{E}[w(Y- < \mathbf{B},\mathbf{T} >)^2]) \mid . \qquad (18)
\end{aligned}$$

We first show that $\sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^n(\hat{w}_i - w_i)(Y_i - <\mathbf{B},\mathbf{T}_i>)^2 \mid$ is $o_p(1)$. Using the Causchy-Schwarz inequality and the fact that $\hat{w} \to^{L^2} w$, we have

$$\begin{aligned}
&\sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^n(\hat{w}_i - w_i)(Y_i - <\mathbf{B},\mathbf{T}_i>)^2 \mid \\
&\le \{\frac{1}{n}\sum_{i=1}^n(\hat{w}_i - w_i)^2\}^{1/2}\sup_{\mathbf{B}\in\Theta_1}\{\frac{1}{n}\sum_{i=1}^n(Y_i - <\mathbf{B},\mathbf{T}_i>)^2\}^{1/2} \\
&\le o_p(1)\{\sup_{\mathbf{B}\in\Theta_1}\mathbb{E}[w(Y - <\mathbf{B},\mathbf{T}>)^2] + o_p(1)\}^{1/2} \\
&= o_p(1).
\end{aligned}$$

Thereafter, under Assumption 5, we can conclude that $\sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^n w_i(Y_i - <\mathbf{B},\mathbf{T}_i>)^2 - \mathbb{E}[w(Y - <\mathbf{B},\mathbf{T}>)^2]) \mid$ is also $o_p(1)$ [2, Lemma 2.4].Hence, we complete the proof for Theorem 3(i). Next, we give the proof of Theorem 3(ii). Define

$$\hat{\mathbf{B}}^* = \text{argmin}_{\mathbf{B}}\sum_{i=1}^n w_i(Y_i - <\mathbf{B},\mathbf{T}_i>)^2.$$

Assume that $\frac{1}{n}\sum_{i=1}^n w_i(Y_i - <\hat{\mathbf{B}}^*,\mathbf{T}_i>)h(\mathbf{T}_i;\hat{\mathbf{B}}^*)) = o_p(n^{-1/2})$ holds with probablility to one as $n \to \infty$

By Assumption 5 and the uniform law of large number, one can get that

$$\frac{1}{n}\sum_{i=1}^n w_i(Y_i - <\mathbf{B},\mathbf{T}_i>)^2 \to \mathbb{E}\{w(Y - <\mathbf{B},\mathbf{T}>)^2\} \text{ in probability uniformly over } \mathbf{B},$$

which implies $\| \hat{\mathbf{B}}^* - \mathbf{B}^* \| \to_p 0$. Let

$$r(\mathbf{B}) = 2\mathbb{E}\{w(Y - <\mathbf{B},\mathbf{T}>)h(\mathbf{T};\mathbf{B})\},$$

which is a differentiable function in $\mathbf{B}$ and $r(\mathbf{B}^*) = 0$. By mean value theorem, we have

$$\sqrt{n}r(\hat{\mathbf{B}}^*) - \bigtriangledown_{\mathbf{B}}r(\zeta)\cdot\sqrt{(}n)(\hat{\mathbf{B}}^* - \mathbf{B}^*) = \sqrt{n}r(\mathbf{B}^*) = 0$$

where ζ lies on the line joining $\hat{\mathbf{B}}^*$ and $\mathbf{B}^*$. Since $\bigtriangledown_{\mathbf{B}}r(\mathbf{B})$ is continuous at $\mathbf{B}^*$ and $\| \hat{\mathbf{B}}^* - \mathbf{B}^* \| \to_p 0$, then

$$\sqrt{n}(\text{vec } \hat{\mathbf{B}}^* - \text{vec } \mathbf{B}^*) = \bigtriangledown_{\mathbf{B}}r(\mathbf{B}^*)^{-1}\cdot\sqrt{n}r(\hat{\mathbf{B}}^*) + o_p(1)$$

Define the empirical process

$$G_n(\mathbf{B}) = \frac{2}{\sqrt{n}}\sum_{i=1}^n\{w_i(Y_i - <\mathbf{B},\mathbf{T}_i>)h(\mathbf{T}_i;\mathbf{B}) - \mathbb{E}\{w(Y - <\mathbf{B},\mathbf{T}>)h(\mathbf{T};\mathbf{B})\}\}.$$

Then we have

$$
\begin{aligned}
&\sqrt{n}(\text{vec }\hat{\mathbf{B}}^* - \text{vec }\mathbf{B}^*)\\
&= \bigtriangledown_\beta r(\mathbf{B}^*)^{-1}\cdot\{\sqrt{n}r(\hat{\mathbf{B}}^*) - \frac{2}{\sqrt{n}}\sum_{i=1}^n\{w_i(Y_i - <\hat{\mathbf{B}}^*, \mathbf{T}_i>)h(\mathbf{T}_i;\hat{\mathbf{B}}^*)\\
&+ \frac{2}{\sqrt{n}}\sum_{i=1}^n\{w_i(Y_i - <\hat{\mathbf{B}}^*, \mathbf{T}_i>)h(\mathbf{T}_i;\hat{\mathbf{B}}^*)\}\\
&= -\bigtriangledown_\beta r(\mathbf{B}^*)^{-1}\cdot G_n(\hat{\mathbf{B}}^*) + o_p(1)\\
&= U^{-1}\cdot\{G_n(\hat{\mathbf{B}}^*) - G_n(\mathbf{B}^*) + G_n(\mathbf{B}^*)\} + o_p(1).
\end{aligned}
$$

By Assumption 5, 6, Theorem 4 and 5 of Andrews(1994), we have $G_n(\hat{\mathbf{B}}^*) - G_n(\mathbf{B}^*) \rightarrow_p 0$. Thus,

$$
\sqrt{n}(\text{vec }\hat{\mathbf{B}}^* - \text{vec }\mathbf{B}^*) = U^{-1}\frac{2}{\sqrt{n}}\sum_{i=1}^n\{w_i(Y_i - <\mathbf{B}^*, \mathbf{T}_i>)h(\mathbf{T}_i;\mathbf{B}^*)\} + o_p(1),
$$

then we can get that the asymptotic variance of $\sqrt{n}(\text{vec }\hat{\mathbf{B}}^* - \text{vec }\mathbf{B}^*)$ is V. Therefore, $\sqrt{n}(\text{vec }\hat{\mathbf{B}}^* - \text{vec }\mathbf{B}^*) \rightarrow_d N(0, V)$. Next, we will prove $\hat{\mathbf{B}} \rightarrow_p \hat{\mathbf{B}}^*$. Since

$$
\begin{aligned}
&\sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^n \hat{w}_i(Y_i - <\mathbf{B}, \mathbf{T}_i>)^2 - \frac{1}{n}\sum_{i=1}^n w_i(Y_i - <\mathbf{B}, \mathbf{T}_i>)^2) \mid\\
&\leq \sup_{\mathbf{B}\in\Theta_1} \mid \frac{1}{n}\sum_{i=1}^n(\hat{w}_i - w_i)(Y_i - <\mathbf{B}, \mathbf{T}_i>)^2 \mid\\
&\leq \{\frac{1}{n}\sum_{i=1}^n(\hat{w}_i - w_i)^2\}^{1/2}\sup_{\mathbf{B}\in\Theta_1}\{\frac{1}{n}\sum_{i=1}^n(Y_i - <\mathbf{B}, \mathbf{T}_i>)^2\}^{1/2}\\
&\leq o_p(1)\{\sup_{\mathbf{B}\in\Theta_1}\mathbb{E}[w(Y - <\mathbf{B}, \mathbf{T}>)^2] + o_p(1)\}^{1/2}\\
&= o_p(1),
\end{aligned}
$$

which implies $\hat{\mathbf{B}}^* \rightarrow_p \hat{\mathbf{B}}$. Then by Slutskey's Theorem, we can draw the conclusion that $\sqrt{n}(\text{vec }\hat{\mathbf{B}} - \text{vec }\mathbf{B}^*) \rightarrow_d N(0, V)$. Therefore, we have completed the proof of Theorem 3.

5. Proof of Theorem 4

For convenience, we use a mapping $\Omega: R^{p\times q\times D}\times R \rightarrow R^{p\times q\times D}$ to represent the operator of absorbing the constant into the coefficients of B-spline basis for the first predictor. More precisely, Ω is defined by

$$\mathbf{G}^b = \Omega(\mathbf{G}, c),$$

where $\mathbf{G}_{i_1,i_2,d} = \mathbf{G}_{i_1,i_2,d}$ for $(i_1, i_2) \neq (1,1)$ and $\mathbf{G}_{1,1,d} = \mathbf{G}_{1,1,d} + pqc, d = 1, \dots, D$. It then follows from the property of B-spline functions that

$$c + \frac{1}{pq} < \mathbf{G}, \Phi(\mathbf{T}) >= \frac{1}{pq} < \mathbf{G}^b, \Phi(\mathbf{T}) > .$$

We also write $\mathbf{G}_0 = \sum_{r=1}^{R_0} \mathbf{B}_{0r} \circ \ _{0r}, r = 1, \dots, R_0$. Suppose $\hat{\mathbf{G}}, \hat{c})$ is a solution to (19) and

$$\hat{\mathbf{G}} = \sum_{r=1}^{R} \hat{\ }_1^{(r)} \circ \hat{\ }_2^{(r)} \circ \hat{\ }_r,$$

then by [[**?**], Lemma B.1], there exists $\check{c} \in R$ and

$$\check{\mathbf{G}} = \sum_{r=1}^{R} \hat{\ }_1^{(r)} \circ \hat{\ }_2^{(r)} \circ \check{\ }_r,$$

such that

$$\check{c} + \frac{1}{pq} < \check{\mathbf{G}}, \Phi(\mathbf{T}) >= \hat{c} + \frac{1}{pq} < \hat{\mathbf{G}}, \tilde{\Phi}(\mathbf{T}) >, \tag{19}$$

where $\check{\ }_r = (\check{\alpha}_{r,1}, \dots, \check{\alpha}_{r,D})'$ satisfying

$$\sum_{d=1}^{D} \check{\alpha}_{r,d} u_d = 0$$

with $u_d = \int_0^1 b_d(x)dx$. Using (27), we have

$$\sum_{i=1}^{n} (\hat{w}_i y_i - \check{c} - \frac{1}{pq} < \check{\mathbf{G}}, \Phi(\mathbf{T}_i) >)^2 \leq \sum_{i=1}^{n} (\hat{w}_i y_i - c_0 - \frac{1}{pq} < \mathbf{G}_0, \Phi(\mathbf{T}_i) >)^2. \tag{20}$$

Let $\check{\mathbf{G}}^b = \Omega(\check{\ }, \check{c})$ and $\mathbf{G}_0^b = \Omega(\mathbf{G}_0, c_0)$, then

$$\sum_{i=1}^{n} (\hat{w}_i y_i - \frac{1}{pq} < \check{\mathbf{G}}^b, \Phi(\mathbf{T}_i) >)^2 \leq \sum_{i=1}^{n} (\hat{w}_i y_i - \frac{1}{pq} < \mathbf{G}_0^b, \Phi(\mathbf{T}_i) >)^2. \tag{21}$$

Therefore, we have

$$\sum_{i=1}^{n} ((\hat{w}_i - w_i + w_i) y_i - \frac{1}{pq} < \check{\mathbf{G}}^b, \Phi(\mathbf{T}_i) >)^2 \leq \sum_{i=1}^{n} ((\hat{w}_i - w_i + w_i) y_i - \frac{1}{pq} < \mathbf{G}_0^b, \Phi(\mathbf{T}_i) >)^2, \tag{22}$$

which leads to

$$\sum_{i=1}^{n}(w_i y_i - \frac{1}{pq} < \check{\mathbf{G}}^b, \Phi(\mathbf{T}_i) >)^2 \leq \sum_{i=1}^{n}(w_i y_i - \frac{1}{pq} < \mathbf{G}_0^b, \Phi(\mathbf{T}_i) >)^2 + 2\sum_{i=1}^{n}(\hat{w}_i - w_i) y_i (\frac{1}{pq} < \check{\mathbf{G}}^b - \mathbf{G}_0, \Phi(\mathbf{T}_i)). \tag{23}$$

Let $\mathbf{G}^{\#} = \check{\mathbf{G}}^b - \mathbf{G}_0^b$, $\mathbf{a}^{\#} = \mathrm{vec}(\mathbf{G}^{\#})$, $\mathbf{a}_0^b = \mathrm{vec}(\mathbf{G}_0^b)$, $\check{\mathbf{a}}^b = \mathrm{vec}(\check{\mathbf{G}}^b)$ and $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_n)^{'} \in R^{n \times pqD}$, where $\mathbf{z}_i = \mathrm{vec}(\Phi(\mathbf{T}_i)), i = 1, \dots, n$. Let $y_{\hat{w}} = (\hat{w}_1 y_1, \dots, \hat{w}_n y_n)^{'}$ and $y_w = (w_1 y_1, \dots, w_n y_n)^{'}$, then using (31) and working out the squares, we obtain

$$\begin{aligned}\frac{1}{p^2q^2} \| \mathbf{Z}\mathbf{a}^{\#} \|^2 \leq\; & 2 < \frac{1}{pq}\mathbf{Z}\check{\mathbf{a}}^b, y_w > -2 < \frac{1}{pq}\mathbf{Z}\mathbf{a}_0^b, y_w > \\ & - 2\frac{1}{p^2q^2} < \mathbf{Z}\mathbf{a}^{\#}, \mathbf{Z}\mathbf{a}_0^b > +2 < \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_{\hat{w}} - y_w > \\ & = 2 < \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_{\hat{w}} - y_w > +2 < \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, \quad > +2 < \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_w - \quad - \frac{1}{pq}\mathbf{Z}\mathbf{a}_0^b > \end{aligned} \tag{24}$$

First, we show the upper bound of $< \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_{\hat{w}} - y_w >$. Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} & < \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_{\hat{w}} - y_w > \\ & \leq \| \hat{w} - w \|_2 \cdot \| \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#} \|_2 \\ & \leq \frac{C_1\sqrt{nh_n}}{pq} \cdot \| \hat{w} - w \|_2 \cdot \| \mathbf{a}^{\#} \|_2 \end{aligned} \tag{25}$$

By the conclusion of Theorem 2(ii), we have

$$\| \hat{w} - w \|_2 = O_p(\frac{K^{1/2}\log(K)}{n^{3/2}} + K^{1/2-\alpha}) \tag{26}$$

Applying (37) to (36), we can obtain that

$$< \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_{\hat{w}} - y_w > \leq O_p(\frac{K^{1/2}\log(K)}{n^{3/2}} + K^{1/2-\alpha})\frac{\sqrt{nh_n}}{pq} \| \mathbf{a}^{\#} \|_2 . \tag{27}$$

Second, by the conclusion of [4, (A.18) and (A.20)], we can obtain the upper bound of $< \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, \quad >$ and $< \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_w - \quad - \frac{1}{pq}\mathbf{Z}\mathbf{a}_0^b >$, which are

$$< \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, \quad > \leq \frac{C_3}{pq} \| \mathbf{a}^{\#} \|_2 \{nh_n(R^3 + R(p+q) + RD)\}^{1/2}. \tag{28}$$

and

$$< \frac{1}{pq}\mathbf{Z}\mathbf{a}^{\#}, y_w - \; - \frac{1}{pq}\mathbf{Z}\mathbf{a}_0^b > \leq \frac{C_4}{pq} \mid\mid \mathbf{a}^{\#} \mid\mid_2 \{\frac{\sum_{r=1}^{R_0} \mid\mid \text{vec}(\mathbf{B}_{0r}) \mid\mid_1}{pq}\}\frac{n\sqrt{h_n}}{D^{\tau}} \quad (29)$$

Therefore, applying (38), (39) and (40) to (35), we have

$$\frac{C_5}{pq} \mid\mid \mathbf{a}^{\#} \mid\mid_2^2 \leq R_1 \mid\mid \mathbf{a}^{\#} \mid\mid_2, \quad (30)$$

where

$$\begin{aligned} R_1 = & C_6(\frac{K^{1/2}\log(K)}{n^{3/2}} + K^{1/2-\alpha})\sqrt{\frac{D}{n}} \\ & + C_7\{\frac{D(R^3 + R(p+q) + RD)}{n}\}^{1/2} \\ & + C_8\{\frac{\sum_{r=1}^{R_0} \mid\mid \text{vec}(\mathbf{B}_{0r}) \mid\mid_1}{pq}\}\frac{1}{D^{\tau-1/2}}. \end{aligned}$$

By solving the second order inequality (41), we have

$$\frac{C_5}{pq} \mid\mid \mathbf{a}^{\#} \mid\mid_2 \leq R_1 \quad (31)$$

Further, by Assumption 6 and [4, (A.38) of Lemma A.2], we have

$$\begin{aligned} \mid\mid \hat{s}(\mathbf{T}) - s(\mathbf{T}) \mid\mid^2 \leq & C_9 h_n \frac{1}{p^2q^2} \mid\mid \mathbf{a}^{\#} \mid\mid^2 \\ = & \frac{C_{10}R_1^2}{D} \\ = & O_p((\frac{K^{1/2}\log(K)}{n^{3/2}} + K^{1/2-\alpha})^2) + O_p(\frac{R^3 + R(p+q) + RD}{n}) \\ & + O_p(\{\frac{\sum_{r=1}^{R_0} \mid\mid \text{vec}(\mathbf{B}_{0r}) \mid\mid_1}{pq}\}^2\frac{1}{D^{2\tau}}). \end{aligned} \quad (32)$$

Hence, the proof of Theorem 4 is completed.

Reference

[1] Mohri, M., Rostamizadeh, A., and Talwalkar, A. (2018). *Foundations of machine learning.* MIT press.

[2] Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypoth- esis testing. *Handbook of econometrics*, 4:2111–2245.

[3] Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

[4] Zhou, Y., Wong, R., and He, K. (2020). Broadcasted nonparametric tensor regression.