Direction Estimation in Single-Index Regressions via Hilbert-Schmidt Independence Criterion

Nan Zhang and Xiangrong Yin

University of Georgia and University of Kentucky

Supplementary Material

This supplementary file provides technical details for the main results in the paper: Proposition 2 and Proposition 3.

S1 Lemmas for Propositions 2 and 3

Here, we introduce some notations in aid of proving Propositions 2 and 3. The Lagrange multiplier is being used for reconstructing the optimization problem.

Let $\mathcal{L}(\zeta) = H(\beta^T\mathbf{X}, Y) + \lambda(\beta^T\Sigma_X\beta - 1)$ and $\mathcal{L}_n(\zeta) = H_n(\beta^T\mathbf{X}, Y) + \lambda(\beta^T\hat{\Sigma}_X\beta - 1)$, where $\zeta = (\beta, \lambda)^T \in \mathbb{R}^{p+1}$, $\beta \in \mathbb{R}^p$, $\lambda \in \mathbb{R}$, Σ_X is the covariance matrix of $\mathbf{X}$, and $\hat{\Sigma}_X$ is the sample estimate of Σ_X.

Let $\eta = \arg\max_{\beta^T\Sigma_X\beta=1} H(\beta^T\mathbf{X}, Y)$ and $\eta_n = \arg\max_{\beta^T\hat{\Sigma}_X\beta=1} H_n(\beta^T\mathbf{X}, Y)$, then there exist λ_0 and λ_n such that $(\eta, \lambda_0)^T$ is a stationary point for $\mathcal{L}(\zeta)$ and $(\eta_n, \lambda_n)^T$ is a stationary point for $\mathcal{L}_n(\zeta)$. For ease of the proof and without loss of generality, η_n and η are selected to keep the first nonzero element positive, respectively.

Let $\theta = (\eta, \lambda_0)^T$ and $\theta_n = (\eta_n, \lambda_n)^T$, then $\theta = \arg\max \mathcal{L}(\zeta)$ and $\theta_n = \arg\max \mathcal{L}_n(\zeta)$, because of the uniqueness of central subspace. Here η and $\eta_n \in \mathbb{R}^p$, λ_0 and $\lambda_n \in \mathbb{R}$.

Lemma 1 *If the support of* $\mathbf{X}$ *is compact, and* $\theta_n \xrightarrow{p} \theta$*, then* $\mathcal{L}_n(\theta_n) - \mathcal{L}_n(\theta) \xrightarrow{p} 0$.

Proof:

$$\begin{aligned}
\mathcal{L}_n(\theta_n) - \mathcal{L}_n(\theta) &= H_n(\eta_n^T\mathbf{X}, Y) + \lambda_n(\eta_n^T\hat{\Sigma}_X\eta_n^T - 1) - H_n(\eta^T\mathbf{X}, Y) - \lambda_0(\eta^T\hat{\Sigma}_X\eta - 1) \\
&= H_n(\eta_n^T\mathbf{X}, Y) - H_n(\eta^T\mathbf{X}, Y) + \lambda_n(\eta_n^T\hat{\Sigma}_X\eta_n^T - 1) - \lambda_0(\eta^T\hat{\Sigma}_X\eta - 1)
\end{aligned}$$

Since $\theta_n \xrightarrow{p} \theta$, therefore $\eta_n \xrightarrow{p} \eta$ and $\lambda_n \xrightarrow{p} \lambda$. Along with, $\hat{\Sigma}_X \xrightarrow{a.s.} \Sigma_X$, we know $\lambda_n\eta_n^T\hat{\Sigma}_X\eta_n \xrightarrow{p} \lambda_0\eta^T\Sigma_X\eta = \lambda_0$, and $\lambda_0\eta^T\hat{\Sigma}\eta \xrightarrow{a.s.} \lambda_0\eta^T\Sigma\eta = \lambda_0$. Therefore, $\lambda_n(\eta_n^T\hat{\Sigma}_X\eta_n - 1) - \lambda_0(\eta^T\hat{\Sigma}_X\eta - 1) = (\lambda_n\eta_n^T\hat{\Sigma}_X\eta_n - \lambda\eta^T\hat{\Sigma}_X\eta) - (\lambda_n - \lambda_0) \xrightarrow{p} 0$.

Next, it's clearly true that $H_n(\eta_n^T\mathbf{X}, Y) - H_n(\eta^T\mathbf{X}, Y) \xrightarrow{p} 0$, since

$$H_n(\eta_n^T\mathbf{X}, Y) = \frac{1}{n^2}\sum_{i,j}\hat{K}_{ij}\hat{L}_{ij} - \frac{2}{n^3}\sum_{i,j,k}\hat{K}_{ij}\hat{L}_{ik} + \frac{1}{n^4}\sum_{i,j,k,l}\hat{K}_{ij}\hat{L}_{kl},$$

where

$$\hat{K}_{ij} := \exp\left(\frac{-(\eta_n^T(\mathbf{X}_i - \mathbf{X}_j))^2}{2\eta_n^T\hat{\Sigma}_X\eta_n}\right) \xrightarrow{p} K_{ij} := \exp\left(\frac{-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2}{2\eta^T\Sigma_X\eta}\right).$$

$$\hat{L}_{ij} := \exp\left(\frac{-((Y_i - Y_j))^2}{2\hat{\sigma}_Y^2}\right) \xrightarrow{p} L_{ij} := \exp\left(\frac{-((Y_i - Y_j))^2}{2\sigma_Y^2}\right).$$

Hence, the conclusion follows. □

Lemma 2 $H_n(\mathbf{X}, \mathbf{Y}) \xrightarrow{a.s} H(\mathbf{X}, \mathbf{Y})$.

Proof: We prove the result in multivariate dimensions of $\mathbf{X}$ and $\mathbf{Y}$, because formulas in (1) and (2) can be straightforwardly extended to multivariate $\mathbf{X}$ and $\mathbf{Y}$, see Gretton et al. (2009).

Recall that $H_n(\mathbf{X}, \mathbf{Y})$ is the empirical estimate of $H(\mathbf{X}, \mathbf{Y})$ after employing a explicit Gaussian weight function (Kankainen (1995); Gretton et al. (2008)), say, $G(t, s)$, where $t \in R^p$ and $s \in R^q$. Based on the development of Kankainen (1995) and Gretton et al. (2009), we have that $H_n(\mathbf{X}, \mathbf{Y}) = \int |f^n_{\mathbf{X},\mathbf{Y}}(t, s) - f^n_{\mathbf{X}}(t)f^n_{\mathbf{Y}}(s)|^2 dG(t, s)$, where $f^n_{\mathbf{X},\mathbf{Y}}(t, s) = \sum_{j=1}^n e^{i<t,\mathbf{X}>+i<s,\mathbf{Y}>}$, $f^n_{\mathbf{X}}(t) = \sum_{j=1}^n e^{i<t,\mathbf{X}>}$, and $f^n_{\mathbf{Y}}(s) = \sum_{j=1}^n e^{i<s,\mathbf{Y}>}$.

Define $D(\delta) = \{(t, s) : |t| < \delta, |s| < \delta\}$, and for any positive δ, let

$H_{n,\delta}(\mathbf{X}, \mathbf{Y}) = \int_{D(\delta)} |f^n_{\mathbf{X},\mathbf{Y}}(t, s) - f^n_{\mathbf{X}}(t)f^n_{\mathbf{Y}}(s)|^2 dG(t, s)$.

Based on the result of Kankainen (1995, Page 19), together with the fact that the Gaussian weight function $G(t, s)$ is always bounded, we have that

$$\lim_{n\to\infty} H_{n,\delta}(\mathbf{X}, \mathbf{Y}) = \int_{D(\delta)} |f_{\mathbf{X},\mathbf{Y}}(t, s) - f_{\mathbf{X}}(t)f_{\mathbf{Y}}(s)|^2 dG(t, s).$$

On the other hand, for any δ,

$$\begin{aligned}\lim_{n\to\infty} H_n(\mathbf{X}, \mathbf{Y}) &= \lim_{n\to\infty} H_{n,\delta} + \lim_{n\to\infty} \int_{/D(\delta)} |f^n_{\mathbf{X},\mathbf{Y}}(t, s) - f^n_{\mathbf{X}}(t)f^n_{\mathbf{Y}}(s)|^2 dG(t, s) \\ &= H(\mathbf{X}, \mathbf{Y}) - \int_{/D(\delta)} |f_{\mathbf{X},\mathbf{Y}}(t, s) - f_{\mathbf{X}}(t)f_{\mathbf{Y}}(s)|^2 dG(t, s) \\ &\quad + \lim_{n\to\infty} \int_{/D(\delta)} |f^n_{\mathbf{X},\mathbf{Y}}(t, s) - f^n_{\mathbf{X}}(t)f^n_{\mathbf{Y}}(s)|^2 dG(t, s).\end{aligned}$$

Thus, as $\delta \to \infty$, $\lim_{n\to\infty} H_n(\mathbf{X}, \mathbf{Y}) = H(\mathbf{X}, \mathbf{Y})$. □

Lemma 3 *If the support of* $\mathbf{X}$ *is compact, then* $\theta_n \xrightarrow{p} \theta$.

Proof: Suppose θ_n fails to converge to θ with probability 1, then there exists a subsequence, still denoted as θ_n, and $\theta^* = (\eta^*, \lambda^*)^T$, with $\theta^* \neq \theta$, such that $\theta_n \xrightarrow{p} \theta^*$. If so, $\lambda_n \xrightarrow{p} \lambda^*$ and $\eta_n \xrightarrow{p} \eta^*$.

By Lemma 1, if $\theta_n \xrightarrow{p} \theta^*$, then $\mathcal{L}_n(\theta_n) \xrightarrow{p} \mathcal{L}_n(\theta^*)$, where $\mathcal{L}_n(\theta^*) = H_n(\eta^{*T}X, Y) + \lambda^*(\eta^{*T}\hat{\Sigma}_X\eta^* - 1)$.

By Lemma 2, $H_n(\eta^{*T}X, Y) \xrightarrow{a.s.} H(\eta^{*T}X, Y)$. And $\hat{\Sigma}_X \xrightarrow{a.s.} \Sigma$, then $\lambda^*(\eta^{*T}\hat{\Sigma}_X\eta^* - 1) \xrightarrow{a.s.} \lambda^*(\eta^{*T}\Sigma\eta^* - 1)$. Hence, $\mathcal{L}_n(\theta^*) \xrightarrow{p} \mathcal{L}(\theta^*)$.

Together with $\mathcal{L}_n(\theta_n) \xrightarrow{p} \mathcal{L}_n(\theta^*)$, we know $\mathcal{L}_n(\theta_n) \xrightarrow{p} \mathcal{L}(\theta^*)$.

On the other hand, since $\theta_n = \arg\max \mathcal{L}_n(\zeta)$, therefore $\mathcal{L}_n(\theta_n) \geq \mathcal{L}_n(\theta)$. If we take limit on both sides of this inequality, we get $\mathcal{L}(\theta^*) \geq \mathcal{L}(\theta)$, which contradicts with our assumption that $\theta = \arg\max \mathcal{L}(\zeta)$ and the uniqueness of the central subspace. Therefore, $\theta_n \xrightarrow{p} \theta$ has to be true. □

Lemma 4 *Under assumptions in Lemma 1,* $\sqrt{n}(\theta_n - \theta) \xrightarrow{D} N(0, V)$.

Proof: For simplicity, let $H_n(\eta) = H_n(\eta^T\mathbf{X}, Y)$. Note that

$$H_n(\eta^T\mathbf{X}, Y) \quad = \quad \frac{1}{n^2}\sum_{i,j} K_{ij}L_{ij} - \frac{2}{n^3}\sum_{i,j,k} K_{ij}L_{ik} + \frac{1}{n^4}\sum_{i,j,k,l} K_{ij}L_{kl},$$

where

$$K_{ij} := \exp\left(\frac{-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2}{2\eta^T\hat{\Sigma}_X\eta}\right) \quad \text{and} \quad L_{kl} := \exp\left(\frac{-\|Y_k - Y_l\|^2}{2\hat{\sigma}_Y^2}\right).$$

Recall that $\mathcal{L}_n(\theta) = H_n(\eta^T\mathbf{X}, Y) + \lambda_0(\eta^T\hat{\Sigma}_X\eta - 1)$, which is maximized subjecting to the constraint $\eta^T\hat{\Sigma}_X\eta = 1$. Hence, for ease of derivations, in calculations below, K_{ij} is simplified as $K_{ij} := \exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]$.

The Taylor expansion of $\mathcal{L}'_n(\theta_n)$ at θ is $0 = \mathcal{L}'_n(\theta_n) = \mathcal{L}'_n(\theta) + \mathcal{L}''_n(\theta)(\theta_n - \theta) + \mathcal{R}_1(\theta^*_n)$, where $|\theta^*_n - \theta| \leq |\theta_n - \theta|$, and $\theta^*_n = (\eta^*_n, \lambda^*_n)^T$. Next, we give explicit expression of $\mathcal{L}'_n(\theta)$, $\mathcal{L}''_n(\theta)$, and $\mathcal{R}'_1(\theta^*_n)$.

It's easy to show that $\mathcal{L}'_n(\theta) = \begin{pmatrix} H'_n(\eta) + 2\lambda_0\hat{\Sigma}_X\eta \\ \eta^T\hat{\Sigma}_X\eta - 1 \end{pmatrix}$,

while $\mathcal{L}''_n(\theta) = \begin{pmatrix} H''_n(\eta) + 2\lambda_0\hat{\Sigma}_X & 2\hat{\Sigma}_X\eta \\ 2\eta^T\hat{\Sigma}_X & 0 \end{pmatrix}$.

Denote $H_n(\eta) = H_n(\eta^T\mathbf{X}, Y) = S_1(\eta) - 2S_2(\eta) + S_3(\eta)$ where

$$S_1(\eta) = \frac{1}{n^2}\sum_{i,j}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot\exp\left(\frac{-\|Y_i - Y_j\|^2}{2\hat{\sigma}_Y^2}\right),$$

$$S_2(\eta) = \frac{1}{n^3}\sum_{i,j,k}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot\exp\left(\frac{-\|Y_i - Y_k\|^2}{2\hat{\sigma}_Y^2}\right),$$

$$S_3(\eta) = \frac{1}{n^4}\sum_{i,j,k,l}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot\exp\left(\frac{-\|Y_k - Y_l\|^2}{2\hat{\sigma}_Y^2}\right).$$

Then the first derivative is $H_n'(\eta) = S_1'(\eta) - 2S_2'(\eta) + S_3'(\eta)$.

Take $S_1'(\eta)$ as an example,

$$S_1'(\eta) = \frac{1}{n^2}\cdot\frac{\partial}{\partial\eta^T}\sum_{i,j}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot\exp\left(\frac{-\|Y_i - Y_j\|^2}{2\hat{\sigma}_Y^2}\right),$$

where

$$\frac{\partial}{\partial\eta^T}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2] = -\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T\eta].$$

Then, the second derivative $H_n''(\eta) = S_1''(\eta) - 2S_2''(\eta) + S_3''(\eta)$.

Take the first term, $S_1''(\eta)$, as an example,

$$S_1''(\eta) = \frac{1}{n^2}\cdot\frac{\partial^2}{\partial\eta^T\partial\eta}\sum_{i,j}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot\exp\left(\frac{-\|Y_i - Y_j\|^2}{2\hat{\sigma}_Y^2}\right),$$

where

$$\frac{\partial^2}{\partial\eta^T\partial\eta}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2] =$$
$$\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T\eta\cdot\eta^T(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T - (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T].$$

Thus,

$$S_1''(\eta) =$$
$$\frac{1}{n^2}\sum_{i,j}\exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2]\cdot\exp\left(\frac{-\|Y_i - Y_j\|^2}{2\hat{\sigma}_Y^2}\right)$$
$$\cdot[(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T\eta\cdot\eta^T(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T - (\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T].$$

$S_2''(\eta)$ and $S_3''(\eta)$ can be calculated in a similar fashion.

Apparently, $S_1''(\eta)$ as a function of $(\mathbf{X}_i, Y_i)$ and $(\mathbf{X}_j, Y_j)$ $(i, j = 1, \ldots, n)$, can be written as a U-statistic (in a similar fashion as $S_1'(\eta)$ which is demonstrated in later discussions). Similarly as $S_2''(\eta)$ and $S_3''(\eta)$. By Strong Law of Large Number (SLLN) for U-statistics, $H_n''(\eta)$ converges to its population version $H''(\eta)$ almost surely, where $H''(\eta)$ is the Hessian matrix of $H(\eta)$,

$$\begin{aligned} H(\eta) =& \mathrm{E}\left[\exp\left(\frac{-(\eta^T(\mathbf{X}-\mathbf{X}'))^2}{2\eta^T\Sigma_X\eta}\right)\exp\left(\frac{-\|Y-Y'\|^2}{2\sigma_Y^2}\right)\right] \\ &+ \mathrm{E}\left[\exp\left(\frac{-(\eta^T(\mathbf{X}-\mathbf{X}'))^2}{2\eta^T\Sigma_X\eta}\right)\right]\mathrm{E}\left[\exp\left(\frac{-\|Y-Y'\|^2}{2\sigma_Y^2}\right)\right] \\ &- 2\mathrm{E}\Big\{\mathrm{E}\left[\exp\left(\frac{-(\eta^T(\mathbf{X}-\mathbf{X}'))^2}{2\eta^T\Sigma_X\eta}\right)\Big|X\right]\mathrm{E}\left[\exp\left(\frac{-\|Y-Y'\|^2}{2\sigma_Y^2}\right)\Big|Y\right]\Big\}. \end{aligned}$$

Thus, we obtain that $\mathcal{L}_n''(\theta) = \begin{pmatrix} H_n''(\eta)+2\lambda_0\hat{\Sigma}_X & 2\hat{\Sigma}_X\eta \\ 2\eta^T\hat{\Sigma}_X & 0 \end{pmatrix}$ converges almost surely to $\begin{pmatrix} H''(\eta)+2\lambda_0\Sigma_X & 2\Sigma_X\eta \\ 2\eta^T\Sigma_X & 0 \end{pmatrix}$. If $H''(\eta)=0$, then $\begin{vmatrix} 2\lambda_0\Sigma_X & 2\Sigma_X\eta \\ 2\eta^T\Sigma_X & 0 \end{vmatrix} = -\lambda_0^{p-1}|\Sigma_X| \neq 0$, i.e. $\begin{pmatrix} 2\lambda_0\hat{\Sigma}_X & 2\hat{\Sigma}_X\eta \\ 2\eta^T\hat{\Sigma}_X & 0 \end{pmatrix}$ is invertible when n is large.

On the other hand, if $H''(\eta)\neq 0$, it's possible that $\begin{vmatrix} H''(\eta)+2\lambda_0\Sigma_X & 2\Sigma_X\eta \\ 2\eta^T\Sigma_X & 0 \end{vmatrix}$ vanished. In that case, in spirit of von Mises' proposition (Serfling (1980, Section 6.1)), $\sqrt{n}$ or higher order-consistency can be achieved. For the following derivation, without loss of generality, we assume $\begin{vmatrix} H''(\eta)+2\lambda_0\Sigma_X & 2\Sigma_X\eta \\ 2\eta^T\Sigma_X & 0 \end{vmatrix} \neq 0$.

As for $\mathcal{R}_1(\theta_n^*)$, let $T_n = \mathcal{L}_n'''(\theta_n^*)$, then T_n is a $(p+1)\times(p+1)\times(p+1)$ array. Each $T_n(j,:,:)$, $j=1,\dots,p+1$ is a $(p+1)\times(p+1)$ matrix.

Let $\hat{\Sigma}_X = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1p} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{p1} & \hat{\sigma}_{p2} & \cdots & \hat{\sigma}_{pp} \end{pmatrix}$, then we can write

$$T_n(j,:,:) = \begin{pmatrix} \frac{\partial H_n''(\eta)_{11}}{\partial\eta_j} & \frac{\partial H_n''(\eta)_{12}}{\partial\eta_j} & \cdots & \frac{\partial H_n''(\eta)_{1p}}{\partial\eta_j} & 2\hat{\sigma}_{j1} \\ \frac{\partial H_n''(\eta)_{21}}{\partial\eta_j} & \frac{\partial H_n''(\eta)_{22}}{\partial\eta_j} & \cdots & \frac{\partial H_n''(\eta)_{2p}}{\partial\eta_j} & 2\hat{\sigma}_{j2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\partial H_n''(\eta)_{p1}}{\partial\eta_j} & \frac{\partial H_n''(\eta)_{p2}}{\partial\eta_j} & \cdots & \frac{\partial H_n''(\eta)_{pp}}{\partial\eta_j} & 2\hat{\sigma}_{jp} \\ 2\hat{\sigma}_{j1} & 2\hat{\sigma}_{j2} & \cdots & 2\hat{\sigma}_{jp} & 0 \end{pmatrix}, \ j=1,2,\dots,p, \text{ and}$$

$$T_n(p+1,:,:) = 2\begin{pmatrix} \hat\sigma_{11} & \hat\sigma_{21} & \cdots & \hat\sigma_{p1} & 0 \\ \hat\sigma_{12} & \hat\sigma_{22} & \cdots & \hat\sigma_{p2} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \hat\sigma_{1p} & \hat\sigma_{2p} & \cdots & \hat\sigma_{pp} & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

The form of $\mathcal{R}_1(\theta_n^*)$ can be written as

$$\mathcal{R}_1(\theta_n^*) = \frac{1}{2}\begin{pmatrix} (\theta_n-\theta)^T T_n(1,:,:)(\theta_n-\theta) \\ (\theta_n-\theta)^T T_n(2,:,:)(\theta_n-\theta) \\ \vdots \\ (\theta_n-\theta)^T T_n(p+1,:,:)(\theta_n-\theta) \end{pmatrix}.$$

Therefore, the Taylor expansion of $\mathcal{L}_n'(\theta_n)$ at θ can be written as

$$0 = \begin{pmatrix} H_n'(\eta) + 2\lambda_0\hat\Sigma_X\eta \\ \eta^T\hat\Sigma_X\eta - 1 \end{pmatrix} + \begin{pmatrix} H_n''(\eta) + 2\lambda_0\hat\Sigma_X & 2\hat\Sigma_X\eta \\ 2\eta^T\hat\Sigma_X & 0 \end{pmatrix}\begin{pmatrix} \eta_n - \eta \\ \lambda_n - \lambda_0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} (\theta_n-\theta)^T T_n(1,:,:)(\theta_n-\theta) \\ (\theta_n-\theta)^T T_n(2,:,:)(\theta_n-\theta) \\ \vdots \\ (\theta_n-\theta)^T T_n(p+1,:,:)(\theta_n-\theta) \end{pmatrix}.$$

Since, we assume $\begin{pmatrix} H_n''(\eta) + 2\lambda_0\hat\Sigma_X & 2\hat\Sigma_X\eta \\ 2\eta^T\hat\Sigma_X & 0 \end{pmatrix}$ is invertible, from the Taylor expansion above, we obtain that

$$-\begin{pmatrix} H_n''(\eta) + 2\lambda_0\hat\Sigma_X & 2\hat\Sigma_X\eta \\ 2\eta^T\hat\Sigma_X & 0 \end{pmatrix}^{-1}\sqrt{n}\begin{pmatrix} H_n'(\eta) + 2\lambda_0\hat\Sigma_X\eta \\ \eta^T\hat\Sigma_X\eta - 1 \end{pmatrix} = \left[I_{p+1} + \frac{1}{2}\begin{pmatrix} H_n''(\eta) + 2\lambda_0\hat\Sigma_X & 2\hat\Sigma_X\eta \\ 2\eta^T\hat\Sigma_X & 0 \end{pmatrix}^{-1}\begin{pmatrix} (\theta_n-\theta)^T T_n(1,:,:) \\ (\theta_n-\theta)^T T_n(2,:,:) \\ \vdots \\ (\theta_n-\theta)^T T_n(p+1,:,:) \end{pmatrix}\right]\sqrt{n}(\theta_n-\theta).$$

Next, we are going to prove two parts:

Part 1: $\begin{pmatrix} H_n''(\eta) + 2\lambda_0\hat\Sigma_X & 2\hat\Sigma_X\eta \\ 2\eta^T\hat\Sigma_X & 0 \end{pmatrix}^{-1}\sqrt{n}\begin{pmatrix} H_n'(\eta) + 2\lambda_0\hat\Sigma_X\eta \\ \eta^T\hat\Sigma_X\eta - 1 \end{pmatrix} \xrightarrow{D} N(0,V).$

Part 2:

$$\sqrt{n}(\theta_n - \theta) \stackrel{D}{=}$$

$$\left[I_{p+1} + \frac{1}{2} \begin{pmatrix} H_n''(\eta) + 2\lambda_0 \hat{\Sigma}_X & 2\hat{\Sigma}_X \eta \\ 2\eta^T \hat{\Sigma}_X & 0 \end{pmatrix}^{-1} \begin{pmatrix} (\theta_n - \theta)^T T_n(1, :, :) \\ (\theta_n - \theta)^T T_n(2, :, :) \\ \vdots \\ (\theta_n - \theta)^T T_n(p+1, :, :) \end{pmatrix} \right] \sqrt{n}(\theta_n - \theta).$$

To show Part 1, we use the asymptotic properties for U-statistics and argue that $\eta^T \hat{\Sigma}_X \eta$ is a linear combination of U-statistics and $H_n'(\eta) + 2\lambda_0 \hat{\Sigma}_X \eta$ converges to a U-statistic as $n \longrightarrow \infty$.

As defined, $H_n'(\eta) + 2\lambda \hat{\Sigma}_X \eta = S_1'(\eta) - 2S_2'(\eta) + S_3'(\eta) + 2\lambda \hat{\Sigma}_X \eta$, where

$$S_1'(\eta) = -\frac{1}{n^2} \sum_{i,j} \exp[-(\eta^T (\mathbf{X}_i - \mathbf{X}_j))^2/2] \cdot [(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \eta] \cdot \exp\left(\frac{-\|Y_i - Y_j\|^2}{2\hat{\sigma}_Y^2} \right),$$

$$S_2'(\eta) = -\frac{1}{n^3} \sum_{i,j,k} \exp[-(\eta^T (\mathbf{X}_i - \mathbf{X}_j))^2/2] \cdot [(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \eta] \cdot \exp\left(\frac{-\|Y_i - Y_k\|^2}{2\hat{\sigma}_Y^2} \right),$$

$$S_3'(\eta) = -\frac{1}{n^4} \sum_{i,j,k,l} \exp[-(\eta^T (\mathbf{X}_i - \mathbf{X}_j))^2/2] \cdot [(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \eta] \cdot \exp\left(\frac{-\|Y_k - Y_l\|^2}{2\hat{\sigma}_Y^2} \right).$$

$$\hat{\Sigma}_X = \frac{1}{n} \sum_i \mathbf{X}_i \mathbf{X}_i^T - \bar{\mathbf{X}} \bar{\mathbf{X}}^T = \frac{1}{n} \sum_i \mathbf{X}_i \mathbf{X}_i^T - \frac{1}{n^2} \sum_{i,j} \mathbf{X}_i \mathbf{X}_j^T.$$

It can be shown as following that $\hat{\Sigma}_X$ is a U-statistic, since it can be written as a linear combination of U-statistics.

$$\begin{aligned} \hat{\Sigma}_X &= \frac{1}{n} \sum_i \mathbf{X}_i \mathbf{X}_i^T - \frac{1}{n^2} \sum_{i,j} \mathbf{X}_i \mathbf{X}_j^T \\ &= \frac{1}{n} \sum_i \mathbf{X}_i \mathbf{X}_i^T - \frac{1}{n^2} \sum_i \mathbf{X}_i \mathbf{X}_i^T - \frac{1}{n^2} \sum_{i \neq j} \mathbf{X}_i \mathbf{X}_j^T \\ &= \frac{n-1}{n} [\frac{1}{n} \sum_i \mathbf{X}_i \mathbf{X}_i^T] - \frac{2}{n^2} (\sum_{i<j} \frac{1}{2} (\mathbf{X}_i \mathbf{X}_j^T + \mathbf{X}_j \mathbf{X}_i^T)) \\ &= \frac{n-1}{n} [\frac{1}{n} \sum_i \mathbf{X}_i \mathbf{X}_i^T] - \frac{n-1}{n} [\binom{n}{k}^{-1} \sum_{i<j} \frac{1}{2} (\mathbf{X}_i \mathbf{X}_j^T + \mathbf{X}_j \mathbf{X}_i^T)]. \end{aligned}$$

$S_1'(\eta)$, $S_2'(\eta)$, and $S_3'(\eta)$ are U-statistics as well, which are demonstrated in following discussions.

Let

$$U_{1n} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2] \cdot [(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T \eta] \cdot \exp\left(\frac{-\|Y_i - Y_j\|^2}{2\sigma_Y^2}\right),$$

$$U_{2n} = \binom{n}{3}^{-1} \sum_{1\le i<j<k\le n} \Big[\frac{1}{6}\sum_{P_1} \exp[-(\eta^T(\mathbf{X}_{i_1} - \mathbf{X}_{i_2}))^2/2] \cdot [(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})^T \eta] \cdot \exp\left(\frac{-\|Y_{i_1} - Y_{i_3}\|^2}{2\sigma_Y^2}\right)\Big],$$

$$U_{3n} = \binom{n}{4}^{-1} \sum_{1\le i<j<k<l\le n} \Big[\frac{1}{24}\sum_{P_2} \exp[-(\eta^T(\mathbf{X}_{i_1} - \mathbf{X}_{i_2}))^2/2] \cdot [(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})^T \eta] \cdot \exp\left(\frac{-\|Y_{i_3} - Y_{i_4}\|^2}{2\sigma_Y^2}\right)\Big],$$

$$U_{4n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i\mathbf{X}_i^T\eta,$$

$$U_{5n} = \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{2}(\mathbf{X}_i\mathbf{X}_j^T + \mathbf{X}_j\mathbf{X}_i^T)\eta,$$

$$U_{6n} = \frac{1}{n}\sum_{i=1}^{n} \eta^T\mathbf{X}_i\mathbf{X}_i^T\eta,$$

$$U_{7n} = \binom{n}{2}^{-1} \sum_{i<j} \frac{1}{2}\eta^T(\mathbf{X}_i\mathbf{X}_j^T + \mathbf{X}_j\mathbf{X}_i^T)\eta.$$

Here $\sum_{P_1}$ denotes summation over 3! permutations (i_1, i_2, i_3) of (i, j, k), and $\sum_{P_2}$ denotes summation over 4! permutations (i_1, i_2, i_3, i_4) of (i, j, k, l). Therefore,

$$
\begin{aligned}
S_1'(\eta) &= -\frac{2}{n^2}\binom{n}{2}\Big[\binom{n}{2}^{-1}\sum_{1\le i<j\le n}\exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\\
&\quad\cdot\exp\left(\frac{-\|Y_i-Y_j\|^2}{2\sigma_Y^2}\right)\Big]\\
&= -\frac{n-1}{n}U_{1n},
\end{aligned}
$$

$$
\begin{aligned}
S_2'(\eta) &= -\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\\
&\quad\cdot\exp\left(\frac{-\|Y_i-Y_k\|^2}{2\sigma_Y^2}\right)\\
&= -\frac{1}{n^3}\Big[\sum_{i\ne j,i\ne k,j\ne k}\exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\cdot\exp\left(\frac{-\|Y_i-Y_k\|^2}{2\sigma_Y^2}\right)\\
&\quad+\sum_{i\ne j,i\ne k,j=k}\exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\cdot\exp\left(\frac{-\|Y_i-Y_k\|^2}{2\sigma_Y^2}\right)\Big]\\
&= -\frac{6}{n^3}\binom{n}{3}\Big\{\binom{n}{3}^{-1}\sum_{1\le i<j<k\le n}\Big[\frac{1}{6}\sum_{P_1}\exp[-(\eta^T(\mathbf{X}_{i_1}-\mathbf{X}_{i_2}))^2/2]\cdot[(\mathbf{X}_{i_1}-\mathbf{X}_{i_2})(\mathbf{X}_{i_1}-\mathbf{X}_{i_2})^T\eta]\\
&\quad\cdot\exp\left(\frac{-\|Y_{i_1}-Y_{i_3}\|^2}{2\sigma_Y^2}\right)\Big]\Big\}\\
&\quad-\frac{2}{n^2}\binom{n}{2}\Big\{\binom{n}{2}^{-1}\sum_{1\le i<j=k\le n}\exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\\
&\quad\cdot\exp\left(\frac{-\|Y_i-Y_j\|^2}{2\sigma_Y^2}\right)\Big\}\\
&= -\frac{(n-1)(n-2)}{n^2}U_{2n}-\frac{n-1}{n^2}U_{1n},
\end{aligned}
$$

$$
\begin{aligned}
S_3'(\eta) &= -\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\cdot\exp\left(\frac{-\|Y_k-Y_l\|^2}{2\sigma_Y^2}\right)\\
&= -\frac{1}{n^4}\Big\{\sum_{(i,j,k,l)all\neq}+\sum_{i=k,j\neq l}+\sum_{i=l,j\neq k}+\sum_{j=l,i\neq k}+\sum_{j=k,i\neq l}+\sum_{i=k,j=l}+\sum_{i=l,j=k}\Big\}\\
&\quad \exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\cdot\exp\left(\frac{-\|Y_k-Y_l\|^2}{2\sigma_Y^2}\right)\\
&= -\frac{1}{n^4}\Big\{\sum_{(i,j,k,l)all\neq}+4\sum_{i=k,j\neq l}+2\sum_{i=k,j=l}\Big\}\\
&\quad \exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\cdot\exp\left(\frac{-\|Y_k-Y_l\|^2}{2\sigma_Y^2}\right)\\
&= -\frac{24}{n^4}\binom{n}{4}\Big\{\binom{n}{4}^{-1}\sum_{1\le i<j<k<l\le n}\Big[\frac{1}{24}\sum_{P_2}\exp[-(\eta^T(\mathbf{X}_{i_1}-\mathbf{X}_{i_2}))^2/2]\\
&\quad \cdot[(\mathbf{X}_{i_1}-\mathbf{X}_{i_2})(\mathbf{X}_{i_1}-\mathbf{X}_{i_2})^T\eta]\cdot\exp\left(\frac{-\|Y_{i_3}-Y_{i_4}\|^2}{2\sigma_Y^2}\right)\Big]\Big\}\\
&\quad -\frac{4\times 6}{n^4}\binom{n}{3}\Big\{\binom{n}{3}^{-1}\sum_{1\le i<j<k<l\le n}\Big[\frac{1}{24}\sum_{P_2}\exp[-(\eta^T(\mathbf{X}_{i_1}-\mathbf{X}_{i_2}))^2/2]\\
&\quad \cdot[(\mathbf{X}_{i_1}-\mathbf{X}_{i_2})(\mathbf{X}_{i_1}-\mathbf{X}_{i_2})^T\eta]\cdot\exp\left(\frac{-\|Y_{i_3}-Y_{i_4}\|^2}{2\sigma_Y^2}\right)\Big]\Big\}\\
&\quad -\frac{2\times 2}{n^4}\binom{n}{2}\Big\{\binom{n}{2}^{-1}\sum_{1\le i<j\le n}\exp[-(\eta^T(\mathbf{X}_i-\mathbf{X}_j))^2/2]\cdot[(\mathbf{X}_i-\mathbf{X}_j)(\mathbf{X}_i-\mathbf{X}_j)^T\eta]\\
&\quad \cdot\exp\left(\frac{-\|Y_i-Y_j\|^2}{2\sigma_Y^2}\right)\Big\}.\\
&= -\frac{(n-1)(n-2)(n-3)}{n^3}U_{3n}-\frac{4(n-1)(n-2)}{n^3}U_{2n}-\frac{2(n-1)}{n^3}U_{1n}.
\end{aligned}
$$

That is

$$
\begin{aligned}
S_1'(\eta) &= -\frac{n-1}{n}U_{1n},\\
S_2'(\eta) &= -\frac{(n-1)(n-2)}{n^2}U_{2n}-\frac{n-1}{n^2}U_{1n},\\
S_3'(\eta) &= -\frac{(n-1)(n-2)(n-3)}{n^3}U_{3n}-\frac{4(n-1)(n-2)}{n^3}U_{2n}-\frac{2(n-1)}{n^3}U_{1n},\\
\hat{\Sigma}_X\eta &= \frac{n-1}{n}U_{4n}-\frac{n-1}{n}U_{5n},\\
\eta^T\hat{\Sigma}_X\eta &= \frac{n-1}{n}U_{6n}-\frac{n-1}{n}U_{7n}.
\end{aligned}
$$

Therefore,

$$H_n'(\eta) + 2\lambda\hat{\Sigma}_X\eta = \\ -\frac{(n-1)(n^2-2n+2)}{n^3}U_{1n} + \frac{2(n-1)(n-2)^2}{n^3}U_{2n} - \frac{(n-1)(n-2)(n-3)}{n^3}U_{3n} \\ + 2\lambda\frac{n-1}{n}U_{4n} - 2\lambda\frac{n-1}{n}U_{5n}.$$

Let

$$\phi^{(1)}((\mathbf{X}_i, Y_i), (\mathbf{X}_j, Y_j)) = \exp[-(\eta^T(\mathbf{X}_i - \mathbf{X}_j))^2/2] \cdot [(\mathbf{X}_i - \mathbf{X}_j)(\mathbf{X}_i - \mathbf{X}_j)^T\eta] \cdot \exp\left(\frac{-\|Y_i - Y_j\|^2}{2\sigma_Y^2}\right)$$

$$\phi^{(2)}((\mathbf{X}_i, Y_i), (\mathbf{X}_j, Y_j), (\mathbf{X}_k, Y_k)) = \frac{1}{6}\sum_{P_1}\exp[-(\eta^T(\mathbf{X}_{i_1} - \mathbf{X}_{i_2}))^2/2] \\ \cdot [(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})^T\eta] \cdot \exp\left(\frac{-\|Y_{i_1} - Y_{i_3}\|^2}{2\sigma_Y^2}\right),$$

$$\phi^{(3)}((\mathbf{X}_i, Y_i), (\mathbf{X}_j, Y_j), (\mathbf{X}_k, Y_k), (\mathbf{X}_l, Y_l)) = \frac{1}{24}\sum_{P_2}\exp[-(\eta^T(\mathbf{X}_{i_1} - \mathbf{X}_{i_2}))^2/2] \\ \cdot [(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})(\mathbf{X}_{i_1} - \mathbf{X}_{i_2})^T\eta] \cdot \exp\left(\frac{-\|Y_{i_3} - Y_{i_4}\|^2}{2\sigma_Y^2}\right),$$

$$\phi^{(4)}((\mathbf{X}_i, Y_i)) = \mathbf{X}_i\mathbf{X}_i^T\eta,$$

$$\phi^{(5)}((\mathbf{X}_i, Y_i), (\mathbf{X}_j, Y_j)) = \frac{1}{2}(\mathbf{X}_i\mathbf{X}_j^T + \mathbf{X}_j\mathbf{X}_i^T)\eta,$$

$$\phi^{(6)}((\mathbf{X}_i, Y_i)) = \eta^T\mathbf{X}_i\mathbf{X}_i^T\eta,$$

$$\phi^{(7)}((\mathbf{X}_i, Y_i), (\mathbf{X}_j, Y_j)) = \frac{1}{2}\eta^T(\mathbf{X}_i\mathbf{X}_j^T + \mathbf{X}_j\mathbf{X}_i^T)\eta.$$

Also, let

$$\mu_1 = \mathrm{E}\Big\{\exp[-(\eta^T(\mathbf{X} - \mathbf{X}'))^2/2] \cdot [(\mathbf{X} - \mathbf{X}')(\mathbf{X} - \mathbf{X}')^T\eta] \cdot \exp\left(\frac{-\|Y - Y'\|^2}{2\sigma_Y^2}\right)\Big\},$$

$$\mu_2 = \mathrm{E}\Big\{\exp[-(\eta^T(\mathbf{X} - \mathbf{X}'))^2/2] \cdot [(\mathbf{X} - \mathbf{X}')(\mathbf{X} - \mathbf{X}')^T\eta] \cdot \exp\left(\frac{-\|Y - Y''\|^2}{2\sigma_Y^2}\right)\Big\},$$

$$\mu_3 = \mathrm{E}\Big\{\exp[-(\eta^T(\mathbf{X} - \mathbf{X}'))^2/2] \cdot [(\mathbf{X} - \mathbf{X}')(\mathbf{X} - \mathbf{X}')^T\eta] \cdot \exp\left(\frac{-\|Y'' - Y'''\|^2}{2\sigma_Y^2}\right)\Big\},$$

$$\mu_4 = \mathrm{E}\mathbf{X}\mathbf{X}^T\eta,$$

$$\mu_5 = (\mathrm{E}\mathbf{X})(\mathrm{E}\mathbf{X})^T\eta,$$

$$\mu_6 = \eta^T(\mathrm{E}\mathbf{X}\mathbf{X}^T)\eta,$$

$$\mu_7 = \eta^T(\mathrm{E}\mathbf{X})(\mathrm{E}\mathbf{X})^T\eta,$$

where $(\mathbf{X}, Y), (\mathbf{X}', Y'), (\mathbf{X}'', Y''), (\mathbf{X}''', Y''')$ are i.i.d. copies.

Assumption: $\mathrm{Var}[\phi^{(1)}((\mathbf{X}_i,Y_i),(\mathbf{X}_j,Y_j))]$, $\mathrm{Var}[\phi^{(2)}((\mathbf{X}_i,Y_i),(\mathbf{X}_j,Y_j),(\mathbf{X}_k,Y_k))]$, $\mathrm{Var}[\phi^{(3)}((\mathbf{X}_i,Y_i),(\mathbf{X}_j,Y_j),(\mathbf{X}_k,Y_k),(\mathbf{X}_l,Y_l))]$, $\mathrm{Var}[\phi^{(4)}((\mathbf{X}_i,Y_i))]$, $\mathrm{Var}[\phi^{(5)}((\mathbf{X}_i,Y_i),(\mathbf{X}_j,Y_j))]$, $\mathrm{Var}[\phi^{(6)}((\mathbf{X}_i,Y_i))]$, $\mathrm{Var}[\phi^{(7)}((\mathbf{X}_i,Y_i),(\mathbf{X}_j,Y_j))]$ are all $<\infty$.

This assumption is similar to the assumed conditions of Theorem 6.1.6 (Lehmann (1999, Chapter 6)) so that in the spirit of von Mises propositions (Serfling, (1980, Section 6.1)), the first non-vanishing term of our Taylor expansion is the linear term. Hence root-n consistency can be achieved. If this term is vanished, then n or higher order-consistency would be achieved.

By Theorem 6.1.6 (Lehmann (1999, Chapter 6)), under above Assumption,

$$\sqrt{n}\begin{pmatrix} U_{1n}-\mu_1 \\ U_{2n}-\mu_2 \\ U_{3n}-\mu_3 \\ U_{4n}-\mu_4 \\ U_{5n}-\mu_5 \\ U_{6n}-\mu_6 \\ U_{7n}-\mu_7 \end{pmatrix} \xrightarrow{D} N(0,\Sigma), \text{where } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} & \Sigma_{14} & \Sigma_{15} & \Sigma_{16} & \Sigma_{17} \\ \cdot & \Sigma_{22} & \Sigma_{23} & \Sigma_{24} & \Sigma_{25} & \Sigma_{26} & \Sigma_{27} \\ \cdot & \cdot & \Sigma_{33} & \Sigma_{34} & \Sigma_{35} & \Sigma_{36} & \Sigma_{37} \\ \cdot & \cdot & \cdot & \Sigma_{44} & \Sigma_{45} & \Sigma_{46} & \Sigma_{47} \\ \cdot & \cdot & \cdot & \cdot & \Sigma_{55} & \Sigma_{56} & \Sigma_{57} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \Sigma_{66} & \Sigma_{67} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \Sigma_{77} \end{pmatrix}.$$

Using Hoeffding's result (1948, Section 6), let $(\mathbf{X}_1,Y_1)$, $(\mathbf{X}_2,Y_2)$, $(\mathbf{X}_2',Y_2')$, $(\mathbf{X}_3,Y_3)$, $(\mathbf{X}_3',Y_3')$, $(\mathbf{X}_4,Y_4)$, $(\mathbf{X}_4',Y_4')$ be i.i.d. copies, one could obtain that

$$\Sigma_{11} = 4\mathrm{Cov}(\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2)),\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'))),$$
$$\Sigma_{12} = 6\mathrm{Cov}(\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2)),\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'),(\mathbf{X}_3',Y_3'))),$$
$$\Sigma_{13} = 8\mathrm{Cov}(\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2)),\phi^{(3)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'),(\mathbf{X}_3',Y_3'),(\mathbf{X}_4',Y_4'))),$$
$$\Sigma_{14} = 2\mathrm{Cov}(\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2)),\phi^{(4)}((\mathbf{X}_1,Y_1))),$$
$$\Sigma_{15} = 4\mathrm{Cov}(\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2)),\phi^{(5)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'))),$$
$$\Sigma_{16} = 2\mathrm{Cov}(\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2)),\phi^{(6)}((\mathbf{X}_1,Y_1))),$$
$$\Sigma_{17} = 4\mathrm{Cov}(\phi^{(1)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2)),\phi^{(7)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'))),$$
$$\Sigma_{22} = 9\mathrm{Cov}(\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2),(\mathbf{X}_3,Y_3)),\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'),(\mathbf{X}_3',Y_3'))),$$
$$\Sigma_{23} = 12\mathrm{Cov}(\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2),(\mathbf{X}_3,Y_3)),\phi^{(3)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'),(\mathbf{X}_3',Y_3'),(\mathbf{X}_4',Y_4'))),$$
$$\Sigma_{24} = 3\mathrm{Cov}(\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2),(\mathbf{X}_3,Y_3)),\phi^{(4)}((\mathbf{X}_1,Y_1))),$$
$$\Sigma_{25} = 6\mathrm{Cov}(\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2),(\mathbf{X}_3,Y_3)),\phi^{(5)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'))),$$
$$\Sigma_{26} = 3\mathrm{Cov}(\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2),(\mathbf{X}_3,Y_3)),\phi^{(6)}((\mathbf{X}_1,Y_1))),$$
$$\Sigma_{27} = 6\mathrm{Cov}(\phi^{(2)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2,Y_2),(\mathbf{X}_3,Y_3)),\phi^{(7)}((\mathbf{X}_1,Y_1),(\mathbf{X}_2',Y_2'))),$$

$$\Sigma_{33} = 16\text{Cov}(\phi^{(3)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), (\mathbf{X}_3, Y_3), (\mathbf{X}_4, Y_4)), \phi^{(3)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'), (\mathbf{X}_3', Y_3'), (\mathbf{X}_4', Y_4'))),$$
$$\Sigma_{34} = 4\text{Cov}(\phi^{(3)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), (\mathbf{X}_3, Y_3), (\mathbf{X}_4, Y_4)), \phi^{(4)}((\mathbf{X}_1, Y_1))),$$
$$\Sigma_{35} = 8\text{Cov}(\phi^{(3)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), (\mathbf{X}_3, Y_3), (\mathbf{X}_4, Y_4)), \phi^{(5)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))),$$
$$\Sigma_{36} = 4\text{Cov}(\phi^{(3)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), (\mathbf{X}_3, Y_3), (\mathbf{X}_4, Y_4)), \phi^{(6)}((\mathbf{X}_1, Y_1))),$$
$$\Sigma_{37} = 8\text{Cov}(\phi^{(3)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2), (\mathbf{X}_3, Y_3), (\mathbf{X}_4, Y_4)), \phi^{(7)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))),$$
$$\Sigma_{44} = \text{Cov}(\phi^{(4)}((\mathbf{X}_1, Y_1)), \phi^{(4)}((\mathbf{X}_1, Y_1))),$$
$$\Sigma_{45} = 2\text{Cov}(\phi^{(4)}((\mathbf{X}_1, Y_1)), \phi^{(5)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))),$$
$$\Sigma_{46} = \text{Cov}(\phi^{(4)}((\mathbf{X}_1, Y_1)), \phi^{(6)}((\mathbf{X}_1, Y_1))),$$
$$\Sigma_{47} = 2\text{Cov}(\phi^{(4)}((\mathbf{X}_1, Y_1)), \phi^{(7)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))),$$
$$\Sigma_{55} = 4\text{Cov}(\phi^{(5)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2)), \phi^{(5)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))),$$
$$\Sigma_{56} = 2\text{Cov}(\phi^{(5)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2)), \phi^{(6)}((\mathbf{X}_1, Y_1))),$$
$$\Sigma_{57} = 4\text{Cov}(\phi^{(5)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2)), \phi^{(7)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))),$$
$$\Sigma_{66} = \text{Cov}(\phi^{(6)}((\mathbf{X}_1, Y_1)), \phi^{(6)}((\mathbf{X}_1, Y_1))),$$
$$\Sigma_{67} = 2\text{Cov}(\phi^{(6)}((\mathbf{X}_1, Y_1)), \phi^{(7)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))),$$
$$\Sigma_{77} = 4\text{Cov}(\phi^{(7)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2, Y_2)), \phi^{(7)}((\mathbf{X}_1, Y_1), (\mathbf{X}_2', Y_2'))).$$

Let $\hat{\mathbf{A}} = \begin{pmatrix} H_n''(\eta) + 2\lambda_0 \hat{\Sigma}_X & 2\hat{\Sigma}_X \eta \\ 2\eta^T \hat{\Sigma}_X & 0 \end{pmatrix}^{-1}$, $\mathbf{A} = \begin{pmatrix} H''(\eta) + 2\lambda_0 \Sigma_X & 2\Sigma_X \eta \\ 2\eta^T \Sigma_X & 0 \end{pmatrix}^{-1}$, and

$\mathbf{B} = \begin{pmatrix} -\mathbf{I}_p & \mathbf{I}_p \otimes 2 & -\mathbf{I}_p & \mathbf{I}_p \otimes 2\lambda_0 & \mathbf{I}_p \otimes (-2\lambda_0) & 0_p & 0_p \\ 0_p^T & 0_p^T & 0_p^T & 0_p^T & 0_p^T & 1 & -1 \end{pmatrix}$, where 0_p is a $p \times 1$ zero vector, then by the definition of μ_i, $i = 1, \ldots, 7$, for instance $\mu_6 - \mu_7 = \eta^T \Sigma_X \eta = 1$, we have

$$\sqrt{n}\mathbf{B} \begin{pmatrix} U_{1n} - \mu_1 \\ U_{2n} - \mu_2 \\ U_{3n} - \mu_3 \\ U_{4n} - \mu_4 \\ U_{5n} - \mu_5 \\ U_{6n} - \mu_6 \\ U_{7n} - \mu_7 \end{pmatrix} = \sqrt{n} \begin{pmatrix} -U_{1n} + 2U_{2n} - U_{3n} + 2\lambda_0 U_{4n} - 2\lambda_0 U_{5n} \\ U_{6n} - U_{7n} - 1 \end{pmatrix}.$$

Note that $\sqrt{n} \begin{pmatrix} H_n'(\eta) + 2\lambda_0 \hat{\Sigma}_X \eta \\ \eta^T \hat{\Sigma}_X \eta - 1 \end{pmatrix} =$

$$\sqrt{n} \begin{pmatrix} -\frac{(n-1)(n^2-2n+2)}{n^3} U_{1n} + \frac{2(n-1)(n-2)^2}{n^3} U_{2n} - \frac{(n-1)(n-2)(n-3)}{n^3} U_{3n} + 2\lambda_0 \frac{n-1}{n} U_{4n} - 2\lambda_0 \frac{n-1}{n} U_{5n} \\ \frac{n-1}{n} U_{6n} - \frac{n-1}{n} U_{7n} - 1 \end{pmatrix}.$$

And $\sqrt{n}\begin{pmatrix} H_n'(\eta)+2\lambda_0\hat{\Sigma}_X\eta \\ \eta^T\hat{\Sigma}_X\eta-1 \end{pmatrix} - \sqrt{n}\mathbf{B}\begin{pmatrix} U_{1n}-\mu_1 \\ U_{2n}-\mu_2 \\ U_{3n}-\mu_3 \\ U_{4n}-\mu_4 \\ U_{5n}-\mu_5 \\ U_{6n}-\mu_6 \\ U_{7n}-\mu_7 \end{pmatrix} =$

$$\sqrt{n}\begin{pmatrix} -\frac{(n-1)(n^2-2n+2)}{n^3}U_{1n}+\frac{2(n-1)(n-2)^2}{n^3}U_{2n}-\frac{(n-1)(n-2)(n-3)}{n^3}U_{3n}+2\lambda_0\frac{n-1}{n}U_{4n}-2\lambda_0\frac{n-1}{n}U_{5n} \\ \frac{n-1}{n}U_{6n}-\frac{n-1}{n}U_{7n}-1 \end{pmatrix}$$

$$-\sqrt{n}\begin{pmatrix} -U_{1n}+2U_{2n}-U_{3n}+2\lambda_0U_{4n}-2\lambda_0U_{5n} \\ U_{6n}-U_{7n}-1 \end{pmatrix} \xrightarrow{p} 0 \text{ by Assumption.}$$

Therefore by Slutsky's theorem, $\sqrt{n}\begin{pmatrix} H_n'(\eta)+2\lambda_0\hat{\Sigma}_X\eta \\ \eta^T\hat{\Sigma}_X\eta-1 \end{pmatrix} \overset{D}{=} \sqrt{n}\mathbf{B}\begin{pmatrix} U_{1n}-\mu_1 \\ U_{2n}-\mu_2 \\ U_{3n}-\mu_3 \\ U_{4n}-\mu_4 \\ U_{5n}-\mu_5 \\ U_{6n}-\mu_6 \\ U_{7n}-\mu_7 \end{pmatrix}.$

Hence, $\begin{pmatrix} H_n''(\eta)+2\lambda_0\hat{\Sigma}_X & 2\hat{\Sigma}_X\eta \\ 2\eta^T\hat{\Sigma}_X & 0 \end{pmatrix}^{-1}\sqrt{n}\begin{pmatrix} H_n'(\eta)+2\lambda_0\hat{\Sigma}_X\eta \\ \eta^T\hat{\Sigma}_X\eta-1 \end{pmatrix} \overset{D}{=} \sqrt{n}\mathbf{AB}\begin{pmatrix} U_{1n}-\mu_1 \\ U_{2n}-\mu_2 \\ U_{3n}-\mu_3 \\ U_{4n}-\mu_4 \\ U_{5n}-\mu_5 \\ U_{6n}-\mu_6 \\ U_{7n}-\mu_7 \end{pmatrix} \xrightarrow{D}$ $N(0,V)$ where $V=\mathbf{AB}\Sigma\mathbf{B}^T\mathbf{A}^T$.

For Part 2, as shown before, all elements of $T_n(i,:,:)$, $i=1,\dots,p+1$ are bounded. $\hat{\mathbf{A}}^{-1}\longrightarrow\mathbf{A}^{-1}$, thus all elements of $\hat{\mathbf{A}}^{-1}$ are bounded as well.

Since

$\hat{\mathbf{A}}^{-1}=\begin{pmatrix} H_n''(\eta)+2\lambda_0\hat{\Sigma}_X & 2\hat{\Sigma}_X\eta \\ 2\eta^T\hat{\Sigma}_X & 0 \end{pmatrix}\longrightarrow\begin{pmatrix} H''(\eta)+2\lambda_0\Sigma_X & 2\Sigma_X\eta \\ 2\eta^T\Sigma_X & 0 \end{pmatrix}=\mathbf{A}^{-1}$, all elements of $\hat{\mathbf{A}}^{-1}$ are bounded as well.

Lemma 3 has shown that $\theta_n\xrightarrow{p}\theta$, so

$$\left[I_{p+1}+\frac{1}{2}\begin{pmatrix} H_n''(\eta)+2\lambda_0\hat{\Sigma}_X & 2\hat{\Sigma}_X\eta \\ 2\eta^T\hat{\Sigma}_X & 0 \end{pmatrix}^{-1}\begin{pmatrix} (\theta_n-\theta)^TT_n(1,:,:) \\ (\theta_n-\theta)^TT_n(2,:,:) \\ \vdots \\ (\theta_n-\theta)^TT_n(p+1,:,:) \end{pmatrix}\right]\xrightarrow{p} I_{p+1}.$$

Then, by Slutsky's theorem,

$$\sqrt{n}(\theta_n - \theta) \overset{D}{=} \left[I_{p+1} + \frac{1}{2} \begin{pmatrix} H_n''(\eta) + 2\lambda_0 \hat{\Sigma}_X & 2\hat{\Sigma}_X \eta \\ 2\eta^T \hat{\Sigma}_X & 0 \end{pmatrix}^{-1} \begin{pmatrix} (\theta_n - \theta)^T T_n(1, :, :) \\ (\theta_n - \theta)^T T_n(2, :, :) \\ \vdots \\ (\theta_n - \theta)^T T_n(p+1, :, :) \end{pmatrix} \right] \sqrt{n}(\theta_n - \theta).$$

Therefore,

$$\sqrt{n}(\theta_n - \theta) \overset{D}{=} \begin{pmatrix} H_n''(\eta) + 2\lambda_0 \hat{\Sigma}_X & 2\hat{\Sigma}_X \eta \\ 2\eta^T \hat{\Sigma}_X & 0 \end{pmatrix}^{-1} \sqrt{n} \begin{pmatrix} H_n'(\eta) + 2\lambda_0 \hat{\Sigma}_X \eta \\ \eta^T \hat{\Sigma}_X \eta - 1 \end{pmatrix} \xrightarrow{D} N(0, V).$$

In conclusion, θ_n is $\sqrt{n}$-consistent estimate of θ. □

References

Gretton, A., Fukumizu, K., and Spiperumbudur, B. K. (2009). Discussion of: Brownian distance covariance. *The Annals of Applied Statistics* **3**, 1285-1294.

Gretton, A., Fukumizu, K., Teo, C. H., Song, L., Schëlkopf, B., and Smola, A. J. (2008). A kernel statistical test of independence. *Advances in Neural Information Processing Systems* **20**, 585-592.

Kankainen, A. (1995). Consistency testing of total independence based on the empirical characteristic function. Ph.D. thesis, University of Jyväskylä.

Lehmann, E. L. (1999). Elements of Large-Sample Theory. New York: Springer-Verlag.

Serfling, R. J. (1980). Approximation Theorems of Mathematical Statistics. John Wiley & Sons, Inc.