Intrinsic Minimum Average Variance Estimation for Dimension Reduction with Symmetric Positive Definite Matrices and Beyond

Baiyu Chen, Shuang Dai and Zhou Yu

East China Normal University

Supplementary Material

This supplementary material contains: 1) algorithms for iOPG and iMAVE; 2) expressions of asymptotic covariance matrices in Theorem 2 and 3 in the main manuscript; 3) convergence results of iOPG and iMAVE on a general manifold; 4) a simulation study testing the CV procedure of choosing the structural dimension d and a simulation study under the general manifold case; 5) details of data collection and processing in the New York taxi network application; 6) all proofs of theoretical results that appear in the main manuscript.

1. Algorithms of iMAVE and iOPG

First we introduce three operators in matrix algebra. "vec(·)" is the common matrix vec operator that vectorize an $m \times n$ matrix by column into an $mn \times 1$ vector. For an $m \times m$ symmetric matrix $A = (a_{ij})$, define $\mathrm{vecs}(A) = (a_{11}, a_{21}, a_{22}, ..., a_{m1}, ..., a_{mm})^{\mathsf{T}}$. That is, "vecs(·)" vectorize the lower triangle part of a symmetric matrix by row. For C_0 in (3.8) in the main manuscript, define $\text{vecss}(C_0) = (c_{11}^\top(X_0), c_{21}^\top(X_0), c_{22}^\top(X_0), ..., c_{m1}^\top(X_0), ..., c_{mm}^\top(X_0))^\top$. We will frequently use $\text{vec}(B)$, $\text{vecs}(a_j)$ and $\text{vecss}(b_j)$.

The intrinsic MAVE under the log-Euclidean metric is formulated as

$$\min_{\substack{B:B^\top B=I\\ a_j,b_j}} \sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}\|a_j + b_j\left[I_m \otimes \{B^\top(X_i - X_j)\}\right] - \text{log}Y_i\|_F^2, \qquad \text{(S1.1)}$$

We show how to solve a_j and b_j from (S1.1). Similar to classic MAVE, the alternating iterative optimization approach can be adopted here. We first fix B, differentiate (S1.1) w.r.t a_j, b_j, set the derivative to 0 and solve out a_j, b_j. Then fix a_j, b_j to similarly get B.

Now suppose B is known. Since $||A||_F^2 = \text{tr}(A^{\otimes 2})$ where $A^{\otimes 2} = AA^\top$ and the minimizer of $\text{tr}(A^{\otimes 2})$ is the same as that of $\text{tr}\{\text{vecs}(A)^{\otimes 2}\}$ when A is symmetric, we rewrite the Frobenius norm as the matrix trace since differentiating the trace w.r.t. a matrix or a vector is convenient. So optimizing (S1.1) is equivalent to optimizing

$$\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij}\cdot\text{tr}\left[\{\text{vecs}(a_j) + \text{vecs}\{b_j \cdot I_m \otimes \left(B^\top(X_i - X_j)\right)\} - \text{vecs}(\text{log}Y_i)\}^{\otimes 2}\right]. \qquad \text{(S1.2)}$$

Recall that $b_j = \left(c_{kl}(X_j)\right)_{kl}$ where $c_{kl}(X_j) = c_{lk}(X_j) \in R^d$. Thus

$$
\begin{aligned}
&\text{vecs}\left\{b_j \cdot I_m \otimes \left(B^\top (X_i - X_j)\right)\right\} \\
=&\left(c_{11}^\top B^\top (X_i - X_j), c_{21}^\top B^\top (X_i - X_j), c_{22}^T B^\top (X_i - X_j), ..., c_{mm}^\top B^\top (X_i - X_j)\right)^\top \\
=&\left\{\text{vecss}(b_j)^\top \cdot I_q \otimes \left(B^\top (X_i - X_j)\right)\right\}^\top,
\end{aligned}
$$

where $q = m(m+1)/2$. Here and hereafter we drop X_j from $c_{kl}^\top(X_j)$ for simplicity. Define

$$
\chi_i(B^\top X_j) = (I_q, I_q \otimes (X_i - X_j)^\top B), \quad \alpha_j = \begin{pmatrix} \text{vecs}(a_j) \\ \text{vecss}(b_j) \end{pmatrix}.
$$

Then (S1.2) can be written as

$$
\sum_{j=1}^{n} \sum_{i=1}^{n} w_{ij} \cdot \text{tr}\left[\left\{\chi_i(B^\top X_j)\alpha_j - \text{vecs}(\text{log}Y_i)\right\}^{\otimes 2}\right]. \tag{S1.3}
$$

Differentiate (S1.3) w.r.t α_j and we get the expression of α_j which we present in the algorithm later to avoid redundancy.

Now we fix a_j and b_j and differentiate (S1.2) w.r.t B. We write

$$
b_j \cdot I_m \otimes \left(B^\top (X_i - X_j)\right) = \begin{pmatrix} c_{11}^\top B^\top (X_i - X_j) & \cdots & c_{1m}^\top B^\top (X_i - X_j) \\ \vdots & & \vdots \\ c_{m1}^\top B^\top (X_i - X_j) & \cdots & c_{mm}^\top B^\top (X_i - X_j) \end{pmatrix}.
$$

Since $c_{kl}^\top B^\top (X_i - X_j)$ is a scalar,

$$
\begin{aligned}
c_{kl}^\top B^\top (X_i - X_j) =&\text{vec}\{c_{kl}^\top B^\top (X_i - X_j)\} = \{c_{kl} \otimes (X_i - X_j)\}^\top \text{vec}(B) \\
:=&a_{kl}^\top \cdot \text{vec}(B), \quad k, l = 1, ..., m.
\end{aligned}
$$

Hence,

$$b_j \cdot I_m \otimes \big(B^{\mathsf{T}}(X_i - X_j)\big) = \begin{pmatrix} a_{11}^{\mathsf{T}}\mathrm{vec}(B) & \cdots & a_{1m}^{\mathsf{T}}\mathrm{vec}(B) \\ \vdots & & \vdots \\ a_{m1}^{\mathsf{T}}\mathrm{vec}(B) & \cdots & a_{mm}^{\mathsf{T}}\mathrm{vec}(B) \end{pmatrix},$$

and

$$\begin{aligned} &\mathrm{vecss}\left\{B_j \cdot I_m \otimes B^{\mathsf{T}}\big((X_i - X_j)\big)\right\} \\ =& \big(a_{11}^{\mathsf{T}}\mathrm{vec}(B), a_{21}^{\mathsf{T}}\mathrm{vec}(B), a_{22}^{\mathsf{T}}\mathrm{vec}(B), ..., a_{m1}^{\mathsf{T}}\mathrm{vec}(B), ..., a_{mm}^{\mathsf{T}}\mathrm{vec}(B)\big)^{\mathsf{T}} \\ =&(a_{11}, a_{21}, a_{22}, ..., a_{m1}, ..., a_{mm})^{\mathsf{T}}\mathrm{vec}(B) \\ \triangleq& A_{ij} \cdot \mathrm{vec}(B), \end{aligned}$$

where i and j in A_{ij} indicate A_{ij} varies according to X_i and X_j.

Consequently we can rewrite (S1.2) as

$$\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} \cdot \mathrm{tr}\left[\{\mathrm{vecs}(a_j - \mathrm{log}Y_i) + A_{ij}\mathrm{vec}(B)\}^{\otimes 2}\right]. \qquad \text{(S1.4)}$$

Differentiate (S1.4) w.r.t vec(B) to get the expression of vec(B). We place it in the algorithm as well. The optimization procedure of iOPG can be derived similaryly and is thus omitted.

Now we are ready to state algorithms for iMAVE and iOPG. In order to reduce the dimension of the kernel form p to d, we adopt the approach of the refined MAVE (Xia, 2007). That is, replace $K_{h_t}(X_i - X_j)$ with

$K_{h_t}(\hat{B}_{(t)}^{\mathsf{T}}(X_i - X_j))$ in each iteration. Denote $q = m(m+1)/2$ and

$$w_{ij} = \frac{K_h(B^{\mathsf{T}}(X_i - X_j))}{\sum_{i=1}^n K_h(B^{\mathsf{T}}(X_i - X_j))}, \quad \alpha_j = \begin{pmatrix} \text{vecs}(a_j) \\ \text{vecss}(b_j) \end{pmatrix},$$

$$\chi_i(X_j) = \Big(I_q, I_q \otimes (X_i - X_j)^{\mathsf{T}}\Big)^{\mathsf{T}}, \quad \chi_i(B^{\mathsf{T}} X_j) = \Big(I_q, I_q \otimes \{(X_i - X_j)^{\mathsf{T}} B\}\Big)^{\mathsf{T}},$$

$$A_{ij} = \Big(c_{11}(X_j), c_{21}(X_j), c_{22}(X_j), ..., c_{m1}(X_j), ..., c_{mm}(X_j)\Big) \otimes (X_i - X_j),$$

where $c_{kl}(X_j)$ $(1 \le l \le k \le m)$ are components of C_j.

Algorithm 1: refined iMAVE under the log-Euclidean metric

Step 1. Marginally standardize $X_1, ..., X_n$ when necessary. Set the bandwidth $h_0 = c_0 n^{-1/(p_0+6)}$, where $c_0 = 2.34$ and $p_0 = \max(p, 3)$. Let $\hat{B}_{(0)}$ be an initial estimator. Set $t = 1$.

Step 2. Compute

$$\begin{aligned} \hat{\alpha}_j^{(t)} =& \Big\{ \sum_{i=1}^n w_{ij}^{(t-1)} \chi_i(\hat{B}_{(t-1)}^{\mathsf{T}} X_j) \chi_i(\hat{B}_{(t-1)}^{\mathsf{T}} X_j)^{\mathsf{T}} \Big\}^{-1} \\ & \times \sum_{i=1}^n w_{ij}^{(t-1)} \chi_i(\hat{B}_{(t-1)}^{\mathsf{T}} X_j) \text{vecs}(\log Y_i), j = 1, ..., n. \end{aligned}$$

Read off $\text{vecs}(\hat{a}_j^{(t)})$ and $\text{vecss}(\hat{b}_j^{(t)})$ respectively from the first q and the remaining qd components of $\hat{\alpha}_j^{(t)}$.

Step 3. Compute

$$\text{vec}(\hat{B}_{(t)}) = \Big\{ \sum_{j=1}^n \sum_{i=1}^n w_{ij}^{(t-1)} A_{ij}^{(t)} (A_{ij}^{(t)})^{\mathsf{T}} \Big\}^{-1} \sum_{j=1}^n \sum_{i=1}^n w_{ij}^{(t-1)} A_{ij}^{(t)} \text{vecs}(\log Y_i - \hat{a}_j^{(t)}).$$

Step 4. If $t < 30$, reset $h_{t+1} = \max(r_n h_t, c_0 n^{-1/(d+4)})$, where $r_n = n^{-1/2(p_0+6))}$.

Set $t = t + 1$ and go back to step 2. Otherwise, get the iMAVE estimator $\hat{B}_{(t)}$.

The choice of bandwidth in iOPG is the same as iMAVE and the estimated $\hat{B}_{iOPG}$ by iOPG can usually be used as the initial value for $\hat{B}_{(0)}$ in iMAVE.

Algorithm 2: refined iOPG under the log-Euclidean metric

Step1. Marginally standardize $X_1, ..., X_n$ when necessary. Set the bandwidth $h_0 = c_0 n^{-1/(p_0+6)}$, where $c_0 = 2.34$ and $p_0 = \max(p, 3)$. Set $\hat{B}_{(0)} = I_p$. Set iteration time $t = 1$.

Step2. Compute

$$\hat{\alpha}_j^{(t)} = \left\{ \sum_{i=1}^{n} w_{ij}^{(t-1)} \chi_i(X_j) \chi_i(X_j)^\mathsf{T} \right\}^{-1} \sum_{i=1}^{n} w_{ij}^{(t-1)} \chi_i(X_j) \text{vecs}(\log Y_i), j = 1, ..., n.$$

Read off $\text{vecss}(\hat{b}_j^{(t)})$ from the last qd components of $\hat{\alpha}_j^{(t)}$.

Step3. Recover $\hat{b}_j^{(t)}, j = 1, ..., n$ from $\text{vecss}(\hat{b}_j^{(t)})$ in step 2 as

$$\hat{b}_j^{(t)} = \begin{pmatrix} c_{11}^\mathsf{T} & & & \\ c_{21}^\mathsf{T} & c_{22}^\mathsf{T} & & \\ \vdots & \vdots & \ddots & \\ c_{m1}^\mathsf{T} & c_{m2}^\mathsf{T} & \cdots & c_{mm}^\mathsf{T} \end{pmatrix}, j = 1, ..., n,$$

with the symmetric part omitted. Rearrange the lower triangle part of $\hat{b}_j^{(t)}$ to get $\hat{\beta}_j^{(t)} = (c_{11}, c_{21}, c_{22}, ..., c_{m1}, ..., c_{mm})^\mathsf{T} \in R^{q \times p}, j = 1, ..., n.$

Step 4. Compute

$$\hat{\Lambda}^{(t)} = \frac{1}{n}\sum_{j=1}^{n}(\hat{\beta}_j^{(t)})^{\mathsf{T}}\hat{\beta}_j^{(t)}.$$

Perform eigen-decomposition for $\hat{\Lambda}^{(t)}$ and get the d eigenvectors $\hat{v}_1, ..., \hat{v}_d$ corresponding to its largest d eigenvalues. Let $\hat{B}_{(t)} = (\hat{v}_1, ..., \hat{v}_d)$.

Step 5. If $t < 30$, reset $h_{t+1} = \max(r_n h_t, c_0 n^{-1/(d+4)})$, where $r_n = n^{-1/2(p_0+6))}$. Set $t = t + 1$ and go back to step 2. Otherwise, get the iOPG estimator $\hat{B}_{(t)}$.

2. Asymptotic Properties of iMAVE and iOPG

Recall our transformed model under the log-Euclidean metric is

$$\log Y = \log\{g(B_0^{\mathsf{T}}X)\} + \log\varepsilon. \tag{S2.5}$$

Denote $h(B_0^{\mathsf{T}}X) = \log\{g(B_0^{\mathsf{T}}X)\}$ and $\zeta = \log\varepsilon$. Since $h(B_0^{\mathsf{T}}X)$ and ζ are $m \times m$ symmetric matrices, denote their (k, l)-th component as h_{kl} and ζ_{kl} $(1 \le l \le k \le m)$. Let $\mu_B(u) = E(X \mid B^{\mathsf{T}}X = u)$, $w_B(u) = E(XX^{\mathsf{T}} \mid B^{\mathsf{T}}X = u)$, $v_B(u) = \mu_B(B^{\mathsf{T}}u) - u$, and $\bar{w}_B(u) = w_B(B^{\mathsf{T}}u) - \mu_B(B^{\mathsf{T}}u)\mu_B^{\mathsf{T}}(B^{\mathsf{T}}u)$ which will be frequently encountered in proofs. For any square matrix A, A^{-1} and A^{+} denote the inverse (if it exists) and the Moore-Penrose inverse matrix.

Define

$$W_{\text{SPD}} = E\left[\left\{\sum_{k=1}^{m}\sum_{l=1}^{k} h_{kl}^{(1)}(B_0^{\mathsf{T}}X)h_{kl}^{(1)}(B_0^{\mathsf{T}}X)^{\mathsf{T}}\right\} \otimes \left\{v_{B_0}(X)v_{B_0}^{\mathsf{T}}(X)\right\}\right],$$

$$\Sigma_{\text{SPD}} = \text{var}\left[\left\{\sum_{k=1}^{m}\sum_{l=1}^{k} h_{kl}^{(1)}(B_0^{\mathsf{T}}X)\zeta_{kl}\right\} \otimes v_{B_0}(X)\right],$$

and $W_0^{\text{SPD}} = \text{var}\left[\left\{M_{\text{SPD}}^{-1}\sum_{k=1}^{m}\sum_{l=1}^{k} h_{kl}^{(1)}(B_0^{\mathsf{T}}X)\zeta_{kl}\right\} \otimes \left\{\bar{w}_{B_0}^{+}(X)v_{B_0}(X)\right\}\right]$.

Under several assumptions listed in the main manuscript, we have

$$\sqrt{n}\left\{\text{vec}(\hat{B}_{\text{iMAVE}}\hat{B}_{\text{iMAVE}}^{\mathsf{T}}B_0) - \text{vec}(B_0)\right\} \xrightarrow{d} N(0, W_{\text{SPD}}^{+}\Sigma_{\text{SPD}}W_{\text{SPD}}^{+}),$$

$$\sqrt{n}\left\{\text{vec}(\hat{B}_{\text{iOPG}}\hat{B}_{\text{iOPG}}^{\mathsf{T}}B_0) - \text{vec}(B_0)\right\} \xrightarrow{d} N(0, W_0^{\text{SPD}}).$$

Next we extend our model to a general Riemannian manifold other than $\text{Sym}^{+}(m)$. As noted in the main manuscript, we assume the model as

$$\text{Log}_{\mu}Y = h(B_0^{\mathsf{T}}X) + \zeta, \tag{S2.6}$$

where Y belongs to a general Riemannian manifold $\mathcal{M}$ and μ is the Fréchet mean of Y. Since $\text{Log}_{\mu}Y \in T_{\mu}\mathcal{M}$, $\text{Log}_{\mu}Y \in R^{s}$ where s is the dimension of $\mathcal{M}$. Denote the k-th component of h as h_k ($k = 1, ..., s$). Substitute y_k, h_k for y_{kl}, h_{kl} in conditions (A1)-(A5) in the manuscript. Replace the matrix M_{SPD} in condition (A4) with $M_0 = E\{h^{(1)}(B_0^{\mathsf{T}}X)^{\mathsf{T}}h^{(1)}(B_0^{\mathsf{T}}X)\}$ where $h^{(1)} = \nabla h(B_0^{\mathsf{T}}X) \in R^{s\times d}$. Denote the modified conditions as (A1')-(A5').

Define

$$W_{B_0} = E\left[\left\{h^{(1)}(B_0^{\mathsf{T}}X)^{\mathsf{T}}h^{(1)}(B_0^{\mathsf{T}}X)\right\} \otimes \left\{v_{B_0}(X)v_{B_0}^{\mathsf{T}}(X)\right\}\right],$$

$$\Sigma_0 = \text{var}\left[\left\{h^{(1)}(B_0^{\mathsf{T}}X)^{\mathsf{T}} \otimes v_{B_0}(X)\right\}\zeta\right],$$

---

and $W_0 = \text{var}\left[\left\{M_0^{-1} h^{(1)}(B_0^\mathsf{T} X)^\mathsf{T} \zeta\right\} \otimes \left\{\bar{w}_{B_0}^{+}(X) v_{B_0}(X)\right\}\right]$.

Theorem S1. *Under (A1')-(A5') and (C1)-(C6), the estimated* $\hat{B}_{\text{iMAVE}}$ *from (S2.6) satisfies*

$$\|\hat{B}_{\text{iMAVE}}\hat{B}_{\text{iMAVE}}^\mathsf{T} - B_0 B_0^\mathsf{T}\|_F = O(h^3 + h\delta_{dh} + \delta_{dh}^2/h + n^{-1/2})$$

in probability as $n \to \infty$, *where* $\delta_{dh} = (nh^d/\text{logn})^{-1/2}$. *If* $h^3 + h\delta_{dh} + \delta_{dh}^2/h = o(n^{-1/2})$, *then*

$$\sqrt{n}\left\{\text{vec}(\hat{B}_{\text{iMAVE}}\hat{B}_{\text{iMAVE}}^\mathsf{T} B_0) - \text{vec}(B_0)\right\} \xrightarrow{d} N(0, W_{B_0}^{+}\Sigma_0 W_{B_0}^{+}).$$

Theorem S2. *Under (A1')-(A5') and (C1)-(C6), the estimated* $\hat{B}_{\text{iOPG}}$ *from (S2.6) satisfies*

$$\|\hat{B}_{\text{iOPG}}\hat{B}_{\text{iOPG}}^\mathsf{T} - B_0 B_0^\mathsf{T}\|_F = O(h^3 + h\delta_{dh} + n^{-1/2})$$

in probability as $n \to \infty$, *where* $\delta_{dh} = (nh^d/\text{logn})^{-1/2}$. *If* $h^3 + h\delta_{dh} = o(n^{-1/2})$, *then*

$$\sqrt{n}\left\{\text{vec}(\hat{B}_{\text{iOPG}}\hat{B}_{\text{iOPG}}^\mathsf{T} B_0) - \text{vec}(B_0)\right\} \xrightarrow{d} N(0, W_0).$$

Results in Theorem S1 and S2 are consistent with those in Xia et al (2002), Xia (2007) and Zhang (2021) as well. The discrepancy between $\text{Log}_{\hat{\mu}} Y_i$ and $\text{Log}_{\mu} Y_i$ does not affect the convergence results. Actually Theorem 2 and 3 in the main manuscript can be seen as corollaries of Theorem

S1 and S2 here since $m \times m$ symmetric matrices in (S2.5) are equivalent to their lower triangle parts which are $m(m+1)/2$-dimensional vectors and Theorem S1 and S2 can be directly awakened.

3. More Simulation Studies

3.1 Simulation Study III: Spherical Data

Since the proposed iMAVE and iOPG can be extended to general manifolds, we test the performance of our methods in a general manifold. We generate $Y \in S^2$ according to the following model:

III: Let $p_0 = (0, 0, 1)^\intercal$ and the tangent vector at p_0 be

$$l(X_i) = \left(\exp(X_{i1}) \sin X_{i1} + \epsilon_{i1}, \frac{\exp(X_{i1} + X_{i2}) - 1}{\exp(X_{i1} + X_{i2}) + 1} + \epsilon_{i2}, 0 \right)^\intercal.$$

We generate i.i.d. observations $X_1, ..., X_n$ from the uniform distribution on $[-1, 1]$ and i.i.d. $\epsilon_{i1}, \epsilon_{i2} \sim N(0, 0.1^2)$. Then Y_i is generated by

$$Y_i = \text{Exp}_{p_0}\{l(X_i)\} = \cos(||l(X_i)||)p_0 + \sin(||l(X_i)||)l(X_i)/||l(X_i)||,$$

where $|| \cdot ||$ is the Euclidean norm.

We set $n = 100, 200$ and $p = 20, 30$. The simulation results are listed in Table S1. Our iMAVE or iOPG perform better than others in all scenarios.

Model	(p, n)	WIRE	fOPG	fMAVE	iOPG	iMAVE
III	(20,100)	0.5247	1.2053	1.6644	0.4471	0.4849
		±0.1077	±0.2424	±0.0665	±0.0520	±0.3619
	(20,200)	0.3651	0.7008	1.6491	0.2782	0.2365
		±0.0550	±0.2036	±0.1306	±0.0417	±0.0406
	(30,100)	0.6930	1.2803	1.6929	0.7287	0.6036
		±0.0873	±0.1488	±0.1141	±0.1569	±0.3239
	(30,200)	0.4431	1.0343	1.6810	0.3711	0.4619
		±0.0360	±0.1799	±0.1070	±0.0557	±0.3811

Table S1: Mean (± standard deviation) of estimation errors for different methods in model III.

3.2 Simulation Study IV: CV Procedure Tested

We assume now the structural dimension d is unknown. We generate data from the five models in simulation study I, II (in the main manuscript) and III and use the proposed CV procedure to estimate d. We use iOPG to estimate B. We set $p = 10$, $n = 200$, repeat 100 times for each model and list the counts of correct and false estimates in 100 times when $\sigma = 0.1$ and 0.2, which is shown in Figure S1.

Except model I-1 with $\sigma = 0.2$, the CV procedure always gives satisfying estimations, reaching an accuracy greater than 80% and even approaching

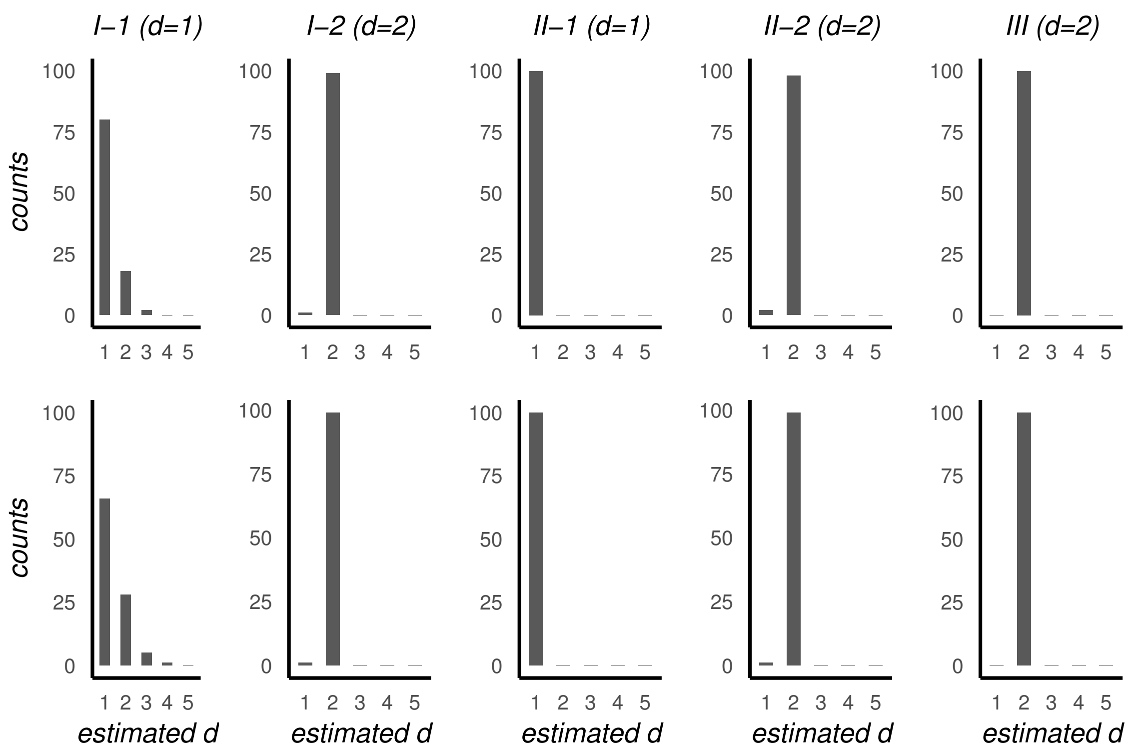

Figure S1: Bar charts: counts of correct and false estimates in 100 replications for five models with $(p, n) = (10, 200)$. The upper and the lower row correspond respectively to $\sigma = 0.1$ and 0.2.

100% in most cases. And if we increase the sample size to 300, the result corresponding to model I-1 with $\sigma = 0.2$ becomes: $(\hat{d} < d) : 0$, $(\hat{d} = d) : 92$, $(\hat{d} > d) : 8$. Such improvement validates Theorem 1 in the main manuscript.

4. New York Taxi Network Data

The New York City Taxi and Limousine Commission (TLC) provides records on pick-up and drop-off dates and times, pick-up and drop-off locations, trip distances, itemized fares, payment types and other information for yellow taxis (Tucker *et al.*, 2021). The data are available from

`https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page`

Similar to Tucker *et al.* (2021), we transform the raw data into network data (adjacent matrices), where zones are nodes and edges are weighted by the number of taxi rides which picked up in one zone and dropped off in another within a single hour. After proper mapping, these adjacent matrices can lie in the space of SPD matrices. We do the following to collect SPD matrices together with several prediction variables:

1. We only choose the data of January and February, 2019 (59 days) due to resource restrictions.

2. We further filter on observations with both pick-up and drop-off occurring in Manhattan (islands excluded).

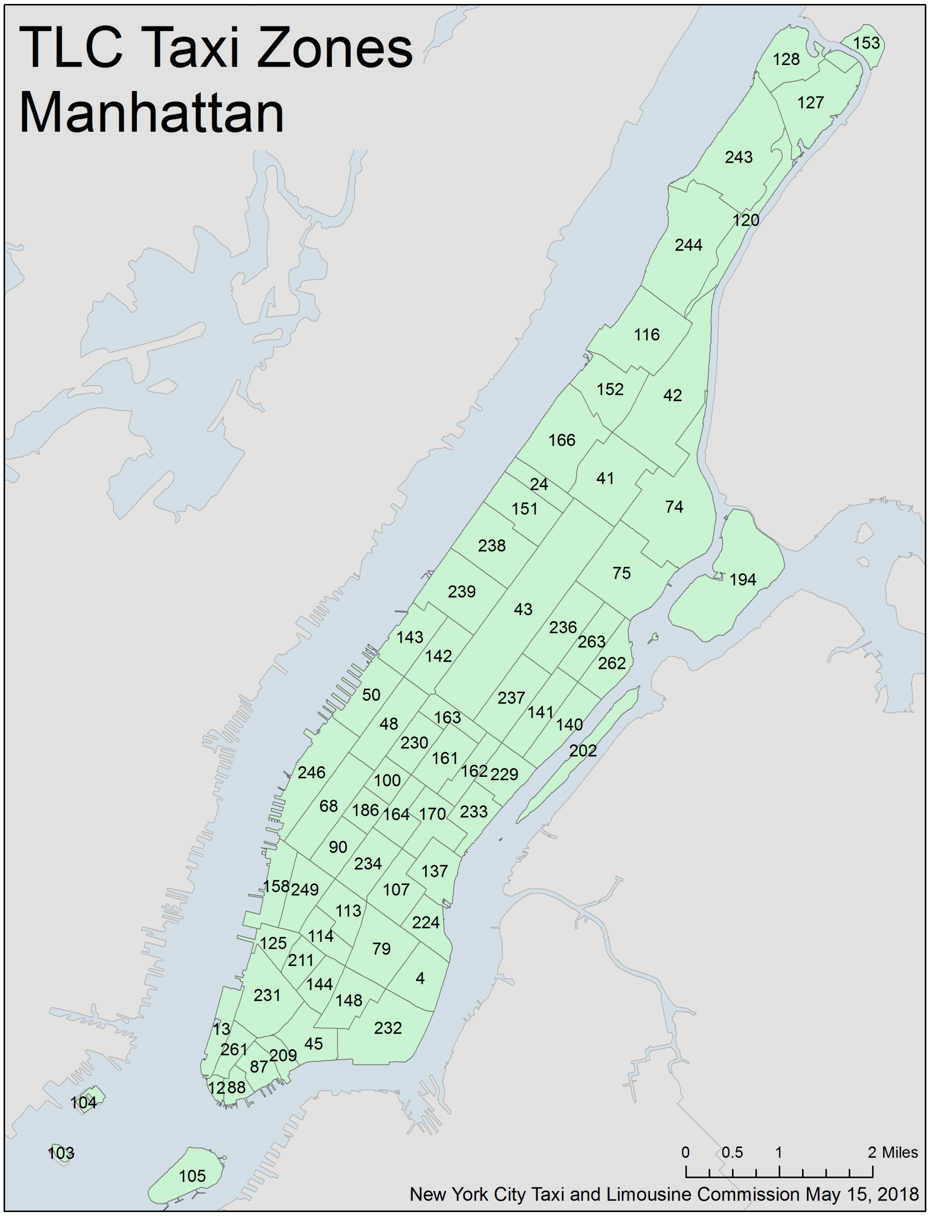

Figure S2: The map of the TLC taxi zones in Manhattan. This map is downloaded from https:/ /www1.nyc.gov/assets/tlc/images/content/pages /about/taxi_zone_map_manhattan.jpg

Bigger Zones	Original Taxi Zones
1	128, 127, 243, 244, 120, 116, 152, 42, 41, 74, 75, 166, 24, 151, 238, 239, 43, 236, 237, 263, 262, 141, 140
2	143, 142, 50, 48, 246, 68, 90, 186, 230, 100, 163, 161, 164, 234, 162, 170, 107, 229, 233, 137, 224
3	158, 249, 125, 113, 114, 211, 144, 79, 4, 148, 232, 231, 13, 261, 12, 88, 87, 209, 45

Table S2: Grouped zones of Manhattan.

3. We then group zones in Manhattan into 3 zones and label them similar to Dubey and Müller (2020). To be specific, Figure S2 shows the map of the TLC taxi zones in Manhattan. We group these zones (islands excluded) into three bigger zones according to Table S2. That is, each network has 3 nodes.

4. For each hour, we collected the number of pairwise connections between nodes based on pick-ups and drop-offs. These correspond to weights between nodes. We then further normalize the weights by the maximum edge weight in each hour so that they lie in $[0, 1]$.

By doing so, we collected 1416 (59×24) weighted adjacent matrices of 3×3 describing the taxi movements between zones in Manhattan. To ensure that they are SPD matrices, we apply $\exp(\cdot)$ to these symmetric matrices.

From the taxi data, we also collect the following 9 potential predictors, with values averaged over each hour:

Ave.Distance: mean distance traveled, standardized

Ave.Fare: mean total fare, standardized

Ave.Passengers: mean number of passengers, standardized

Ave.tip: mean tip, standardized

Cash: sum of cash indicators for type of payment, standardized

Credit: sum of credit indicators for type of payment, standardized

Dispute: sum of dispute indicators for type of payment, standardized

Free: sum of free indicators for type of payment, standardized

LateHour: indicator for the hour being between 11pm and 5am

Apart from these, we also collect New York City weather history for January and February 2019 from

`https://www.wunderground.com/history/daily/us/ny/new-york-city/KLGA/date`

The following 5 weather variables are included as potential predictors:

Ave.temp: daily mean temperatire, standardized

Ave.humid: daily mean humidity, standardized

Ave.wind: daily mean wind speed, standardized

Ave.press: daily mean barometric pressure, standardized

Precip: daily total precipitation, standardized

This then yields a total of 14 potential predictors.

5. Proof

5.1 Proof of Theorem 1

All the proofs of lemmas needed for theoretical proof can be found in section 5.6. Here we prove the CV result of iMAVE or iOPG for $\mathrm{Sym}^+(m)$ endowed with the log-Euclidean metric. As we have mentioned, the log-Cholesky case only replaces $\log Y_i$ by $\mathrm{chol}(Y_i)$ and thus the conclusion together with the proof is the same and is omitted. Little modification is needed to derive the CV procedure for the general Riemannian manifolds, which is also shown in Zhang (2021). To prove Theorem 1, we need the following lemma.

Lemma S1. *Suppose* $\log Y_i = g(X_i) + \varepsilon_i, i = 1, ..., n$ *where* $g(\cdot) : R^p \to \mathrm{Sym}(m)$ *. Suppose* $g_{kl}(\cdot)$ *has fifth derivatives. Let* $(\hat{a}_j, \hat{b}_j)$ *minimize*

$$\sum_{i \neq j} \|\log Y_i - a_j - b_j \cdot I_m \otimes (X_i - X_j)\| K_h(X_i - X_j). \tag{S5.7}$$

Then

$$\frac{1}{n}\sum_{j=1}^{n}\|\text{vecs}(\log Y_j)-\text{vecs}(\hat{a}_j)\|^2=\sum_{l\le k}\sigma_{kl}^2+\frac{1}{4}h^4\sum_{l\le k}\int\{\text{tr}(g_{kl}^{(2)})(x)\}^2 f(x)dx$$
$$+\sum_{l\le k}\frac{\alpha_p\sigma_{kl}^2}{nh^p}\{1+o_P(1)\}+O_P(h^5+n^{-1/2})$$

where $\sum_{l\le k}$ *is short for* $\sum_{1\le l\le k\le m}$, σ_{kl}^2 *is the variance of* ε_{kl} *and* $\alpha_p=\{\int K^2(u)du\}^p$.

Proof of Theorem 1: The proof follows almost the same line as that of Xia et al. (2002). Let $(B_0,\bar{B}_0):p\times p$ satisfy $(B_0,\bar{B}_0)^\mathsf{T}(B_0,\bar{B}_0)=I$ and B_l be the first l columns of $(B_0,\bar{B}_0)$. Define

$$\tilde{f}_{l,j}(x)=\frac{1}{n}\sum_{i\ne j}K_h(B_l^\mathsf{T}(X_i-x)),$$
$$\tilde{a}_{l0,j}(x)=\{n\tilde{f}_{l,j}(x)\}^{-1}\sum_{i\ne j}K_h(B_l^\mathsf{T}(X_i-x))\text{vecs}(\log Y_i),$$
$$\hat{f}_{l,j}(x)=\frac{1}{n}\sum_{i\ne j}K_h(\hat{B}_l^\mathsf{T}(X_i-x)),$$
$$\hat{a}_{l0,j}(x)=\{n\hat{f}_{l,j}(x)\}^{-1}\sum_{i\ne j}K_h(\hat{B}_l^\mathsf{T}(X_i-x))\text{vecs}(\log Y_i).$$

Suppose $B_d=(\beta_1,...,\beta_d)$. That is, $\log Y=g(\beta_1X,...,\beta_dX)+\varepsilon$. If $d<p$, nominally extend the number of directions to p, say $\{\beta_1,...,\beta_d,...,\beta_p\}$, such that they are perpendicular to one another. Now the problem becomes the selection of covariates among $\{\beta_1X,...,\beta_pX\}$, which is just the focus of Yao and Tong (1994). However, since $\beta_1,...,\beta_p$ are unknown, we must replace

β_k by their estimate $\hat{\beta}_k$. So to take advantage of the proof in Yao and Tong (1994), we need $\tilde{f}_{l,j}(x)$ and $\tilde{a}_{l0,j}(x)$.

Let $\mathrm{CV}_0(l) = n^{-1}\sum_{j=1}^{n} ||\mathrm{vecs}(\log Y_j) - \tilde{a}_{l0,j}(X_j)||^2$ and $\mathrm{CV}(l) = n^{-1}\sum_{j=1}^{n} ||\mathrm{vecs}(\log Y_j) - \hat{a}_{l0,j}(X_j)||^2$. Suppose we have shown that

$$\mathrm{CV}(d) - \mathrm{CV}_0(d) = o_P(h^4). \tag{S5.8}$$

Following the proof of Yao and Tong (1994), there is a constant $\delta > 0$ such that for the working dimension $l < d$,

$$\lim_{n\to\infty} P\{\mathrm{CV}(l) > \mathrm{CV}_0(d) + \delta\} = 1.$$

Hence $\lim_{n\to\infty} P\{\mathrm{CV}(l) > \mathrm{CV}(d)\} = 1$.

For $l > d$, by Lemma S1, we have $\mathrm{CV}(l) > \mathrm{CV}(d) + O(h^4)$. Consequently,

$$\lim_{n\to\infty} P\{\mathrm{CV}(l) > \mathrm{CV}(d)\} = 1.$$

Therefore, all that is left is to prove (S5.8).

$$\begin{aligned}
\mathrm{CV}(d) =& \frac{1}{n}\sum_{j=1}^{n} ||\mathrm{vecs}(\log Y_j) - \hat{a}_{d0,j}(X_j)||^2 \\
=& \frac{1}{n}\sum_{j=1}^{n} ||\mathrm{vecs}(\log Y_j) - \tilde{a}_{l0,j}(X_j) + \tilde{a}_{l0,j}(X_j) - \hat{a}_{d0,j}(X_j)||^2 \\
=& \frac{1}{n}\sum_{j=1}^{n} ||\mathrm{vecs}(\log Y_j) - \tilde{a}_{l0,j}(X_j)||^2 + \frac{1}{n}\sum_{j=1}^{n} ||\tilde{a}_{l0,j}(X_j) - \hat{a}_{d0,j}(X_j)||^2 \\
&+ \frac{2}{n}\sum_{j=1}^{n} \left(\mathrm{vecs}(\log Y_j) - \tilde{a}_{l0,j}(X_j)\right)^T \left(\tilde{a}_{l0,j}(X_j) - \hat{a}_{d0,j}(X_j)\right).
\end{aligned}$$

Let $w_{ij}(B_l) = K_h(B_l^{\mathsf{T}}(X_i - X_j))/\sum_{i\neq j} K_h(B_l^{\mathsf{T}}(X_i - X_j))$ and denote Y_j^{kl} as the (k,l) element of the $m \times m$ matrix $\log Y_j$. We have

$$\begin{aligned}\mathrm{CV}(d) = \sum_{l\leq k}\Bigg\{ & \frac{1}{n}\sum_{j=1}^{n}\left(Y_j^{kl} - \sum_{i\neq j} w_{ij}(\hat{B}_d)Y_j^{kl}\right)^2 + \frac{1}{n}\sum_{j=1}^{n}\left(\sum_{i\neq j} w_{ij}(B_d)Y_j^{kl} - \sum_{i\neq j} w_{ij}(\hat{B}_d)Y_j^{kl}\right)^2 \\ & + \frac{1}{n}\sum_{j=1}^{n}\left(Y_j^{kl} - \sum_{i\neq j} w_{ij}(\hat{B}_d)Y_j^{kl}\right)\left(\sum_{i\neq j} w_{ij}(B_d)Y_j^{kl} - \sum_{i\neq j} w_{ij}(\hat{B}_d)Y_j^{kl}\right)\Bigg\}.\end{aligned}$$

That is, $\mathrm{CV}(d)$ is the summation of the CV values of the case where Y is a scalar, which is the focus of Xia et al. (2002). Directly apply the results of Xia et al. (2002) and get the result.

5.2 Proof of Proposition 1

Proof of Proposition 1: Our model $Y = g(B_0^T X) \oplus \varepsilon$ is equivalent to

$$Y = e \oplus g(B_0^T X) \oplus \varepsilon, \tag{S5.9}$$

where e is the identity element of group $(\mathrm{Sym}^+(m), \oplus)$. According to Lin *et al.* (2022), we have $\mathrm{Log}_\mu(\mu \oplus z) = \phi_{e,\mu}\mathfrak{log}(z)$ for $\mu, z \in \mathrm{Sym}^+(m)$. Applying this to (S5.9) with $\mu = e$ and $z = g(B_0^T X) \oplus \varepsilon$, we have $\mathrm{Log}_e Y = \mathfrak{log}(g(B_0^T X) \oplus \varepsilon)$. Use another equation in Lin *et al.* (2020): $\mathfrak{log}(u \oplus v) = \mathfrak{log}(u) + \mathfrak{log}(v), u, v \in \mathrm{Sym}^+(m)$ and we have $\mathrm{Log}_e Y = \mathfrak{log}(g(B_0^T X)) + \mathfrak{log}\varepsilon$. Based on the bi-invariance of the Log-Euclidean metric, $\mathrm{Log}_e = \mathfrak{log} = \log$ which helps us arrive at the conclusion.

5.3 Proof of Theorem 2 and 3

As pointed out below Theorem S1 and S2, Theorem 2 and 3 in the main manuscript can be seen as corollaries of Theorem S1 and S2. So we only present proofs for Theorem S1 and S2 here.

Recall that our model is

$$\log_{\mu} Y_i = h(B_0^{\top} X_i) + \zeta_i \quad (i = 1, ..., n),$$

where $\log_{\mu} Y_i, \zeta_i \in R^s$ and $h(\cdot) : R^d \to R^s$.

Expand $\log Y_i$ at $B_0^{\top} x$ by Taylor expansion, we have

$$\begin{aligned} \log Y_i =& h(B_0^{\top} x) + \begin{pmatrix} h_1^{(1)}(B_0^{\top} x)^{\top} B_0^{\top}(X_i - x) \\ \vdots \\ h_s^{(1)}(B_0^{\top} x)^{\top} B_0^{\top}(X_i - x) \end{pmatrix} + \frac{1}{2} \begin{pmatrix} (X_i - x)^{\top} B_0 h_1^{(2)}(B_0^{\top} x) B_0^{\top}(X_i - x) \\ \vdots \\ (X_i - x)^{\top} B_0 h_s^{(2)}(B_0^{\top} x) B_0^{\top}(X_i - x) \end{pmatrix} \\ &+ O(||B_0^{\top}(X_i - x)||^3) + \zeta_i. \end{aligned}$$

Here $h_k^{(1)}(B_0^{\top} x)$ is a $d \times 1$ vector and is the coefficient of the first-order term in the Taylor expansion series at $B_0^{\top} x$ of the kth component of $\log Y_i$. Similarly, $h_k^{(2)}(B_0^{\top} x)$ is a $d \times d$ matrix and is the second-order derivative matrix. We collect these derivatives and form into two matrices for later use:

$$h^{(1)}(B_0^\top x) = \begin{pmatrix} h_1^{(1)}(B_0^\top x)^\top \\ \vdots \\ h_s^{(1)}(B_0^\top x)^\top \end{pmatrix}_{s\times d} \qquad h^{(2)}(B_0^\top x) = \begin{pmatrix} h_1^{(2)}(B_0^\top x) \\ \vdots \\ h_s^{(2)}(B_0^\top x) \end{pmatrix}_{sd\times d}.$$

We first provide some lemmas needed for the proofs. We denote $\mu_B(u) = E(X \mid B^\top X = u)$, $w_B(u) = E(XX^\top \mid B^\top X = u)$, $v_B(x) = \mu_B(B^\top x) - x$, $\tilde{w}_B(x) = w_B(B^\top x) - \mu_B(B^\top X)x^\top - x\mu_B^\top(B^\top x) + xx^\top$ and $\bar{w}_B(x) = w_B(B^\top x) - \mu_B(B^\top x)\mu_B^\top(B^\top x) = \tilde{w}_B(x) - v_B(x)v_B^\top(x)$. Additionally, we denote $\delta_n = (n/\log n)^{-1/2}$, $\delta_{dh} = (nh^d/\log n)^{-1/2}$ and $\tau_n = h^2 + \delta_{dh}$. Assume A_n is a matrix. We say $A_n = O(a_n)$ (or $o(a_n)$) for simplicity if all elements in A_n are $O(a_n)$ (or $o(a_n)$) almost surely.

Lemma S2 (Kernel smoother in OPG)**.** *Let*

$$S_n^B(x) = \frac{1}{n}\sum_{i=1}^n K_h(B^\top(X_i - x)) \begin{pmatrix} I_s \\ I_s \otimes (X_i - x) \end{pmatrix} \begin{pmatrix} I_s \\ I_s \otimes (X_i - x) \end{pmatrix}^\top$$

and

$$\begin{pmatrix} a_x \\ \mathrm{vecs}(b_x) \end{pmatrix} = (nS_n^B(x))^{-1} \sum_{i=1}^n K_h(B^\top(X_i - x)) \begin{pmatrix} I_s \\ I_s \otimes (X_i - x) \end{pmatrix} \mathrm{Log}_\mu Y_i.$$

Under assumptions (A1')-(A3'), if $h \to 0$, $\delta_B/h \to 0$ and $nh^d/\log n \to 0$, *then*

$$\begin{aligned}\text{vecs}(b_x) =&(I_s \otimes B_0)\text{vecs}(h^{(1)}(B_0^\top x)) \\ &+ \{nf_B(B^\top x)\}^{-1}\sum_{i=1}^n K_h(B^\top(X_i - x))I_s \otimes \left[\bar{w}_B^+(x)\{X_i - \mu_B(x)\}\right]\zeta_i + O(\epsilon_{dh}),\end{aligned}$$

where $\epsilon_{dh} = h^3 + h\delta_{dh} + h\delta_B$.

Lemma S3 (Kernel smoother in MAVE)**.** *Let*

$$\Sigma_n^B(x) = \frac{1}{n}\sum_{i=1}^n K_h(B^\top(X_i - x))\begin{pmatrix} I_s \\ I_s \otimes B^\top(X_i - x)/h \end{pmatrix}\begin{pmatrix} I_s \\ I_s \otimes B^\top(X_i - x)/h \end{pmatrix}^\top$$

and

$$\begin{pmatrix} a_x \\ \text{vecs}(b_x)h \end{pmatrix} = (\Sigma_n^B(x))^{-1}\sum_{i=1}^n K_h(B^\top(X_i - x))\begin{pmatrix} I_s \\ I_s \otimes B^\top(X_i - x)/h \end{pmatrix}\text{Log}_\mu Y_i.$$

Under assumptions (A1')-(A3'), if $h \to 0$*,* $\delta_B/h \to 0$ *and* $nh^d/\text{log}n \to 0$*, then*

$$a_x = h(B_0^\top x) + h^{(1)}(B_0^\top x)(B_0 - B)^\top v_B(x) + \text{tr}(x)h^2/2 + V_{1n}^B(x) + O(h^3 + h\delta_{dh} + h\delta_B),$$

$$h\text{vecs}(b_x) = h\text{vecs}\{h^{(1)}(B_0^\top x)\} + V_{2n}^B(x) + O(h^3 + h\delta_{dh} + h\delta_B),$$

where

$$\begin{aligned}V_{1n}^B(x) =&\mathcal{E}_{n,1}^B(x) + M_{2n}^B(x)\mathcal{E}_{n,2}^B(x)h, \\ V_{2n}^B(x) =&M_{3n}^B(x)\mathcal{E}_{n,1}^B(x)h + \mathcal{E}_{n,2}^B(x).\end{aligned}$$

Above $\mathrm{tr}(x), M_{kn}^B(x)$ $(k = 1, 2, 3)$ *are matrices whose components are all bounded and continuous functions (explicit forms can be found in the proofs) and*

$$\begin{aligned}\mathcal{E}_{n,1}^B(x) =&\{nf_B(B^\mathsf{T}x)\}^{-1}\sum_{i=1}^n K_h(B^\mathsf{T}(X_i - x))\zeta_i,\\ \mathcal{E}_{n,2}^B(x) =&\{nf_B(B^\mathsf{T}x)\}^{-1}\sum_{i=1}^n K_h(B^\mathsf{T}(X_i - x))I_s \otimes \{B^\mathsf{T}(X_i - x)/h\}\zeta_i.\end{aligned}$$

Lemma S4 (Denominator of MAVE)**.** *Define* $A_{ij} = b_j^\mathsf{T} \otimes (X_i - X_j)$*. Under assumptions (A1')-(A3'), if* $h \to 0$*,* $\delta_B/h \to 0$ *and* $nh^{d+3}/\mathrm{log}n \to 0$*, then*

$$\left\{\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n K_h(B^\mathsf{T}(X_i - X_j))A_{ij}A_{ij}^\mathsf{T}\bigg/\frac{1}{n}\sum_{i=1}^n K_h(B^\mathsf{T}(X_i - X_j))\right\}^{-1}$$

$$=(I_d \otimes B)L_1^B(I_d \otimes B^T) - (I_d \otimes B)L_2^B - L_3^B(I_d \otimes B^T) + W_B^+/2 + O(\tau_n/h + \delta_B),$$

where

$$\begin{aligned}L_1^B =& \left[E\{G(B_0^\mathsf{T}X) \otimes I_d\}\right]^{-1}/h^2,\\ L_2^B =&(L_3^B)^T = \left[E\{G(B_0^TX) \otimes I_d\}\right]^{-1} F^T W_B^+/2,\\ F =&E\left[G(B^\mathsf{T}X) \otimes \left\{v_B(X)\nabla^\mathsf{T} f_B(B^\mathsf{T}X) + f_B(B^\mathsf{T}X)\nabla v_B(X)\right\}/f_B(B^\mathsf{T}X)\right],\\ W_B =&E\left[G(B^\mathsf{T}X) \otimes \{v_B(X)v_B^\mathsf{T}(X)\}\right],\end{aligned}$$

and $G(B^\mathsf{T}X) = h^{(1)}(B^\mathsf{T}X)^\mathsf{T}h^{(1)}(B^\mathsf{T}X)$*.*

Lemma S5 (Numerator of MAVE)**.** *Define* $A_{ij} = b_j^\mathsf{T} \otimes (X_i - X_j)$*. Under* *assumptions (A1')-(A3'), if* $h \to 0$, $\delta_B/h \to 0$ *and* $nh^{d+3}/\log n \to 0$*, then*

$$\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n} K_h(B^\top(X_i - X_j))A_{ij}\{\mathrm{Log}_\mu Y_i - a_j - A_{ij}^\top \mathrm{vec}(B_0)\}/\frac{1}{n}\sum_{i=1}^{n} K_h(B^\top(X_i - X_j))$$

$$= W_B \mathrm{vec}(B - B_0) + \Phi_n(B_0) + O(h^3 + h\delta_{dh} + h\delta_B + \delta_{dh}^2/h + \delta_{dh}\delta_B/h),$$

where

$$W_B = E\left[G(B^\top X) \otimes \{v_B(X)v_B^\top(X)\}\right],$$
$$\Phi_n(B_0) = -\frac{1}{n}\sum_{i=1}^{n}\left\{h^{(1)}(B_0^\top X_i)^\top \otimes v_{B_0}(X_i)\right\}\zeta_i.$$

Proof of Theorem S1: Define $\mathscr{M}$ as the inverse of "vec". In one iteration of the algorithm,

$$\Lambda_{t+1} = [\mathscr{M}\{\mathrm{vec}(B_{t+1})\}]^\top \mathscr{M}\{\mathrm{vec}(B_{t+1})\},$$
$$B_{t+1} = \mathscr{M}\{\mathrm{vec}(B_{t+1})\}\Lambda_{t+1}^{-1/2}.$$

In our algorithms, we use $\mathrm{Log}_{\hat\mu} Y_i$ instead of $\mathrm{Log}_\mu Y_i$. So replacing $\mathrm{Log}_\mu Y_i$ by $\mathrm{Log}_\mu Y_i + (\phi \mathrm{Log}_{\hat\mu} Y_i - \mathrm{Log}_\mu Y_i)$ in Lemma S3 and we have

$$\tilde{a}_x = a_x + D_1(x)\frac{1}{n}\sum_{i=1}^{n} K_h(B^\top(X_i - x))(\phi \mathrm{Log}_{\hat\mu} Y_i - \mathrm{Log}_\mu Y_i),$$
$$h\mathrm{vecs}(\tilde{b}_x) = h\mathrm{vecs}(b_x) + D_2(x)\frac{1}{n}\sum_{i=1}^{n} K_h(B^\top(X_i - x))B^\top(X_i - x)(\phi \mathrm{Log}_{\hat\mu} Y_i - \mathrm{Log}_\mu Y_i),$$

where a_x and $h\mathrm{vecs}(b_x)$ are exactly what have been listed in Lemma S3 and $D_1(x), D_2(x)$ are bounded matrices. As in Lin and Yao (2019), we write $\phi \mathrm{Log}_{\hat\mu} Y_i - \mathrm{Log}_\mu Y_i = \{-H_i(\mu) + \Delta_i(\hat\mu)\}\mathrm{Log}_\mu \hat\mu$ and apply the theoretical result that $||\mathrm{Log}_\mu \hat{\mu}|| = O(n^{-1/2})$ to see under conditions (A1')-(A3') and (C1)-(C6),

$$\begin{aligned}
&||\frac{1}{n}\sum_{i=1}^n K_h(B^\mathsf{T}(X_i - x))(\phi \mathrm{Log}_{\hat{\mu}} Y_i - \mathrm{Log}_\mu Y_i)|| \\
=&||\frac{1}{n}\sum_{i=1}^n K_h(B^\mathsf{T}(X_i - x))\{-H_i(\mu) + \Delta_i(\hat{\mu})\}\mathrm{Log}_\mu \hat{\mu}|| \\
\leq&\mathrm{const}\left\{\frac{1}{n}\sum_{i=1}^n K_h^2(B^\mathsf{T}(X_i - x))\right\}^{1/2} ||\mathrm{Log}_\mu \hat{\mu}|| \\
=&O_P(n^{-1/2}).
\end{aligned}$$

Similarly $n^{-1}\sum_{i=1}^n K_h(B^\mathsf{T}(X_i - x))B^\mathsf{T}(X_i - x)(\phi \mathrm{Log}_{\hat{\mu}} Y_i - \mathrm{Log}_\mu Y_i) = O_P(n^{-1/2})$. Thus we can write $\tilde{a}_x = a_x + R_a(x)$ and $h\mathrm{vecs}(\tilde{b}_x) = \mathrm{vecs}(b_x) + R_b(x)$ where $R_a(x), R_b(x) = O_P(n^{-1/2})$.

Replacing $a_j, b_j, \mathrm{Log}_\mu Y_i$ in Lemma S4 and Lemma S5 by $\tilde{a}_j, \tilde{b}_j, \mathrm{Log}_\mu Y_i + (\phi \mathrm{Log}_{\hat{\mu}} Y_i - \mathrm{Log}_\mu Y_i)$, calculations show the extra $R_a(X_j)$ and $R_b(X_j)$ have no effects on the results of the denominator and the numerator of MAVE. Thus we have

$$\begin{aligned}
\mathrm{vec}(B) =&\mathrm{vec}(B_0) \\
&+ \left\{(I_d \otimes B)L_1^B(I_d \otimes B^\mathsf{T}) - (I_d \otimes B)L_2^B - L_3^B(I_d \otimes B^\mathsf{T}) + W_B^+/2 + O_P(\tau_n/h + \delta_B)\right\} \\
&\times \left\{W_B \mathrm{vec}(B - B_0) + \Phi_n(B_0) + O_P(h^3 + h\delta_{dh} + h\delta_B + \delta_{dh}^2/h + \delta_{dh}\delta_B/h)\right\}.
\end{aligned}$$

Since $(I_d \otimes B)^\mathsf{T} W_B = 0$, $(I_d \otimes B_0)^\mathsf{T}\Phi_n(B_0) = 0$, $I_d \otimes B = I_d \otimes B_0 + O(\delta_B)$, $W_B^+ W_B = I_d \otimes (\tilde{B}\tilde{B}^\mathsf{T})$, $(I_d \otimes B)L_2^B W_B = 0$ and $(I_d \otimes B)L_2^B \Phi_n(B) = 0$, we have

$$\begin{aligned}\text{vec}(B) =&(I_d \otimes B_0)\text{vec}(I_d) + (I_d \otimes \tilde{B}\tilde{B}^{\mathsf{T}})\text{vec}(B - B_0)/2 + W_B^{+}\Phi_n(B_0)/2 \\ &+ (I_d \otimes B_0)/h^2 O_P\{h^3 + \delta_{dh}^2/h + h\delta_{dh} + (\delta_{dh}/h + h)\delta_B\} \\ &+ O_P\{(\delta_{dh}/h + h)\delta_B + h^3 + \delta_{dh}^2 + \delta_{dh}^2/h\}.\end{aligned}$$

For any $B_t \in \mathscr{B} = \{B : ||B - B_0|| \leq \delta_{B_t}\}$, if $\delta_{B_t}/h \to 0$, then $I_d \otimes \tilde{B}_t\tilde{B}_t^{\mathsf{T}} = I_d \otimes \tilde{B}_0\tilde{B}_0^{\mathsf{T}} + O(\delta_{B_t})$. Here and hereafter we use the subscript t to indicate the tth iteration.

$$\begin{aligned}\text{vec}(B_{t+1}) =&(I_d \otimes B_0)\text{vec}(I_d) + (I_d \otimes \tilde{B}_t\tilde{B}_t^{\mathsf{T}})\text{vec}(B_t - B_0)/2 + W_{B_t}^{+}\Phi_n(B_0)/2 \\ &+ (I_d \otimes B_0)/h_t^2 O_P\{h_t^3 + \delta_{dh_t}^2/h_t + h_t\delta_{dh_t} + (\delta_{dh_t}/h_t + h_t)\delta_{B_t}\} \\ &+ O_P\{(\delta_{dh_t}/h_t + h_t)\delta_{B_t} + h_t^3 + \delta_{dh_t}^2 + \delta_{dh_t}^2/h_t\} \\ =&(I_d \otimes B_0)\{\text{vec}(I_d) + O_P(C_t)\} + \Psi_t\text{vec}(B_t - B_0)/2 + W_{B_t}^{+}\Phi_n(B_0)/2 \\ &+ O_P\{\triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\},\end{aligned}$$

where $\triangle_t = h_t^3 + \delta_{dh_t}^2/h_t + h_t\delta_{dh_t}$, $C_t = \{\triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}/h_t^2$, $\Psi_t = I_d \otimes (\tilde{B}_t\tilde{B}_t^{\mathsf{T}}) = I_d \otimes (\tilde{B}_0\tilde{B}_0^{\mathsf{T}}) + O(\delta_{B_t})$.

$$\begin{aligned}B_{t+1} =&\mathscr{M}\,[(I_d \otimes B_0)\{\text{vec}(I_d) + O_P(C_t)\}] + \mathscr{M}\{\Psi_t\text{vec}(B_t - B_0)\}/2 + \mathscr{M}\{W_{B_t}^{+}\Phi_n(B_0)\}/2 \\ &+ O_P\{\triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\} \\ =&B_0\Omega_t + \mathscr{M}\{\Psi\text{vec}(B_t - B_0)\}/2 + \mathscr{M}\{W_{B_t}^{+}\Phi_n(B_0)\}/2 + O_P\{\triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\},\end{aligned}$$

where $\Omega_t = I_d + O_P(C_t)$. Since $\Psi\text{vec}(B_t - B_0) = O(\delta_{B_t})$, $\Phi_n(B_0) = O(\delta_n)$, we have $B_{t+1} = B_0\Omega_t + O_P\{\delta_{B_t} + \delta_n + \triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}$. Then

$$\begin{aligned}\Lambda_{t+1} =&[\mathscr{M}\{\text{vec}(B_{t+1})\}]^\mathsf{T}\mathscr{M}\{\text{vec}(B_{t+1})\}\\ =&(B_0\Omega_t)^\mathsf{T}(B_0\Omega_t) + O_P\{\delta_{B_t} + \delta_n + \triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}\\ =&\Omega_t^2 + O_P\{\delta_{B_t} + \delta_n + \triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\},\end{aligned}$$

and

$$\begin{aligned}B_{t+1} =&\mathscr{M}\{\text{vec}(B_{t+1})\}\Lambda_{t+1}^{-1/2}\\ =&\left[B_0\Omega_t + \mathscr{M}\{\Psi\text{vec}(B_t - B_0)\}/2 + \mathscr{M}\{W_{B_t}^+\Phi_n(B_0)\}/2 + O_P\{\triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}\right]\\ &\times\left[\Omega_t^2 + O_P\{\delta_{B_t} + \delta_n + \triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}\right]^{-1/2}.\end{aligned}$$

We can show that $C(t) = o(1)$ and thus

$$\begin{aligned}B_{t+1} =&B_0 + \mathscr{M}\{\Psi\text{vec}(B_t - B_0)\}/2 + \mathscr{M}\{W_{B_t}^+\Phi_n(B_0)\}/2 + O_P\{\triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}\\ =&B_0 + \mathscr{M}\{\Psi\text{vec}(B_t - B_0)\}/2 + O_P\{\delta_n + \triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}.\end{aligned}$$

Since $\Psi^2 = \Psi$, we have

$$\begin{aligned}\text{vec}(B_{t+1} - B_0) =&\Psi\text{vec}(B_t - B_0)/2 + O_P\{\delta_n + \triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}\\ =&\Psi\text{vec}(B_t - B_0)/2 + O_P(\delta_n + \triangle_t)\\ =&\Psi\left\{\Psi\text{vec}(B_{t-1} - B_0)/2 + O_P(\delta_n + \triangle_{t-1})\right\}/2 + O_P(\delta_n + \triangle_t)\\ =&\Psi\text{vec}(B_{t-1} - B_0)/2^2 + O_P(\delta_n + \triangle_t)\\ =&...\\ =&\Psi\text{vec}(B_1 - B_0)/2^t + O_P(\delta_n + \triangle_t).\end{aligned}$$

Now let $t \to \infty$, we have $\delta_{B_\infty} = O_P(\delta_n + \triangle_\infty) = O_P(\delta_n + h^3 + \delta_{dh}^2/h + h\delta_{dh}) = O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh} + n^{-1/2})$. This is the first part of the conclusions. From $B_{t+1} = B_0 + \mathscr{M}\{\Psi \mathrm{vec}(B_t - B_0)\}/2 + \mathscr{M}\{W_{B_t}^+ \Phi_n(B_0)\}/2 + O_P\{\triangle_t + (h_t + \delta_{dh_t}/h_t)\delta_{B_t}\}$, we know that when $t \to \infty$,

$$B_\infty - B_0 = \mathscr{M}\{\Psi \mathrm{vec}(B_\infty - B_0)\}/2 + \mathscr{M}\{W_{B_\infty}^+ \Phi_n(B_0)\}/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh}). \tag{S5.10}$$

Then multiplying (S5.10) by $B_0^\top$ from left, we have

$$\begin{aligned}
&B_0^\top B_\infty - B_0^\top B_0\\
=&B_0^\top B_\infty - I_d\\
=&B_0^\top \mathscr{M}\{\Psi \mathrm{vec}(B_\infty - B_0)\}/2 + B_0^\top \mathscr{M}\{W_{B_\infty}^+ \Phi_n(B_0)\}/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&\mathscr{M}\{(I_d \otimes B_0^\top)\Psi \mathrm{vec}(B_\infty - B_0)\}/2 + \mathscr{M}\{(I_d \otimes B_0^\top) W_{B_\infty}^+ \Phi_n(B_0)\}/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&\mathscr{M}\{I_d \otimes B_0^\top \tilde{B}_0 \tilde{B}_0{}^\top \mathrm{vec}(B_\infty - B_0)\}/2\\
&+ \mathscr{M}\{(I_d \otimes B_0^\top)(I_d \otimes B_\infty)[(I_d \otimes B_\infty^\top) W_{B_\infty} (I_d \otimes B_\infty)]^{-1} (I_d \otimes B_\infty^\top)\Phi_n(B_0)\}/2\\
&+ O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&O_P(\delta_{B_\infty} + \delta_n + h^3 + \delta_{dh}^2/h + h\delta_{dh}) = O_P(\delta_n + h^3 + \delta_{dh}^2/h + h\delta_{dh}).
\end{aligned}$$

From above calculations we can also see that $B_0^\top B_\infty = I_d + O_P(\delta_n + h^3 + \delta_{dh}^2/h + h\delta_{dh})$, which also implies $B_\infty^\top B_0 = I_d + O_P(\delta_n + h^3 + \delta_{dh}^2/h + h\delta_{dh})$. Multiply (S5.10) from right by $B_\infty^T B_0$, we have

$$B_\infty B_\infty^\top B_0 - B_0 = \mathscr{M}\{\Psi \mathrm{vec}(B_\infty - B_0)\}/2 + \mathscr{M}\{W_{B_\infty}^+ \Phi_n(B_0)\}/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh}).$$

Then we have

$$
\begin{aligned}
&\mathrm{vec}(B_\infty B_\infty^\top B_0) - \mathrm{vec}(B_0)\\
=&\Psi \mathrm{vec}(B_\infty - B_0)/2 + W_{B_\infty}^{+}\Phi_n(B_0)/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&\Psi\left\{\Psi \mathrm{vec}(B_\infty - B_0)/2 + W_{B_\infty}^{+}\Phi_n(B_0)/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\right\}/2\\
&+ W_{B_\infty}^{+}\Phi_n(B_0)/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&\Psi \mathrm{vec}(B_\infty - B_0)/2^2 + \Psi W_{B_\infty}^{+}\Phi_n(B_0)/2^2 + W_{B_\infty}^{+}\Phi_n(B_0)/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&\Psi \mathrm{vec}(B_\infty - B_0)/2^2 + W_{B_\infty}^{+}\Phi_n(B_0)/2^2 + W_{B_\infty}^{+}\Phi_n(B_0)/2 + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&...\\
=&\Psi \mathrm{vec}(B_\infty - B_0)/2^t + W_{B_\infty}^{+}\Phi_n(B_0)/(2^t + ... + 2) + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&W_{B_0}^{+}\Phi_n(B_0) + O_P(\delta_{B_\infty}\delta_n + h^3 + \delta_{dh}^2/h + h\delta_{dh})\\
=&W_{B_0}^{+}\Phi_n(B_0) + O_P(h^3 + \delta_{dh}^2/h + h\delta_{dh}).
\end{aligned}
$$

where the fourth equality comes from $\Psi W_{B_\infty}^{+} = W_{B_\infty}^{+} + O(\delta_{B_\infty})$.

If $h^3 + \delta_{dh}^2/h + h\delta_{dh} = o(n^{-1/2})$, $\mathrm{vec}(B_\infty B_\infty^\top B_0) - \mathrm{vec}(B_0) = W_{B_0}^{+}\Phi_n(B_0) + o_P(n^{-1/2})$. Here $W_{B_0}^{+}$ is a non-random matrix and $\Phi_n(B_0) = -n^{-1}\sum_{i=1}^{n}\left\{h^{(1)}(B_0^\top X_i)^\top \otimes v_B(X_i)\right\}\zeta_i$. We can calculate $E\{\Phi_n(B_0)\} = 0$ and $\mathrm{var}\{\Phi_n(B_0)\} = \mathrm{var}\left[\left\{h^{(1)}(B_0^\top X)^\top \otimes v_{B_0}(X)\right\}\zeta\right] = \Sigma_0$. So we have

$$
\sqrt{n}\left\{\mathrm{vec}(B_\infty B_\infty^\top B_0) - \mathrm{vec}(B_0)\right\} \xrightarrow{d} N(0, W_{B_0}^{+}\Sigma_0 W_{B_0}^{+}).
$$

Proof of Theorem S2: Replacing $\mathrm{Log}_\mu Y_i$ by $\mathrm{Log}_\mu Y_i + (\phi \mathrm{Log}_{\hat\mu} Y_i - \mathrm{Log}_\mu Y_i)$ in Lemma S2, we have

$$
\begin{aligned}
b_{xk} =& B_0 h_k^{(1)}(B_0^\intercal x) + \{n f_B(B^\intercal x)\}^{-1} \sum_{i=1}^n K_h(B^\intercal(X_i - x)) \bar w_B^+(x)(X_i - \mu_B(B^\intercal x))\zeta_{ik} \\
&+ R(x) + O(\epsilon_{dh}) \\
=& (B_0, \tilde B_0) \begin{pmatrix} h_k^{(1)}(B_0^\intercal x) \\ O(\epsilon_{dh}) \end{pmatrix} + \{n f_B(B^\intercal x)\}^{-1} \sum_{i=1}^n K_h(B^\intercal(X_i - x)) \bar w_B^+(x)(X_i - \mu_B(B^\intercal x))\zeta_{ik} \\
&+ R(x) + O(\epsilon_{dh}).
\end{aligned}
$$

where $R(x) = n^{-1}\sum_{i=1}^n K_h(B^\intercal(X_i - x))\{\gamma_B(x) + \bar w_B^+(x)\}(X_i - x)(\hat y_{ik} - y_{ik})$ and $\hat y_{ik} - y_{ik}$ is the kth component in $\phi \mathrm{Log}_{\hat\mu} Y_i - \mathrm{Log}_\mu Y_i$. Similar reasoning as in the proof of Theorem S1 gives $R(x) = O_P(n^{-1/2})$. Denote $\{n f_B(B^\intercal x)\}^{-1}\sum_{i=1}^n K_h(B^\intercal(X_i - x))\bar w_B^+(x)(X_i - \mu_B(B^\intercal x))\zeta_{ik} = \mathcal{E}_k^B(x)$, then

$$
\begin{aligned}
b_{xk} b_{xk}^\intercal =& (B_0, \tilde B_0) \begin{pmatrix} h_k^{(1)}(B_0^\intercal x) h_{kl}^{(1)}(B_0^\intercal x)^\intercal & O(\epsilon_{dh}) \\ O(\epsilon_{dh}) & O(\epsilon_{dh}^2) \end{pmatrix} \begin{pmatrix} B_0^\intercal \\ \tilde B_0^\intercal \end{pmatrix} \\
&+ B_0 h_k^{(1)}(B_0^\intercal x)\mathcal{E}_k^B(x)^\intercal + \mathcal{E}_k^B(x) h_k^{(1)}(B_0^\intercal x)^\intercal B_0^\intercal + O_P(n^{-1/2}) + O(\epsilon_{dh}\delta_{dh}).
\end{aligned}
$$

We use the subscript t in letters to indicate the tth iteration and denote

b_{xk} in the tth iteration as $\hat{b}_k^{(t)}(x)$. Then

$$\begin{aligned}
\hat{\Sigma}_{(t+1)} =& \frac{1}{n}\sum_{j=1}^{n} \hat{b}_k^{(t)}(X_j)^\top \hat{b}_k^{(t)}(X_j) \\
=& (B_0, \tilde{B}_0) \begin{pmatrix} n^{-1}\sum_{j=1}^{n} h^{(1)}(B_0^\top X_j)^\top h^{(1)}(B_0^\top X_j) & O(\epsilon_{dh_t}) \\ O(\epsilon_{dh_t}) & O(\epsilon_{dh_t}^2) \end{pmatrix} \begin{pmatrix} B_0^\top \\ \tilde{B}_0^\top \end{pmatrix} \\
&+ \frac{1}{n}\sum_{j=1}^{n}\left\{S_j^{(t)} + (S_j^{(t)})^\top\right\} + O(n^{-1/2}) + O(\epsilon_{dh_t}\delta_{dh_t}) \\
=& (B_0, \tilde{B}_0) \begin{pmatrix} \Lambda_n^{(t)} & O(\epsilon_{dh_t}) \\ O(\epsilon_{dh_t}) & O(\epsilon_{dh_t}^2) \end{pmatrix} \begin{pmatrix} B_0^\top \\ \tilde{B}_0^\top \end{pmatrix} + \frac{1}{n}\sum_{j=1}^{n}\left\{S_j^{(t)} + (S_j^{(t)})^\top\right\} \\
&+ O_P(n^{-1/2}) + O(\epsilon_{dh_t}\delta_{dh_t}),
\end{aligned}$$

where

$$S_j^{(t)} = B_0 \sum_{k=1}^{s} h_k^{(1)}(B_0^\top X_j)\mathcal{E}_k^B(X_j)^\top.$$

Now we calculate $n^{-1}\sum_{j=1}^{n}\left\{S_j^{(t)} + (S_j^{(t)})^\top\right\}$.

$$\begin{aligned}
&\frac{1}{n}\sum_{j=1}^{n}\left\{S_j^{(t)} + (S_j^{(t)})^\top\right\} \\
=& (B_0, \tilde{B}_0)\left[\begin{pmatrix} B_0^\top \\ \tilde{B}_0^\top \end{pmatrix} \frac{1}{n}\sum_{j=1}^{n}\left\{S_j^{(t)} + (S_j^{(t)})^\top\right\}(B_0, \tilde{B}_0)\right]\begin{pmatrix} B_0^\top \\ \tilde{B}_0^\top \end{pmatrix} \\
=& (B_0, \tilde{B}_0)\begin{pmatrix} 0 & C_{12,n}^{(t)} \\ (C_{12,n}^{(t)})^\top & 0 \end{pmatrix}\begin{pmatrix} B_0^\top \\ \tilde{B}_0^\top \end{pmatrix} + O(\delta_{dh_t}\delta_{B_t})
\end{aligned}$$

In above calculation, we write $S_j = B_0 A_j$ where A_j is the summation shown in the definition of S_j. Note that $B^\top \mu_B(B^\top x) = 0$, which implies

$B^\top \bar{w}_B^+(x) = 0$ and further $B^\top \mathcal{E}_k^B(x) = 0$, $B^\top A_j = 0$. Additionally, one can show that $\mathcal{E}_{kl}^B(x) = O(\delta_{dh})$. These relationships help us derive the above equation where

$$C_{12,n} = \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{s} h_k^{(1)}(B_0^\top X_j)\mathcal{E}_k^B(X_j)^\top \tilde{B}_0 = M_n^{(t)}\tilde{B}_0.$$

We can show that the matrix $M_n = \sum_{k=1}^s n^{-1}\sum_{i=1}^n n^{-1}\sum_{j=1}^n f_B^{-1}(B^\top X_j)K_h(B^\top(X_i - X_j))h_k^{(1)}(B_0^\top X_j)(X_i - \mu_B(B^\top X_j)^\top \bar{w}_B^+(x)\zeta_{ik} = O(\delta_n)$. So

$$\begin{aligned}
\hat{\Sigma}_{(t+1)} =& B_0\Lambda_n B_0^\top + O(\epsilon_{dh_t}) + B_0 M_n^{(t)}\tilde{B}_0\tilde{B}_0^\top + \tilde{B}_0\tilde{B}_0^\top (M_n^{(t)})^\top B_0^\top \\
& + O(\delta_{dh_t}\delta_{B_t}) + O(\epsilon_{dh_t}\delta_{dh_t}) + O_P(n^{-1/2}) \\
=& B_0\Lambda_n B_0^\top + B_0 M_n^{(t)}\tilde{B}_0\tilde{B}_0^\top + \tilde{B}_0\tilde{B}_0^\top (M_n^{(t)})^\top B_0^\top + O_P(h_t^3 + h_t\delta_{dh_t} + h_t\delta_{B_t}) \\
=& B_0\Lambda_n^{(t)} B_0^\top + O_P(\delta_n + h_t^3 + h_t\delta_{dh_t} + h_t\delta_{B_t}).
\end{aligned} \tag{S5.11}$$

By the same argument used by Xia (2007), we have

$$B_{t+1}B_{t+1}^\top - B_0B_0^\top = O_P(\delta_n + h_t^3 + h_t\delta_{dh_t} + h_t\delta_{B_t}).$$

Let $t \to \infty$, we have $B_\infty B_\infty^\top - B_0 B_0^\top = O_P(h^3 + h\delta_{dh} + \delta_n)$ and this is the first part of the conclusions.

By (S5.11), we have

$$\hat{\Sigma}_\infty = (B_0 + \eta_n)\Lambda_n(B_0 + \eta_n)^\top + O_P(h^3 + h\delta_{dh}),$$

where $\eta_n = (M_n^{(t)})^\mathsf{T} \Lambda_n^{-1} = O(\delta_n)$ and $(B_0 + \eta_n)^\mathsf{T}(B_0 + \eta_n) = I_d + O(\delta_n^2)$. By the same reasoning as Xia (2007), we have

$$\begin{aligned}\hat{B}_\infty \hat{B}_\infty^\mathsf{T} - B_0 B_0^\mathsf{T} =& (B_0 + \eta_n)(B_0 + \eta_n)^\mathsf{T} - B_0 B_0^\mathsf{T} + O_P(h^3 + h\delta_{dh})\\ =& B_0 \eta_n^\mathsf{T} + \eta_n B_0^\mathsf{T} + O_P(h^3 + h\delta_{dh}).\end{aligned}$$

Since $B_\infty^\mathsf{T} \eta_n = 0$ and $B_0 - B_\infty = \delta_{B_\infty} = O_P(h^3 + h\delta_{dh} + \delta_n)$, we have $\hat{B}_\infty \hat{B}_\infty^\mathsf{T} B_0 - B_0 = \eta_n + B_0 \eta_n^T B_0 + O_P(h^3 + h\delta_{dh}) = \eta_n + O_P(h^3 + h\delta_{dh})$. When $h^3 + h\delta_{dh} = o(n^{-1/2})$, all left is to calculate the variance of $\mathrm{vec}(\eta_n)$.

5.4 Proof of Theorem 4: DOPG

We first prove the conclusion about DOPG. Before our proof, we introduce some notations. Under the log-Euclidean metric, our model $Y = g(B_0^{\mathsf{T}} X) \oplus \varepsilon$ can be transformed into $\log Y = \log g(B_0^T X) + \log \varepsilon$ which we rewrite as

$$\log Y = h(B_0^{\mathsf{T}} X) + \zeta.$$

Expand $\log Y_i$ at x by Taylor expansion, we have

$$\begin{aligned}\log Y_i =& h(x) + \begin{pmatrix} h_{11}^{(1)}(x)^T(X_i - x) & \dots & h_{1m}^{(1)}(x)^T(X_i - x) \\ \vdots & & \vdots \\ h_{m1}^{(1)}(x)^T(X_i - x) & \dots & h_{mm}^{(1)}(x)^T(X_i - x) \end{pmatrix} \\ &+ \frac{1}{2} \begin{pmatrix} (X_i - x)^T h_{11}^{(2)}(x)(X_i - x) & \dots & (X_i - x)^T h_{1m}^{(2)}(x)(X_i - x) \\ \vdots & & \vdots \\ (X_i - x)^T h_{m1}^{(2)}(x)(X_i - x) & \dots & (X_i - x)^T h_{mm}^{(2)}(x)(X_i - x) \end{pmatrix} \\ &+ O(|(X_i - x)|^3) + \zeta_i.\end{aligned}$$

That is, we expand every element of the $m \times m$ matrix $\log Y_i$ at x. Above $h_{kl}^{(1)}(x)$ is a $p \times 1$ vector and is the coefficient of the first-order term in the Taylor expansion series at x of the (k, l)-th element of $\log Y_i$. Similarly, $h_{kl}^{(2)}(x)$ is a $p \times p$ matrix and is the second-order derivative matrix.

For further simplicity, we denote the (k, l)-th element of the $m \times m$ symmetric matrix $h(x)$ as $a_{kl}(x)$, $1 \leq l, k \leq m$ and denote $h_{kl}^{(1)}(x)$ as $b_{kl}(x)$.

Proof of DOPG: By the same argument of Cai et al. (2022), we have almost surely for $1 \le l \le k \le m$

$$\sup_{x\in D} |\hat{a}_{kl}(x) - a_{kl}(x)| = O\left(\left(\frac{p_n \log n}{n h_n^{|\alpha|}}\right)^{1/2} + \omega_n^2\right)$$

$$\sup_{x\in D} |\hat{b}_{kl}^{[j]}(x) - b_{kl}^{[j]}(x)| = O\left(\left(\frac{p_n \log n}{n h_n^{|\alpha|+2\alpha_j}}\right)^{1/2} + \frac{\omega_n^2}{h_n^{\alpha_j}}\right), j = 1, ..., p.$$

By condition (1), we denote the support of X by D which is a compact set in R^p. Then for every $x \in D$,

$$\hat{b}_{kl}(x) = b_{kl}(x) + \Delta b_{kl}^n(x) = B_0 h_{kl}^{(1)}(B_0^\top x) + \Delta b_{kl}^n(x),$$

where $\Delta b_{kl}^n(x) = ((b_{kl}^n)^{[1]}(x), ..., (b_{kl}^n)^{[p]}(x))^\top$ is a p-dimensional vector. If $\alpha_j = 0$, we have $(b_{kl}^n)^{[j]}(x) = O_P(\sqrt{p_n/n})$. If $\alpha_j \neq 0$, we have $(b_{kl}^n)^{[j]}(x) = O(c_n^{[j]})$. Let $(B_0, \tilde{B}_0)$ be a $p\times p$ orthogonal matrix such that $(B_0, \tilde{B}_0)(B_0, \tilde{B}_0)^\top = I_p$ and $(B_0, \tilde{B}_0)^\top (B_0, \tilde{B}_0) = I_p$. We can write

$$\hat{b}_{kl}(x) = (B_0, \tilde{B}_0)\begin{pmatrix} h_{kl}^{(1)}(B_0^\top x) + B_0^\top \Delta b_{kl}^n(x) \\ \tilde{B}_0^\top \Delta b_{kl}^n(x) \end{pmatrix}.$$

By the algorithm of OPG,

$$\hat{\Sigma} = \frac{1}{n}\sum_{j=1}^{n}\sum_{l\le k} \hat{b}_{kl}\hat{b}_{kl}^\top = (B_0, \tilde{B}_0)\begin{pmatrix} \Lambda_n^{(1)} & \Lambda_n^{(2)} \\ \Lambda_n^{(3)} & \Lambda_n^{(4)} \end{pmatrix}(B_0, \tilde{B}_0)^\top, \tag{S5.12}$$

where

$$\Lambda_n^{(1)} = \sum_{l\le k}\frac{1}{n}\sum_{j=1}^{n}\{h_{kl}^{(1)}(B_0^\top X_j)(h_{kl}^{(1)}(B_0^\top X_j))^\top + 2B_0^\top \Delta b_{kl}^n(X_j)(h_{kl}^{(1)}(B_0^\top X_j))^\top + B_0^\top \Delta b_{kl}^n(X_j)(\Delta b_{kl}^n(X_j))^\top B_0\},$$

$$\Lambda_n^{(2)} = (\Lambda_n^{(3)})^\mathsf{T} = \sum_{l\le k}\frac{1}{n}\sum_{j=1}^{n}\{h_{kl}^{(1)}(B_0^\mathsf{T}X_j)(\Delta b_{kl}^n(X_j))^\mathsf{T}\tilde{B}_0 + B_0^\mathsf{T}\Delta b_{kl}^n(X_j)(\Delta b_{kl}^n(X_j))^\mathsf{T}\tilde{B}_0\},$$

$$\Lambda_n^{(4)} = \sum_{l\le k}\frac{1}{n}\sum_{j=1}^{n}\tilde{B}_0^\mathsf{T}\Delta b_{kl}^n(X_j)(\Delta b_{kl}^n(X_j))^\mathsf{T}\tilde{B}_0.$$

Note that the p-dimensional vector β_k in B_0 satisfies $\|\beta_k\|_2^2 = 1$ and the remainder $\Delta b_{kl}^n(x)$ satisfies $\|\Delta b_{kl}^n(x)\|^2 = \sum_{j=1}^p\{(b_{kl}^n)^{[1]}\}^2 = \sigma_n^2$ with $\sigma_n = \{\sum_{\alpha_j\neq 0}(c_n^{[j]})^2 + \sum_{\alpha_j=0}p_n/n\}^{1/2}$. Then it can be shown that for a p-dimensional unit vector β,

$$\beta\Delta b_{kl}^n(x) \le \|\beta\|\cdot\|\Delta b_{kl}^n(x)\| \le \sigma.$$

This together with assumption (A2) results in $B_0^\mathsf{T}\Delta b_{kl}^n(X_j)(h_{kl}^{(1)}(B_0^\mathsf{T}X_j))^\mathsf{T} = O_P(\sigma_n)$. By the same discussion, we have $B_0^\mathsf{T}\Delta b_{kl}^n(X_j)(\Delta b_{kl}^n(X_j))^\mathsf{T}B_0 = O_P(\sigma_n^2)$.

By the central limit theorem, it is easy to see that

$$\Lambda_n^{(1)} = E\{h_{kl}^{(1)}(B_0^\mathsf{T}X)(h_{kl}^{(1)}(B_0^\mathsf{T}X))^\mathsf{T}\} + O_P(1/\sqrt{n}) + O_P(\sigma_n) := \Sigma_0 + O_P(\sigma_n + 1/\sqrt{n}).$$

Using Lemma 6 in Cai et al. (2022), we conclude that the eigenvalues of $\Lambda_n^{(1)}$ is asymptotically converge to eigenvalues of Σ_0 in probability with order $O(d(\sigma_n + 1/\sqrt{n}))$.

Similarly, we obtain that $\Lambda_n^{(2)} = (\Lambda_n^{(3)})^\top = O_P(\sigma_n + 1/\sqrt{n})$ and $\Lambda_n^{(4)} = O_P(\sigma_n^2 + 1/\sqrt{n})$ with eigenvalues of $\Lambda_n^{(4)}$ being $O(\sigma_n^2 + (p_n - d)/\sqrt{n})$. Let $\lambda_1 \geq ... \geq \lambda_p$ be the eigenvalues of $\hat{\Sigma}$ and $\hat{\beta}_1, ..., \hat{\beta}_p$ be their corresponding eigenvectors. By the Eigenvalue Interlacing Theorem and assumption 3 and 4, we have $\min\{\lambda_1, ..., \lambda_d\} > c > 0$ and $\max\{\lambda_{d+1}, ..., \lambda_p\} = O(\sigma_n^2 + (p_n - d)/\sqrt{n} = o(1))$. Therefore, the top-$d$ eigenvalues can be distinguished from others asymptotically. By (S5.12) we have in probability

$$\hat{\Sigma} = B_0 \Sigma_0 B_0^\top + O_P(\sigma_n + 1/\sqrt{n}). \tag{S5.13}$$

Let $\hat{B}_{\text{DOPG}} = (\hat{\beta}_1, ..., \hat{\beta}_d)$. Using (ii) of Lemma 6 in Cai et al. (2022) and under assumptions (B1) and (B3), we obtained

$$\hat{B}_{\text{DOPG}} \hat{B}_{\text{DOPG}}^\top - B_0 B_0^\top = O_P(p_n \sigma_n).$$

5.5 Proof of Theorem 4: DMAVE

The target function of DMAVE is equivalent to

$$\sum_{j=1}^{n} \sum_{i=1}^{n} \text{tr}\Big[\{\text{vecs}(a_j) + I_q \otimes (X_{ij}^\top B)\text{vecss}(b_j) - \text{vecs}(\log Y_i)\}^{\otimes 2}\Big] K_h(X_{ij}; \hat{\alpha})$$

which can be rewritten as

$$\sum_{1 \leq l \leq k \leq m} \sum_{j=1}^{n} \sum_{i=1}^{n} \{y_{kl}^i - a_{kl}(X_j) - c_{kl}^\top(X_j) B^\top X_{ij}\}^2 K_h(X_{ij}; \hat{\alpha})$$

where y^i_{kl} is the (k,l)-th element of the matrix $\log Y_i$. So it is safe to only inspect $a_{kl}(x)$ and $c_{kl}(x)$ for some $1 \leq l \leq k \leq m$. To simplify our notations, we now fix (k,l) and write y^i_{kl} as Y_i, $a_{kl}(x)$ as $a(x)$ and $c_{kl}(x)$ as $c(x)$. From our model $\log Y_i = \log g(B_0^\top X_i) + \log\varepsilon_i$, we have $y^i_{kl} = (\log g(B_0^\top X_i))_{kl} + \zeta_{kl}$ and we denote it as $Y_i = h(B_0^\top X_i) + \epsilon_i$ since (k,l) is fixed. That is, we now only need to consider the model $Y = m(X) + \epsilon = h(B_0^\top X) + \epsilon$ where $Y \in R$ and $X \in R^p$. And the DMAVE for $Y = h(B_0^\top X) + \epsilon$ estimates B_0 by minimizing the following objective function

$$\frac{1}{n^2}\sum_{j=1}^{n}\sum_{i=1}^{n}\{Y_i - a_j - c_j^\top B^\top(X_i - X_j)\}\}^2 K_h(X_i - X_j;\hat{\alpha}),$$

Before proof, we introduce the following notations as Cai et al. (2022) did. A local approximation of $m(z)$ by a polynomial of total order r is given by

$$m(z) \approx \sum_{0\leq|k|\leq r}\frac{1}{k!}(D^k m)(z)(z-x)^k,$$

where

$$k = (k^{[1]},...,k^{[p]}), k! = k^{[1]}! \times ... \times k^{[p]}!, |k| = \sum_{j=1}^{p} k^{[j]};$$

$$x^k = (x^{[1]})^{k^{[1]}} \times ... \times (x^{[p]})^{k^{[p]}}, \sum_{0\leq|k|\leq r} = \sum_{j=0}^{r}\sum_{k^{[1]}=0}^{j}\cdots\sum_{\substack{k^{[p]}=0\\ k^{[1]}+...+k^{[p]}=j}}^{j};$$

and

$$(D^k m)(x) = \frac{\partial m(y)}{\partial(y^{[1]})^{k^{[1]}}\cdots\partial(y^{[p]})^{k^{[p]}}}\Big|_{y=x}.$$

With samples $(X_i, Y_i), i = 1, ..., n$, the problem of local linear regression can be written as minimizing

$$\sum_{i=1}^{n}\{Y_i - \sum_{0\le|k|\le 1} b_k(x)(X_i - x)^k\}^2 K_h(X_i - x; \hat{\alpha}) \tag{S5.14}$$

w.r.t. $b_k(x)$. Denote the minimizer of (S5.14) by $\hat{b}_k(x)$, then we have estimation $\widehat{(D^k m)}(x) = k!\hat{b}_k(x)$. The minimization of (S5.14) leads to the set of equations

$$t_j(x) = \sum_{0\le|k|\le 1} h^{k\cdot\alpha}\hat{b}_k(x)s_{j+k}(x), \quad 0 \le |j| \le 1, \tag{S5.15}$$

where

$$\begin{aligned} t_j(x) &= \frac{1}{n}\sum_{i=1}^{n} Y_i[Z_i(h;\alpha) - z(h;\alpha)]^j K_h(X_i - x; \hat{\alpha}), \\ s_j(x) &= \frac{1}{n}\sum_{i=1}^{n}[Z_i(h;\alpha) - z(h;\alpha)]^j K_h(X_i - x; \hat{\alpha}), \end{aligned} \tag{S5.16}$$

with

$$z(h;\alpha) = (\frac{x^{[1]}}{h^{\alpha_1}}, ..., \frac{x^{[p]}}{h^{\alpha_p}}).$$

Define $\tau(x) = (\tau_0(x), ..., \tau_p(x))^\mathsf{T}$, where $\tau_0(x) = t_{(0,...,0)}(x)$, $\tau_1(x) = t_{(1,...,0)}(x)$,...,$\tau_p(x) = t_{(0,...,1)}(x)$. Arranging $h^{k\cdot\alpha}\hat{b}_k(x), 0 \le |k| \le 1$ in the same order, we can obtain $\hat{\theta}$ as an estimator of column vector $\theta(x) = (\theta_0(x), ..., \theta_p(x))^\mathsf{T} := (m(x), h^{\alpha_1}m^{[1]}(x), ..., h^{\alpha_p}m^{[p]}(x))^\mathsf{T}$. Then define $S(x)$

as

$$S(x) = \begin{pmatrix} s_{(0,0,\dots,0)}(x) & s_{(1,0,\dots,0)}(x) & \cdots & s_{(0,0,\dots,1)}(x) \\ s_{(1,0,\dots,0)}(x) & s_{(2,0,\dots,0)}(x) & \cdots & s_{(1,0,\dots,1)}(x) \\ \vdots & \vdots & \ddots & \vdots \\ s_{(0,0,\dots,1)}(x) & s_{(1,0,\dots,1)}(x) & \cdots & s_{(0,0,\dots,2)}(x) \end{pmatrix}.$$

Then the set of equations in (S5.15) can be written in matrix as

$$\tau(x) = S(x)\hat{\theta}(x).$$

By assumption (B4), $S(x)$ is invertible and we can henceforth write

$$\hat{\theta}(x) = S^{-1}(x)\tau(x),$$

as the solution of the set of equations (S5.15).

A fundamental decomposition for the error $\hat{\theta} - \theta$ is provided next. Firstly, let

$$t_j^*(x) = \frac{1}{n}\sum_{i=1}^{n}[Y_i - m(X_i)][Z_i(h;\alpha) - z(h;\alpha)]^j K_h(X_i - x;\hat{\alpha}),$$

and we have

$$t_j(x) - t_j^*(x) = \frac{1}{n}\sum_{i=1}^{n} m(X_i)[Z_i(h;\alpha) - z(h;\alpha)]^j K_h(X_i - x;\hat{\alpha}). \quad \text{(S5.17)}$$

The Taylor series of $m(X_i)$ at x with a mean-value form of remainder is

$$m(X_i) = \sum_{0\le|k|\le 1} \frac{1}{k!}(D^k m)(x)(X_i - x)^k + \sum_{|k|=2}(D^k m)(\tilde{x}_i)(X_i - x)^k, \quad \text{(S5.18)}$$

where $\tilde{x}_i$ is a point between x and X_i. Substituting (S5.18) and (S5.16) to (S5.17), we find

$$t_j(x) - t_j^*(x) = \sum_{0 \le |k| \le 1} \frac{1}{k!} h^{k \cdot \alpha} (D^k m)(x) s_{j+k}(x) + e_j(x),$$

where

$$e_j(x) = \frac{1}{n} \sum_{|k|=2} \frac{h^{k \cdot \alpha}}{k!} \sum_{i=1}^{n} (D^k m)(\tilde{x}_i)[Z_i(h;\alpha) - z(h;\alpha)]^j K_h(X_i - x; \hat{\alpha}).$$

By (S5.15) and $(D^k m)(x) = k! b_k(x)$, we obtain

$$t_j^*(x) = \sum_{0 \le |k| \le 1} h^{k \cdot \alpha} [\hat{b}_k(x) - b_k(x)] s_{j+k}(x) - e_j(x). \tag{S5.19}$$

For $0 \le |j| \le 1$, using the same arrangement as for $\tau(x)$, we can define the $(p+1)$ column vector $\tau^*(x)$ and $e(x)$ as

$$\tau^*(x) = \begin{pmatrix} t^*_{(0,\ldots,0)}(x) \\ t^*_{(1,\ldots,0)}(x) \\ \vdots \\ t^*_{(0,\ldots,1)}(x) \end{pmatrix}, e(x) = \begin{pmatrix} e_{(0,\ldots,0)}(x) \\ e_{(1,\ldots,0)}(x) \\ \vdots \\ e_{(0,\ldots,1)}(x) \end{pmatrix}.$$

The vector form of (S5.19) is

$$\tau^*(x) = S(x)(\hat{\theta}(x) - \theta(x)) - e(x).$$

Thus

$$\hat{\theta}(x) - \theta(x) = S^{-1}(x)\tau^*(x) + S^{-1}(x)e(x).$$

We next prove the following two lemmas.

Lemma S6 (Kernel smoother in DMAVE)**.** *Let*

$$S(B^\top x) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{ix};\hat{\alpha}) \begin{pmatrix} 1 \\ B^\top X_{ix} \end{pmatrix} \begin{pmatrix} 1 \\ B^\top X_{ix} \end{pmatrix}^\top$$

and

$$\begin{pmatrix} \hat{a}(x) \\ \hat{c}(x) \end{pmatrix} = \{nS(B^\top x)\}^{-1} \sum_{i=1}^{n} K_h(X_{ix};\hat{\alpha}) \begin{pmatrix} 1 \\ B^\top X_{ix} \end{pmatrix} Y_i.$$

Suppose assumptions (B1)-(B5), (A2) and (A4) hold, then we have almost surely

$$\sup_{x\in D} |\hat{a}(x) - a(x)| = O\left(\left(\frac{p_n \log n}{n h_n^{|\alpha|}}\right)^{1/2} + \omega_n^2 \right)$$
$$\sup_{x\in D} |\hat{c}^{[j]}(x) - c^{[j]}(x)| = O\left(\left(\frac{p_n \log n}{n h_n^{|\alpha|+2\alpha_j}}\right)^{1/2} + \frac{\omega_n^2}{h_n^{\alpha_j}} \right), j = 1, ..., p.$$

Proof. Note that $\omega_n = \sum_{\alpha_j \neq 0} h^{\alpha_j} \|\beta^{[j]}\|_2$ and $\sum_{j=1}^{p} \|\beta^{[j]}\|_2^2 = d$. By the Cauchy-Schwarz inequality: $(\sum_{i=1}^{n} x_i)^2 \leq n \sum_{i=1}^{n} x_i^2$ and $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ for $a, b \geq 0$, it can be easily seen that

$$\begin{aligned}
\Big\{ \sum_{j=1}^{p} (h^{\alpha_j} \|\beta^{[j]}\|_2)^2 \Big\}^{1/2} &= \Big\{ \sum_{\alpha_j \neq 0} (h^{\alpha_j} \|\beta^{[j]}\|_2)^2 + \sum_{\alpha_j = 0} \|\beta^{[j]}\|_2^2 \Big\}^{1/2} \\
&\leq \Big\{ \sum_{\alpha_j \neq 0} (h^{\alpha_j} \|\beta^{[j]}\|_2)^2 \Big\}^{1/2} + \Big\{ \sum_{\alpha_j = 0} \|\beta^{[j]}\|_2^2 \Big\}^{1/2} \\
&\leq \sum_{\alpha_j \neq 0} h^{\alpha_j} \|\beta^{[j]}\|_2 + \Big\{ \sum_{j=1}^{p} \|\beta^{[j]}\|_2^2 \Big\}^{1/2} \\
&\leq \omega_n + d^{1/2},
\end{aligned} \tag{S5.20}$$

and

$$\sum_{j=1}^{p} h^{\alpha_j}\|\beta^{[j]}\|_2 = \sum_{\alpha_j\neq 0} h^{\alpha_j}\|\beta^{[j]}\|_2 + \sum_{\alpha_j=0}\|\beta^{[j]}\|_2$$
$$\leq \omega_n + \sum_{j=1}^{p}\|\beta^{[j]}\|_2 \leq \omega_n + \left\{p\sum_{j=1}^{p}\|\beta^{[j]}\|_2^2\right\}^{1/2} \qquad \text{(S5.21)}$$
$$\leq \omega_n + d^{1/2}p_n^{1/2}.$$

The above two inequalities are widely employed in the following analysis.

By the Taylor's expansion of $h(B_0^\top X_i)$ at x and $Y_i = h(B_0^\top X_i) + \epsilon_i$, we have

$$Y_i = h(B_0^\top x) + (Dh)^\top(B_0^\top x)B_0^\top X_{ix} + \sum_{|k|=2}\frac{1}{k!}(D^k m)(\tilde{x}_i)X_{ix}^k + \epsilon_i$$
$$= \begin{pmatrix} 1 \\ B_0^\top X_{ix}\end{pmatrix}^\top \begin{pmatrix} h(B_0^\top x) \\ (Dh)(B_0^\top x)\end{pmatrix} + \sum_{|k|=2}\frac{1}{k!}h^{\alpha\cdot k}(D^k m)(\tilde{x}_i)[X_i(h;\alpha) - x(h;\alpha)]^k + \epsilon_i,$$

where $(Dh)(B_0^\top x)$ is the derivative vector of $h(\cdot)$ at $B_0^\top x$.

Then $(\hat{a}(x), \hat{c}(x))$ can be written as

$$\begin{pmatrix} \hat{a}(x) \\ \hat{c}(x)\end{pmatrix} = \begin{pmatrix} h(B_0^\top x) \\ (Dh)(B_0^\top x)\end{pmatrix} + \{nS(B^\top x)\}^{-1}\sum_{i=1}^{n} K_h(X_{ix};\hat{\alpha})\begin{pmatrix} 1 \\ B^\top X_{ix}\end{pmatrix}\epsilon_i$$
$$+ \{nS(B^\top x)\}^{-1}\sum_{i=1}^{n} K_h(X_{ix};\hat{\alpha})\begin{pmatrix} 1 \\ B^\top X_{ix}\end{pmatrix}\sum_{|k|=2}\frac{1}{k!}h^{\alpha\cdot k}(D^k m)(\tilde{x}_i)[X_i(h;\alpha) - x(h;\alpha)]^k$$
$$:= \begin{pmatrix} h(B_0^\top x) \\ (Dh)(B_0^\top x)\end{pmatrix} + A_1 + A_2.$$

Firstly the numerator in A_1:

$$\sup_{x\in D}\Big|\frac{1}{n}\sum_{i=1}^{n}K_h(X_{ix};\hat{\alpha})\begin{pmatrix}1\\ B^{\mathsf{T}}X_{ix}\end{pmatrix}\epsilon_i\Big|$$

$$=\sup_{x\in D}\Big|\begin{pmatrix}1 & 0 & \cdots & 0\\ 0_d & h^{\alpha_1}(\beta^{[1]})^{\mathsf{T}} & \cdots & h^{\alpha_p}(\beta^{[p]})^{\mathsf{T}}\end{pmatrix}\frac{1}{n}\sum_{i=1}^{n}K_h(X_{ix};\hat{\alpha})\begin{pmatrix}1\\ X_{ix}^{[1]}/h^{\alpha_1}\\ \vdots\\ X_{ix}^{[p]}/h^{\alpha_p}\end{pmatrix}\Big|$$

$$=\begin{pmatrix}\sup_{x\in D}|t^*_{(0,0,\ldots,0)}(x)|\\ \sup_{x\in D}|\sum_{j=1}^{p}h^{\alpha_j}(\beta^{[j]})^{\mathsf{T}}t^*_j(x)|\end{pmatrix}$$

where $t^*_j(x)=n^{-1}\sum_{i=1}^{n}K_h(X_{ix};\hat{\alpha})[X_i(h;\alpha)-x(h;\alpha)]^j\epsilon_i$ for each j with $0\le|j|\le 1$. By an argument similar to Lemma 5 in Cai et al. (2021), we have $\sup_{x\in D}|t^*_j(x)|=O[\{p_n\mathrm{log}n/(nh_n^{|\alpha|})\}^{1/2}]$. And

$$\begin{aligned}
\sup_{x\in D}|\sum_{j=1}^{p}h^{\alpha_j}(\beta^{[j]})^{\mathsf{T}}t^*_j(x)| &\le \sup_{x\in D}|\sum_{j=1}^{p}h^{\alpha_j}(\beta^{[j]})^{\mathsf{T}}|\cdot\sup_{x\in D}|\sum_{j=1}^{p}t^*_j(x)|\\
&\le \sup_{x\in D}\|\sum_{j=1}^{p}h^{\alpha_j}(\beta^{[j]})^{\mathsf{T}}\|_2\cdot\sup_{x\in D}|\sum_{j=1}^{p}t^*_j(x)|\cdot 1_d\\
&= \sup_{x\in D}\{\sum_{j=1}^{p}\sum_{s=1}^{p}h^{\alpha_j+\alpha_s}\beta^{[j]}(\beta^{[l]})^{\mathsf{T}}\}^{1/2}\cdot\sup_{x\in D}|\sum_{j=1}^{p}t^*_j(x)|\cdot 1_d\\
&\le \sup_{x\in D}\{\sum_{j=1}^{p}\sum_{s=1}^{p}h^{\alpha_j+\alpha_s}\|\beta^{[j]}\|_2\|\beta^{[l]}\|_2\}^{1/2}\cdot\sup_{x\in D}|\sum_{j=1}^{p}t^*_j(x)|\cdot 1_d\\
&= \sup_{x\in D}\sum_{j=1}^{p}h^{\alpha_j}\|\beta^{[j]}\|_2\cdot\sup_{x\in D}|\sum_{j=1}^{p}t^*_j(x)|\cdot 1_d.
\end{aligned}$$

This together with (S5.20) results in

$$\sup_{x\in D}|\sum_{j=1}^{p} h^{\alpha_j}(\beta^{[j]})^\top t_j^*(x)| = O\Big((\frac{p_n^2\log n}{nh_n^{|\alpha|}})^{1/2}\Big).$$

And this implies for the numerator in A_1:

$$\sup_{x\in D}\Big|\frac{1}{n}\sum_{i=1}^{n} K_h(X_{ix};\hat{\alpha})\begin{pmatrix}1\\ B^\top X_{ix}\end{pmatrix}\epsilon_i\Big| = \begin{pmatrix}O((p_n\log n/(nh_n^{|\alpha|}))^{1/2})\\ O((p_n^2\log n/(nh_n^{|\alpha|}))^{1/2})\end{pmatrix}.$$

Consequently we obtain

$$\sup_{x\in D}|A_1| = \begin{pmatrix}O((p_n\log n/(nh_n^{|\alpha|}))^{1/2})\\ O((p_n^2\log n/(nh_n^{|\alpha|}))^{1/2})\end{pmatrix}.$$

Secondly, the numerator in A_2 is

$$\frac{1}{n}\sum_{i=1}^{n} K_h(X_{ix};\hat{\alpha})\begin{pmatrix}1\\ B^\top X_{ix}\end{pmatrix}\sum_{|k|=2}\frac{1}{k!}h^{\alpha\cdot k}(D^k m)(\tilde{x}_i)[X_i(h;\alpha)-x(h;\alpha)]^k$$

$$= \begin{pmatrix}e_{(0,0,\dots,0)}(x)\\ \sum_{j=1}^{p} h^{\alpha_j}(\beta^{[j]})^\top e_j(x)\end{pmatrix},$$

where $e_j(x) = \sum_{|k|=2}\frac{1}{k!}h^{\alpha\cdot k}n^{-1}\sum_{i=1}^{n}(D^k m)(\tilde{x}_i)[X_i(h;\alpha)-x(h;\alpha)]^{k+j}$ with $\sup_{x\in D}|e_j(x)| = O(\omega_n^2)$ for each $0\le|j|\le 1$. Follow the same steps used for A_1, we get

$$\sup_{x\in D}|A_2| = \begin{pmatrix}O(\omega_n^2)\\ O(p_n^{1/2}\omega_n^2)\end{pmatrix}.$$

In conclusion, we have

$$\begin{pmatrix} \sup_{x\in D}|\hat{a}(x)-h(B_0^\top x)| \\ \sup_{x\in D}|\hat{c}(x)-(Dh)(B_0^\top x)| \end{pmatrix} = \sup_{x\in D}|A_1+A_2| \le \sup_{x\in D}|A_1| + \sup_{x\in D}|A_2|$$

$$= \begin{pmatrix} O((p_n \log n/(nh_n^{|\alpha|}))^{1/2}+\omega_n^2) \\ O((p_n^2 \log n/(nh_n^{|\alpha|}))^{1/2}+p_n^{1/2}\omega_n^2) \end{pmatrix}.$$

□

Lemma S7. *Suppose assumptions (B1)-(B5), (A2) and (A4) hold, we have*

$$\hat{\beta}^{[j]}-\beta^{[j]} = O(c_n^{[j]}), \quad j=1,...,p,$$

where $c_n^{[j]} = p_n \log n/(nh_n^{|\alpha|+2\alpha_j}))^{1/2}+\omega_n^2/h_n^{\alpha_j}$.

Proof. Observe that $(Dh)^\top(B_0^\top x)B_0^\top X_{ix} = (B_0^\top X_{ix})^\top(Dh)(B_0^\top x) = \sum_{j=1}^p \beta^{[j]}X_{ix}^{[j]}(Dh)(B_0^\top x)$. Let $\beta^{[1]}=\beta_{(1,0,...,0)},...,\beta^{[p]}=\beta_{(0,0,...,1)}$. For a p-dimensional vector k satisfying $|k|=\sum_{j=1}^p k_j=1$, we further have $(Dh)^\top(B_0^\top x)B_0^\top X_{ix} = \sum_{|k|=1}\beta_k X_{ix}^k(Dh)(B_0^\top x)$. Then the expression of Y_i can be written as

$$\begin{aligned} Y_i &= h(B_0^\top x)+\sum_{|k|=1}\beta_k X_{ix}^k(Dh)(B_0^\top x) \\ &\quad + \sum_{|k|=2}\frac{1}{k!}h^{\alpha\cdot k}D^k m(\tilde{x}_i)[X_i(h;\alpha)-x(h;\alpha)]^k+\epsilon_i \\ &= a(x)+\sum_{|k|=1}h^{\alpha\cdot k}\beta_k[X_i(h;\alpha)-x(h;\alpha)]^k c(x) \\ &\quad + \sum_{|k|=2}\frac{1}{k!}h^{\alpha\cdot k}D^k m(\tilde{x}_i)[X_i(h;\alpha)-x(h;\alpha)]^k+\epsilon_i \end{aligned}$$

Given $(a(X_j), c(X_j)), 1 \le l \le k \le m, j = 1, ..., n$, the analytic solution of B equals to

$$\hat{\beta}_k = \underset{\beta_k \beta_k^{\mathsf{T}}=1}{\arg\min} \frac{1}{n^2} \sum_{j=1}^{n} \sum_{i=1}^{n} \{Y_i - a(X_j) - \sum_{|k|=1} h^{\alpha \cdot k} \beta_k [X_i(h;\alpha) - X_j(h;\alpha)]^k c(X_j)\}^2 K_h(X_{ij}; \hat{\alpha}) \tag{S5.22}$$

where $k \in R^p$ with $|k| = 1$. Denote

$$L(\beta_j) = \frac{1}{n} \sum_{i=1}^{n} \{Y_i - a(x) - \sum_{|k|=1} h^{\alpha \cdot k} \beta_k [X_i(h;\alpha) - x(h;\alpha)]^k c(x)\}^2 K_h(X_{ix}; \hat{\alpha}), |j| = 1.$$

The minimizer $\hat{\beta}_j$ satisfying $DL(\hat{\beta}_j) = 0$, which is equivalent to

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} \{Y_i - a(x)\}[X_i(h;\alpha) - x(h;\alpha)]^j c(x) K_h(X_{ix}; \hat{\alpha}) \\ =&\frac{1}{n} \sum_{i=1}^{n} \sum_{|k|=1} h^{\alpha \cdot k} \hat{\beta}_k [X_i(h;\alpha) - x(h;\alpha)]^k c(x) [X_i(h;\alpha) - x(h;\alpha)]^j c(x) K_h(X_{ix}; \hat{\alpha}). \end{aligned}$$

Recall the definitions of $t_j(x), t_j^*(x), e_j(x), s_j(x)$ with $|j| = 1$. By the decomposition of Y_i, the above equation can be written in the form of a row vector as

$$\sum_{l \le k} \sum_{|k|=1} h^{\alpha \cdot k} (\hat{\beta}_k - \beta_k) s_{k+j}(x) c(x) c^{\mathsf{T}}(x) = \sum_{l \le k} \Big\{ t_j^*(x) c^{\mathsf{T}}(x) + e_j(x) c^{\mathsf{T}}(x) \Big\}.$$

Its matrix form is

$$\sum_{l \le k} S_{-1,-1}(x) \mathrm{diag}(h^{\alpha_1}, ..., h^{\alpha_p})(\hat{B} - B) c(x) c^{\mathsf{T}}(x) = \sum_{l \le k} \{\tau_{-1}^*(x) c^{\mathsf{T}}(x) + e_{-1}^*(x) c^{\mathsf{T}}(x)\},$$

where $\tau_{-1}^*(x)$ and $e_{-1}^*(x)$ are p-dimensional vectors after removing the first element in $\tau(x)$ and $e(x)$, respectively , and $S_{-1,-1}(x) \in R^{p \times p}$ is a $(1, 1)$-

minor of matrix $S(x)$. Replacing x with X_j, we have

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}S_{-1,-1}(X_j)\mathrm{diag}(h^{\alpha_1},...,h^{\alpha_p})(\hat{B}-B)c(X_j)c^{\mathsf{T}}(X_j)\\=&\frac{1}{n}\sum_{j=1}^{n}\{\tau^*_{-1}(X_j)c^{\mathsf{T}}(X_j)+e^*_{-1}(X_j)c^{\mathsf{T}}(X_j)\}\end{aligned}$$

Applying the vectorization operator vec$(\cdot)$ with properties: $\mathrm{vec}(A_1+A_2)=\mathrm{vec}(A_1)+\mathrm{vec}(A_2)$ and $\mathrm{vec}(A_1BA_2)=(A_2^{\mathsf{T}}\otimes A_1)\cdot\mathrm{vec}(B)$, we have

$$\begin{aligned}&\frac{1}{n}\sum_{j=1}^{n}\Big[\{c(X_j)c^{\mathsf{T}}(X_j)\}\otimes S_{-1,-1}(X_j)\Big]\mathrm{vec}\{\mathrm{diag}(h^{\alpha_1},...,h^{\alpha_p})(\hat{B}-B)\}\\=&\frac{1}{n}\sum_{j=1}^{n}\mathrm{vec}\Big[\{\tau^*_{-1}(X_j)+e^*_{-1}(X_j)\}c^{\mathsf{T}}(X_j)\Big].\end{aligned}$$

By Lemma S6, we have $c(x)=(Dh)(B_0^{\mathsf{T}}x)+O_P((p_n^2\mathrm{log}n/(nh_n^{|\alpha|}))^{1/2}+p_n^{1/2}\omega_n^2)$. Simple calculation yields that

$$\begin{aligned}&\Big[\frac{1}{n}\sum_{j=1}^{n}\{c(X_j)c^{\mathsf{T}}(X_j)\}\otimes S_{-1,-1}(X_j)\Big]^{-1}\\\leq&\frac{1}{n}\sum_{j=1}^{n}\Big[\{(Dh)(B_0^{\mathsf{T}}X_j)(Dh)^{\mathsf{T}}(B_0^{\mathsf{T}}X_j)\}\otimes S_{-1,-1}(X_j)\Big]^{-1}+O_P((p_n^2\mathrm{log}n/(nh_n^{|\alpha|}))^{1/2}+p_n^{1/2}\omega_n^2)\\=&E\Big[\{(Dh)(B_0^{\mathsf{T}}X)(Dh)^{\mathsf{T}}(B_0^{\mathsf{T}}X)\}^{-1}\otimes S_{-1,-1}^{-1}(X)\Big]+O_P((p_n^2\mathrm{log}n/(nh_n^{|\alpha|}))^{1/2}+p_n^{1/2}\omega_n^2),\end{aligned}$$

where the inequality comes from the Jensen's inequality applied to the convex function $x\to x^{-1}$ for $x\geq 0$.

By $\sup_{x\in D}|t_j^*(x)|=O((p_n\mathrm{log}n/(nh_n^{|\alpha|}))^{1/2})$ and $\sup_{x\in D}|e_j(x)|=O(\omega_n^2)$, it can be seen that

$$\frac{1}{n}\sum_{j=1}^{n}\mathrm{vec}\Big[\{\tau^*_{-1}(X_j)+e^*_{-1}(X_j)\}c^{\mathsf{T}}(X_j)\Big]=O((p_n\mathrm{log}n/(nh_n^{|\alpha|}))^{1/2}+\omega_n^2).$$

From

$$\begin{aligned}&\text{vec}\{\text{diag}(h^{\alpha_1},...,h^{\alpha_p})(\hat{B}-B)\}\\=&\Big[\sum_{l\le k}\frac{1}{n}\sum_{j=1}^{n}\{c(X_j)c^{\mathsf{T}}(X_j)\}\otimes S_{-1,-1}(X_j)\Big]^{-1}\cdot\sum_{l\le k}\frac{1}{n}\sum_{j=1}^{n}\text{vec}\Big[\{\tau^*_{-1}(X_j)+e^*_{-1}(X_j)\}c^{\mathsf{T}}(X_j)\Big],\end{aligned}$$

we know that for each $j=1,...,p$,

$$h^{\alpha_j}(\hat{\beta}^{[j]}-\beta^{[j]})=O((p_n\text{log}n/(nh_n^{|\alpha|}))^{1/2}+\omega_n^2),$$

which completes the proof.

□

Now we are ready to prove results of DMAVE.

Proof of DMAVE: Let $\hat{B}_{\text{DMAVE}}=((\hat{\beta}^{[1]})^{\mathsf{T}},...,(\hat{\beta}^{[p]})^{\mathsf{T}})^{\mathsf{T}}$ be the DMAVE estimation of B_0 and $\hat{B}_{\text{DMAVE}}\hat{B}^{\mathsf{T}}_{\text{DMAVE}}=(\hat{\beta}^{[j]}(\hat{\beta}^{[l]})^{\mathsf{T}})_{j,l=1,...,p}$. By the Cauchy-Schwarz inequality, we can show that

$$\begin{aligned}(\hat{\beta}^{[j]}-\beta^{[j]})\cdot(\beta^{[l]})^{\mathsf{T}}&=\|(\hat{\beta}^{[j]}-\beta^{[j]})\cdot(\beta^{[l]})\|_2\le\|\hat{\beta}^{[j]}-\beta^{[j]}\|_2\cdot\|\beta^{[l]}\|_2\\&\le\sum_{j=1}^{p}\|\hat{\beta}^{[j]}-\beta^{[j]}\|_2\cdot\sum_{l=1}^{p}\|\beta^{[l]}\|_2\\&\le\{p\sum_{j=1}^{p}\|\hat{\beta}^{[j]}-\beta^{[j]}\|_2^2\}^{1/2}\cdot\{p\sum_{l=1}^{p}\|\beta^{[l]}\|_2^2\}^{1/2}\\&=p_n\{\sum_{j=1}^{p}\|\hat{\beta}^{[j]}-\beta^{[j]}\|_2^2\}^{1/2}\cdot d^{1/2}.\end{aligned}$$

Note that for $\alpha_j\neq 0$ we have $\hat{\beta}^{[j]}-\beta^{[j]}=O(c_n^{[j]})$ where $c_n^{[j]}=p_n\text{log}n/(nh_n^{|\alpha|+2\alpha_j}))^{1/2}+\omega_n^2/h_n^{\alpha_j}$ and for $\alpha_j=0$ we have $\hat{\beta}^{[j]}-\beta^{[j]}=O_P(\sqrt{p_n/n})$. Hence $\sum_{j=1}^{p}\|\hat{\beta}^{[j]}-$

$\beta^{[j]}\|_2^2 = \sum_{\alpha_j \neq 0} (c_n^{[j]})^2 + \sum_{\alpha_j = 0} p_n/n = \sigma_n^2$. This implies

$$(\hat{\beta}^{[j]} - \beta^{[j]}) \cdot (\beta^{[l]})^{\mathsf{T}} = O_P(p_n \sigma_n), \quad j, l = 1, ..., p.$$

Similarly $(\hat{\beta}^{[j]} - \beta^{[j]})(\hat{\beta}^{[j]} - \beta^{[j]})^{\mathsf{T}} = O_P(p_n \sigma_n^2)$. Then it follows that

$$\hat{\beta}^{[j]}(\hat{\beta}^{[l]})^{\mathsf{T}} = \beta^{[j]}(\beta^{[l]})^{\mathsf{T}} + O_P(p_n \sigma_n), \quad j, l = 1, ..., p,$$

which completes the proof.

5.6 Proof of Lemmas

Here we collect proofs of Lemmas S1-S5 in the previous section.

Proof of Lemma S1: Denote the solution to S5.7) as $\hat{\delta}_{-j}(X_j)$. We have

$$\begin{aligned}
&\text{vecs}\{\hat{\delta}_{-j}(X_j)\} \\
=&\left\{\sum_{i\neq j} w_{ij}\begin{pmatrix} I_q & I_q\otimes(X_i-X_j)^{\mathsf{T}} \\ I_q\otimes(X_i-X_j) & I_q\otimes(X_i-X_j)(X_i-X_j)^{\mathsf{T}}\end{pmatrix}\right\}^{-1} \\
&\cdot\sum_{i\neq j} w_{ij}\begin{pmatrix} I_q \\ I_q\otimes(X_i-X_j)\end{pmatrix}\text{vecs}(\log Y_i) \\
\triangleq& A_{2j}^{-1}\cdot\frac{1}{n}\sum_{i\neq j} w_{ij}\begin{pmatrix} I_q \\ I_q\otimes(X_i-X_j)\end{pmatrix}\text{vecs}(\log Y_i),
\end{aligned}$$

where $w_{ij}=K_h(X_i-X_j)$. Apply Taylor expansion at X_j:

$$\log Y_i = g(X_j)+g^{(1)}(X_j)\cdot I_m\otimes(X_i-X_j)+R_{ij}+\varepsilon_i,$$

where R_{ij} is the remaining term of the Taylor series. Thus

$$\begin{aligned}
&\text{vecs}\{\hat{\delta}_{-j}(X_j)\} \\
=&\begin{pmatrix}\text{vecs}\{g(X_j)\} \\ \text{vecs}\{g^{(1)}(X_j)\}\end{pmatrix}+A_{2j}^{-1}\cdot\frac{1}{n}\sum_{i\neq j} w_{ij}\begin{pmatrix} I_q \\ I_q\otimes(X_i-X_j)\end{pmatrix}\{\text{vecs}(R_{ij})+\text{vecs}(\varepsilon_i)\} \\
\triangleq&\delta(X_j)+A_{2j}^{-1}A_{1j}.
\end{aligned}$$

Denote $\hat{f}_j=n^{-1}\sum_{i\neq j}w_{ij}$, $B_{1j}=n^{-1}\sum_{i\neq j}w_{ij}(X_i-X_j)$, $B_{2j}=n^{-1}\sum_{i\neq j}w_{ij}(X_i-X_j)(X_i-X_j)^{\mathsf{T}}$. Then $e_1^{\mathsf{T}}A_{2j}^{-1}=(I_q\otimes(\hat{f}_j+C_{1j}),I_q\otimes(-C_{2j}))$ where e_1 is a

$(q+qd)$-dimensional vector with the first q elements equaling 1 and the remaining equaling 0, $C_{1j} = \hat{f}_j^{-2} B_{1j}^T (B_{2j} - \hat{f}_j^{-1} B_{1i} B_{1i}^\mathsf{T})$ and $C_{2j} = \hat{f}_j^{-1} B_{1j}^\mathsf{T} (B_{2j} - \hat{f}_j^{-1} B_{1i} B_{1i}^\mathsf{T})^{-1}$. Then $\text{vecs}(\hat{a}_j) = e_1^\mathsf{T} \text{vecs}\{\hat{\delta}_{-j}(X_j)\} = \text{vecs}\{g(X_j)\} + e_1^\mathsf{T} A_{2j}^{-1} A_{1j}$.

Denote $D_j^{kl} = n^{-1} \sum_{i \neq j} w_{ij} (R_{ij}^{kl} + \varepsilon_j^{kl}) / \hat{f}_j,\ M_j^{kl} = n^{-1} \sum_{i \neq j} w_{ij} (R_{ij}^{kl} + \varepsilon_j^{kl}) \{C_{1j} - C_{2j}(X_i - X_j)\}$ $(1 \leq l \leq k \leq m)$, then

$$\text{vecs}(\hat{a}_j) = \begin{pmatrix} g_{11}(X_j) + D_j^{11} + M_j^{11} \\ g_{21}(X_j) + D_j^{21} + M_j^{21} \\ \vdots \\ g_{mm}(X_j) + D_j^{mm} + M_j^{mm} \end{pmatrix}.$$

Thus

$$\begin{aligned}
&\frac{1}{n} \sum_{j=1}^{n} ||\text{vecs}(\log Y_j) - \text{vecs}(\hat{a}_j)||^2 \\
=&\frac{1}{n} \sum_{j=1}^{n} ||\text{vecs}(g(X_j)) + \text{vecs}(\varepsilon_j) - \text{vecs}(\hat{a}_j)||^2 \\
=&\frac{1}{n} \sum_{j=1}^{n} \left\| \begin{pmatrix} g_{11}(X_j) + \varepsilon_j^{11} - (g_{11}(X_j) + D_j^{11} + M_j^{11}) \\ g_{21}(X_j) + \varepsilon_j^{21} - (g_{21}(X_j) + D_j^{21} + M_j^{21}) \\ \vdots \\ g_{mm}(X_j) + \varepsilon_j^{mm} - (g_{mm}(X_j) + D_j^{mm} + M_j^{mm}) \end{pmatrix} \right\|^2 \\
=&\frac{1}{n} \sum_{j=1}^{n} \sum_{l \leq k} (\varepsilon_j^{kl} - D_j^{kl} - M_j^{kl})^2 \\
=&\sum_{l \leq k} \frac{1}{n} \sum_{j=1}^{n} (\varepsilon_j^{kl} - D_j^{kl} - M_j^{kl})^2,
\end{aligned}$$

which is just a summation of the CV values of the case in Li and Racine (2004) where Y is a scalar. Directly applying the results of Li and Racine (2004), we reach the conclusion.

Proof of Lemma S2: The proof is similar to that of Lemma S3 and more details can be found in the proof of Lemma S3.

First

$$S_n^B(x) = f_B(B^\top x)\begin{pmatrix} I_s & I_s \otimes v_B^\top(x) \\ I_s \otimes v_B^\top(x) & I_s \otimes \tilde{w}_B(x) \end{pmatrix} + O(h^2 + \delta_{dh}),$$

where $\tilde{w}_B(x) = w_B(x) - \mu_B(x)x^\top - x\mu_B(x)^\top + xx^\top$.

And we can calculate

$$\{S_n^B(x)\}^{-1} = f_B^{-1}(B^\top x)\begin{pmatrix} I_s \otimes C_B(x) & I_s \otimes \gamma_B^\top(x) \\ I_s \otimes \gamma_B(x) & I_s \otimes \bar{w}_B^+(x) \end{pmatrix} + O(h^2 + \delta_{dh}),$$

where

$$\begin{aligned} C_B(x) =& \{1 - v_B^\top(x)\tilde{w}_B^+(x)v_B(x)\}^{-1}, \\ \gamma_B(x) =& -\bar{w}_B^+(x)v_B(x). \end{aligned}$$

Expanding $\mathrm{Log}_\mu Y_i$ at $B_0^\top x$ by Taylor expansion, we have

$$\begin{aligned} \mathrm{Log}_\mu Y_i =& \left\{h(B_0^\top x) + h^{(1)}(B_0^T x)B_0^\top(X_i - x)\right\} \\ &+ \frac{1}{2} I_s \otimes \{(X_i - x)^\top B\} h^{(2)}(B_0^\top x) B^\top (X_i - x) + \zeta_i \\ &+ O\left\{||B_0^\top(X_i - x)||^3 + ||B_0^\top(X_i - x)|| \times ||X_i - x||\delta_B + ||X_i - x||^2\delta_B^2\right\} \\ =&(1) + (2) + (3) + (4), \end{aligned}$$

and we denote

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} K_h(B^{\mathsf{T}}(X_i - x)) \begin{pmatrix} I_s \\ I_s \otimes (X_i - x) \end{pmatrix} \mathrm{Log}_{\mu} Y_i \\
=&\frac{1}{n}\sum_{i=1}^{n} K_h(B^{\mathsf{T}}(X_i - x)) \begin{pmatrix} I_s \\ I_s \otimes (X_i - x) \end{pmatrix} \{(1) + (2) + (3) + (4)\} \\
\triangleq&(1') + (2') + (3') + (4').
\end{aligned}
$$

We can show

$$
\{S_n^B(x)\}^{-1} \times (1') = \begin{pmatrix} h(B_0^{\mathsf{T}} x) \\ (I_s \otimes B_0)\mathrm{vecs}\{h^{(1)}(B_0^{\mathsf{T}} x)\} \end{pmatrix};
$$

$$
\{S_n^B(x)\}^{-1} \times (2') = \begin{pmatrix} I_s \otimes \{C_B(x) + \gamma_B^{\mathsf{T}}(x) v_B(x)\} \\ I_s \otimes \{\gamma_B(x) + \bar{w}_B^{+}(x) v_B(x)\} \end{pmatrix} \mathrm{tr}(x) h^2 + O(h^4 + h^2 \delta_{dh}),
$$

where

$$
\mathrm{tr}(x) = \Big(\mathrm{tr}\{h_1^{(2)}(B_0^{\mathsf{T}} x)\}, ..., \mathrm{tr}\{h_s^{(2)}(B_0^{\mathsf{T}} x)\}\Big)^{\mathsf{T}};
$$

$$
\begin{aligned}
&\{S_n^B(x)\}^{-1} \times (3') \\
=&\{n f_B(B^{\mathsf{T}} x)\}^{-1} \sum_{i=1}^{n} K_h(B^{\mathsf{T}}(X_i - x)) \begin{pmatrix} I_s \otimes \{C_B(x) + \gamma_B^{\mathsf{T}}(x)(X_i - x)\} \\ I_s \otimes \{\gamma_B(x) + \bar{w}_B^{+}(x)(X_i - x)\} \end{pmatrix} \zeta_i \\
&+ O\{(h^2 + \delta_{dh})\delta_{dh}\};
\end{aligned}
$$

and

$$\{S_n^B(x)\}^{-1} \times (4') = O(h^3 + \delta_{dh} + h\delta_B + \delta_B^2).$$

Putting everything together, we reach the conclusion.

Proof of Lemma S3: First we can calculate that

$$\Sigma_n^B(x) = \begin{pmatrix} f_B(B^\top x) I_s & h I_s \otimes \nabla^\top f_B(B^\top x) \\ h I_s \otimes \nabla f_B(B^\top x) & f_B(B^\top x) I_{sd} \end{pmatrix} + O(h^2 + \delta_{dh}),$$

and

$$\begin{aligned}\left\{\Sigma_n^B(x)\right\}^{-1} =& f_B^{-1}(B^\top x)\left\{ I - h f_B^{-1}(B^\top x) \begin{pmatrix} 0 & I_s \otimes \nabla^\top f_B(B^\top x) \\ I_s \otimes \nabla f_B(B^\top x) & 0 \end{pmatrix}\right\} \\ &+ O(h^2 + \delta_{dh}).\end{aligned}$$

Next expand $\mathrm{Log}_\mu Y_i$ at $B_0^\top x$ by Taylor expansion, we have

$$\begin{aligned}\mathrm{Log} Y_i =& \left\{h(B_0^\top x) + h^{(1)}(B_0^\top x) B^\top (X_i - x)\right\} + h^{(1)}(B_0^\top x)(B_0 - B)^\top (X_i - x) \\ &+ \frac{1}{2} I_s \otimes \{(X_i - x)^\top B\} h^{(2)}(B_0^\top x) B^\top (X_i - x) + \zeta_i \\ &+ \frac{1}{2}\Big[I_s \otimes \{(X_i - x)^T (B_0 - B)\} h^{(2)}(B_0^\top x) B^\top (X_i - x) \\ &\quad + I_s \otimes \{(X_i - x)^\top B\} h^{(2)}(B_0^T x)(B_0 - B)^\top (X_i - x) \\ &\quad + I_s \otimes \{(X_i - x)^\top (B_0 - B)\} h^{(2)}(B_0^\top x)(B_0 - B)^\top (X_i - x)\Big] \\ &+ O(||B_0^\top (X_i - x)||^3) \\ =& (1) + (2) + (3) + (4) + (5) + (6).\end{aligned}$$

Then we calculate

$$\frac{1}{n}\sum_{i=1}^{n} K_h(B^{\mathsf{T}}(X_i - x)) \begin{pmatrix} I_s \\ I_s \otimes B^{\mathsf{T}}(X_i - x)/h \end{pmatrix} \mathrm{Log}_{\mu} Y_i$$

$$= \frac{1}{n}\sum_{i=1}^{n} K_h(B^{T}(X_i - x)) \begin{pmatrix} I_s \\ I_s \otimes B^{\mathsf{T}}(X_i - x)/h \end{pmatrix} \{(1) + (2) + (3) + (4) + (5) + (6)\}$$

$$= (1') + (2') + (3') + (4') + (5') + (6').$$

By the definition of $\Sigma_n^B(x)$, we have

$$\{\Sigma_n^B(x)\}^{-1} \times (1') = \begin{pmatrix} h(B_0^{\mathsf{T}} x) \\ h\mathrm{vecs}\{h^{(1)}(B_0^{\mathsf{T}} x)\} \end{pmatrix}.$$

Next we turn to $(2')$.

$$(2') = \frac{1}{n}\sum_{i=1}^{n} K_h(B^{\mathsf{T}}(X_i - x)) \begin{pmatrix} h^{(1)}(B_0^{\mathsf{T}} x)(B_0 - B)^{\mathsf{T}}(X_i - x) \\ I_s \otimes \{B^{\mathsf{T}}(X_i - x)/h\} h^{(1)}(B_0^{\mathsf{T}} x)(B_0 - B)^{\mathsf{T}}(X_i - x) \end{pmatrix}$$

For example, we can calculate the first term in the upper half of the vector $(2')$:

$$\frac{1}{n}\sum_{i=1}^{n} K_h(B^{\mathsf{T}}(X_i - x)) h_1^{(1)}(B_0^{\mathsf{T}} x)^{\mathsf{T}}(B_0 - B)^{\mathsf{T}}(X_i - x)$$

$$= f_B(B^{\mathsf{T}} x) h_1^{(1)}(B_0^{\mathsf{T}} x)(B_0 - B)^{\mathsf{T}} v_B(x) + O(h^2 \delta_B + \delta_{dh}),$$

and the first term in the lower half of $(2')$:

$$\frac{1}{n}\sum_{i=1}^{n} K_h(B^{\mathsf{T}}(X_i - x))\{B^{\mathsf{T}}(X_i - x)/h\} h_1^{(1)}(B_0^{\mathsf{T}} x)^{\mathsf{T}}(B_0 - B)^{\mathsf{T}}(X_i - x)$$

$$= h_1^{(1)}(B_0^{\mathsf{T}} x)^{\mathsf{T}}(B_0 - B)^{\mathsf{T}} v_B(x) \nabla^{\mathsf{T}} f_B(B^{\mathsf{T}} x) h + O(\delta_{dh}) = O(h\delta_B).$$

So

$$(2') = \begin{pmatrix} f_B(B^\top x)h^{(1)}(B_0^\top x)(B_0 - B)^\top v_B(x) + O(h^2\delta_B) \\ O(h\delta_B) \end{pmatrix} + O(\delta_{dh}\delta_B)$$

and

$$\{\Sigma_n^B(x)\}^{-1} \times (2') = \begin{pmatrix} h^{(1)}(B_0^\top x)(B_0 - B)^\top v_B(x) + O(h^2\delta_B + \delta_{dh}\delta_B) \\ O(h\delta_B) \end{pmatrix}.$$

Similarly, $(3')$ equals

$$\frac{1}{n}\sum_{i=1}^{n} K_h(B^\top(X_i - x)) \begin{pmatrix} I_s \otimes \{(X_i - x)^\top B\} h^{(2)}(B_0^\top x) B^\top (X_i - x) \\ I_s \otimes \{B^\top(X_i - x)(X_i - x)^\top B/h\} h^{(2)}(B_0^\top x) B^\top (X_i - x) \end{pmatrix}$$

In the upper half, the first component, for example, is

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} K_h(B^\top(X_i - x))(X_i - x)^\top B h^{(2)}(B_0^\top x) B^\top (X_i - x) \\
=& f_B(B^\top x)\mathrm{tr}\{h_1^{(2)}(B_0^\top x)\}h^2 + O(h^4 + h^2\delta_{dh}) \\
=& f_B(B^\top x)\mathrm{tr}_1(x)h^2 + O(h^4 + h^2\delta_{dh}).
\end{aligned}$$

The first component in the lower half is

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} K_h(B^\top(X_i - x))B^\top(X_i - x)/h(X_i - x)^\top B h_1^{(2)}(B_0^\top x) B^\top (X_i - x) \\
=& h^3 \int K(u)uu^\top h_1^{(2)}(B_0^\top x)u \triangledown^\top f_B(B^\top x)u\mathrm{d}u + O(h^2\delta_{dn}) \\
=& M_{11}^B(x)h^3 + O(h^2\delta_{dh}).
\end{aligned}$$

Hence

$$(3') = \frac{1}{2}\begin{pmatrix} f_B(B^\top x)\mathrm{tr}(x)h^2 + O(h^4) \\ M_{1n}^B(x)h^3 \end{pmatrix} + O(h^2\delta_{dh})$$

and

$$\{\Sigma_n^B(x)\}^{-1} \times (3') = \begin{pmatrix} \mathrm{tr}(x)h^2 \\ M_{1n}^B(x)h^3 \end{pmatrix} + O(h^4 + h^2\delta_{dh}).$$

We next move to $(4')$.

$$\begin{aligned}
&\{\Sigma_n^B(x)\}^{-1} \times (4') \\
=&\begin{pmatrix} \mathcal{E}_{n,1}^B(x) - hf_B^{-1}(B^\top x)I_s \otimes \nabla^\top f_B(B^\top x)\mathcal{E}_{n,2}^B(x) \\ -hf_B^{-1}(B^\top x)I_s \otimes \nabla f_B(B^\top x)\mathcal{E}_{n,1}^B(x) + \mathcal{E}_{n,2}^B(x) \end{pmatrix} + O(h^2 + \delta_{dh})\begin{pmatrix} f_B(B^\top x)\mathcal{E}_{n,1}^B(x) \\ f_B(B^\top x)\mathcal{E}_{n,2}^B(x) \end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{E}_{n,1}^B(x) =& (nf_B(B^\top x))^{-1}\sum_{i=1}^n K_h(B^\top(X_i - x))\zeta_i, \\
\mathcal{E}_{n,2}^B(x) =& (nf_B(B^\top x))^{-1}\sum_{i=1}^n K_h(B^\top(X_i - x))I_s \otimes \{B^\top(X_i - x)/h\}\zeta_i.
\end{aligned}$$

It can be shown that $\mathcal{E}_{n,1}^B(x) = \mathcal{E}_{n,2}^B(x) = O(\delta_{dh})$. Hence

$$\begin{aligned}
\{\Sigma_n^B(x)\}^{-1} \times (4') =& \begin{pmatrix} \mathcal{E}_{n,1}^B(x) + M_{2n}^B(x)\mathcal{E}_{n,2}^B(x)h \\ M_{3n}^B(x)\mathcal{E}_{n,1}^B(x)h + \mathcal{E}_{n,2}^B(x) \end{pmatrix} + O(h^2\delta_{dh} + \delta_{dh}^2) \\
=& \begin{pmatrix} V_{1n}^B(x) \\ V_{2n}^B(x) \end{pmatrix} + O(h^2\delta_{dh} + \delta_{dh}^2)
\end{aligned}$$

where

$$M^B_{2n}(x) = -\, h f_B^{-1}(B^\mathsf{T} x) I_s \otimes \nabla^\mathsf{T} f_B(B^\mathsf{T} x),$$

$$M^B_{3n}(x) = -\, h f_B^{-1}(B^\mathsf{T} x) I_s \otimes \nabla f_B(B^\mathsf{T} x) = \{M^B_{2n}(x)\}^\mathsf{T}.$$

In the end, we similarly calculate that

$$\{\Sigma^B_n(x)\}^{-1} \times \{(5') + (6')\} = O(h^3 + h\delta_B + \delta_B^2).$$

Putting everything together, we get the result.

Proof of Lemma S4:

$$\begin{aligned}
&\mathrm{vec}(B) \\
=&\left\{ \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n K_h(B^\mathsf{T}(X_i - X_j))(b_j^\mathsf{T} b_j) \otimes \{(X_i - X_j)(X_i - X_j)^\mathsf{T}\} \Big/ \frac{1}{n} \sum_{i=1}^n K_h(B^\mathsf{T}(X_i - X_j)) \right\}^{-1} \\
&\times \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1}^n K_h(B^\mathsf{T}(X_i - X_j)) b_j^\mathsf{T} \otimes (X_i - X_j)(\mathrm{Log}_\mu Y_i - a_j) \Big/ \frac{1}{n} \sum_{i=1}^n K_h(B^T(x_i - x_j)).
\end{aligned}$$

From Lemma S3, we have

$$\begin{aligned}
&(b_j^\mathsf{T} b_j) \otimes \{(X_i - X_j)(X_i - X_j)^\mathsf{T}\} \\
=&h^2 \left\{ G(B_0^\mathsf{T} X_j) + O(h^4 + h\delta_{dh} + h^2\delta_B) \right\} \otimes \left\{ (X_i - X_j)(X_i - X_j)^\mathsf{T}/h^2 \right\},
\end{aligned}$$

where $G(B_0^\mathsf{T} x) = h^{(1)}(B_0^\mathsf{T} x)^\mathsf{T} h^{(1)}(B_0^\mathsf{T} x)$. Thus the denominator equals

$$\Big[\frac{1}{n} \sum_{j=1}^n \left\{ G(B_0^\mathsf{T} X_j) + O(h^4 + h\delta_{dh} + h^2\delta_B) \right\} \Big] \otimes \Big[\frac{h^2}{n} \sum_{i=1}^n K_h(B^\mathsf{T}(X_i - X_j)) \{(X_i - X_j)(X_i - X_j)^T/h^2\}$$
$$\Big/ \frac{1}{n^2} \sum_{i=1}^n K_h(B^\mathsf{T}(X_i - X_j)) \Big] \triangleq (1) \otimes (2).$$

Suppose $\tilde{B}$ is a $p\times(p-d)$ matrix such that $(B,\tilde{B})$ is a $p\times p$ orthogonal matrix. Then $(I_d\otimes B, I_d\otimes\tilde{B})$ is a $dp\times dp$ orthogonal matrix and we have the denominator equaling

$$
\begin{aligned}
&(I_d\otimes B, I_d\otimes\tilde{B})\begin{pmatrix} I_d\otimes B^{\top} \\ I_d\otimes\tilde{B}^{\top}\end{pmatrix}\Big\{(1)\otimes(2)\Big\}(I_d\otimes B, I_d\otimes\tilde{B})\begin{pmatrix} I_d\otimes B^{\top} \\ I_d\otimes\tilde{B}^{\top}\end{pmatrix}\\
=&(I_d\otimes B, I_d\otimes\tilde{B})\begin{pmatrix} (1)\otimes\{B^{\top}(2)B\} & (1)\otimes\{B^{\top}(2)\tilde{B}\} \\ (1)\otimes\{\tilde{B}^{\top}(2)B\} & (1)\otimes\{\tilde{B}^{\top}(2)\tilde{B}\}\end{pmatrix}\begin{pmatrix} I_d\otimes B^{\top} \\ I_d\otimes\tilde{B}^{\top}\end{pmatrix}.
\end{aligned} \tag{S5.23}
$$

First we notice in (S5.23) that

$$
\begin{aligned}
&\begin{pmatrix} B^{\top}(2)B & B^{\top}(2)\tilde{B} \\ \tilde{B}^{\top}(2)B & \tilde{B}^{\top}(2)\tilde{B}\end{pmatrix}\\
=&\begin{pmatrix} h^2 I_d + O(h^2\tau_n) & h^2 F_B^{\top}(B^{\top}x)/f_B(B^{\top}x)\tilde{B} + O(h\tau_n) \\ \tilde{B}^{\top}F_B(B^{\top}x)/f_B(B^{\top}x)h^2 + O(h\tau_n) & C_B(B^{\top}x) + O(\tau_n)\end{pmatrix},
\end{aligned}
$$

where $F_B(B^{\top}x) = v_B(x)\nabla^{\top}f_B(B^{\top}x) + f_B(B^{\top}x)\nabla v_B(x)$ and $C_B(B^{\top}x) = \tilde{B}^{\top}\bar{w}_B(x)\tilde{B}$. Here $\nabla v_B(x)$ is the derivative matrix of $v_B(x)$ w.r.t $B^{\top}x$ and $v_B(x)\in R^{p\times d}$.

Now the denominator becomes

$$
(I_d\otimes B, I_d\otimes\tilde{B})\begin{pmatrix} (1') & (2')^{\top} \\ (2') & (3)\end{pmatrix}\begin{pmatrix} I_d\otimes B^{\top} \\ I_d\otimes\tilde{B}^{\top}\end{pmatrix}.
$$

where

$$(1') = (1) \otimes \{h^2 I_d + O(h^2\tau_n)\} = h^2 E\{G(B_0^\top X) \otimes I_d\} + O(h^4 + h^2\delta_{dh} + h^4\delta_B),$$

$$(2') = (1) \otimes \{\tilde{B}^\top F_B(B^\top x)/f_B(B^\top x)h^2 + O(h\tau_n)\} = h^2(I_d \otimes \tilde{B})^\top F + O(h^3 + h\delta_{dh} + h^2\delta_B),$$

$$(3') = (1) \otimes \{C_B(B^\top x) + O(\tau_n)\} = 2(I_d \otimes \tilde{B}_0)^\top W_B(I_d \otimes \tilde{B}_0) + O(h^2 + \delta_{dh} + \delta_B),$$

and $F = E\{G(B_0^\top X) \otimes F_B(B^\top X)/f_B(B^\top X)\}, W_B = E[G(B_0^\top X) \otimes \{v_B(X)v_B^\top(X)\}]$.

Then use $(A + hB)^{-1} = A^{-1} - hA^{-1}BA^{-1} + O(h^2)$ to get

$$\begin{aligned}&1/\text{Denominator}\\ =&(I_d \otimes B)L_1^B(I_d \otimes B^\top) - (I_d \otimes B)L_2^B - L_3^B(I_d \otimes B^\top) + \frac{1}{2}W_B^+ + O(\tau_n/h + \delta_B),\end{aligned}$$

where

$$\begin{aligned} L_1^B &= h^{-2}\left[E\{G(B_0^\top X) \otimes I_d\}\right]^{-1}, \\ L_2^B &= (L_3^B)^\top = \left[E\{G(B_0^T X) \otimes I_d\}\right]^{-1} F^T W_B^+/2, \\ W_B^+ &= (I_d \otimes \tilde{B})\left\{(I_d \otimes \tilde{B}^\top)W_B(I_d \otimes \tilde{B})\right\}^{-1}(I_d \otimes \tilde{B}^\top). \end{aligned}$$

Proof of Lemma S5: Lemma S5 aims to simplify

$$\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n K_h(B^\top(X_i - X_j))b_j^\top \otimes (X_i - X_j)\{\text{Log}_\mu Y_i - a_j - b_j \otimes (X_i - X_j)^\top\}$$

divided by $n^{-1}\sum_{i=1}^n K_h(B^\top(X_i - X_j))$. We have proved in Lemma S3 that

$$a_j = h(B_0^\top X_j) + h^{(1)}(B_0^\top X_j)(B_0 - B)^\top v_B(X_j) + h^2\text{tr}(X_j)/2 + V_{1n}^B(X_j) + O(h^3 + h\delta_{dh} + h\delta_B),$$

$$\text{vecs}(b_j) = \text{vecs}(h^{(1)}(B_0^\top X_j)) + V_{2n}^B(X_j)/h + O(h^2 + \delta_{dh} + \delta_B).$$

Thus by Taylor expansion, we have

$$
\begin{aligned}
&\mathrm{Log}_\mu Y_i - a_j - b_j \otimes (X_i - X_j)^\top \\
=&h(B_0^\top X_j) + h^{(1)}(B_0^\top X_j)B_0^\top(X_i - X_j) + I_s \otimes \{(X_i - X_j)^T B_0\}h^{(2)}(B_0^\top X_j)B_0^\top(X_i - X_j) + \zeta_i \\
&+ O(||B_0^\top(X_i - X_j)||^3) - h(B_0^\top X_j) - h^{(1)}(B_0^\top X_j)(B_0 - B)^\top v_B(X_j) - h^2\mathrm{tr}(X_j)/2 - V_{1n}^B(X_j) \\
&+ O(h^3 + h\delta_{dh} + h\delta_B) - h^{(1)}(B_0^\top X_j)(X_i - X_j) - \mathscr{M}\{V_{2n}^B(X_j)\}B_0^\top(X_i - X_j)/h \\
&+ O(h^2 + \delta_{dh} + \delta_B)B_0^\top(X_i - X_j) \\
=&I_s \otimes \{(X_i - X_j)^\top B\}h^{(2)}(B_0^\top X_j)B^\top(X_i - X_j)/2 + \zeta_i + h^{(1)}(B_0^\top X_j)(B - B_0)^\top v_B(X_j) - h^2\mathrm{tr}(X_j)/2 \\
&- V_{1n}^B(X_j) - \mathscr{M}\{V_{2n}^B(X_j)\}B_0^\top(X_i - X_j)/h + \triangle_n(X_i, X_j, B) \\
&+ O(||B_0^\top(X_i - X_j)||^3) + O(h^3 + h\delta_{dh} + h\delta_B) + O(h^2 + \delta_{dh} + \delta_B)B_0^\top(X_i - X_j) \\
=&(1) + (2) + (3) - (4) - (5) - (6) + (7) + (8) + (9) + (10).
\end{aligned}
$$

where

$$
\begin{aligned}
\triangle_n(X_i, X_j, B) =&\frac{1}{2}\Big[I_s \otimes \{(X_i - X_j)^\top(B_0 - B)\}h^{(2)}(B_0^\top X_j)B^\top(X_i - X_j) \\
&+ I_s \otimes \{(X_i - X_j)^\top B\}h^{(2)}(B_0^\top X_j)(B_0 - B)^\top(X_i - X_j) \\
&+ I_s \otimes \{(X_i - X_j)^\top(B_0 - B)\}h^{(2)}(B_0^\top X_j)(B_0 - B)^\top(X_i - X_j)\Big].
\end{aligned}
$$

Next we calculate

$$
\frac{1}{n^2}\sum_{j=1}^n\sum_{i=1}^n K_h(B^\top(X_i - X_j))b_j^\top \otimes (X_i - X_j)\Big\{(1) \sim (10)\Big\}/\frac{1}{n}\sum_{i=1}^n K_h(B^\top(X_i - X_j))
$$

and denote corresponding results as $(1') \sim (10')$.

$$
\begin{aligned}
(1') =& \frac{h^2}{n}\sum_{j=1}^{n}\sum_{k=1}^{s} h_k^{(1)}(B_0^{\mathsf{T}}X_j) \otimes v_B(X_j)\mathrm{tr}_k(X_j) + O(h^4 + h\delta_{dh} + h^2\delta_B),\\
(2') =& -\frac{1}{n}\sum_{i=1}^{n} h^{(1)}(B_0^{\mathsf{T}}X_i)^{\mathsf{T}} \otimes v_B(X_i)\zeta_i + o(n^{-1/2}) \triangleq \Phi_n(B_0) + o(n^{-1/2}),\\
(3') =& W_B\mathrm{vec}(B - B_0) + O(h^2\delta_B + \delta_{dh}\delta_B/h + \delta_B^2),\\
(4') =& \frac{h^2}{n}\sum_{j=1}^{n}\sum_{k=1}^{s} h_k^{(1)}(B_0^{\mathsf{T}}X_j) \otimes v_B(X_j)\mathrm{tr}_k(X_j) + O(h^4 + h\delta_{dh} + h^2\delta_B),\\
(5') =& O(h^2\delta_{dh} + \delta_{dh}^2/h + \delta_B\delta_{dh}), (6') = O(h^3 + \delta_{dh}\delta_B + h^2\delta_B + \delta_{dh}),\\
(7') =& O(h^2\delta_B + \delta_B^2), (8') = O(h^3), (9') = O(h^3 + h\delta_{dh} + h\delta_B),\\
(10') =& O(h^4 + h^2\delta_{dh} + h^2\delta_B).
\end{aligned}
$$

Collecting above 10 terms shows the numerator equals $W_B\mathrm{vec}(B-B_0)+\Phi_n(B_0) + O(h^3 + h\delta_{dh} + h\delta_B + \delta_{dh}^2/h + \delta_{dh}\delta_B/h)$.

References

Cai, Z., Y. Xia, and W. Hang (2022). An outer-product-of-gradient approach to dimension reduction and its application to classification in high dimensional space. *Journal of the American Statistical Association 118*(543), 1671–1681.

Li, Q. and J. Racine (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica 14*, 485–512.

Lin, Z. and F. Yao (2019). Intrinsic riemannian functional data analysis. *The Annals of Statistics 47*, 3533–3577.

Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory 22*, 1112–1137.

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *The Annals of Statistics 35*, 2654–2690.

Xia, Y., H. Tong, W. K. Li, and L.-X. Zhu (2002). An adaptive estimation of dimension reduction space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*, 363–410.

Yao, Q. and H. Tong (1994). On subset selection in non-parametric stochatic regression. *Statistica Sinica 4*, 51–70.

Zhang, H. (2021). Minimum average variance estimation with group lasso for the multivariate response central mean subspace. *Journal of Multivariate Analysis 184*.

School of Statistics, East China Normal University, Shanghai, China

E-mail: 52214404011@stu.ecnu.edu.cn

E-mail: shuangdai95@163.com

E-mail: zyu@stat.ecnu.edu.cn