Biometrika (xxxx), **xxx**, x, *pp.* 1–23
Printed in Great Britain xxxx

Supplementary Material for "Functional Regression on Manifold with Contamination"

BY ZHENHUA LIN

Department of Statistics and Applied Probability, National University of Singapore, 117546, Singapore

stalz@nus.edu.sg

AND FANG YAO

Department of Probability and Statistics, School of Mathematical Sciences, Center for Statistical Science, Peking University, Beijing, 100871, China

fyao@math.pku.edu.cn

SUMMARY

Section S.1 contains details and simulation studies for sparsely observed functional data, Section S.2 contains auxiliary examples and results, Section S.3 provides proofs to the main theorems, and Section S.4 contains technical lemmas.

S.1. DETAILS AND SIMULATION STUDIES FOR SPARSE DESIGN

When the functional data are sparsely observed, we adopt the procedure proposed by Yao et al. (2005) to recover individual functions, as follows. First, the local linear smoother (Fan, 1993) is adopted to produce an estimate $\hat{\mu}$ of the global mean function and an estimate $\hat{\mathcal{C}}$ of the global covariance function of X by pooling all observed data; see Zhang & Wang (2016) for more details. Then estimates $\hat{\psi}_k$ and $\hat{\lambda}_k$ of the eigenfunctions and eigenvalues, respectively, are obtained by solving the eigen-equations $\int_D \hat{\mathcal{C}}(s,t)\hat{\psi}_k(s)\mathrm{d}s = \hat{\lambda}_k\hat{\psi}_k(t)$. The global principal component scores ξ_{ij} are estimated by $\hat{\xi}_{ik} = \hat{\lambda}_k\hat{\phi}_{ik}^T\widehat{\Sigma}_i^{-1}(a_i - b_i)$, where $\hat{\phi}_{ik}^T = (\hat{\psi}_k(T_{i1}),\ldots,\hat{\psi}_k(T_{im_i}))$, $a_i = (X_{i1},\ldots,X_{im_i})^T$, $b_i = (\hat{\mu}(T_{i1}),\ldots,\hat{\mu}(T_{im_i}))^T$, and $\widehat{\Sigma}_i$ is an $m_i \times m_i$ matrix whose element in the jth row and lth column is $\hat{\mathcal{C}}(T_{ij},T_{il}) + \hat{\sigma}_\zeta^2 1_{j=l}$ with $\hat{\sigma}_\zeta^2$ being the estimate of the variance of the noise ζ. Finally, $X_i(t)$ is estimated by $\hat{X}_i(t) = \hat{\mu}(t) + \sum_{j=1}^{\mathcal{K}}\hat{\xi}_{ij}\hat{\psi}_j(t)$, where $\mathcal{K}$ is a tuning parameter whose selection is discussed in Yao et al. (2005).

To illustrate the numerical performance of the proposed method for sparsely and irregularly observed data, we adopt the same setting from Section 4 for dense data, except that now $m_i \sim 1 + \mathrm{Poisson}(3)$ and $T_{ij} \sim \mathrm{uniform}(0,1)$. In this new setting, the average number of observations per curve is 4 and the observed time points are irregularly scattered. From the results presented in Table S.1, we observe that, the proposed method is comparable to other methods for the $SO(3)$ manifold while exhibits a clear advantage for the other two manifolds. In addition, as the data are rather sparse, the contamination is expected to dominate the convergence rate in (11) and (12). Thus, we observe that the root mean square error decreases slowly with the sample size, in contrast with the fast rate observed in the case of dense data.

Since the contamination is of a high level in this setting, the structure of the $SO(3)$ manifold might be buried by the contamination and thus could not be exploited. This might explain why

Table S.1: Results of simulation studies for sparsely observed data

	$SO(3)$ Manifold			Klein Bottle			Gaussian Mixture		
	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$	$n = 250$	$n = 500$	$n = 1000$
FLR	24.0 (0.27)	23.8 (0.27)	23.6 (0.25)	61.4 (0.51)	61.3 (0.44)	61.1 (0.35)	43.2 (1.64)	42.0 (1.65)	41.1 (1.63)
FNW	23.9 (0.24)	23.9 (0.26)	23.8 (0.24)	60.9 (0.59)	60.2 (0.62)	59.7 (0.62)	49.7 (3.65)	47.2 (1.05)	46.2 (0.76)
FCE	25.0 (0.42)	24.9 (0.33)	24.8 (0.31)	62.5 (0.89)	62.2 (0.75)	61.9 (0.56)	49.5 (1.12)	49.2 (0.83)	48.8 (0.78)
FMO	33.5 (1.61)	32.8 (1.19)	32.0 (1.45)	83.9 (3.70)	82.7 (3.06)	81.5 (2.64)	63.8 (3.06)	62.4 (2.16)	60.2 (2.30)
FCM	25.5 (0.60)	25.1 (0.42)	24.4 (0.30)	65.3 (1.56)	63.4 (1.15)	61.5 (0.74)	50.5 (1.52)	49.4 (1.05)	47.7 (0.76)
MUL	26.3 (0.59)	26.0 (0.45)	25.5 (0.39)	66.3 (1.32)	65.4 (1.12)	64.2 (0.79)	51.6 (1.35)	50.9 (0.99)	49.5 (0.89)
FREM	24.9 (0.96)	24.4 (0.87)	23.9 (0.68)	56.1 (2.62)	52.0 (1.19)	50.1 (0.64)	37.1 (2.11)	34.7 (1.98)	32.7 (1.46)

FLR, functional linear regression; FNW, functional Nadaraya–Watson smoothing; FCE, functional conditional expectation; FMO, functional mode, FCM, functional conditional median; MUL, multi-method; FREM, the proposed functional regression on manifold; MSP, meat spectrometric data; DTI, diffusion tensor imaging data; SBP, systolic blood pressure data. The numbers outside of parentheses are the Monte Carlo average of root mean square error based on 100 independent simulation replicates, and the numbers in parentheses are the corresponding standard error.

the proposed method shares a similar performance with the functional linear regression or is even slightly outperformed by the latter. Also the performance of sophisticated regression methods like nonparametric regression methods is generally more sensitive to the noise level of the predictor, especially when the predictor resides in a space of higher dimension. This might explain why in the setting of the $SO(3)$ manifold, almost all nonparametric regression methods listed in Table S.1 perform no better than the functional linear regression which is perhaps the simplest parametric method in functional regression.

S.2. AUXILIARY EXAMPLE AND RESULTS

Example 1. Let $S^1 = \{v_\omega = (\cos\omega, \sin\omega) : \omega \in [0, 2\pi)\}$ denote the unit circle regarded as a one-dimensional Riemannian manifold. Let $D = [0, 1]$ and denote $\phi_1, \phi_2, \ldots$ a complete orthonormal basis of $\mathcal{L}^2(D)$. Define map $X(v_\omega) = \sqrt{C}\sum_k k^{-c}\{\cos(k\omega)\phi_{2k-1} + \sin(k\omega)\phi_{2k}\}$ with $c > 3/2$ and $C = 1/\sum_k k^{-2c+2} \in (0, \infty)$. According to Proposition S.1, X is an isometric embedding of S^1 into $\mathcal{L}^2(D)$. Then $\mathcal{M} = X(S^1)$ is a submanifold of $\mathcal{L}^2(D)$. Moreover, ***no finite-dimensional*** linear subspace of $\mathcal{L}^2(D)$ fully encompasses $\mathcal{M}$. A consequence of this observation is that, a random process taking samples from such $\mathcal{M}$ might have an infinite number of eigenfunctions, even though $\mathcal{M}$ is merely one-dimensional, as we shall exhibit in the following. Let us treat S^1 as a probability space endowed with the uniform probability measure, and define random variables $\xi_{2k-1}(v_\omega) = \sqrt{C}k^{-c}\cos(k\omega)$ and $\xi_{2k}(v_\omega) = \sqrt{C}k^{-c}\sin(k\omega)$. Then $X = \sum_k \xi_k\phi_k$ can be regarded as a random process with samples from $\mathcal{M}$. It is easy to check that $E(\xi_k\xi_j) = 0$ if $k \neq j$, $E(\xi_k) = 0$, and $E\xi_{2k-1}^2 = E\xi_{2k}^2 = C\pi k^{-2c}$, which implies that $E(\|X\|_{\mathcal{L}^2}^2) < \infty$. One can see that the eigenfunctions of the covariance operator of X are exactly ϕ_k. Therefore, $X = \sum_k \xi_k\phi_k$ is the Karhunen-Loève expansion of the random process X, which clearly includes an infinite number of principal components, while X is intrinsically sampled from the one-dimensional manifold $\mathcal{M}$.

PROPOSITION 1. ***The embedding*** X ***defined in Example*** 1 ***is an isometric embedding. Moreover, there is no finite-dimensional linear subspace of*** $\mathcal{L}^2(D)$ ***that fully contains the image*** $X(S^1)$.

Proof. Let $V = \{(\cos\omega, \sin\omega) : \omega \in (a, b)\}$ be a local neighborhood of v, and let $\psi(v) = \omega \in (a, b)$ for $v = (v_1, v_2) = (\cos\omega, \sin\omega) \in V$. Then ψ is a chart of S^1. Let U be open in $\mathcal{L}^2$ such that $X(v) \in U$. Since $\mathcal{L}^2$ is a linear space, the identity map I serves as a chart.

Let $X_{U,V} : \psi(V) \to \mathcal{L}^2$ denote the map $X \circ \psi^{-1}$. Let $\vartheta = \sqrt{C}\sum_k k^{-c+1}\{-\sin(k\omega)\phi_{2k-1} + \cos(k\omega)\phi_{2k}\}$. It defines a linear map from $\mathbb{R}$ to $\mathcal{L}^2$, denoted by $\Theta(t) = t\vartheta \in \mathcal{L}^2$. Then,

$$
\begin{aligned}
A(t) \equiv& t^{-2}\|X_{U,V}(\omega+t) - X_{U,V}(\omega) - \Theta(t)\|^2 \\
=& C\sum_{k=1}\left\{\frac{k^{-c}\cos(k\omega+kt) - k^{-c}\cos(k\omega) + tk^{-c+1}\sin(k\omega)}{t}\right\}^2 + \\
& C\sum_{k=1}\left\{\frac{k^{-c}\sin(k\omega+kt) - k^{-c}\sin(k\omega) - tk^{-c+1}\cos(k\omega)}{t}\right\}^2 \\
\equiv& CB_1^2(t) + CB_2^2(t).
\end{aligned}
$$

By Lipschitz property of the function

$$B_{1,k}(t) \equiv k^{-c}\cos(k\omega+kt) - k^{-c}\cos(k\omega) + tk^{-c+1}\sin(k\omega),$$

we conclude that $|B_{1,k}(t)| \le t\sup_t |B'_{1,k}(t)| \le 2k^{-c+1}t$. This implies that $\sup_t B_1^2(t) \le \sum_k 4k^{-2c+2} < \infty$. By similar reasoning, $\sup_t B_2^2(t) < \infty$ and hence $\sup_t A(t) < \infty$. We now apply the dominated convergence theorem to conclude that

$$\lim_{t\to 0} A(t) = C\lim_{t\to 0}\{B_1^2(t) + B_2^2(t)\} = C\sum_{k=1}\lim_{t\to 0}\left\{\frac{B_{1,k}(t)}{t}\right\}^2 + C\sum_{k=1}\lim_{t\to 0}\left\{\frac{B_{2,k}(t)}{t}\right\}^2 = 0.$$

By recalling that the tangent space $T_{v_\omega}S^1$ at v_ω is $\mathbb{R}$ and the tangent space $T_{X(v_\omega)}\mathcal{L}^2(D)$ at $X(v_\omega)$ is $\mathcal{L}^2(D)$, the above shows that the differential map $X_{*,v_\omega} : T_{v_\omega}S^1 \to T_{X(v_\omega)}\mathcal{L}^2(D)$ at v_ω is given by the linear map Θ, i.e,

$$X_{*,v_\omega}(t) = \Theta(t) = t\sum_k \sqrt{C}k^{-c+1}\{-\sin(k\omega)\phi_{2k-1} + \cos(k\omega)\phi_{2k}\},$$

and the embedded tangent space at v_ω is $\mathrm{span}\{-\sum_k k^{-c+1}\sin(k\omega)\phi_{2k-1} + \sum_k k^{-c+1}\cos(k\omega)\phi_{2k}\}$. As this differential map is injective at all $v \in S^1$, X is indeed an immersion. Since S^1 is compact, X is also an embedding, and the image $X(S^1)$ is a submanifold of $\mathcal{L}^2(D)$.

To show that X is isometric, note that the tangent space of S^1 at v is the real line $\mathbb{R}$, equipped with the usual inner product $\langle s,t\rangle = st$ for $s,t \in \mathbb{R}$. Let $\langle\!\langle f_1, f_2\rangle\!\rangle = \int_D f_1(t)f_2(t)\mathrm{d}t$ for $f_1, f_2 \in \mathcal{L}^2(D)$ denote the canonical inner product of $\mathcal{L}^2(D)$. Recalling the definition $C = 1/\sum_{k=1} k^{-2c+2}$ in the example, we deduce that

$$
\begin{aligned}
\langle\!\langle X_{*,v_\omega}(s), X_{*,v_\omega}(t)\rangle\!\rangle = \langle\!\langle \Theta(s), \Theta(t)\rangle\!\rangle = st\langle\!\langle \vartheta,\vartheta\rangle\!\rangle \\
= Cst\sum_{k=1}\{k^{-2c+2}\sin^2(k\omega) + k^{-2c+2}\cos^2(k\omega)\} \\
= Cst\sum_{k=1} k^{-2c+2} = st = \langle s,t\rangle,
\end{aligned}
$$

which shows that X is isometric.

Finally, to show that there is no finite-dimensional linear subspace of $\mathcal{L}^2(D)$ that fully contains $X(S^1)$, we take the strategy of "proof by contradiction" to assume that H is a finite-dimensional linear subspace of $\mathcal{L}^2(D)$ such that $X(S^1) \subset H$. Since H is finite-dimensional, there exists $0 \neq \varphi \in \mathcal{L}^2(D)$ such that $\varphi \perp H$ and hence $\varphi \perp X(S^1)$, or more specifically, $\langle\!\langle \varphi, x\rangle\!\rangle = 0$ for each $x \in X(S^1)$. As $\phi_1, \phi_2, \dots$ form a complete orthonormal basis of $\mathcal{L}^2(D)$, we can find real numbers $a_1, a_2, \dots$ such that $\varphi = \sum_k (a_{2k-1}\phi_{2k-1} + a_{2k}\phi_{2k})$. Then, $\langle\langle \varphi, x \rangle\rangle = 0$ for each $x \in X(S^1)$ is equivalent to $\sum_k k^{-c}\{a_{2k-1}\cos(k\omega) + a_{2k}\sin(k\omega)\} = 0$ for all ω. Since $\cos(k\omega)$ and $\sin(k\omega)$, as functions of ω, are orthogonal, it implies that $a_{2k-1} = 0$ and $a_{2k} = 0$ for all k, which indicates that $\varphi = 0$. However, by assumption, $\varphi \neq 0$, and we draw a contradiction. $\square$

The contamination of the predictor X poses substantial challenge on the estimation of the manifold structure. For instance, the quality of the tangent space at x, denoted by $T_x\mathcal{M}$, crucially depends on a bona fide neighborhood around x, while the contaminated neighborhood $\hat{\mathcal{N}}_{\mathcal{L}^2}(h_{pca}, x)$ and the inaccessible true neighborhood $\mathcal{N}_{\mathcal{L}^2}(h_{pca}, x) = \{X_i : \|X_i - x\|_{\mathcal{L}^2} < h_{pca}\}$ might contain different observations. Fortunately we can show that they are not far apart in the sense of Proposition 2. In practice, we suggest to choose $\max(h_{reg}, h_{pca}) < \min\{2/\tau, \mathrm{inj}(\mathcal{M})\}/4$, where τ is the condition number of $\mathcal{M}$ and $\mathrm{inj}(\mathcal{M})$ is the injectivity radius of $\mathcal{M}$ (Cheng & Wu, 2013), so that $\hat{\mathcal{N}}_{\mathcal{L}^2}(h_{pca}, x)$ provides a good approximation of the true neighborhood of x within the manifold.

PROPOSITION 2. *For* $0 < \varrho < \beta$, *define* $h_- = h_{pca} - m^{-(\beta+\varrho)/2}$ *and* $h_+ = h_{pca} + m^{-(\beta+\varrho)/2}$. *Let* $Z_i = 1_{\hat{X}_i \in \mathbb{B}^{\mathcal{L}^2}_{h_{pca}}(x)}$, $V_{i0} = 1_{X_i \in \mathbb{B}^{\mathcal{L}^2}_{h_-}(x)}$ *and* $V_{i1} = 1_{X_i \in \mathbb{B}^{\mathcal{L}^2}_{h_+}(x)}$. *Under the assumption* (B3) *and* $\log m \gtrsim \log n$, $\mathrm{pr}(\forall i : V_{i0} \leq Z_i \leq V_{i1}) \to 1$ *as* $n \to \infty$.

Hence one can always obtain lower and upper bounds for quantities involving Z_i in terms of V_{i0} and V_{i1}, i.e., with large probability, it is equivalent to substitute Z_i with V_{i0} and V_{i1} in our analysis.

Proof. We first bound the following event

$$
\begin{aligned}
\mathrm{pr}(\forall i : Z_i \leq V_{i1}) &= \prod_{i=1}^{n} \{1 - \mathrm{pr}(Z_i > V_{i1})\} = \{1 - \mathrm{pr}(Z > V_1)\}^n \\
&= \{1 - \mathrm{pr}(Z = 1, V_1 = 0)\}^n \geq \left\{1 - \mathrm{pr}(\|\hat{X} - X\| \geq m^{-(\beta+\varrho)/2})\right\}^n \\
&\geq \left(1 - c_1^p m^{p(\beta+\varrho)/2} m^{-p\beta}\right)^n \geq \left(1 - c_1^p m^{-2}\right)^n \to 1,
\end{aligned}
$$

where $c_1 > 0$ is some constant, and $p > 0$ is a constant that is sufficiently large. Similarly, we can deduce that

$$
\mathrm{pr}(\forall i : V_{i0} \leq Z_i) \to 1,
$$

and the conclusion $\mathrm{pr}(\forall i : V_{0i} \leq Z_i \leq V_{1i}) \to 1$ follows. $\square$

We now address the case that the predictor x is not fully observed. It is reasonable to assume that x comes from the same source of the data, in the sense that its smoothed version $\tilde{x}$ has the same contamination level as those $\hat{X}_1, \dots, \hat{X}_n$, as per (B3). To be specific, assume that

(B4) the estimate $\tilde{x}$ is independent of $X_1, \dots, X_n$ and $\mathbb{X}_1, \dots, \mathbb{X}_n$. Also $\{E\|\tilde{x} - x\|^p_{\mathcal{L}^2}\}^{1/p} \leq C_p m^{-\beta}$ for all $p \geq 1$, where C_p is a constant depending on p only.

Note that the independent condition in (B4) is satisfied if $t_1, \dots, t_{m_x}$ are independent of $X_1, \dots, X_n$ and $\mathbb{X}_1, \dots, \mathbb{X}_n$. The second part of (B4) is met if assumptions similar to (A1)–(A4) hold also for x and $t_1, \dots, t_{m_x}$, according to Proposition 1.

THEOREM 4. *With the conditions* (A1), (B1)–(B3)*, and the additional assumption* (B4)*, the equation* (11) *holds when* $x \in \mathcal{M}\backslash\mathcal{M}_{h_{reg}}$*, and the equation* (12) *holds when* $x \in \mathcal{M}_{h_{reg}}$*, both with* $\hat{g}(x)$ *replaced by* $\hat{g}(\tilde{x})$.

Proof. We first observe that

$$E\left[\{\hat{g}(\tilde{x}) - g(x)\}^2 \mid \mathscr{X}\right] \leq 2E\left[\{\hat{g}(\tilde{x}) - \hat{g}(x)\}^2 \mid \mathscr{X}\right] + 2E\left[\{\hat{g}(x) - g(x)\}^2 \mid \mathscr{X}\right].$$

To derive the order for the first term, we shall point out that, with Lemma S.7, S.8 and S.9, by following almost the same lines of argument, the conclusions of Theorem 1 hold for $\tilde{x}$. This means, by working on $\tilde{x}$ instead of x, we still have a consistent estimate of the intrinsic dimension and a good estimate of the tangent space at x. Given this, it is not difficult but somewhat tedious to verify that the argument in the proof of Theorem 2 and 3 still holds for $\tilde{x}$, with care for the discrepancy $\|\tilde{x} - x\|_{\mathcal{L}^2}$ instead of the discrepancy $\|\hat{X}_i - X_i\|_{\mathcal{L}^2}$. This argument also shows that the order of the first term is the same as the second one (this is expected since $\tilde{x}$ has the same contamination level of those $\hat{X}_i$), and hence the conclusion of the theorem follows. □

S.3. PROOFS OF MAIN THEOREMS

To reduce notational burden, $\mathcal{L}^2(D)$ is simplified by $\mathcal{L}^2$, and we shall use $\|\cdot\|$ to denote the norm $\|\cdot\|_{\mathcal{L}^2}$ when no confusion arises.

Proof of Theorem 1. (a) Without loss of generality, assume $x = 0$. Let $\tilde{G}_j = \hat{G}_j + \Delta$ and $\tilde{G}_{(1)}, \tilde{G}_{(2)}, \ldots, \tilde{G}_{(k)}$ be the associated order statistics of $\tilde{G}_1, \tilde{G}_2, \ldots, \tilde{G}_k$. Also note that the estimator in Levina & Bickel (2004) is still consistent if G is replaced with $\check{G} \equiv G + \Delta$. Then,

$$\begin{aligned}
&\left|\frac{1}{k-1}\sum_{j=1}^{k-1}\log\frac{\hat{G}_{(k)}+\Delta}{\hat{G}_{(j)}+\Delta} - \frac{1}{k-1}\sum_{j=1}^{k-1}\log\frac{\check{G}_{(k)}}{\check{G}_{(j)}}\right| \\
&\leq \left|\log\tilde{G}_{(k)} - \log\check{G}_{(k)}\right| + \left|\frac{1}{k-1}\sum_{j=1}^{k}\left(\log\tilde{G}_{(j)} - \log\check{G}_{(j)}\right)\right| \equiv I_1 + I_2. \qquad (1)
\end{aligned}$$

For I_1, let q and p be the indices such that $\tilde{G}_{(k)} = \hat{G}_q$ and $\check{G}_p = \check{G}_{(k)}$, respectively. For the case $q = p$, we have $|\tilde{G}_{(k)} - \check{G}_{(k)}| = |\tilde{G}_p - \check{G}_p| \leq \left|\|\hat{X}_p\| - \|X_p\|\right| + \Delta \leq \|\hat{X}_p - X_p\| + \Delta$ by reverse triangle inequality. When $q \neq p$, it is seen that $\tilde{G}_p < \tilde{G}_{(k)} = \tilde{G}_q$ and $G_q < G_p = G_{(k)}$. If $\tilde{G}_{(k)} > \check{G}_{(k)}$, then $|\tilde{G}_{(k)} - \check{G}_{(k)}| \leq |\tilde{G}_{(k)} - \check{G}_q| = |\tilde{G}_q - G_q| \leq \max_{1\leq j\leq k}\{\|\hat{X}_j - X_j\|\}$. Otherwise, $|\tilde{G}_{(k)} - \check{G}_{(k)}| \leq |\tilde{G}_p - \check{G}_p| \leq \max_{1\leq j\leq k}\{\|\hat{X}_j - X_j\|\}$. Now, $\mathrm{pr}(\forall 1 \leq j \leq k : \|\hat{X}_j - X_j\| > \epsilon) \leq \sum_{j=1}^{k}\mathrm{pr}\left(\|\hat{X}_j - X_j\| > \epsilon\right) \leq kE\|\hat{X}_j - X_j\|^r\epsilon^{-r} = O(km^{-r\beta}) = o(1)$ for a sufficiently large constant r. Therefore, $|\tilde{G}_{(k)} - \check{G}_{(k)}|$ converges to zero in probability, or $\tilde{G}_{(k)}$ converges to $\check{G}_{(k)}$ in probability. By Slutsky's lemma, $\log\tilde{G}_{(k)}$ converges to $\log\check{G}_{(k)}$ in probability and hence $I_1 = o_P(1)$.

For I_2, we first observe that

$$I_2 = \left|\frac{1}{k-1}\sum_{j=1}^{k}\left(\log\tilde{G}_{(j)} - \log\check{G}_{(j)}\right)\right| = \left|\frac{1}{k-1}\sum_{j=1}^{k}\left(\log\tilde{G}_j - \log\check{G}_j\right)\right|.$$

By Markov's inequality, for any fixed $\epsilon > 0$,

$$\mathrm{pr}\,(I_2 > \epsilon) \leq \frac{EI_2}{\epsilon} \leq \frac{kE|\log \tilde{G} - \log \check{G}|}{(k-1)\epsilon} = o(1),$$

where the last equality is obtained by Lemma S.7. We then deduce that $I_2 = o_P(1)$. Together with $I_1 = o_P(1)$ and (1), this implies that

$$\left| \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{\hat{G}_{(k)} + \Delta}{\hat{G}_{(j)} + \Delta} - \frac{1}{k-1} \sum_{j=1}^{k-1} \log \frac{\check{G}_{(k)}}{\check{G}_{(j)}} \right| \to 0 \quad \text{in probability}.$$

Now we apply the argument in Levina & Bickel (2004) to conclude that $\hat{d}$ is a consistent estimator.

(b) Let $h = h_{pca}$, and $\{\tilde{\phi}_k\}_{k=1}^d$ be a orthonormal basis system for $T_x\mathcal{M}$ and $\{\psi_k\}_{k=1}^{\infty}$ be an orthonormal basis of $\mathcal{L}^2$. Without loss of generality, assume that $\mathcal{M}$ is properly rotated and translated so that $\psi_k = \tilde{\phi}_k$ for $k = 1, 2, \ldots, d$, and $x = 0 \in \mathcal{L}^2$. The sample covariance operator based on observations in $\hat{\mathcal{N}}_{\mathcal{L}^2}(h, x)$ is denoted by $\hat{\mathcal{C}}_x$ as in (6). It is seen that $\hat{\mathcal{C}}_x = n^{-1}\sum_{i=1}^n (\hat{X}_i - \hat{\mu}_x)(\hat{X}_i - \hat{\mu}_x)Z_i$, where $Z_i = 1_{\hat{X}_i \in \mathbb{B}_h^{\mathcal{L}^2}(x)}$ and $\hat{\mu}_x = n^{-1}\sum_{i=1}^n \hat{X}_i Z_i$. Let $\mathcal{H}_1 = \mathrm{span}\{\psi_k : k = 1, 2, \ldots, d\}$ and $\mathcal{H}_2$ be the complementary subspace of $\mathcal{H}_1$ in $\mathcal{L}^2$, so that $\mathcal{L}^2 = \mathcal{H}_1 \oplus \mathcal{H}_2$. Let $\mathcal{P}_j : \mathcal{L}^2 \to \mathcal{H}_j$ be projection operators, and we define operator $\mathcal{A} = \mathcal{P}_1\hat{\mathcal{C}}_x\mathcal{P}_1$, $\mathcal{B} = \mathcal{P}_2\hat{\mathcal{C}}_x\mathcal{P}_2$, $\mathcal{D}_{12} = \mathcal{P}_1\hat{\mathcal{C}}_x\mathcal{P}_2$ and $\mathcal{D}_{21} = \mathcal{P}_2\hat{\mathcal{C}}_x\mathcal{P}_1$. Then $\hat{\mathcal{C}}_x = \mathcal{A} + \mathcal{B} + \mathcal{D}_{12} + \mathcal{D}_{21}$. Note that $\mathcal{D}_{12} + \mathcal{D}_{21}$ is self-adjoint. Therefore, if $y = \sum_{k=1}^{\infty} a_k\psi_k \in \mathcal{L}^2$,

$$\begin{aligned}
\|\mathcal{D}_{12} + \mathcal{D}_{21}\|_{op} =& \sup_{\|y\|=1} \langle(\mathcal{D}_{12} + \mathcal{D}_{21})y, y\rangle = \sup_{\|y\|=1} \left(\langle\mathcal{P}_1\hat{\mathcal{C}}_x\mathcal{P}_2 y, y\rangle + \langle\mathcal{P}_2\hat{\mathcal{C}}_x\mathcal{P}_1 y, y\rangle\right) \\
=&2 \sup_{\|y\|=1} \left(\sum_{k=d+1}^{\infty} \sum_{j=1}^{d} a_j a_k \langle\hat{\mathcal{C}}_x\psi_j, \psi_k\rangle \right) \leq 2 \sup_{\|y\|=1} \left(\sum_{k=d+1}^{\infty} \sum_{j=1}^{d} |a_j a_k| \cdot \left|\langle\hat{\mathcal{C}}_x\psi_j, \psi_k\rangle\right| \right) \\
\leq&2 \sup_{j\leq d} \sup_{k\geq d+1} \left|\langle\hat{\mathcal{C}}_x\psi_j, \psi_k\rangle\right| \sup_{\|y\|=1} \left\{ \sum_{k=d+1}^{\infty} \sum_{j=1}^{d} (a_j^2 + a_k^2) \right\} \leq 2 \sup_{j\leq d} \sup_{k\geq d+1} \left|\langle\hat{\mathcal{C}}_x\psi_j, \psi_k\rangle\right|.
\end{aligned}$$

From Lemma S.9, $\|\mathcal{D}_{12} + \mathcal{D}_{21}\|_{op} = O_P\left(h^{d+3} + n^{-1/2}h^{d/2+3} + m^{-\beta}h^{d+1}\right)$. Similarly, we have $\|\mathcal{B}\|_{op} = O_P\left(h^{d+4} + n^{-1/2}h^{d/2+4} + m^{-\beta}h^{d+1}\right)$, and $\mathcal{A} = \pi_{d-1}f(x)d^{-1}h^{d+2}I_d + O_P\left(n^{-1/2}h^{d/2+2} + m^{-\beta}h^{d+1}\right)$, where π_{d-1} is the volume of the $d-1$ dimensional unit sphere, and I_d is the identity operator on $\mathcal{H}_1$.

Let $a_n = n^{-1/2}h^{-d/2}$ and $b_n = m^{-\beta}h^{-1}$. Then we have

$$\hat{\mathcal{C}}_x = \pi_{d-1}f(x)d^{-1}h^{d+2}\{I_d + O_P(a_n + b_n)\tilde{\mathcal{A}} + O_P\left(h^2 + b_n\right)\tilde{\mathcal{B}} + O_P(h + b_n)(\tilde{\mathcal{D}}_{12} + \tilde{\mathcal{D}}_{21})\}$$

where $\tilde{\mathcal{A}}$, $\tilde{\mathcal{B}}$, $\tilde{\mathcal{D}}_{12}$ and $\tilde{\mathcal{D}}_{21}$ are operators with norm equal to one, and $\tilde{\mathcal{D}}_{12}$ is the adjoint of $\tilde{\mathcal{D}}_{21}$. With the choice of ϱ, we have $\hat{\mathcal{C}}_x = \pi_{d-1}f(x)d^{-1}h^{d+2}\left\{I_d + O_P(\sqrt{h})\tilde{\mathcal{A}} + O_P(h)\tilde{\mathcal{B}} + O_P(h)(\tilde{\mathcal{D}}_{12} + \tilde{\mathcal{D}}_{21})\right\}$. The same perturbation argument done in Singer & Wu (2012) leads to the desired result. □

Proof of Theorem 2. To reduce notions, let $h = h_{reg}$ and fix $x \in \mathcal{M}\backslash\mathcal{M}_{h_{reg}}$. Let $\{\hat{\varphi}_k\}_{k=1}^d$ be the orthonormal set determined by local FPCA and $\{\phi_k\}_{k=1}^d$ the associated orthonormal basis of $T_x\mathcal{M}$. Let $\{\psi_k\}_{k=1}^{\infty}$ be an orthonormal basis of $\mathcal{L}^2$. Without loss of generality, assume $\mathcal{M}$

is properly rotated and translated so that $x = 0 \in \mathcal{L}^2$ and $\psi_k = \phi_k$ for $k = 1, 2, \ldots, d$. Let $\mathfrak{g} = (g(X_1), g(X_2), \ldots, g(X_n))^T$. Then we have

$$E\{\hat{g}(x) \mid \mathscr{X}\} = e_1^T(\hat{Q}^T\hat{W}\hat{Q})^{-1}\hat{Q}^T\hat{W}\mathfrak{g}.$$

Take $z = \exp_x(t\theta)$, where $t = O(h)$, $\theta \in T_x\mathcal{M}$, $\|\theta\|_{\mathcal{L}^2} = 1$, and $\exp_x$ denotes the exponential map of $\mathcal{M}$ at x. By Theorem 1, we have $\langle\theta, \hat{\varphi}_k\rangle = \langle\theta, \psi_k\rangle + O_P(h_{pca}^{3/2})$ and $\langle\Pi_x(\theta, \theta), \hat{\varphi}_k\rangle = O_P(h_{pca})$. By Lemma A.2.2. of Cheng & Wu (2013), we have

$$t\theta = y - t^2\Pi_x(\theta, \theta)/2 + O(t^3). \tag{2}$$

Therefore, for $k = 1, 2, \ldots, d$, $\langle t\theta, \psi_k\rangle = \langle t\theta, \hat{\varphi}_k - O_P(h_{pca}^{3/2})u_k\rangle = \langle z, \hat{\varphi}_k\rangle - t^2\langle\Pi_x(\theta, \theta), \hat{\varphi}_k\rangle/2 + O_P(hh_{pca}^{3/2} + h^2h_{pca}) = \langle z, \hat{\varphi}_k\rangle + O_P\left(hh_{pca}^{3/2} + h^2h_{pca}\right)$. Since $\theta \in T_x\mathcal{M}$, we have $\theta = \sum_{k=1}^d \langle\theta, \psi_k\rangle\phi_k$. Let $\mathfrak{z} = (\langle z, \hat{\varphi}_1\rangle, \langle z, \hat{\varphi}_2\rangle, \ldots, \langle z, \hat{\varphi}_d\rangle)^T$. By (2), it is easy to see that

$$\begin{aligned} g(z) - g(x) &= t\theta\nabla g(x) + \text{Hess}\, g(x)(\theta, \theta)t^2/2 + O_P(t^3) \\ &= \sum_{k=1}^d \langle t\theta, \psi_k\rangle\nabla_{\phi_k}g(x) + \frac{1}{2}\sum_{j,k=1}^d \langle t\theta, \psi_j\rangle\langle t\theta, \psi_k\rangle\text{Hess}\, g(x)(\phi_j, \phi_k) + O_P(h^3) \\ &= \mathfrak{z}^T\nabla g(x) + \frac{1}{2}\mathfrak{z}^T\text{Hess}\, g(x)\mathfrak{z} + O_P(h^{5/2}). \end{aligned}$$

Due to the smoothness of g, the compactness of $\mathcal{M}$ and the compact support of K, we have $\mathfrak{g} = Q[g(x)\ \nabla g(x)]^T + H/2 + O_P(h^{5/2})$, where $H = [\xi_1^T\text{Hess}\, g(x)\xi_1, \xi_2^T\text{Hess}\, g(x)\xi_2, \ldots, \xi_n^T\text{Hess}\, g(x)\xi_n]^T$, $\xi_i = (\xi_{i1}, \ldots, \xi_{id})^T$, $\xi_{ij} = \langle X_i, \psi_j\rangle$, and

$$Q = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \xi_1 & \xi_2 & \cdots & \xi_n \end{pmatrix}^T.$$

Then the conditional bias is

$$\begin{aligned} E\{\hat{g}(x) - g(x) \mid \mathscr{X}\} &= e_1^T(\hat{Q}^T\hat{W}\hat{Q})^{-1}\hat{Q}^T\hat{W}\mathfrak{g} - g(x) \\ &= e_1^T\left(\frac{1}{n}\hat{Q}^T\hat{W}\hat{Q}\right)^{-1}\frac{1}{n}\hat{Q}^T\hat{W}(Q - \hat{Q})\begin{bmatrix} g(x) \\ \nabla g(x) \end{bmatrix} & (3) \\ &+ e_1^T\left(\frac{1}{n}\hat{Q}^T\hat{W}\hat{Q}\right)^{-1}\frac{1}{n}\hat{Q}^T\hat{W}\left\{\frac{1}{2}H + O_P(h^{5/2})\right\}. & (4) \end{aligned}$$

Now we analyze the term in (3). Let $Z = 1_{\hat{X} \in \mathbb{B}_h^{\mathcal{L}^2}(x)}$. By Lemma S.8, $EZ \asymp h^d$. Then, by Hölder's inequality, for any fixed $\epsilon > 0$, we choose a constant $q > 1$ and a sufficiently large $p > 0$ so that $1/q + 1/p = 1$ and $EZ\|\hat{X} - X\| = (EZ)^{1/q}(E\|\hat{X} - X\|^p)^{1/p} = O(h^{d-\epsilon d}m^{-\beta})$. Therefore,

$$\frac{1}{n}\hat{Q}^T\hat{W}(Q - \hat{Q})\begin{bmatrix} g(x) \\ \nabla g(x) \end{bmatrix} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|)(\xi_i - \hat{\xi}_i)^T\nabla g(x) \\ \frac{1}{n}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|)(\xi_i - \hat{\xi}_i)^T\nabla g(x)\hat{\xi}_i \end{bmatrix} = O_P(h^{-1-\epsilon d}m^{-\beta}), \tag{5}$$

since

$$\left|\frac{1}{n}\sum_{i=1}^{n} K_h(\|\hat{X}_i - x)\|)(\xi_i - \hat{\xi}_i)^T \nabla g(x)\right| \leq \frac{1}{n}\sum_{i=1}^{n} K_h(\|\hat{X}_i - x\|)\|\xi_i - \hat{\xi}_i\|_{\mathbb{R}^d}\|\nabla g(x)\|$$

$$\leq \{\sup_v |K(v)|\}\|\nabla g(x)\| \left(\frac{1}{n}\sum_{i=1}^{n} Z_i \|\xi_i - \hat{\xi}_i\|_{\mathbb{R}^d}\right) = O_P(h^{\ 1\ \ \epsilon d} m^{\ \ \beta}),$$

and similarly, $n^{\ 1}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|)(\xi_i - \hat{\xi}_i)^T\nabla g(x)\hat{\xi}_i = O_P(h^{\ 1\ \ \epsilon d} m^{\ \ \beta})\mathbb{1}_{d\ \ 1}$.

For $\hat{Q}^T\hat{W}\hat{Q}$, a direct calculation shows that

$$\frac{1}{n}\hat{Q}^T\hat{W}\hat{Q} = \begin{bmatrix} n^{\ 1}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|) & n^{\ 1}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|)\hat{\xi}_i^T \\ n^{\ 1}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|)\hat{\xi}_i & n^{\ 1}\sum_{i=1}^n \hat{\xi}_i^T K_h(\|\hat{X}_i - x)\|)\hat{\xi}_i \end{bmatrix}.$$

It is easy to check that $n^{\ 1}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|) = n^{\ 1}\sum_{i=1}^n K_h(\|X_i - x\|) + O_P(h^{\ 1\ \ \epsilon d} m^{\ \ \beta})$, and note that the choice of h ensures that $h^{1+\epsilon d} \gg m^{\ \ \beta}$. Similar calculation shows that $\frac{1}{n}\sum_{i=1}^n K_h(\|\hat{X}_i - x\|)\hat{\xi}_i^T = \frac{1}{n}\sum_{i=1}^n K_h(\|X_i - x\|)\xi_i^T + O_P(h^{\ 1\ \ \epsilon d} m^{\ \ \beta})$ and also

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\xi}_i^T K_h(\|\hat{X}_i - x)\|)\hat{\xi}_i = \frac{1}{n}\sum_{i=1}^{n} \xi_i^T K_h(\|X_i - x\|)\xi_i + O_P(h^{\ 1\ \ \epsilon d} m^{\ \ \beta}).$$

Therefore,

$$\frac{1}{n}\hat{Q}^T\hat{W}\hat{Q} = \frac{1}{n}Q^T W Q + O_P(h^{\ 1\ \ \epsilon d} m^{\ \ \beta})\mathbb{1}_{d\ \ 1}\mathbb{1}_{d\ \ 1}^T, \tag{6}$$

with

$$\frac{1}{n}Q^T W Q = \begin{bmatrix} n^{\ 1}\sum_{i=1}^n K_h(\|X_i - x\|) & n^{\ 1}\sum_{i=1}^n K_h(\|X_i - x\|)\xi_i^T \\ n^{\ 1}\sum_{i=1}^n K_h(\|X_i - x\|)\xi_i & n^{\ 1}\sum_{i=1}^n \xi_i K_h(\|X_i - x\|)\xi_i^T \end{bmatrix}.$$

By Lemma S.4, S.5 and S.6, we have

$$\frac{1}{n}Q^T W Q$$

$$= \Big[\begin{matrix} f(x) & h^2 u_{1,2} d^{\ 1}\nabla f(x)_1^T . \ \ x)_x \ TTf\ Td\ Td\ P\kappa \\ h^2 u_{1,2} d^{\ 1} \ (x) & h^2 u_{1\ 2} d \end{matrix}$$

Then, by the binomial inverse theorem and matrix blockwise inversion, we have

$$\left(\frac{1}{n}\hat{Q}^T\hat{W}\hat{Q}\right)^{\ 1} = \begin{bmatrix} \frac{1}{f(x)} & -\frac{\ ^T f(x)}{f(x)^2} \\ -\frac{f(x)}{f(x)^2} & h^{\ 2}\frac{d}{u_{1,2}f(x)}I_d \end{bmatrix}$$
$$+ \begin{bmatrix} O_P(h^2 + h^{\ 1\ \epsilon d}m^{\ \beta} + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}}) & O_P(h + h^{\ 3\ \epsilon d}m^{\ \beta} + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}\ 1}) \\ O_P(h + h^{\ 3\ \epsilon d}m^{\ \beta} + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}\ 1}) & O_P(h^{\ \frac{1}{2}} + h^{\ 5\ \epsilon d}m^{\ \beta} + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}\ 2}) \end{bmatrix}. \tag{7}$$

Together with (5), it implies that

$$e_1^T(\hat{Q}^T\hat{W}\hat{Q})^{\ 1}\hat{Q}^T\hat{W}\left\{(Q-\hat{Q})\begin{bmatrix} g(x) \\ \nabla g(x)\end{bmatrix}\right\} = O_P(h^{\ 1\ \epsilon d}m^{\ \beta}). \tag{8}$$

Now we analyze (4) with a focus on the term $\hat{Q}^T\hat{W}H$. A calculation similar to those in Lemma S.5 and S.6 shows that $\frac{1}{n}\sum_{i=1}^n K_h(\|X_i - x\|)\,\xi_i^T\text{Hess}\,g(x)\xi_i = h^2u_{1,2}d^{\ 1}f(x)\Delta g(x) + O_P(h^{7/2} + n^{\ 1/2}h^{\ d/2+2})$ and $\frac{1}{n}\sum_{i=1}^n K_h(\|X_i - x\|)\,\xi_i^T\text{Hess}\,g(x)\xi_i\xi_i = O_P(h^4 + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}+3})$. Therefore,

$$\frac{1}{n}Q^TWH = \begin{bmatrix} h^2u_{1,2}d^{\ 1}f(x)\Delta g(x) + O_P(h^{7/2} + n^{\ 1/2}h^{\ d/2+2}) \\ h^4 + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}+3} \end{bmatrix}$$

and hence

$$\frac{1}{n}\hat{Q}^T\hat{W}H = \begin{bmatrix} h^2u_{1,2}d^{\ 1}f(x)\Delta g(x) + O_P(h^{7/2} + n^{\ 1/2}h^{\ d/2+2} + h^{\ 1\ \epsilon d}m^{\ \beta}) \\ O_P(h^4 + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}+3} + h^{\ 1\ \epsilon d}m^{\ \beta}) \end{bmatrix}.$$

The condition on h implies that $n^{\ \frac{1}{2}}h^{\ \frac{d}{2}} \ll 1$. With (7), we conclude that

$$e_1^T(\hat{Q}^T\hat{W}\hat{Q})^{\ 1}\hat{Q}^T\hat{W}\left\{\frac{1}{2}H + O_P(h^3)\right\} = \frac{1}{2d}h^2u_{1,2}\Delta g(x) + O_P(h^3 + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}+2} + h^{\ 1\ \epsilon d}m^{\ \beta}).$$

Combining this equation with (4) and (8) , we immediately see that the conditional bias is

$$E\{\hat{g}(x) - g(x) \mid \mathscr{X}\} = \frac{1}{2d}h^2u_{1,2}\Delta g(x) + O_P(h^3 + n^{\ \frac{1}{2}}h^{\ \frac{d}{2}+2} + h^{\ 1\ \epsilon d}m^{\ \beta}). \tag{9}$$

Now we analyze the conditional variance. Simple calculation shows that

$$\text{var}\{\hat{g}(x) \mid \mathscr{X}\} = n^{\ 1}\sigma_\varepsilon^2 e_1^T(n^{\ 1}\hat{Q}^T\hat{W}\hat{Q})^{\ 1}(n^{\ 1}\hat{Q}^T\hat{W}\hat{W}\hat{Q})(n^{\ 1}\hat{Q}^T\hat{W}\hat{Q})^{\ 1}e_1^T \tag{10}$$

and

$$\frac{1}{n}\hat{Q}^T\hat{W}\hat{W}\hat{Q} = \frac{1}{n}Q^TWWQ + O_P(m^{\ \beta}h^{\ d\ 1\ \epsilon d})\mathbb{1}_{(d+1)\ (d+1)}. \tag{11}$$

In addition,

$$\frac{1}{n}Q^TWWQ = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n K_h^2(\|X_i - x\|) & \frac{1}{n}\sum_{i=1}^n K_h^2(\|X_i - x\|)\,\xi_i^T \\ \frac{1}{n}\sum_{i=1}^n K_h^2(\|X_i - x\|)\,\xi_i & \frac{1}{n}\sum_{i=1}^n K_h^2(\|X_i - x\|)\,\xi_i\xi_i^T \end{bmatrix}.$$

With Lemma S.4, S.5 and S.6, we can show that

$$\begin{aligned}&\frac{h^d}{n}Q^TWWQ\\ &=\begin{bmatrix}u_{2,0}\sigma^2 f(x) & h^2d^{-1}u_{2,2}\sigma^2\nabla f(x)\\ h^2d^{-1}u_{2,2}\sigma^2\nabla^T f(x) & h^2d^{-1}u_{2,2}\sigma^2 f(x)I_d\end{bmatrix}+\begin{bmatrix}O(h^2)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}}) & O_P(h^3+n^{-\frac{1}{2}}h^{-\frac{d}{2}+1})\\ O_P(h^3+n^{-\frac{1}{2}}h^{-\frac{d}{2}+1}) & O_P(h^{\frac{7}{2}}+n^{-\frac{1}{2}}h^{-\frac{d}{2}+2})\end{bmatrix}.\end{aligned}$$

Combined with (7), the above equation implies that

$$n^{-1}\sigma_\varepsilon^2 e_1^T(n^{-1}\hat{Q}^T\hat{W}\hat{Q})^{-1}(n^{-1}Q^TWWQ)(n^{-1}\hat{Q}^T\hat{W}\hat{Q})^{-1}e_1^T=O_P(n^{-1}h^{-d}). \qquad (12)$$

Also,

$$\begin{aligned}&n^{-1}\sigma_\varepsilon^2 e_1^T(n^{-1}\hat{Q}^T\hat{W}\hat{Q})^{-1}\mathbb{1}_{(d+1)\times(d+1)}(n^{-1}\hat{Q}^T\hat{W}\hat{Q})^{-1}e_1^T O_P(h^{-d-1-\epsilon d}m^{-\beta})\\ &=O_P\left(m^{-\beta}n^{-1}h^{-d-1-\epsilon d}(1+h^{-2-\epsilon d}m^{-\beta}+h^{-4-2\epsilon d}m^{-2\beta})\right). \qquad (13)\end{aligned}$$

Combining the above result with (10), (11), (12), (13) and the condition on h, gives the following conditional variance

$$\operatorname{var}\{\hat{g}(x)\mid\mathscr{X}\}=\frac{1}{nh^d}\frac{u_{2,0}\sigma_\varepsilon^2}{f(x)}+O_P\left(n^{-1}h^{-d}(h+n^{-\frac{1}{2}}h^{-\frac{d}{2}})\right). \qquad (14)$$

Finally, the rate for $E[\{\hat{g}(x)-g(x)\}^2\mid\mathscr{X}]$ is derived from (9) and (14). □

Proof of Theorem 3. The proof is similar to the proof for Theorem 2. Below we shall only discuss those that are different. Let $h=h_{reg}$ to reduce notational burden. We first have

$$\begin{aligned}\frac{1}{n}Q^TWQ&=f(x)v\kappa_1 v\\ &\quad+\begin{bmatrix}O(h)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}}) & O(h^2)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}+1})\\ O(h^2)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}+1}) & O(h^3)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}+2})\end{bmatrix}\end{aligned}$$

where

$$\kappa_q=\begin{bmatrix}\kappa_{11,q} & \kappa_{12,q}\\ \kappa_{12,q}^T & \kappa_{22,q}\end{bmatrix},\ \kappa_{22,q}=(\kappa_{22,q,j,k})_{j,k=1}^d,\ v=\begin{bmatrix}1 & 0\\ 0 & hI_d\end{bmatrix}.$$

Then,

$$\begin{aligned}&\frac{1}{n}\hat{Q}^T\hat{W}\hat{Q}=f(x)v\kappa_1 v\\ &+\begin{bmatrix}O(h)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}})+O_P(h^{-1-\epsilon d}m^{-\beta}) & O(h^2)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}+1})+O_P(h^{-1-\epsilon d}m^{-\beta})\\ O(h^2)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}+1})+O_P(h^{-1-\epsilon d}m^{-\beta}) & O(h^3)+O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}+2})+O_P(h^{-1-\epsilon d}m^{-\beta})\end{bmatrix}\end{aligned}$$

and also

$$\begin{aligned}\left(\frac{1}{n}\hat{Q}^T\hat{W}\hat{Q}\right)^{-1}&=v^{-1}\kappa_1^{-1}v^{-1}/f(x)\\ &\quad+\begin{bmatrix}O_P(h+h^{-2-\epsilon d}m^{-\beta}+n^{-\frac{1}{2}}h^{-\frac{d}{2}}) & O_P(1+h^{-4-\epsilon d}m^{-\beta}+n^{-\frac{1}{2}}h^{-\frac{d}{2}-1})\\ O_P(1+h^{-4-\epsilon d}m^{-\beta}+n^{-\frac{1}{2}}h^{-\frac{d}{2}-1}) & O_P(h^{-1}+h^{-5-\epsilon d}m^{-\beta}+n^{-\frac{1}{2}}h^{-\frac{d}{2}-2})\end{bmatrix}.\end{aligned}$$

This implies that

$$e_1^T(\hat{Q}^T\hat{W}\hat{Q})^{\ 1}\hat{Q}^T\hat{W}\left\{(Q-\hat{Q})\begin{bmatrix} g(x) \\ \nabla g(x)\end{bmatrix}\right\} = e_1^T\begin{bmatrix} 1 & h^{\ 1} \\ h^{\ 1} & h^{\ 2}\end{bmatrix} O_P(h^{\ 1\ \ \epsilon d} m^{\ \beta})\mathbb{1}_{d\ 1}$$
$$= O_P(h^{\ 1\ \ \epsilon d} m^{\ \beta}).$$

By Lemma (4), (5) and (6), one can(

S.4. TECHNICAL LEMMAS

Here we collect some technical lemmas that will be used in the proofs of the main theorems. The lemma below is used to establish Proposition 1. Its proof depends on Lemma 2 and 3.

LEMMA 1. *Suppose $h_0 \to 0$, and $mh_0 \to \infty$. For any $p \geq 1$, assume $E|\zeta|^p < \infty$. Under the assumptions (A1)–(A3), for the estimate $\hat{X}$ in (3) with a proper choice of δ,*

$$\{E(\|\hat{X} - X\|^p_{\mathcal{L}^2} \mid X)\}^{1/p} = O\left(m^{-1/2}h_0^{-1/2}\right)\left\{\sup_t |X(t)| + L_X\right\} + O(h_0^{\nu})L_X, \quad (17)$$

where $O(\cdot)$ does not depend on X.

Proof. In order to reduce notations, let $h = h_0$. Denoting $\Delta = \delta 1_{S_0S_2-S_1^2<\delta}$ with $\delta = m^{-2}$, according to (3), we have

$$\hat{X}(t) - X(t) = \frac{S_2(R_0 - S_0X)}{S_0S_2 - S_1^2 + \Delta} - \frac{S_1(R_1 - S_1X)}{S_0S_2 - S_1^2 + \Delta} - \frac{\Delta X}{S_0S_2 - S_1^2 + \Delta} \equiv I_1 + I_2 + I_3.$$

Therefore,

$$\|\hat{X} - X\|^p \leq c_p(\|I_1\|^p + \|I_2\|^p + \|I_3\|^p) \quad (18)$$

for some constant c_p depending on p only.

For I_1, we have

$$\begin{aligned}\|I_1\|^p &= \left[\int_D \left\{\frac{S_2(R_0 - S_0X)}{S_0S_2 - S_1^2 + \Delta}\right\}^2 \mathrm{d}t\right]^{p/2} \leq \left[\int_D \{S_2(R_0 - S_0X)\}^4 \mathrm{d}t \int_D \left(\frac{1}{S_0S_2 - S_1^2 + \Delta}\right)^4\right]^{p/4} \\ &\leq \left\{\int_D S_2^8 \mathrm{d}t \int_D (R_0 - S_0X)^8 \mathrm{d}t\right\}^{p/8} \left\{\int_D \left(\frac{1}{S_0S_2 - S_1^2 + \Delta}\right)^4\right\}^{p/4}.\end{aligned}$$

This also shows that, for $p \geq 2$,

$$\begin{aligned}&E(\|I_1\|^p \mid X) \\ &\leq \left(E\left\{\left[\int_D S_2^8 \mathrm{d}t \int_D (R_0 - S_0X)^8 \mathrm{d}t\right]^{p/4} \mid X\right\}\right)^{1/2} \left(E\left\{\left[\int_D \left(\frac{1}{S_0S_2 - S_1^2 + \Delta}\right)^4 \mathrm{d}t\right]^{p/2} \mid X\right\}\right)^{1/2} \\ &\leq \left(E\left[\int_D S_2^{4p} \mathrm{d}t\right] E\left[\int_D (R_0 - S_0X)^{4p} \mathrm{d}t \mid X\right]\right)^{1/4} \left(E\left[\int_D \left\{\frac{1}{S_0S_2 - S_1^2 + \Delta}\right\}^{2p} \mathrm{d}t\right]\right)^{1/2} \\ &= \left\{\int_D E(S_2^{4p}) \mathrm{d}t\right\}^{1/4} \left[\int_D E\{(R_0 - S_0X)^{4p} \mid X\} \mathrm{d}t\right]^{1/4} \left[\int_D E\left\{\left(\frac{1}{S_0S_2 - S_1^2 + \Delta}\right)^{2p}\right\} \mathrm{d}t\right]^{1/2}.\end{aligned} \quad (19)$$

Let $E_{0,X} = E(R_0 - S_0 X \mid X)$ and ℓ be the largest integer strictly less than ν. By Taylor expansion,

$$E_{0,X} = E\left[\frac{1}{mh}\sum_{j=1}^{m} K\left(\frac{T_j - t}{h}\right)\{X(T_j) + \zeta_j - X(t)\} \mid X\right]$$
$$= \frac{1}{mh}\sum_{j=1}^{m} E\left[K\left(\frac{T_j - t}{h}\right)\left\{\frac{X^{(\ell)}(t + \tau_j(T_j - t)) - X^{(\ell)}(t)}{\ell!}(T_j - t)^{\ell}\right\} \mid X\right],$$

where $\tau_j \in [0,1]$. Hence, with u_ν denoting $\int_{-1}^{1} K(s)|s|^{\nu}\mathrm{d}s$, we have

$$|E_{0,X}| \le \frac{1}{mh}\sum_{j=1}^{m} E\left\{K\left(\frac{T_j - t}{h}\right)\left|\frac{X^{(\ell)}(t + \tau_j(T_j - t)) - X^{(\ell)}(t)}{\ell!}(T_j - t)^{\ell}\right| \mid X\right\},$$
$$\le \frac{1}{h\ell!}L_X E\left\{K\left(\frac{T - t}{h}\right)|T - t|^{\nu}\right\} \le \frac{L_X C_{T,2}}{\ell!}h^{\nu}u_{\nu}. \tag{20}$$

Let $\sigma_{r,X} = E\left(\left[h^{-1}K\{(T_j - t)/h\}\{X(T_j) + \zeta_j - X(t)\} - E_{0,X}\right]^r \mid X\right)$. Then, for $r \ge 2$,

$$\sigma_{r,X} = E\left(\left[\frac{1}{h}K\left(\frac{T_j - t}{h}\right)\{X(T_j) + \zeta_j - X(t)\} - E_{0,X}\right]^r \mid X\right)$$
$$\le 3^r E\left[\left\{\frac{1}{h}K\left(\frac{T_j - t}{h}\right)|X(T_j) - X(t)|\right\}^r \mid X\right] + 3^r E\left[\left\{\frac{1}{h}K\left(\frac{T_j - t}{h}\right)|\zeta_j|\right\}^r \mid X\right] + 3^r|E_{0,X}|^r$$
$$\le 2\cdot 3^r\left\{\sup_t |X(t)|^r\right\}E\left[\left\{\frac{1}{h}K\left(\frac{T_j - t}{h}\right)\right\}^r\right] + 3^r E\left[\left\{\frac{1}{h}K\left(\frac{T_j - t}{h}\right)\right\}^r|\zeta_j|^r\right] + 3^r|E_{0,X}|^r$$
$$\le 2\cdot 3^r\left\{\sup_t |X(t)|^r\right\}h^{1-r}C_{T,2}^r + 3^r h^{1-r}C_{T,2}^r E|\zeta|^r + 3^r L_X^r C_{T,2}^r h^{r\nu}u_\nu^r. \tag{21}$$

With (20) and (21), by Lemma 2, conditioning on X, we have $E\left\{(R_0 - S_0X - E_{0,X})^{4p} \mid X\right\} \le c_1(p)C_{T,2}^{4p}\{\sup_t |X(t)|^{4p} + L_X^{4p}u_\nu^{4p}\}m^{-2p}h^{-2p}$, where $c_j(p)$ denote a constant depending on p only for any j. This implies that

$$E\left\{(R_0 - S_0X)^{4p} \mid X\right\} \le 2^{4p}E\left\{(R_0 - S_0X - E_{0,X})^{4p} \mid X\right\} + 2^{4p}E\left\{(E_{0,X})^{4p} \mid X\right\}$$
$$\le 2^{4p}c_1(p)C_{T,2}^{4p}\{\sup_t |X(t)|^{4p} + L_X^{4p}u_\nu^{4p}\}m^{-2p}h^{-2p} + (2C_{T,2}u_\nu)^{4p}h^{4p\nu}L_X^{4p}. \tag{22}$$

By a similar argument, we can show that $E(S_2 - ES_2)^{4p} \le c_2(p)m^{-2p}h^{-2p}$. Also, it is easy to check that $C_{T,1}u_2 \le ES_2 \le C_{T,2}u_2$ with u_q denoting $\int_{-1}^{1} K(u)|u|^q\mathrm{d}u$ and hence

$$E(S_2^{4p}) \le 2^{4p}E(S_2 - ES_2)^{4p} + 2^{4p}|ES_2|^{4p} \le C_{T,2}^{4p}u_2^{4p} + c_2(p)m^{-2p}h^{-2p} = O(1). \tag{23}$$

The same argument leads to $E\{S_0S_2 - S_1^2 - E(S_0S_2 - S_1^2)\}^{2p} \le c_3(p)m^{-p}h^{-p}$. Note that $\inf_t E(S_0S_2 - S_1^2) > 0$ so that $\{E(S_0S_2 - S_1^2)\}^{-1} = O(1)$. By Lemma 3, this also implies that

$$\int_D E\left\{\left(\frac{1}{S_0S_2 - S_1^2 + \Delta}\right)^{2p}\right\}\mathrm{d}t = O(1) + O(m^{-p}h^{-p} + m^{-4p}) = O(1). \tag{24}$$

Puting (19), (22), (23) and (24) together, we conclude that

$$E(\|I_1\|^p \mid X) = c_4(p) \left[\{\sup_t |X(t)|^p + L_X^p\} m^{\ p/2} h^{\ p/2} + h^{p\nu} L_X^p \right]. \tag{25}$$

The same rate for I_2 can be obtained in a similar fashion. For I_3, we have

$$\|I_3\|^p = \left\{ \int_D \left(\frac{\Delta X}{S_0 S_2 - S_1^2 + \Delta} \right)^2 \mathrm{d}t \right\}^{p/2} \leq \Delta^p \sup_t |X(t)|^p \left\{ \int_D \left(\frac{1}{S_0 S_2 - S_1^2 + \Delta} \right)^2 \mathrm{d}t \right\}^{p/2}.$$

Therefore, with (24),

$$E(\|I_3\|^p \mid X) \leq \Delta^p \sup_t |X \ |X(\ ^2$$

$(a^r + b^r)h^{1\quad r}$ for $r \geq 2$. Therefore,

$$
\begin{aligned}
E\left(\sum_{j=1}^{N} V_j\right)^{2p} &\leq \sum_{q=1}^{p} \binom{N}{q} (2p)^{q\quad 1} \binom{2p}{k_1, k_2, \ldots, k_N} 2^q (a^{2p} h^{q\quad 2p} + b^{2p} h^{q\quad 2p}) \\
&\leq (2p)^{p\quad 1} (2p)! 2^p (a^{2p} + b^{2p}) \sum_{q=1}^{p} \frac{N!}{(N-q)!q!} h^{q\quad 2p} \\
&\leq (2p)^{p\quad 1} (2p)! p 2^p (a^{2p} + b^{2p}) \sum_{q=1}^{p} N^p h^{q\quad 2p} \\
&\leq (2p)^{p\quad 1} (2p)! p 2^p (a^{2p} + b^{2p}) N^p h^{\quad p}.
\end{aligned}
$$

Multiplying both sides by $N^{\quad 2p}$ yields the conclusion of the lemma. □

LEMMA 3. *Suppose $D \subset \mathbb{R}^p$ is compact set and $S_N(t)$ for $t \in D$ is a sequence of random processes defined on D. For $b_N > 0$, define $\delta_N(t) = 2b_N 1_{\ S(t)\quad b_N}$. Suppose for some constant c_0, $0 < c_0 \leq ES_N(t) < \infty$ for all t and sufficiently large N. Also, assume* $\lim_N \quad \inf_{t\quad D} ES_N(t) > 0$. *For a sequence of $a_N \to 0$ and any $p > 0$, if $b_N^{\quad r} a_N^p \leq 1$ for some p and $r > 0$, and*

(1) if $E\left(\sup_{t\quad D} |S_N(t) - ES(t)|^p\right) = O(a_N^p)$, then we have

$$
E \sup_{t\quad D} \left| \frac{1}{S_N(t) + \delta_N(t)} - \frac{1}{ES_N(t)} \right|^r = O(a_N^r + b_N^r);
$$

(2) if $\int_D E\left(|S_N(t) - ES_N(t)|^p\right) \mathrm{d}t = O(a_N^p)$, then we have

$$
E \int \left| \frac{ES_N(t)}{S_N(t) + \delta_N(t)} - 1 \right|^r \mathrm{d}t = O(a_N^r + b_N^r).
$$

Proof. From now on, we shall supress N when there is no confusion raised. Let $\tilde{S}(t) = S(t) + \delta(t)$ and $v(t) = E\tilde{S}(t) - E\delta(t) = ES(t)$. For a fixed t that is suppressed below,

$$
\begin{aligned}
\left| \frac{v}{\tilde{S}} - 1 \right|^r &\leq \left(\frac{|\tilde{S} - v|^r}{|\tilde{S}|^r} \right) 1_{\ \tilde{S}\quad v\quad v/2} + \left(\frac{|\tilde{S} - v|^r}{|\tilde{S}|^r} \right) 1_{\ \tilde{S}\quad v > v/2} \\
&\leq (v/2)^{\quad r} \left| \tilde{S} - v \right|^r + b_N^{\quad r} \left| \tilde{S} - v \right|^r 1_{\ \tilde{S}\quad v > v/2} \\
&\leq c_r (v/2)^{\quad r} |S - ES|^r + c_r (v/2)^{\quad r} \delta^r + b_N^{\quad r} \left| \tilde{S} - v \right|^r 1_{\ \tilde{S}\quad v > v/2} \\
&\equiv I_1 + I_2 + I_3,
\end{aligned}
$$

where $c_r > 0$ is a constant independent of t.

(1) By assumption, $E \sup_t I_1 = O(a_N^r)$. Since $|\delta| \leq 2b_N$, we have $E \sup_t I_2 = O(b_N^r)$. Also,

$$
E \sup_t I_3 \leq c_p b_N^{\quad r} \sup_t (v/2)^{\quad p} E(\sup_t |S - v|^{p+r} + \sup_t |\delta|^{p+r}) = O(b_N^{\quad r} a_N^{p+r}) = O(a_N^r + b_N^r)
$$

for a sufficiently large p, and a constant $c_p > 0$.

(2) By the assumption, $\int I_1 \mathrm{d}t = O(a_N^r)$. Since $|\delta| \leq 2b_N$, we have $\int I_2 \mathrm{d}t = O(b_N^r)$. Also,

$$
\int I_3 \mathrm{d}t \leq b_N^{\quad r} \int \{v(t)/2\}^{\quad p} E|\tilde{S}(t) - v(t)|^{p+r} \mathrm{d}t = O(b_N^{\quad r} a_N^{p+r}) = O(a_N^r + b_N^r).
$$

Therefore, the conclusion of part (2) follows. □

In order to state the following lemmas, we shall first establish some notations and convention. Let $u_{q,k} = \int_{\mathbb{B}_1^{\mathbb{R}}(0)} K^q(\|u\|_{\mathbb{R}^d})\|u\|_{\mathbb{R}^d}^k \mathrm{d}u$. We also identify $T_x\mathcal{M}$ with $\mathbb{R}^d$. Let $\mathbb{D}_{x,h}$ denote the set $\{\theta \in \mathbb{R}^d : \exp_x(\theta) \in \mathbb{B}_h(x)\}$, where $\exp_x$ denotes the exponential map at x. Let $\kappa_{11,q} = \int_{h^{-1}\mathbb{D}_{x,h}} K^q(\|u\|_{\mathbb{R}^d})\mathrm{d}u$, $\kappa_{12,q,j} = \int_{h^{-1}\mathbb{D}_{x,h}} K^q(\|u\|_{\mathbb{R}^d})\theta_j \mathrm{d}u$ and $\kappa_{22,q,j,k} = \int_{h^{-1}\mathbb{D}_x} K^q\left(h^{-1}\|\theta\|_{\mathbb{R}^d}\right)\theta_j\theta_k \mathrm{d}\theta$, where θ_j denotes the jth component of θ. Let π_{d-1} denote the volume of the unit sphere S^{d-1}. The following three lemmas are based on Lemma A.2.5 of Cheng & Wu (2013) and hence their proofs are omitted.

LEMMA 4. *Suppose K is a kernel function compactly supported in $[-1,1]$ and continuously differentiable in $[0,1]$. Let $h \geq h_{pca}$.*

1. *If $x \in \mathcal{M}\backslash\mathcal{M}_h$, then*

$$n^{-1}\sum_{i=1}^{n} h^{-d}K^q\left(\frac{\|X_i - x\|_{\mathcal{L}^2}}{h}\right) = u_{q,0}f(x) + O(h^2) + O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}})$$

2. *If $x \in \mathcal{M}_h$, then*

$$n^{-1}\sum_{i=1}^{n} h^{-d}K^q\left(\frac{\|X_i - x\|_{\mathcal{L}^2}}{h}\right) = f(x)\kappa_{11,q} + O(h) + O_P(n^{-\frac{1}{2}}h^{-\frac{d}{2}}).$$

LEMMA 5. *Suppose K is a kernel function compactly supported in $[-1,1]$ and continuously differentiable in $[0,1]$. Let $h \geq h_{pca}$ and $\hat{\varphi}_k$ be the estimate in Theorem 1. Then,*

(1) if $x \in \mathcal{M}\backslash\mathcal{M}_h$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} h^{-d}K^q\left(\frac{\|X_i - x\|_{\mathcal{L}^2}}{h}\right)\langle X_i - x, \hat{\varphi}_k\rangle \\ &= h^2 u_{q,1} d^{-1}\nabla_{\phi_k} f(x) + O_P(h^3 + n^{-\frac{1}{2}}h^{-\frac{d}{2}+1} + h^2 h_{pca}^{3/2} + h^3 h_{pca}).\end{aligned}$$

(2) if $x \in \mathcal{M}_h$,

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} h^{-d}K^q\left(\frac{\|X_i - x\|_{\mathcal{L}^2}}{h}\right)\langle X_i - x, \hat{\varphi}_k\rangle \\ &= h\kappa_{12,q,k}(x) + O_P(h^2 + n^{-\frac{1}{2}}h^{-\frac{d}{2}+1} + h h_{pca}^{3/2} + h^2 h_{pca}).\end{aligned}$$

LEMMA 6. *Suppose K is a kernel function compactly supported in $[-1,1]$ and continuously differentiable in $[0,1]$. Let $h \geq h_{pca}$.*

(1) If $x \in \mathcal{M}\backslash\mathcal{M}_h$, then

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n} h^{-d}K^q\left(\frac{\|X_i - x\|_{\mathcal{L}^2}}{h}\right)\langle X_i - x, \hat{\varphi}_j\rangle\langle X_i - x, \hat{\varphi}_k\rangle \\ &= \begin{cases} h^2 u_{q,2} d^{-1} f(x) + O_P(h^{\frac{7}{2}} + n^{-\frac{1}{2}}h^{-\frac{d}{2}+2}) & \textit{if } 1 \leq j = k \leq d \\ O_P(h^{\frac{7}{2}} + n^{-\frac{1}{2}}h^{-\frac{d}{2}+2}) & \textit{otherwise}.\end{cases}\end{aligned}$$

(2) If $x \in \mathcal{M}_h$, then

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n} h^{-d}K^q\left(\frac{\|X_i - x\|_{\mathcal{L}^2}}{h}\right)\langle X_i - x, \hat{\varphi}_j\rangle\langle X_i - x, \hat{\varphi}_k\rangle \\
&= h^2 f(x)\kappa_{22,q,j,k}(x) + O_P(h^3 + n^{-\frac{1}{2}}h^{-\frac{d}{2}+2} + h^2 h_{pca}^{3/2}).
\end{aligned}
$$

In order to prove Theorem 1, we establish the following auxiliary lemmas.

LEMMA 7. *Let $\tilde{G} = \hat{G} + \Delta$ and $\check{G} = G + \Delta$ with $\Delta = 1/\log m$. Then $E|\log\tilde{G} - \log\check{G}| = o(1)$. This result also holds for $\hat{\mathcal{C}}_{\tilde{x}}$, if $\tilde{x}$ is independent of $\hat{X}_1, \ldots, \hat{X}_n$, and that $\{E\|\tilde{x} - x\|^p\}^{1/p} = O(m^{-\beta})$ for all $p \geq 1$.*

Proof. By Jensen's inequality and the concavity of $\log(\cdot)$,

$$
\begin{aligned}
E(\log\check{G} - \log\tilde{G}) &\leq \log E\frac{\check{G}}{\tilde{G}} = \log E\frac{\|X - x\| + \Delta}{\|\hat{X} - x\| + \Delta} \leq \log E\frac{\|\hat{X} - x\| + \|\hat{X} - X\| + \Delta}{\|\hat{X} - x\| + \Delta} \\
&\leq \log(1 + \Delta^{-1}E\|\hat{X} - X\|) \equiv a_n
\end{aligned}
$$

with $a_n \geq 0$ and $a_n \to 0$. For the other direction, we first observe that

$$
\begin{aligned}
E\left(\frac{\tilde{G}}{\check{G}}\right)^{1/4} &= E\left\{\frac{1}{\check{G}^{1/4}}E(\tilde{G}^{1/4} \mid X)\right\} \leq E\left[\frac{1}{\check{G}^{1/4}}E\{\|\hat{X} - X\|^{1/4} + (\|X - x\| + \Delta)^{1/4} \mid X\}\right] \\
&\leq E\left[\frac{C_1^{1/4}m^{-\beta/4}\{\eta(X)\}^{1/4}}{\check{G}^{1/4}} + 1\right] \leq 1 + \Delta^{-1/4}E\left[C_1^{1/4}m^{-\beta/4}\{\eta(X)\}^{1/4}\right] \\
&= 1 + O\left(m^{-\beta/4}(\log m)^{1/4}\right)
\end{aligned}
$$

where $C_1 > 0$ is some constant. This implies that

$$
\frac{1}{4}E(\log\tilde{G} - \log\check{G}) = E\log\left(\frac{\tilde{G}}{\check{G}}\right)^{1/4} \leq \log E\left(\frac{\tilde{G}}{\check{G}}\right)^{1/4} \equiv \frac{b_n}{4}
$$

with $b_n \geq 0$ and $b_n \to 0$, or equivalently,

$$
E(\log\check{G} - \log\tilde{G}) \geq -b_n.
$$

Therefore $E|\log\check{G} - \log\tilde{G}| \leq a_n + b_n = o(1)$. Following almost the same lines, we can deduce the same result for $\tilde{G}(\tilde{x})$, i.e., the quantity $\tilde{G}(x)$ when x is replaced with $\tilde{x}$. □

LEMMA 8. *Let $\tilde{x}$ be an estimate of x such that $\tilde{x}$ is independent of X and $\hat{X}$, and that $\{E\|\tilde{x} - x\|^p\}^{1/p} = O(m^{-\beta})$ for all $p \geq 1$. Suppose $0 < a < \beta$, $h \gtrsim m^{-a}$ and $h_+ = h + m^{-(\beta+a)/2}$. Let $\tilde{Z} = 1_{\hat{X} \in \mathbb{B}_h^{\mathcal{L}^2}(\tilde{x})}$ and $V = 1_{X \in \mathbb{B}_{h_+}^{\mathcal{L}^2}(x)}$. If F is a positive functional of X and $\hat{X}$ such that $E\{F(X, \hat{X})V\} = O(h^b)$ for some $b \geq 0$, and $E\{F(X, \hat{X})\}^q < \infty$ for some $q > 1$, then we have $E\{F(X, \hat{X})|\tilde{Z} - V|\} = O(h^b)$ and $E\{F(X, \hat{X})\tilde{Z}\} = O(h^b)$. Also, $E\{F(X, \hat{X})|Z - \tilde{Z}|\} = o(h^b)$ with Z denoting $1_{\hat{X} \in \mathbb{B}_h^{\mathcal{L}^2}(x)}$.*

Proof. Let $\kappa = m^{-(\beta+a)/2}$ and $\tilde{V} = 1_{X \in \mathbb{B}_{h_+}^{\mathcal{L}^2}(\tilde{x})}$. Choose $r > 1$ such that $r^{-1} + q^{-1} = 1$. To reduce notational burden, we simply use F to denote $F(X, \hat{X})$.

We shall first establish that $E(F|V - \tilde{V}|) = O(h^b)$. To this end, we observe that

$$E(F|V - \tilde{V}|) \leq E(FV1_{\tilde{V}=0}) + E(F\tilde{V}1_{V=0}).$$

For the first term, for any fixed $s \geq 2rab/(\beta - a)$, we have

$$\begin{aligned}
E(FV1_{\tilde{V}=0}) &= E\{F1_{\ X \quad x \quad h,\ X \quad \tilde{x} \quad h}\} \\
&\leq E\{F1_{\ \tilde{x} \quad x \quad \kappa,\ X \quad x \quad h} + F1_{h \quad \kappa \quad X \quad x \quad h}\} \\
&\leq E\{F1_{\ \tilde{x} \quad x \quad \kappa}\} + E\{FV\} \\
&\leq \{EF^q\}^{1/q} \mathrm{pr}\left(\|\tilde{x} - x\| \geq \kappa\right)^{1/r} + O(h^b) \\
&\leq \{EF^q\}^{1/q} \left(m^{s(\beta+a)/2} E\|\tilde{x} - x\|^s\right)^{1/r} + O(h^b) \\
&= O(m^{s(\beta+a)/(2r) \quad s\beta/r} + h^b) = O(h^b).
\end{aligned}$$

Similar result can be derived for the second term. Thus, we prove that $E(F|V - \tilde{V}|) = O(h^b)$.

Define $h \quad = h - \kappa$, $\tilde{U} = 1_{\ X \quad \mathbb{B}^{\mathcal{L}^2}_{h_-}(\tilde{x})}$. Note that $\tilde{U} \leq \tilde{V}$. Then, by Hölder inequality, we have

$$\begin{aligned}
E(F|\tilde{Z} - \tilde{V}|) &= E(F1_{\tilde{Z}=1}1_{\tilde{V}=0}) + E(F1_{\tilde{Z}=0}1_{\tilde{U}=1}) + E(F1_{\tilde{Z}=0}1_{\tilde{V}=1}1_{\tilde{U}=0}) \\
&\leq E(F1_{\tilde{Z}=1}1_{\tilde{V}=0}) + E(F1_{\tilde{Z}=0}1_{\tilde{U}=1}) + E(F\tilde{V}1_{\tilde{U}=0}) \\
&\leq (EF^q)^{1/q} \left\{(E1_{\tilde{Z}=1}1_{\tilde{V}=0})^{1/r} + (E1_{\tilde{Z}=0}1_{\tilde{U}=1})^{1/r}\right\} + O(h^b) \\
&\leq 2(EF^q)^{1/q}\{\mathrm{pr}(\|X - \hat{X}\| \geq m^{\quad (\beta+a)/2})\}^{1/r} + O(h^b) \\
&\leq 2(EF^q)^{1/q} \left(m^{s(\beta+a)/2} E\|X - \hat{X}\|^s\right)^{1/r} + O(h^b) \\
&= O\left(m^{s(\beta+a)/(2r) \quad s\beta/r}\right) + O(h^b) = O(h^b).
\end{aligned}$$

Then $E(F|\tilde{Z} - V|) \leq E(F|\tilde{Z} - \tilde{V}|) + E(F|\tilde{V} - V|) = O(h^b)$. Since $|E(F\tilde{Z}) - E(FV)| \leq E(F|\tilde{Z} - V|)$, the result $E(F\tilde{Z}) = O(h^b)$ follows. $\square$

LEMMA 9. *Suppose $\{\psi_k\}_{k=1}$ is an orthonormal basis of $\mathcal{H}$ and $x \in \mathcal{M}$ is fixed. Assume that $\psi_1, \ldots, \psi_d$ span the tangent space $T_x\mathcal{M}$. Let $\pi_{d \quad 1}$ be the volume of the $d - 1$ dimensional unit sphere $S^{d \quad 1}$ and $\hat{\mathcal{C}}_x$ the sample covariance operator based on $\hat{\mathcal{N}}_{\mathcal{L}^2}(h, x)$ for some $h \gtrsim m^{\quad a}$ with $0 < a < \beta$. Then,*

$$\begin{aligned}
\sup_{j \quad d} \sup_{k \quad d+1} \left|\langle \hat{\mathcal{C}}_x \psi_j, \psi_k \rangle\right| &= O_P\left(h^{d+4} + n^{\quad 1/2} h^{d/2+3} + m^{\quad \beta} h^{d+1}\right), \\
\sup_{j,k \quad d+1} \left|\langle \hat{\mathcal{C}}_x \psi_j, \psi_k \rangle\right| &= O_P\left(h^{d+4} + n^{\quad 1/2} h^{d/2+4} + m^{\quad \beta} h^{d+1}\right), \\
\sup_{1 \quad j=k \quad d} \left|\langle \hat{\mathcal{C}}_x \psi_j, \psi_k \rangle\right| &= O_P\left(h^{d+3} + n^{\quad 1/2} h^{d/2+3} + m^{\quad \beta} h^{d+1}\right),
\end{aligned}$$

for $1 \leq k \leq d:$ $\qquad \langle \hat{\mathcal{C}}_x \psi_k, \psi_k \rangle = \pi_{d \quad 1} f(x) d^{\quad 1} h^{d+2} + O_P\left(n^{\quad 1/2} h^{d/2+2} + m^{\quad \beta} h^{d+1}\right).$

The above results hold also for $\hat{\mathcal{C}}_{\tilde{x}}$, if $\tilde{x}$ is independent of $\hat{X}_1, \ldots, \hat{X}_n$, and that $\{E\|\tilde{x} - x\|^p\}^{1/p} = O(m^{\quad \beta})$ for all $p \geq 1$.

Proof. Denote $Z_i = 1_{\hat{X}_i \in \mathbb{B}_h^{\mathcal{L}^2}(x)}$. Then $\hat{\mathcal{C}}_x$ can be written as $\hat{\mathcal{C}}_x = n^{-1}\sum_{i=1}^n (\hat{X}_i - \hat{\mu}_x) \otimes (\hat{X}_i - \hat{\mu}_x) Z_i$, where $\hat{\mu}_x = \sum_{i=1}^n \hat{X}_i Z_i$. For any y, z such that $\|y\|_{\mathcal{L}^2} = \|z\|_{\mathcal{L}^2} = 1$, we have

$$
\begin{aligned}
\langle \hat{\mathcal{C}}_x y, z\rangle =& \langle n^{-1}\sum_{i=1}^n Z_i(\hat{X}_i - \hat{\mu}_x) \otimes (\hat{X}_i - \hat{\mu}_x) y, z\rangle = n^{-1}\sum_{i=1}^n \langle \hat{X}_i - \hat{\mu}_x, y\rangle \langle \hat{X}_i - \hat{\mu}_x, z\rangle Z_i \\
=& n^{-1}\sum_{i=1}^n \langle X_i - \mu_x, y\rangle\langle X_i - \mu_x, z\rangle Z_i + n^{-1}\sum_{i=1}^n \langle (\hat{X}_i - X_i) - (\hat{\mu}_x - \mu_x), y\rangle \langle \hat{X}_i - \hat{\mu}_x, z\rangle Z_i \\
&+ n^{-1}\sum_{i=1}^n \langle \hat{X}_i - \hat{\mu}_x, y\rangle \langle (\hat{X}_i - X_i) - (\hat{\mu}_x - \mu_x), z\rangle Z_i \\
&+ n^{-1}\sum_{i=1}^n \langle (\hat{X}_i - X_i) - (\hat{\mu}_x - \mu_x), y\rangle \langle (\hat{X}_i - X_i) - (\hat{\mu}_x - \mu_x), z\rangle Z_i \\
\equiv& I_1 + I_2 + I_3 + I_4,
\end{aligned}
$$

where $\mu_x = \sum_{i=1}^n X_i Z_i$. Before we proceed to analyze I_1, I_2, I_3 and I_4, we prepare some calculations.

First, it can be checked that

$$
\|\hat{\mu}_x - \mu_x\| = \left\| \frac{1}{n}\sum_{i=1}^n (\hat{X}_i - \mu_x) \right.
$$

from triangle inequality, and the fifth is based on (28). Now, let $h_1 = h + m^{(\beta+a)/2}$ and $V_i = 1_{X_i \ \mathbb{B}_{h_1}^{\mathcal{L}^2}(x)}$. Based on the assumption (B3), $E($

for $1 \le j \le d$,

$$\begin{aligned}
E \sup_{k\quad d+1} |\Xi_{j,k}| \le & E \sup_{k\quad d+1} \left| n^{-1} \sum_{i=1}^{n} \langle X_i - \mu_x, \psi_j \rangle \langle X_i - \mu_x, \psi_k \rangle Z_i \right| \\
\le & E \sup_{k\quad d+1} \left| n^{-1} \sum_{i=1}^{n} \langle X_i, \psi_j \rangle \langle X_i, \psi_k \rangle Z_i \right| \\
& + E \sup_{k\quad d+1} \left| n^{-1} \sum_{i=1}^{n} \langle \mu_x, \psi_j \rangle \langle \mu_x, \psi_k \rangle Z_i \right| \\
\equiv & I_1 + I_2. \qquad (36)
\end{aligned}$$

It is seen that I_1 is the dominant term, which we evaluate below (utilizing the fact that $\Pi_x(\theta, \theta) \perp T_x\mathcal{M}$):

$$\begin{aligned}
I_1 &= E \sup_{k\quad d+1} \left| n^{-1} \sum_{i=1}^{n} \langle X_i, \psi_j \rangle \langle X_i, \psi_k \rangle Z_i \right| \le \frac{1}{n} E \sum_{i=1}^{n} \sup_{k\quad d+1} |\langle X_i, \psi_j \rangle \langle X_i, \psi_k \rangle Z_i| \\
&= E \sup_{k\quad d+1} |\langle X_i, \psi_j \rangle \langle X_i, \psi_k \rangle Z_i| = E \sup_{k\quad d+1} |\langle X_i, \psi_j \rangle \langle \mathcal{P}_2 X_i, \psi_k \rangle Z_i| . \\
&\le E \left(|\langle X_i, \psi_j \rangle| \|\mathcal{P}_2 X_i\| Z_i \right). \qquad (37)
\end{aligned}$$

Since by Lemma A.2.4 of Cheng & Wu (2013),

$$\begin{aligned}
& E\, |\langle X_i, \psi_j \rangle| \|\mathcal{P}_2 X_i\| V_i| \\
& \le \int_{S^{d-1}} \int_0^{h_1} \langle t\theta, \psi_j \rangle \|t^2 \Pi_x(\theta, \theta)\|_{\mathcal{L}^2} f(\exp(t\theta)) t^{d-1} \mathrm{d}t \mathrm{d}\theta + O(h^{d+5}) \\
& = O(h^{d+3}),
\end{aligned}$$

we can apply Lemma 8 to conclude $E \left(|\langle X_i, \psi_j \rangle| \|\mathcal{P}_2 X_i\| Z_i \right) = O(h^{d+3})$, and hence with (37), we assert that $I_1 = O(h^{d+3})$. This proves (31). The result (32) is obtained in a similar way.

For (33), by the same argument that leads to (36), we can show that for $1 \le j \ne k \le d$, $E\Xi_{j,k}$ is dominated by

$$E n^{-1} \sum_{i=1}^{n} \langle X_i, \psi_j \rangle \langle X_i, \psi_k \rangle Z_i = E \langle X_i, \psi_j \rangle \langle X_i, \psi_k \rangle Z_i.$$

Now, because

$$\begin{aligned}
& E \langle X_i, \psi_j \rangle \langle X_i, \psi_k \rangle V_i \\
& = \int_{S^{d-1}} \int_0^{h_1} \langle t\theta, \psi_j \rangle \langle t\theta, \psi_k \rangle [f(x) + t \nabla_\theta f(x)] t^{d-1} \mathrm{d}t \mathrm{d}\theta + O(h^{d+4}) \\
& = O(h^{d+3}),
\end{aligned}$$

where the second equality is based on the fact that the second fundamental form is self-adjoint, by Lemma 8, (33) follows. The result (34) is dervied in a similar fashion.

Let $\chi_{i,k} = n^{-1}\langle X_i, \psi_j\rangle\langle X_i, \psi_k\rangle Z_i$. Then $\Xi_{j,k} = \sum_{i=1}^n \chi_{i,k}$. Then by Theorem 11.1 of Boucheron et al. (2016), we have

$$\operatorname{var}\left(\sup_{k \le d+1} \Xi_{j,k}\right) = \operatorname{var}\left(\sup_{k \le d+1} \sum_{i=1}^n \chi_{i,k}\right) \le \sum_{i=1}^n E \sup_{k \le d+1} \chi_{i,k}^2 = nE \sup_{k \le d+1} \chi_{i,k}^2. \tag{38}$$

The term $E\sup_{k \le d+1} \chi_{i,k}^2$ can be computed as follows:

$$\begin{aligned} E \sup_{k \le d+1} \chi_{i,k}^2 &= E \sup_{k \le d+1} \left[n^{-1}\langle X_i, \psi_j\rangle\langle \mathcal{P}_2 X_i, \psi_k\rangle Z_i\right]^2 \\ &\le n^{-2} E \sup_{k \le d+1} \|X_i\|^2\|\mathcal{P}_2 X_i\|^2\|\psi_j\|^2\|\psi_k\|^2 Z_i = n^{-2} E\|X_i\|^2\|\mathcal{P}_2 X_i\|^2 Z_i. \end{aligned}$$

Since

$$E\|X_i\|^2\|\mathcal{P}_2 X_i\|^2 V_i = \int_{S^{d-1}} \int_0^{h_1} \|t\theta\|^2\|t^2\Pi_x(\theta,\theta)\|^2 f(\exp(t\theta))t^{d-1}\mathrm{d}t\mathrm{d}\theta + O(h^{d+8}) = O\left(h^{d+6}\right),$$

we apply Lemma 8 to conclude $E\|X_i\|^2\|\mathcal{P}_2 X_i\|^2 Z_i = O(h^{d+6})$. Therefore, $E\sup_{k \le d+1} \chi_{i,k}^2 = O(n^{-2}h^{d+6})$. With (38), we show that

$$\operatorname{var}\left(\sup_{k \le d+1} \Xi_{j,k}\right) = O\left(n^{-1}h^{d+6}\right).$$

Other results are dervied in the same way. □

REFERENCES

BOUCHERON, S., LUGOSI, G. & MASSART, P. (2016). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, United Kingdom: Oxford University Press.

CHENG, M. & WU, H. (2013). Local linear regression on manifolds and its geometric interpretation. *Journal of the American Statistical Association* **108**, 1421–1434.

FAN, J. (1993). Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics* **21**, 196–216.

LEVINA, E. & BICKEL, P. (2004). Maximum likelihood estimation of intrinsic dimension. *Advances in Neural Information* **17**, 777–784.

SINGER, A. & WU, H.-T. (2012). Vector diffusion maps and the connection laplacian. *Communications on Pure and Applied Mathematics* **65**, 1067–1144.

YAO, F., MÜLLER, H.-G. & WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100**, 577–590.

ZHANG, X. & WANG, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **44**, 2281–2321.