Supplement to "Function-on-Function Geometric Quantile Regression with a Distributed Extension"

Bingchen Li[1], Alan T. K. Wan[2], Miaomiao Yu[1], Yong Zhou[1]

[1]*Key Laboratory of Advanced Theory and Application in Statistics and Data Science-MOE School of Statistics and Academy of Statistics and Interdisciplinary Sciences East China Normal University*

[2]*Department of Decision Analytics and Operations, City University of Hong Kong*

Supplementary Material

This supplementary material provides additional details and supporting content for the paper "Function-on-Function Geometric Quantile Regression with a Distributed Extension".

This supplementary material is organized as follows. Section S1 provides supplementary preliminary knowledge. Specifically, this section first presents a concrete example related to the concept of data clouds, further discusses the relationship between the geometric quantile index and condi-

tional spatial depth and introduces practical choices of the index used in the literature, then provides a brief introduction to Functional Principal Component Analysis (FPCA), and further offers a more detailed explanation of the surrogate loss function. Additionally, it introduces the benchmark methods used for comparison in Subsection 5.2 of the main text. Section S2 presents the detailed algorithms used to compute the proposed estimators. Section S3 discusses the differences between the proposed distributed FPCA (DFPCA) algorithm and existing distributed principal component analysis (PCA) methods. Section S4 provides implementation details for the simulation studies. It first describes how to select the truncation dimensions and the smoothing parameter using five-fold cross-validation. It then presents a procedure for choosing the geometric quantile indices so that the proposed and benchmark methods are evaluated against the same conditional quantile target. Section S5 provides detailed discussions and a simulation study of the consequences of link function misspecification and the feasibility of nonparametrically estimating the link functions. Section S6 states the convergence rates for the induced operator estimators. Section S7 presents additional simulation results. Section S8 presents additional details of the real case study. The remaining sections provide the detailed proofs for the theorems and corollaries. Section S9 introduces several key lemmas along

with their detailed proofs. These lemmas play a crucial role in establishing the main theoretical results of the paper. Section S10 presents the proofs of the theorems, while Section S11 contains the proofs of the corollaries.

S1 Preliminaries

S1.1 Illustrative Example of Data Cloud and Spatial Depth

Given a data cloud, the statistical depth of a point quantifies its relative position within the cloud. Points closer to the center are considered deeper, while those near the outer region are regarded as shallower. Direction vectors from the observations to a given point tend to cancel each other out when the point is located near the center of the data cloud, but fail to do so when the point lies closer to the boundary. Therefore, the norm of the average direction vector tends to be small for points deep inside the data cloud, while it becomes relatively large for points near the boundary.

To more clearly illustrate the concepts of data clouds and geometric quantiles introduced in Subsection 2.2 of the main text, we provide a concrete example. Specifically, we generated a data cloud consisting of 20 observations drawn from a bivariate normal distribution. Figure S1.1 presents two representative cases in the generated data cloud, where point A is lo-

cated at the center and point B is situated at the periphery. In the left sub-plot, direction vectors from data points toward a given point A, in the central region of the data cloud are shown. These vectors tend to cancel each other out, resulting in an average direction vector with a length close to zero. According to Equation (2.2) in the main text, this implies that point A has a high spatial depth, close to 1. Conversely, in the right sub-plot, the direction vectors for point B, located at the periphery of the data cloud, point in similar directions and therefore do not cancel out, resulting in an average direction vector with a length close to 1. As a result, the spatial depth of point B is low, approaching 0.

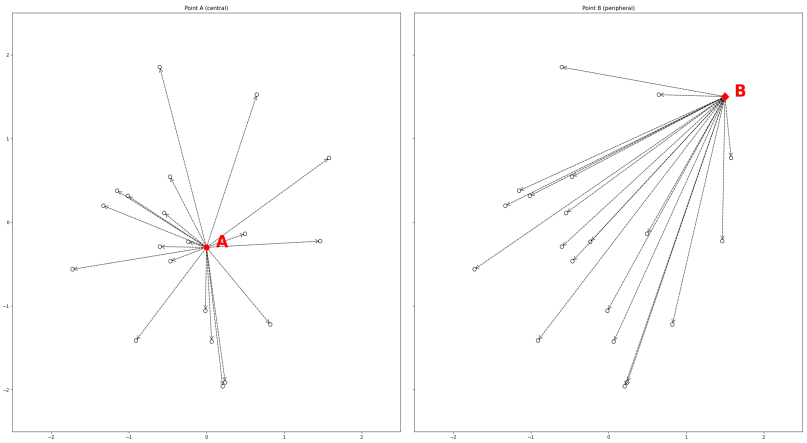


Figure S1.1: Spatial depth obtained using direction vectors.

S1.2 Geometric Quantile Index and Conditional Spatial Depth

A geometric quantile index $\tau \in \mathcal{H}_Y$, with $\|\tau\| < 1$, identifies the target conditional geometric quantile $Q_{Y|X=x}(\tau)$ through Equation (2.4). For a fixed $x \in \mathcal{H}_X$, we write $Q_{Y|X=x}(\tau)$ for the corresponding quantile. If the conditional distribution of Y given $X = x$ is not supported on any line in $\mathcal{H}_Y$ (Assumption A_4), then the conditional geometric quantile $Q_{Y|X=x}(\tau)$ defined in Equation (2.4) exists and is unique. Thus, each geometric quantile index τ uniquely determines the corresponding conditional geometric quantile $Q_{Y|X=x}(\tau)$.

Further, under Assumption A_3, the conditional loss in Equation (2.4) is Fréchet differentiable at $Q_{Y|X=x}(\tau)$ for almost every x. Hence, we have

$$\mathbb{E}\left[\frac{Q_{Y|X=x}(\tau) - Y}{\left\|Q_{Y|X=x}(\tau) - Y\right\|}\mathbb{I}\left\{\left\|Q_{Y|X=x}(\tau) - Y\right\| \neq 0\right\} \,\middle|\, X = x\right] = \tau. \tag{S1.1}$$

We define the conditional spatial distribution and conditional spatial depth of Y given $X = x$ at a point $y \in \mathcal{H}_Y$ as

$$S_{Y|X=x}(y) = \mathbb{E}\left[\frac{y - Y}{\|y - Y\|}\mathbb{I}\{\|y - Y\| \neq 0\} \mid X = x\right]$$

and $SD_{Y|X=x}(y) = 1 - \|S_{Y|X=x}(y)\|$, respectively. Hence, for $\tau = ru$, Equation (S1.1) gives $SD_{Y|X=x}\{Q_{Y|X=x}(ru)\} = 1 - r$. Thus, r directly

determines the conditional spatial depth of the target geometric quantile $Q_{Y|X=x}(ru)$. When $r = 0$, $Q_{Y|X=x}(0)$ is the conditional geometric median with the maximum conditional spatial depth of one. For a fixed u, larger values of r give smaller spatial depths and hence index more extreme conditional geometric quantiles in that direction.

In summary, the geometric quantile index τ can be specified according to the research objective. Specifically, we first choose the quantile order r. If more extreme conditional geometric quantiles are of interest, a larger value of r may be chosen. We then select a direction u of interest and set $\tau = ru$. For example, when $\mathcal{H}_Y = L_2([0,1])$, possible choices of u include a constant function and one or more leading eigenfunctions of the response covariance operator (Chaouch and Goga, 2010; Chakraborty and Chaudhuri, 2014; Chowdhury and Chaudhuri, 2019). Each direction u may also be paired with its opposite direction $-u$ to examine conditional geometric quantiles in two opposite directions.

S1.3 Functional Principal Component Analysis (FPCA)

For a random element $Z \in \mathcal{H}_Z$, its covariance operator is defined as $\Sigma = \mathbb{E}(Z \otimes Z)$. Since this operator is a self-adjoint and compact operator, it admits the following spectral decomposition $\Sigma = \sum_{i=1}^{\infty} \lambda_i \phi_i \otimes \phi_i$, where

$\lambda_1 \geq \lambda_2 \geq \cdots$ are ordered nonnegative eigenvalues and $\phi_i \in \mathcal{H}_Z$ are the corresponding orthogonal eigenfunctions with unit norms.

When the global dataset is accessible, one can estimate the covariance operator of Z as $\hat{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} Z_i \otimes Z_i$. Then, perform spectral decomposition on it and obtain the estimates of the eigenfunctions $\{\hat{\phi}_i\}_{i=1}^{\infty}$, i.e., $\hat{\Sigma} = \sum_{i=1}^{\infty} \hat{\lambda}_i \hat{\phi}_i \otimes \hat{\phi}_i$.

S1.4 Surrogate Loss Function

This subsection introduces the communication-efficient distributed algorithm developed by Jordan et al. (2019). Let $\mathcal{L}_N(\boldsymbol{\theta})$ denote the global loss function, and let $\mathcal{L}_1(\boldsymbol{\theta})$ denote the local loss function computed on the first machine. The global loss function can be approximated by Taylor expansion around a suitably chosen initial estimate $\bar{\boldsymbol{\theta}}$,

$$\mathcal{L}_N(\boldsymbol{\theta}) = \mathcal{L}_N(\bar{\boldsymbol{\theta}}) + \langle \nabla \mathcal{L}_N(\bar{\boldsymbol{\theta}}), \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \rangle + \sum_{j=2}^{\infty} \frac{1}{j!} \nabla^j \mathcal{L}_N(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^{\otimes j}. \quad \text{(S1.2)}$$

Since the data are distributed across different computers, the computation of $\nabla^j \mathcal{L}_N(\bar{\boldsymbol{\theta}})$ requires the transportation of the higher-order derivatives information between computers, which incurs substantial communication costs. To reduce communication costs, the global loss function $\mathcal{L}_N$ is approximated by retaining only the global gradient and replacing higher-order derivatives

with local estimates, which results in the following surrogate expansion:

$$\widetilde{\mathcal{L}}(\boldsymbol{\theta}) = \mathcal{L}_N(\bar{\boldsymbol{\theta}}) + \langle \nabla \mathcal{L}_N(\bar{\boldsymbol{\theta}}), \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \rangle + \sum_{j=2}^{\infty} \frac{1}{j!} \nabla^j \mathcal{L}_1(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^{\otimes j}. \tag{S1.3}$$

If the higher-order derivatives $\nabla^j \mathcal{L}_N(\bar{\boldsymbol{\theta}})$ exhibit superior properties, this approximation can still achieve high accuracy. According to Jordan et al. (2019), the approximation error satisfies $\mathcal{L}_N(\boldsymbol{\theta}) - \widetilde{\mathcal{L}}(\boldsymbol{\theta}) = O(\frac{1}{\sqrt{n}}\|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\|_2^2 + \|\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}\|_2^3)$. When the initial estimate is sufficiently close to the true parameter, such as one obtained by M-estimation on the first machine, then the approximation error becomes negligible relative to the overall estimation error. Furthermore, according to the Taylor expansion, the infinite series of higher-order derivatives in Equation (S1.3) can be replaced by $\mathcal{L}_1(\boldsymbol{\theta}) - \mathcal{L}_1(\bar{\boldsymbol{\theta}}) - \langle \nabla \mathcal{L}_1(\bar{\boldsymbol{\theta}}), \boldsymbol{\theta} - \bar{\boldsymbol{\theta}} \rangle$. And by omitting additive constants that are independent of $\boldsymbol{\theta}$, we redefine $\widetilde{\mathcal{L}}$ as

$$\widetilde{\mathcal{L}}(\boldsymbol{\theta}) := \mathcal{L}_1(\boldsymbol{\theta}) - \langle \boldsymbol{\theta}, \nabla \mathcal{L}_1(\bar{\boldsymbol{\theta}}) - \nabla \mathcal{L}_N(\bar{\boldsymbol{\theta}}) \rangle, \tag{S1.4}$$

which is referred to as the surrogate loss function.

S1.5 Benchmark Methods for Comparison

This section introduces the benchmark methods used in the numerical studies of the main text.

1. Nonparametric depth and quantile regression for functional data (NDQR)

   Given a predictor $X \in \mathcal{H}_X$ and a geometric quantile index $\tau \in \mathcal{H}_Y$ with $\|\tau\| < 1$, the response and τ are projected onto $\mathcal{Z}_n$, yielding $\mathbf{Y}^{(n)}$ and $\tau^{(n)}$, respectively. The conditional τ-quantile $\widehat{\mathbf{Q}}_n(\tau \mid \mathbf{x})$ is defined as the minimizer of the truncated objective function

   $$\widehat{g}_n(\mathbf{Q} \mid \mathbf{x}) = \frac{\sum_{i=1}^n \|\mathbf{Q} - \mathbf{Y}_i^{(n)}\| K\left(h_n^{-1} d(\mathbf{x}, \mathbf{X}_i)\right)}{\sum_{i=1}^n K\left(h_n^{-1} d(\mathbf{x}, \mathbf{X}_i)\right)} - \langle \tau^{(n)}, \mathbf{Q} \rangle.$$

   Here, $K(\cdot) = \mathbb{I}(\cdot)$ denotes the indicator kernel function, $d(\cdot, \cdot)$ is a distance metric on $\mathcal{H}_X$, and h_n is a bandwidth parameter selected via cross-validation. $\widehat{\mathbf{Q}}_n(\tau \mid \mathbf{x})$ can be computed using iterative methods (Chaudhuri, 1996).

2. Penalized function-on-function linear quantile regression (PFLQR)

   A penalized function-on-function (FoF) linear quantile regression model is considered in Beyaztas et al. (2025), where both the response and predictor are square-integrable stochastic processes. The conditional quantile of the response Y given the entire trajectory of the predic-

tor X is modeled as Equation (5.22) in the main text. Estimation of (β_0, β) is achieved by solving the following penalized optimization problem:

$$\arg\min_{(\beta_0,\beta)} \sum_{i=1}^{n} \rho_q \left[Y_i(t) - \beta_0(t) - \int X_i(s)\beta(t,s)ds \right] + \frac{\lambda_1}{2}\mathcal{J}_1(\beta_0) + \frac{\lambda_2}{2}\mathcal{J}_2(\beta),$$

where $q \in (0,1)$ is the scalar quantile level, $\rho_q(e) = qe - \mathbf{1}(e < 0)e$ is the check loss function, and $\mathcal{J}_1$ and $\mathcal{J}_2$ are roughness penalties on the functional parameters β_0 and β, respectively.

3. Function-on-function partial quantile regression (FPQR)

Mutis et al. (2025) developed FPQR methods for estimating a function-on-function linear quantile regression model. For $q \in (0,1)$, the function-on-function linear quantile regression model is expressed as

$$Q_q\{Y_i(t) \mid X_i\} = \int_0^1 X_i(s)\beta_q(s,t)\, ds,$$

where $\beta_q(s,t)$ is the regression coefficient function. Estimation is based on minimizing the check-loss function $\rho_q(e) = e\{q - \mathbb{I}(e < 0)\}$, yielding the optimization problem

$$\arg\min_{\beta_q} \sum_{i=1}^{n} \rho_q \left[Y_i(t) - \int_0^1 X_i(s)\beta_q(s,t)\, ds \right].$$

In our simulations, we estimate this model using the FPQR method of Mutis et al. (2025) via the "fpqr" function in the R package "ffpqr". The numbers of predictor and response basis functions and retained FPQR components are selected by five-fold cross-validation using the mean validation BIC described by Mutis et al. (2025).

S2 Algorithms

S2.1 Optimization Algorithm for the Sieve FoF Geometric Quantile Regression (FGQR) Estimator

This subsection presents the optimization algorithm used to compute the sieve FGQR estimator defined in Equation (2.8) of the main text.

To compute the sieve estimator within this structured class, we employ Newton's method. We therefore derive the first- and second-order Fréchet derivatives of the objective function on the right-hand side of Equation (2.8) with respect to the parameter $\boldsymbol{\theta}$. Define $T_X(\boldsymbol{\theta}) = \sum_{i=1}^{L_Y(n)} f_i(X;\boldsymbol{\theta}_i)\phi_i^{(Y)}$, and let the operators D and D^2 represent the first- and second-order Fréchet derivatives, respectively. We calculate the first- and second-order Fréchet derivatives of the check function as follows:

$$D\rho_\tau(Y - T_X(\boldsymbol{\theta}))$$

$$= \sum_{i=1}^{L_Y(n)} \frac{\nabla f_i(X;\boldsymbol{\theta}_i)\langle T_X(\boldsymbol{\theta}) - Y, \phi_i^{(Y)}\rangle}{\|T_X(\boldsymbol{\theta}) - Y\|} - \sum_{i=1}^{L_Y(n)} \nabla f_i(X;\boldsymbol{\theta}_i)\langle \tau, \phi_i^{(Y)}\rangle, \quad \text{(S2.5)}$$

$$\begin{aligned} & D^2\rho_\tau(Y - T_X(\boldsymbol{\theta})) \\ &= \sum_{i=1}^{L_Y(n)} \frac{\langle T_X(\boldsymbol{\theta}) - Y, \phi_i^{(Y)}\rangle}{\|T_X(\boldsymbol{\theta}) - Y\|} \nabla^2 f_i(X;\boldsymbol{\theta}_i) + \sum_{i=1}^{L_Y(n)} \frac{\nabla f_i(X;\boldsymbol{\theta}_i)\nabla^T f_i(X;\boldsymbol{\theta}_i)}{\|T_X(\boldsymbol{\theta}) - Y\|} \\ &- \frac{\left(\sum_{i=1}^{L_Y(n)} \langle T_X(\boldsymbol{\theta}) - Y, \phi_i^{(Y)}\rangle \nabla f_i(X;\boldsymbol{\theta}_i)\right)\left(\sum_{i=1}^{L_Y(n)} \langle T_X(\boldsymbol{\theta}) - Y, \phi_i^{(Y)}\rangle \nabla f_i(X;\boldsymbol{\theta}_i)\right)^T}{\|T_X(\boldsymbol{\theta}) - Y\|^3} \\ &- \sum_{i=1}^{L_Y(n)} \langle \tau, \phi_i^{(Y)}\rangle \nabla^2 f_i(X;\boldsymbol{\theta}_i), \end{aligned} \quad \text{(S2.6)}$$

where $\nabla f_i(\cdot)$ and $\nabla^2 f_i(\cdot)$ denote the gradient and Hessian matrix of $f_i(\cdot)$, respectively.

Newton's method is a second-order optimization technique that uses both the gradient and the Hessian of the objective function to iteratively update parameters. When the initial estimate is sufficiently close to the optimum, Newton's method converges rapidly. To ensure its efficiency, we first apply backtracking line search gradient descent to obtain a suitable initial value (Boyd and Vandenberghe, 2004). The backtracking line search gradient descent is an adaptive step size adjustment strategy aimed at selecting a suitable step size that ensures each iteration effectively reduces the value of the objective function, thereby enhancing the efficiency and

stability of the gradient descent method.

We begin with a given dataset $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$, an initial parameter $\boldsymbol{\theta}^{(0)}$, and an initial step size $t^{(0)}$. At the m_1-th iteration ($m_1 = 0, \ldots, M_1 - 1$), the parameters are updated according to

$$\boldsymbol{\theta}^{(m_1+1)} = \boldsymbol{\theta}^{(m_1)} - t^{(m_1)} \cdot g^{(m_1)},$$

where the gradient is given by

$$g^{(m_1)} = \mathbb{P}_n[D\rho_\tau(Y - T_X(\boldsymbol{\theta}^{(m_1)}))],$$

and the step size $t^{(m_1)}$ is determined via backtracking line search. Specifically, for $m_1 \geq 1$, $t^{(m_1)} = \alpha^j t^{(m_1-1)}$, where $\alpha \in (0, 1)$ is a real number, and j is the smallest nonnegative integer such that

$$\begin{aligned} j = \min \Big\{ \tilde{j} \in \mathbb{N} : \mathbb{P}_n[\rho_\tau(Y - T_X(\boldsymbol{\theta}^{(m_1)} - \alpha^{\tilde{j}} t^{(m_1-1)} \cdot g^{(m_1)}))] \\ \leq \mathbb{P}_n[\rho_\tau(Y - T_X(\boldsymbol{\theta}^{(m_1)}))] - \alpha^{\tilde{j}} t^{(m_1-1)} \cdot \langle g^{(m_1)}, g^{(m_1)} \rangle \Big\}. \end{aligned} \tag{S2.7}$$

Assuming that the final estimate obtained from the iterative procedure is $\boldsymbol{\theta}^{\text{init}}$, we set $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}^{\text{init}}$ and then apply Newton's method to further refine the parameters and achieve rapid convergence. At each iteration m_2

$(m_2 = 0, \ldots, M_2 - 1)$, the parameters are updated as

$$\hat{\boldsymbol{\theta}}^{(m_2+1)} = \hat{\boldsymbol{\theta}}^{(m_2)} - \left[H^{(m_2)}\right]^{-1} g^{(m_2)},$$

where $H^{(m_2)} = \mathbb{P}_n[D^2\rho_\tau(Y - T_X(\hat{\boldsymbol{\theta}}^{(m_2)}))]$ and $g^{(m_2)} = \mathbb{P}_n[D\rho_\tau(Y - T_X(\hat{\boldsymbol{\theta}}^{(m_2)}))]$. The iteration continues until the gradient norm falls below a pre-specified tolerance or a maximum number of iterations is reached. The detailed algorithm is presented in Algorithm 1.

S2.2 Distributed FPCA (DFPCA) algorithm

Let $\mathcal{D}_k^{(Z)}$ denote the local sample of Z stored on machine k, and the detailed algorithm is provided in Algorithm 2. When applied to X and Y, take $L = L_X(n)$ and $L = L_Y(n)$, respectively.

S2.3 Smoothed Distributed FoF Geometric Quantile Regression Algorithm

This subsection presents the detailed implementation of the Smoothed Distributed FoF Geometric Quantile Regression (SDFGQR) algorithm introduced in Subsection 4.2 of the main text.

We first apply Algorithm 2 separately to the predictor sample $\{X_i\}_{i=1}^N$ and the response sample $\{Y_i\}_{i=1}^N$ to obtain the estimated eigenfunctions

Algorithm 1: Newton's Method for FGQR

Input: Dataset $\mathcal{D} = \{(X_i, Y_i)\}_{i=1}^n$, Initial parameter $\boldsymbol{\theta}^{(0)}$, initial step size $t^{(0)}$, step size decay rate α, Maximum Iterations M_1 and M_2, Gradient Convergence Criterion ϵ.

$\boldsymbol{\theta}^{\text{init}} \leftarrow \boldsymbol{\theta}^{(0)}$;
for $m_1 = 0, 1, \cdots, M_1 - 1$ **do**
  $g^{(m_1)} \leftarrow \mathbb{P}_n D\rho_\tau(Y - T_X(\boldsymbol{\theta}^{(m_1)}))$;
  while $\mathbb{P}_n\rho_\tau(Y - T_X(\boldsymbol{\theta}^{(m_1)} - t^{(m_1)} \cdot g^{(m_1)})) >$ $\mathbb{P}_n\rho_\tau(Y - T_X(\boldsymbol{\theta}^{(m_1)})) - \alpha \cdot t^{(m_1)} \cdot \langle g^{(m_1)}, g^{(m_1)} \rangle$ **do**
    $t^{(m_1)} \leftarrow \alpha \cdot t^{(m_1)}$;
  end
  $\boldsymbol{\theta}^{(m_1+1)} \leftarrow \boldsymbol{\theta}^{(m_1)} - t^{(m_1)} \cdot g^{(m_1)}$;
  $\boldsymbol{\theta}^{\text{init}} \leftarrow \boldsymbol{\theta}^{(m_1+1)}$;
  $t^{(m_1+1)} \leftarrow t^{(m_1)}$;
  if $\|g^{(m_1)}\| < \epsilon$ **then**
    Break;
  end
end
$\hat{\boldsymbol{\theta}}^{(0)} \leftarrow \boldsymbol{\theta}^{\text{init}}$;
$\hat{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^{(0)}$;
for $m_2 = 0, 1, \cdots, M_2 - 1$ **do**
  $g^{(m_2)} \leftarrow \mathbb{P}_n(D\rho_\tau(Y - T_X(\hat{\boldsymbol{\theta}}^{(m_2)})))$;
  $H^{(m_2)} \leftarrow \mathbb{P}_n(D^2\rho_\tau(Y - T_X(\hat{\boldsymbol{\theta}}^{(m_2)})))$;
  $\hat{\boldsymbol{\theta}}^{(m_2+1)} \leftarrow \hat{\boldsymbol{\theta}}^{(m_2)} - [H^{(m_2)}]^{-1} g^{(m_2)}$;
  $\hat{\boldsymbol{\theta}} \leftarrow \hat{\boldsymbol{\theta}}^{(m_2+1)}$;
  if $\|g^{(m_2)}\| < \epsilon$ **then**
    Break;
  end
end

Result: Estimated parameter $\hat{\boldsymbol{\theta}}$.

$\{\tilde{\phi}_j^{(X)}\}_{j=1}^{L_X(n)}$ and $\{\tilde{\phi}_i^{(Y)}\}_{i=1}^{L_Y(n)}$. Using these estimated basis functions, we then obtain an initial estimate on $\mathcal{M}_1$ by applying Algorithm 1, denoted as $\tilde{\boldsymbol{\theta}}^{(0)}$, and set $\bar{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}^{(0)}$. At the m-th iteration ($m = 0, \ldots, M-1$), we set $\bar{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}^{(m)}$ and transmit the current estimate $\tilde{\boldsymbol{\theta}}^{(m)}$ from $\mathcal{M}_1$ to all machines. Each machine $\mathcal{M}_k$ then computes the local gradient $\mathbb{P}_{n_k} D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}}$ based on

Algorithm 2: DFPCA

Input: Datasets $\mathcal{D}_1^{(Z)}, \ldots, \mathcal{D}_K^{(Z)}$; number of components L
Output: Estimated global eigenfunctions $\{\tilde{\phi}_i\}_{i=1}^L$

for *machine* $k \in [K]$ **do**
  Compute local FPCA to obtain $\{\hat{\phi}_i^{(k)}\}_{i=1}^L$;
  Send $\{\hat{\phi}_i^{(k)}\}_{i=1}^L$ to a central machine;
end
for $m \leftarrow 1$ **to** L **do**
  Compute the operator: $\Sigma^{(m)} = \frac{1}{K}\sum_{k=1}^K \hat{\phi}_m^{(k)} \otimes \hat{\phi}_m^{(k)}$;
  Compute the first eigenfunction $V^{(m)}$ of $\Sigma^{(m)}$;
  if $m = 1$ **then**
    Set $\tilde{\phi}_1 \leftarrow V^{(1)}$;
  else
    Orthogonalize and normalize: $\tilde{\phi}_m \leftarrow \frac{\left(I - \sum_{i=1}^{m-1} \tilde{\phi}_i \otimes \tilde{\phi}_i\right) V^{(m)}}{\left\| \left(I - \sum_{i=1}^{m-1} \tilde{\phi}_i \otimes \tilde{\phi}_i\right) V^{(m)} \right\|}$;
  end
end
return $\{\tilde{\phi}_i\}_{i=1}^L$

its local data $\mathcal{D}_k$, and transmits it back to the central server. Upon receiving the gradients from all machines, the central server aggregates them to form the global gradient estimate:

$$\mathbb{P}_N D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}} = \frac{1}{K}\sum_{k=1}^K \mathbb{P}_{n_k} D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}}.$$

Using this aggregated gradient and the local empirical risk from $\mathcal{D}_1$, the central machine constructs the surrogate loss function in (4.15), employs Algorithm 1 to solve the following smoothed surrogate optimization prob-

lem, and updates the parameter estimate:

$$\tilde{\boldsymbol{\theta}}^{(m+1)} = \arg\min_{\boldsymbol{\theta}\in\Theta_n} \mathbb{P}_{n_1}\tilde{m}_{n,\boldsymbol{\theta}} - \left\langle \boldsymbol{\theta}, \mathbb{P}_{n_1} D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}} - \mathbb{P}_N D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}} \right\rangle. \tag{S2.8}$$

The detailed algorithm is presented in Algorithm 3.

Algorithm 3: Smoothed Distributed FoF Geometric Quantile Regression (SDFGQR)

Input: Local datasets $\{\mathcal{D}_k\}_{k=1}^K$, number of iterations M

Estimate the eigenfunctions $\{\tilde{\phi}_j^{(X)}\}_{j=1}^{L_X(n)}$ and $\{\tilde{\phi}_i^{(Y)}\}_{i=1}^{L_Y(n)}$ by applying Algorithm 2 to the predictor and response respectively;

Obtain the initial estimator $\tilde{\boldsymbol{\theta}}^{(0)}$ from $\mathcal{M}_1$ using Algorithm 1;

for $m = 0, \cdots, M-1$ **do**

  Transmit $\tilde{\boldsymbol{\theta}}^{(m)}$ to all machines $\mathcal{M}_k$;

  for $k = 1, \cdots K$ **do**

    Compute the local gradient $\mathbb{P}_{n_k} D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}}$ at machine $\mathcal{M}_k$;

    Transmit the local gradient $\mathbb{P}_{n_k} D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}}$ to machine $\mathcal{M}_1$;

  end

  Calculate the global gradient $\mathbb{P}_N D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}} = \frac{1}{K}\sum_{k=1}^K \mathbb{P}_{n_k} D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}}$;

  Solve the smoothed surrogate optimization problem with $\bar{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}}^{(m)}$ using Algorithm 1:

$$\tilde{\boldsymbol{\theta}}^{(m+1)} = \arg\min_{\boldsymbol{\theta}\in\Theta_n} \mathbb{P}_{n_1}\tilde{m}_{n,\boldsymbol{\theta}} - \left\langle \boldsymbol{\theta}, \mathbb{P}_{n_1} D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}} - \mathbb{P}_N D\tilde{m}_{n,\tilde{\boldsymbol{\theta}}^{(m)}} \right\rangle$$

end

return *Final estimator* $\tilde{\boldsymbol{\theta}}^{(M)}$

S3 Differences Between DFPCA and Existing Distributed Principal Component Analysis Methods

In this section, we discuss the differences between DFPCA and existing distributed PCA methods. Fan et al. (2019) proposed a one-shot distributed PCA procedure. Specifically, each machine first computes its leading eigenvectors and sends them to a central server, where the associated projection matrices are averaged. The server then extracts the leading eigenvectors of the averaged matrix to obtain the estimated global principal subspace. The method therefore focuses on recovering this subspace as a whole, without distinguishing among different orthonormal bases that span it. As in Fan et al. (2019), Shen et al. (2026) focus on recovering the leading eigenspace. However, the method proposed by Shen et al. (2026) is designed for settings in which both the sample size and the ambient dimension are ultra-large. To make the computation scalable, Fast Distributed (FADI) PCA applies multiple Gaussian random sketches to the matrices constructed from the local data. For each sketch, the local results are first aggregated across sites, after which the corresponding principal directions are computed in parallel. The resulting eigenspace estimates are then combined at the central server by averaging their projection matrices.

In contrast to Fan et al. (2019) and Shen et al. (2026), DFPCA is designed to recover the eigenfunctions individually, rather than only the subspace they span. This is necessary because the subsequent SDFGQR procedure uses their estimates in place of the unknown basis functions in Equation (2.5). More specifically, any orthonormal basis of the leading functional principal subspace defines the same orthogonal projection operator onto that subspace. However, the same subspace can be represented by different orthonormal bases that are not necessarily given by the ordered eigenfunctions of the covariance operator. Thus, recovery of the leading functional principal subspace alone does not ensure the individual eigenfunction estimation targeted by DFPCA. To achieve this individual recovery, DFPCA estimates the eigenfunctions one at a time. At the mth step, it averages the local operators $\hat{\phi}_m^{(k)} \otimes \hat{\phi}_m^{(k)}$ across machines, extracts the leading eigenfunction of the averaged operator, and then orthogonalizes and normalizes it against the eigenfunctions estimated in the previous steps. Under Assumption A_{11}, the errors in the individual eigenfunction estimates obtained by DFPCA enter the convergence rate of SDFGQR given in Theorem 4.

Chen et al. (2022) also compute the leading directions sequentially, but use a multi-round distributed procedure. To estimate each direction, their

method performs shift-and-invert iterations. During these iterations, the current estimate is repeatedly sent to the local machines, which compute and return local gradients for approximate Newton updates at the central machine. Once a direction has been obtained, it is orthogonalized and sent back to the local machines to deflate their data matrices. The same iterative procedure is then repeated to estimate the next direction. In contrast, DFPCA does not require iterative communication between the central and local machines. Instead, once the local eigenfunction estimates have been transmitted, the central machine completes all subsequent estimation steps. Furthermore, Fan et al. (2019), Chen et al. (2022), and Shen et al. (2026) all consider distributed PCA for multivariate data, whereas DFPCA is developed for functional data.

For large-scale functional data, Yan and He (2026) propose a Poisson subsampling method for covariance operator estimation and extend it to a decentralized setting. Each site constructs a covariance-operator estimate from a Poisson subsample of its local data, and the sites repeatedly exchange and average these estimates with their neighbors. The eigenvalue and eigenfunction estimates are then obtained from the resulting covariance operator estimator. Although their method also yields eigenfunction estimates, its subsampling and communication steps are used to construct a

decentralized covariance operator estimate, from which the eigenfunctions are subsequently obtained. DFPCA is designed to estimate the individual eigenfunctions directly and does not require local covariance operators to be transmitted.

S4 Implementation Details for the Simulation Studies

S4.1 Selection of Truncation Dimensions and the Smoothing Parameter

We use five-fold cross-validation to select the truncation dimensions for FGQR and SDFGQR. For SDFGQR, the smoothing parameter is selected jointly with the truncation dimensions. For each Monte Carlo replication, tuning parameters for both methods are selected by minimizing the average geometric check loss ρ_τ on the validation observations.

For FGQR, we randomly divide the sample of size N into five folds, with index sets $I_1, \ldots, I_5$. For each choice of L_X and L_Y, let $\widehat{\mathcal{O}}^{(-I_j)}_{L_X,L_Y}$ denote the FGQR operator estimator fitted using all observations outside I_j, for $j = 1, \ldots, 5$. The predictor and response FPCA bases are estimated using only the remaining four folds. We select the dimensions by minimizing the

average validation loss:

$$(\widehat{L}_X, \widehat{L}_Y) = \underset{L_X, L_Y}{\arg\min} \frac{1}{N} \sum_{j=1}^{5} \sum_{i \in I_j} \rho_\tau \left\{ Y_i - \widehat{\mathcal{O}}^{(-I_j)}_{L_X, L_Y}(X_i) \right\}.$$

For SDFGQR, we divide the observations on each machine into five folds and denote the index set of fold j on machine k by I_{kj}. In this distributed setting, let $I_j = \{(k, i) : 1 \leq k \leq K,\ i \in I_{kj}\}$ collect the indices of the validation observations across all machines. At each step, fold j is excluded from every machine, and both DFPCA and SDFGQR are fitted using the remaining local observations. Let $\widetilde{\mathcal{O}}^{(-I_j)}_{L_X, L_Y, h}$ denote the resulting operator estimator for the specified dimensions and smoothing parameter. We select these parameters by minimizing the average validation loss across all machines:

$$(\widehat{L}_X, \widehat{L}_Y, \widehat{h}) = \underset{L_X, L_Y, h}{\arg\min} \frac{1}{N} \sum_{j=1}^{5} \sum_{k=1}^{K} \sum_{i \in I_{kj}} \rho_\tau \left\{ Y_{ki} - \widetilde{\mathcal{O}}^{(-I_j)}_{L_X, L_Y, h}(X_{ki}) \right\},$$

where N is the total sample size across the K machines. After selecting the parameters, we refit the corresponding model using the full sample.

In the simulation studies with joint dimension selection, each of L_X and L_Y is chosen from $\{2, 4, 6, 8, 10\}$, and the smoothing parameter h for SDFGQR is chosen from $\{0.1, 0.2, 0.3\}$.

S4.2 Calibration of Geometric Quantile Indices

We now describe this calibration procedure in detail. Under Equation (5.22), Y is generated as the sum of the regression operator applied to X and an error process $\mathcal{E}$ independent of X. The independence of $\mathcal{E}$ and X, together with the translation equivariance of geometric quantiles, implies that the true conditional geometric quantile operator indexed by τ_q is

$$\mathcal{O}_{\tau_q}(X)(t) = \beta_0(t) + \int_0^1 \beta(t,s)X(s)\,ds + \delta_q(t),$$

where δ_q denotes the unconditional geometric quantile of $\mathcal{E}$ indexed by τ_q. Thus, to obtain an explicit true conditional geometric quantile operator, we construct an error quantile δ_q and determine its corresponding geometric quantile index τ_q.

In this comparison, we consider five unconditional geometric quantiles of the error process $\mathcal{E}$, denoted by $\boldsymbol{\delta}_q = (\delta_{q,1}, \ldots, \delta_{q,10})^\top$, whose components are given by $\delta_{q,j} = \Phi^{-1}(q)\sqrt{\mathrm{Var}\{\mathcal{E}(t_j)\}}$, where Φ denotes the standard normal distribution function and $q \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. We determine the geometric quantile index corresponding to $\boldsymbol{\delta}_q$ as

$$\tau_q = \mathbb{E}\left\{\frac{\boldsymbol{\delta}_q - \mathcal{E}}{\|\boldsymbol{\delta}_q - \mathcal{E}\|}\right\}.$$

For each q, FGQR, SDFGQR, and NDQR estimate the conditional geometric quantile indexed by the τ_q, whereas PFLQR and FPQR estimate the corresponding pointwise conditional quantile at level q. Since these estimands coincide by construction, all five methods are evaluated against the common target $\mathcal{O}_{\tau_q}(X)$. To simplify notation, let $\widehat{\mathcal{O}}_{\tau_q}(X)$ generically denote the conditional quantile estimate produced by any one of the five methods. In each Monte Carlo replication, we evaluate each estimator on an independent test sample using the mean relative error (MRE),

$$\mathrm{MRE} = \frac{1}{N_0} \sum_{\ell=1}^{N_0} \frac{\left\| \widehat{\mathcal{O}}_{\tau_q}(X_\ell) - \mathcal{O}_{\tau_q}(X_\ell) \right\|}{\left\| \mathcal{O}_{\tau_q}(X_\ell) \right\|}, \tag{S4.9}$$

where N_0 denotes the test-sample size.

S5 Link Function Misspecification and Nonparametric Link Modeling

The main analysis assumes that the link functions are specified in advance. This section examines the consequences of link function misspecification and the feasibility of estimating the link functions nonparametrically. The first subsection studies the additional error caused by misspecified link functions and presents the corresponding error decomposition. The second sub-

section introduces a practical extension that represents the link functions using B spline basis functions and estimates their coefficients together with the index parameters. The third subsection presents a simulation study that evaluates the effect of link function misspecification and examines the practical feasibility of this extension.

S5.1 Consequences of Link Function Misspecification

To study the consequences of fixed-link misspecification, we explicitly distinguish the structured operator classes associated with different fixed collections of link functions. Let $\boldsymbol{f} = \{f_i\}_{i=1}^{\infty}$ denote the collection of pre-specified link functions in (2.5) of the main text. Specifically, for a given $\boldsymbol{f}$, let $\mathcal{G}_{\boldsymbol{f}}$ denote the corresponding structured operator class induced by $\boldsymbol{f}$:

$$\mathcal{G}_{\boldsymbol{f}} = \left\{ \mathcal{O}_{\boldsymbol{\theta}} : \mathcal{H}_X \to \mathcal{H}_Y \;\middle|\; \mathcal{O}_{\boldsymbol{\theta}}(x) = \sum_{i=1}^{\infty} f_i\left(\sum_{j=1}^{\infty} \theta_{ij} \langle x, \phi_j^{(X)} \rangle \right) \phi_i^{(Y)},\ \boldsymbol{\theta} \in \Theta \right\}.$$

The analysis in the main text is based on the assumption that $\mathcal{O}^* \in \mathcal{G}_{\boldsymbol{f}}$. In this section, we examine the consequences of link function misspecification, i.e., $\mathcal{O}^* \notin \mathcal{G}_{\boldsymbol{f}}$.

Under link function misspecification, that is, when $\mathcal{O}^* \notin \mathcal{G}_{\boldsymbol{f}}$, the three primary operator estimators considered in the main text continue to be constructed using the same pre-specified link functions and their respective

empirical criteria. Specifically, the known-basis FGQR estimator $\widehat{\mathcal{O}}_n$, the estimated-basis FGQR estimator $\widehat{\widehat{\mathcal{O}}}_n$, and the distributed SDFGQR estimator $\widetilde{\mathcal{O}}_n$ are defined through (2.8), (2.9), and (4.16), respectively. The excess risk relative to $\mathcal{O}^*$ therefore consists of the estimation error and an additional bias caused by misspecification. More precisely, let $\widehat{\mathcal{O}}$ denote any one of $\widehat{\mathcal{O}}_n$, $\widehat{\widehat{\mathcal{O}}}_n$, and $\widetilde{\mathcal{O}}_n$. Its excess risk relative to $\mathcal{O}^*$ can be decomposed as follows:

$$\begin{aligned}
&\mathbb{E}\Big[\rho_\tau\{Y-\widehat{\mathcal{O}}(X)\}\Big]-\mathbb{E}[\rho_\tau\{Y-\mathcal{O}^*(X)\}]\\
=&\left[\mathbb{E}\Big[\rho_\tau\{Y-\widehat{\mathcal{O}}(X)\}\Big]-\inf_{\mathcal{O}\in\mathcal{G}_{\boldsymbol{f}}}\mathbb{E}[\rho_\tau\{Y-\mathcal{O}(X)\}]\right]\\
&+\left[\inf_{\mathcal{O}\in\mathcal{G}_{\boldsymbol{f}}}\mathbb{E}[\rho_\tau\{Y-\mathcal{O}(X)\}]-\mathbb{E}[\rho_\tau\{Y-\mathcal{O}^*(X)\}]\right]. \qquad \text{(S5.10)}
\end{aligned}$$

The first and second terms on the right-hand side of (S5.10) represent, respectively, the excess population risk arising from estimation within the structured operator class $\mathcal{G}_{\boldsymbol{f}}$ and the additional population risk induced by link function misspecification.

Assume that the population risk attains its infimum over $\mathcal{G}_{\boldsymbol{f}}$, and let

$$\mathcal{O}^\dagger_{\boldsymbol{f}}\in\arg\min_{\mathcal{O}\in\mathcal{G}_{\boldsymbol{f}}}\mathbb{E}[\rho_\tau\{Y-\mathcal{O}(X)\}]\,.$$

Under this assumption, the decomposition in (S5.10) can be written as

$$\begin{aligned}
&\mathbb{E}\Big[\rho_\tau\{Y-\widehat{\mathcal{O}}(X)\}\Big]-\mathbb{E}[\rho_\tau\{Y-\mathcal{O}^*(X)\}] \\
=&\,\mathbb{E}\Big[\rho_\tau\{Y-\widehat{\mathcal{O}}(X)\}\Big]-\mathbb{E}\Big[\rho_\tau\{Y-\mathcal{O}^\dagger_{\boldsymbol{f}}(X)\}\Big] \\
&+\mathbb{E}\Big[\rho_\tau\{Y-\mathcal{O}^\dagger_{\boldsymbol{f}}(X)\}\Big]-\mathbb{E}[\rho_\tau\{Y-\mathcal{O}^*(X)\}]\,.
\end{aligned}$$

The first term is the estimation error relative to the $\mathcal{O}^\dagger_{\boldsymbol{f}}$, whereas the second term is the additional population risk induced by link function misspecification. At the operator level, the triangle inequality gives

$$\left\|\widehat{\mathcal{O}}-\mathcal{O}^*\right\|_{\mathcal{X}}\le\left\|\widehat{\mathcal{O}}-\mathcal{O}^\dagger_{\boldsymbol{f}}\right\|_{\mathcal{X}}+\left\|\mathcal{O}^\dagger_{\boldsymbol{f}}-\mathcal{O}^*\right\|_{\mathcal{X}}. \tag{S5.11}$$

Suppose that $\mathcal{O}^\dagger_{\boldsymbol{f}}$ is unique and that the conditions required for Supplementary Corollaries 1 and 2 and for Theorem 4 in the main text hold in the misspecified setting. We can then follow the same reasoning as in their proofs to control the first term on the right-hand side of (S5.11) for each of $\widehat{\mathcal{O}}_n$, $\widehat{\widehat{\mathcal{O}}}_n$, and $\widetilde{\mathcal{O}}_n$. The second term on the right-hand side of (S5.11) is the fixed-link approximation error. Consequently, if $\|\widehat{\mathcal{O}}-\mathcal{O}^\dagger_{\boldsymbol{f}}\|_{\mathcal{X}}=o_p(1)$, the remaining asymptotic discrepancy from $\mathcal{O}^*$ is determined by the approximation error $\|\mathcal{O}^\dagger_{\boldsymbol{f}}-\mathcal{O}^*\|_{\mathcal{X}}$. A natural way to reduce the approximation error is to replace the fixed structured operator class $\mathcal{G}_{\boldsymbol{f}}$ with a sequence of

increasingly rich structured operator classes. If the truncation level and the dimension of the link function sieve are allowed to increase with n, and the resulting classes can approximate $\mathcal{O}^*$ arbitrarily well, then the corresponding approximation error can vanish asymptotically. One feasible approach is to model the link functions nonparametrically and estimate them jointly with the index parameters defining the operator. We regard the systematic development and theoretical analysis of such an unknown-link approach as an important direction for future research. As a preliminary step, we briefly discuss its feasibility in Subsection S5.2.

S5.2 B-Spline Modeling and Estimation of Link Functions

In this subsection, we discuss the feasibility of representing the unknown link functions by B-spline expansions and jointly optimizing over their spline coefficients and the index parameters. Equation (2.8) of the main text estimates the index parameters $\boldsymbol{\theta}$ with the link functions $\{f_i\}_{i=1}^{\infty}$ treated as known. To extend this criterion to unknown links, we represent each f_i by a finite B-spline expansion and jointly optimize over its spline coefficients and $\boldsymbol{\theta}$.

For a truncation level L_Y, let $B_1, \ldots, B_{K_f}$ be B-spline basis functions on the index domain. Replacing each fixed link f_i in Equation (2.8) with

$\sum_{k=1}^{K_f} \gamma_{ik} B_k(\cdot)$, where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \ldots, \gamma_{iK_f})^\top$ and $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1^\top, \ldots, \boldsymbol{\gamma}_{L_Y}^\top)^\top$, gives the estimator

$$\left(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Gamma}}\right) \in \arg\min_{\boldsymbol{\theta}, \boldsymbol{\Gamma}} \frac{1}{n} \sum_{\ell=1}^{n} \rho_\tau \left[Y_\ell - \sum_{i=1}^{L_Y} \left\{ \sum_{k=1}^{K_f} \gamma_{ik} B_k \left(\sum_{j=1}^{L_X} \theta_{ij} \langle X_\ell, \phi_j^{(X)} \rangle \right) \right\} \phi_i^{(Y)} \right]. \tag{S5.12}$$

When the functional bases are unknown, replace $\phi_j^{(X)}$ and $\phi_i^{(Y)}$ in (S5.12) with their FPCA estimates $\hat{\phi}_j^{(X)}$ and $\hat{\phi}_i^{(Y)}$, respectively. Conditional on these estimated bases, the criterion in (S5.12) remains a finite-dimensional optimization problem, and the resulting $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Gamma}}$ jointly induce a conditional geometric quantile operator estimator.

However, Algorithm 1 cannot be applied directly to the joint optimization problem in (S5.12), because it treats the link functions as fixed and updates only $\boldsymbol{\theta}$. Hence, joint optimization over $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}$ requires extending both the derivative calculations and the optimization algorithm. A practical approach is to alternate between updating $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}$, holding the other fixed at its current value. Specifically, conditional on the current value of $\boldsymbol{\theta}$, we update the B-spline coefficients $\boldsymbol{\Gamma}$ by minimizing the empirical loss in (S5.12). In turn, conditional on the current value of $\boldsymbol{\Gamma}$, the spline link functions are fixed, and the optimization with respect to $\boldsymbol{\theta}$ reduces to the optimization problem in Equation (2.8). Therefore, a Newton update

similar to that in Algorithm 1 can be applied to $\boldsymbol{\theta}$ within the alternating procedure. Repeating these two conditional updates can produce a numerical solution to (S5.12). In Subsection S5.3, we provide a preliminary numerical illustration of the estimator defined in (S5.12) to examine the feasibility of modeling the unknown link functions nonparametrically.

Remark 1. Without additional identification conditions, different combinations of the link functions and index coefficients may produce the same conditional geometric quantile operator. However, this subsection examines the computational feasibility of the B-spline extension only at the operator level. For a numerical solution $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Gamma}})$ of (S5.12), the resulting operator is

$$\widehat{\mathcal{O}}(x) = \sum_{i=1}^{L_Y} \left\{ \sum_{k=1}^{K_f} \widehat{\gamma}_{ik} B_k \left(\sum_{j=1}^{L_X} \widehat{\theta}_{ij} \langle x, \phi_j^{(X)} \rangle \right) \right\} \phi_i^{(Y)}.$$

This subsection focuses on the operator $\widehat{\mathcal{O}}$ induced by the numerical solution $(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Gamma}})$, rather than on $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Gamma}}$ separately. Establishing suitable conditions for the identification of $\boldsymbol{\theta}$ and $\boldsymbol{\Gamma}$ and developing the corresponding asymptotic theory are left for future research.

S5.3 Simulation Study of Link Function Misspecification and the Nonparametric Approach

To complement the preceding analysis, we conduct a simulation study to assess the practical consequences of fixed-link misspecification and the feasibility of nonparametric link estimation. We first generate data under nonlinear link functions and fit three estimators across multiple geometric quantile indices, namely a correctly specified FGQR estimator using the true link functions, a misspecified linear FGQR estimator, and the B-spline estimator in Equation (S5.12). The misspecified linear estimator imposes a linear model even though the true link functions are nonlinear. Finally, using an independent test sample, we report the mean relative error (MRE) between the estimated and true conditional geometric quantiles for each estimator.

We first introduce the link functions used in this subsection. For $\gamma \in \{0.5, 1\}$, define $h_\gamma(z) = (1-\gamma)z + 2\gamma\{\exp(z)-1\}/\{\exp(z)+1\}$. For $i = 1, \ldots, 10$, let $f_{i,\gamma}(z) = a_i + b_{i,\gamma} h_\gamma(z)$, where a_i and $b_{i,\gamma}$ are fixed coefficients. Larger values of γ correspond to greater nonlinearity in the link function, which leads to more severe misspecification under the linear model. The covariate X and the error process $\mathcal{E}$ are generated as in Section 5 of the main text. With $L_0 = 10$, let $\boldsymbol{\theta}^0 = (\theta^0_{ij})_{i,j=1}^{L_0} \in \mathbb{R}^{L_0 \times L_0}$ denote the fixed true

parameter. We then generate Y according to

$$Y = \sum_{i=1}^{L_0} f_{i,\gamma}\left(\sum_{j=1}^{L_0} \theta_{ij}^0 \langle X, \phi_j^{(X)} \rangle\right) \phi_i^{(Y)} + \mathcal{E}. \tag{S5.13}$$

We next derive the true conditional geometric quantile operators corresponding to the selected geometric quantile indices. Under Equation (S5.13), Y is generated as the sum of the regression operator applied to X and an error process $\mathcal{E}$ independent of X. The independence of $\mathcal{E}$ and X, together with the translation equivariance of geometric quantiles, implies that the true conditional geometric quantile operator indexed by τ_q is

$$\mathcal{O}_{\tau_q}^{(\gamma)}(X) = \sum_{i=1}^{L_0} f_{i,\gamma}\left(\sum_{j=1}^{L_0} \theta_{ij}^0 \langle X, \phi_j^{(X)} \rangle\right) \phi_i^{(Y)} + \delta_q,$$

where δ_q denotes the unconditional geometric quantile of $\mathcal{E}$ indexed by τ_q. Thus, to obtain an explicit true conditional geometric quantile operator, we construct an error quantile δ_q and determine its corresponding geometric quantile index τ_q.

In this subsection, we consider five unconditional geometric quantiles of the error process $\mathcal{E}$, denoted by $\boldsymbol{\delta}_q = (\delta_{q,1}, \ldots, \delta_{q,10})^\top$, whose components are given by $\delta_{q,j} = \Phi^{-1}(q)\sqrt{\mathrm{Var}\{\mathcal{E}(t_j)\}}$, where Φ denotes the standard normal distribution function and $q \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$. We determine the

geometric quantile index corresponding to $\boldsymbol{\delta}_q$ as

$$\tau_q = \mathbb{E}\left\{\frac{\boldsymbol{\delta}_q - \mathcal{E}}{\|\boldsymbol{\delta}_q - \mathcal{E}\|}\right\},$$

and then use the three methods described above to estimate the conditional geometric quantile operator indexed by τ_q.

For each combination of γ and q, the three estimators described above are fitted using the same training sample and truncation dimension. We use a training sample of size 10,000, an independent test sample of size $N_0 = 5{,}000$, and 100 Monte Carlo replications. Here, $\widehat{\mathcal{O}}$ generically denotes the conditional geometric quantile operator obtained from the correctly specified estimator, the misspecified linear estimator, or the B-spline estimator. We evaluate each estimator using the mean relative error (MRE) between the estimated and true conditional geometric quantiles,

$$\mathrm{MRE} = \frac{1}{N_0}\sum_{\ell=1}^{N_0}\frac{\left\|\widehat{\mathcal{O}}(X_\ell) - \mathcal{O}_{\tau_q}^{(\gamma)}(X_\ell)\right\|}{\left\|\mathcal{O}_{\tau_q}^{(\gamma)}(X_\ell)\right\|}.$$

Table S5.1: Mean relative error (MRE) between the estimated and true conditional geometric quantiles.

γ	q	Correctly specified	Misspecified	B-spline
	0.10	0.0223	0.0443	0.0327
	0.25	0.0247	0.0496	0.0296
0.5	0.50	0.0244	0.0494	0.0281
	0.75	0.0215	0.0430	0.0263
	0.90	0.0179	0.0355	0.0262
	0.10	0.0242	0.0883	0.0346
	0.25	0.0267	0.0984	0.0316
1	0.50	0.0265	0.0979	0.0300
	0.75	0.0234	0.0858	0.0281
	0.90	0.0194	0.0706	0.0277

S6 Operator Convergence Rates

The parameter convergence rates established in Theorems 1 and 2 directly imply corresponding convergence rates for the induced operator estimators. We state the formal results below and defer the proofs to Section S11.

Corollary 1. *Suppose Assumptions A_1–A_7 hold, and let $\hat{\mathcal{O}}_n$ denote the*

operator estimator induced by $\hat{\boldsymbol{\theta}}_n$ under known basis functions. Then

$$\|\hat{\mathcal{O}}_n(\cdot) - \mathcal{O}^*(\cdot)\|_{\mathcal{X}} = O_p\left(\sqrt{\frac{L(n)^2}{n} + M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)}\right). \tag{S6.14}$$

Corollary 2. *Suppose Assumptions A_1–A_9 hold, and let $\hat{\hat{\mathcal{O}}}_n$ denote the operator estimator induced by $\hat{\hat{\boldsymbol{\theta}}}_n$ under estimated basis functions. Then*

$$\|\hat{\hat{\mathcal{O}}}_n(\cdot) - \mathcal{O}^*(\cdot)\|_{\mathcal{X}} \lesssim \|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 + \sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\|. \tag{S6.15}$$

Consequently,

$$\|\hat{\hat{\mathcal{O}}}_n(\cdot) - \mathcal{O}^*(\cdot)\|_{\mathcal{X}} = O_p\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \alpha_n^2 + M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)}\right). \tag{S6.16}$$

In both settings, the convergence rate of the operator estimator is of the same order as that of the corresponding parameter estimator.

S7 Additional Simulation Results

This section provides additional simulation results to supplement and extend those in the main text.

S7.1 Predicted Conditional Geometric Quantiles for Different Geometric Quantile Indices

Let $u_0 \in \mathcal{H}_Y = \mathcal{L}_2([0,1])$ denote the normalized constant direction, where $u_0(t) = 1$ for all $t \in [0,1]$ and $\|u_0\| = 1$. In these simulations, we consider three geometric quantile indices:

$$\tau_0 = 0, \qquad \tau_1 = 0.8u_0, \qquad \tau_2 = -0.8u_0.$$

Here, τ_0 corresponds to the conditional geometric median, while τ_1 and τ_2 index conditional geometric quantiles of order 0.8 in the two opposite constant directions u_0 and $-u_0$, respectively. Figure S7.2 displays the predicted conditional geometric quantile functions for these three indices. The predicted curves closely follow the structure of the functional response distribution, and the conditional geometric median curve indexed by τ_0 lies near the center of the data cloud. The curve indexed by τ_1 is shown as a dashed line and lies mainly above the median curve, while that indexed by τ_2 is shown as a dotted line and lies mainly below it. Both curves have an overall shape similar to that of the median curve.

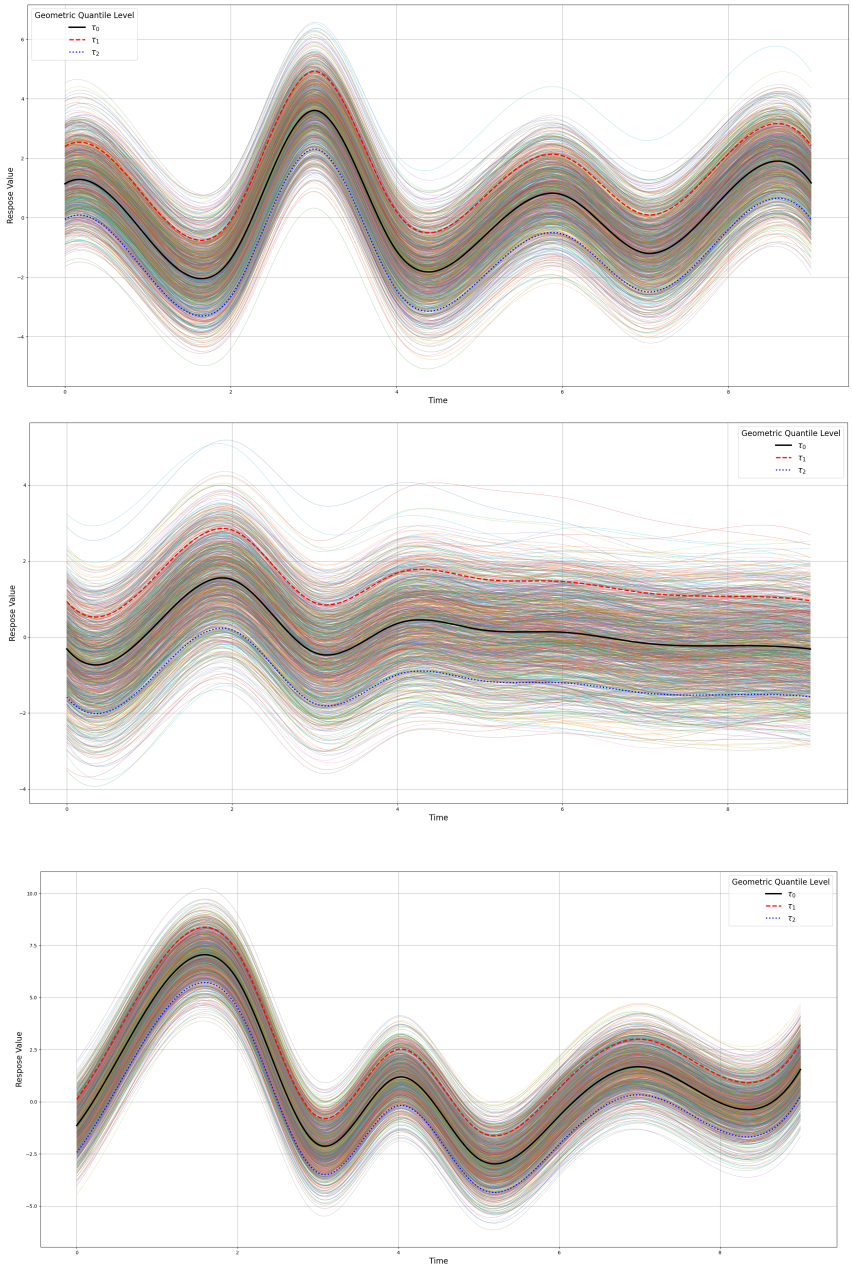


Figure S7.2: Predicted conditional geometric quantiles for different geometric quantile indices

S7.2 Impact of Sample Size on Estimation Accuracy in FoF Geometric Quantile Regression

This subsection examines the influence of sample size on the accuracy of parameter estimation in the proposed model. For brevity, we refer to FoF Geometric Quantile Regression as FGQR.

We consider three different sample sizes N: 100, 1000, and 10000. For each setting, we evaluate the estimation accuracy at the previously defined geometric quantile indices $\tau_0 = 0$, $\tau_1 = 0.8u_0$, and $\tau_2 = -0.8u_0$. We measure the estimation accuracy of the function $\beta(t, s)$ using the mean error (ME), defined as

$$\mathrm{ME} = \frac{1}{500} \sum_{r=1}^{500} \left[\int_0^1 \int_0^1 \left\{ \widehat{\beta}_r(t, s) - \beta(t, s) \right\}^2 ds\, dt \right]^{1/2}. \tag{S7.17}$$

Here, $\widehat{\beta}_r(t, s)$ denotes the estimate of $\beta(t, s)$ obtained in the r-th Monte Carlo replication for the given sample size and geometric quantile index. Figure S7.3 illustrates how the ME changes with the sample size across the three geometric quantile indices.

As shown in Figure S7.3, the ME decreases substantially as the sample size increases across all three geometric quantile indices. For instance, under τ_0, the MEs for sample sizes 100, 1000, and 10000 are 0.4452, 0.1351, and

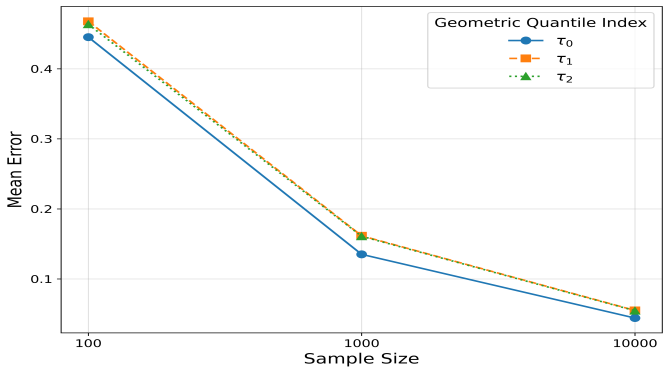


Figure S7.3: Impact of sample size on ME for different geometric quantile indices

0.0442, respectively. Notably, when the sample size increases from 100 to 1000, the ME decreases by approximately 69.7%; a similar reduction of approximately 67.3% is observed when the sample size increases from 1000 to 10000.

The truncation dimensions selected by cross-validation also vary with sample size. At $N = 100$, cross-validation selects $\widehat{L}_X = \widehat{L}_Y = 10$ in 37.2%, 26.6%, and 27.2% of the 500 Monte Carlo simulations for τ_0, τ_1, and τ_2, respectively. In the remaining 62.8%, 73.4%, and 72.8% of simulations under the respective indices, at least one of the selected truncation dimensions is smaller than 10. However, at $N = 1000$ and $N = 10000$, cross-validation selects $\widehat{L}_X = \widehat{L}_Y = 10$ in all 500 Monte Carlo simulations for each of the three geometric quantile indices. These results reflect the bias–variance trade-off, with smaller truncation dimensions helping to control estimation

variance in small samples and larger truncation dimensions helping to reduce approximation error in larger samples.

S7.3 Effect of Local Sample Size on DFPCA Estimation Accuracy

To demonstrate the effectiveness of the proposed DFPCA method, a series of simulation studies are conducted under various configurations. We compare the mean squared error (MSE) of Oracle FPCA and our proposed DFPCA. Figure S7.4 presents the cumulative MSE of the first L eigenfunctions under a fixed number of machines $K = 50$, as the local sample size n increases.

As shown, the MSE of our methods decreases as the local sample size increases. When $L = 5$, our method yields estimates that are highly comparable to those of Oracle FPCA across all settings of n. As L increases, the gap between DFPCA and Oracle FPCA becomes more significant when the local sample size is small. But this discrepancy diminishes as the local sample size increases.

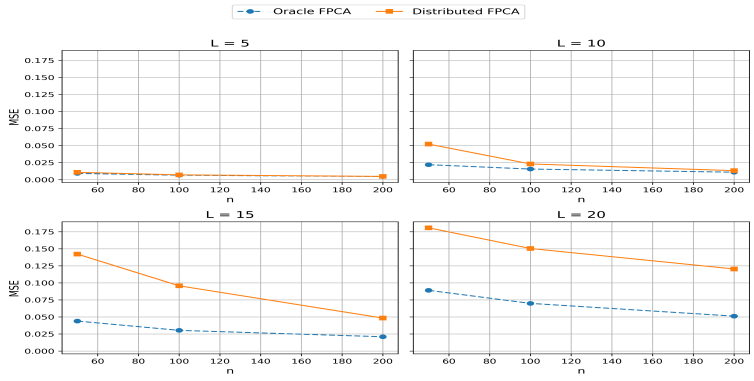


Figure S7.4: Impact of Local Sample Size n on FPCA and DFPCA Estimation Accuracy (Fixed $K = 50$).

S7.4 Effect of Machine Number on DFPCA Estimation Accuracy

Figure S7.5 presents additional results related to Subsection 5.1 of the main text. The figure indicates that, with a fixed local sample size, increasing the number of machines leads to lower MSE for DFPCA. This confirms that our parallel estimation algorithm can effectively aggregate information across machines, thereby significantly reducing the overall estimation error and validating its efficiency and scalability in distributed environments.

S8 Additional Details for the Real Case Study

Our dataset comprises measurements of oil density, pressure, and temperature collected at nine stations in an oil and gas pipeline network in China.

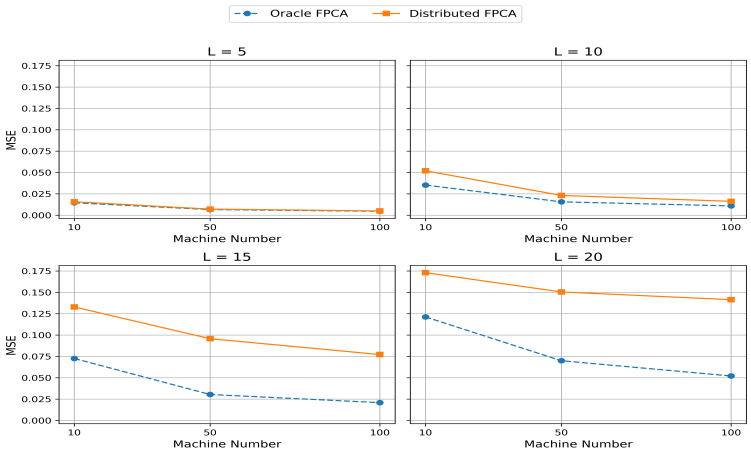


Figure S7.5: Impact of Machine Number K on FPCA and DFPCA Estimation Accuracy (Fixed $n = 100$).

Each station contributes 1,000 functional observations, yielding a total sample size of 9,000. And each functional observation consists of a density response curve and the corresponding pressure and temperature predictor curves, all observed on a common grid of 10 equally spaced time points at six-second intervals. For model training and testing, we divide the 9,000 functional observations into a training set of 8,900 observations and a test set of 100 observations.

Before estimating the conditional geometric median, we center and standardize the density, pressure, and temperature curves at each observation point using the respective means and standard deviations computed from the training data. After estimation, to compare the predicted and observed density curves on the original scale, we reverse the standardization of the

predictions using the density means and standard deviations computed from the training data.

To model the conditional geometric median, we use the linear specification of the structured operator class in the main text, with the following truncated representation:

$$\mathcal{O}(X_1, X_2) = \sum_{i=1}^{L_Y} \left[b_i + \sum_{j=1}^{L_{X_1}} \theta_{ij}^{(1)} \langle X_1, \phi_j^{(X_1)} \rangle + \sum_{j=1}^{L_{X_2}} \theta_{ij}^{(2)} \langle X_2, \phi_j^{(X_2)} \rangle \right] \phi_i^{(Y)}.$$

Here, X_1 and X_2 denote the standardized pressure and temperature curves, respectively. The predictor bases $\{\phi_j^{(X_1)}\}$ and $\{\phi_j^{(X_2)}\}$ are constructed separately, with truncation dimensions L_{X_1} and L_{X_2}, while L_Y is the response truncation dimension.

When estimating the conditional geometric median using the proposed FGQR and SDFGQR methods with the standardized training data, we select the tuning parameters by the cross-validation procedure described in Subsection S4.1, with the same truncation dimension used for the pressure and temperature predictors. Specifically, the selected truncation dimensions are $L_{X_1} = L_{X_2} = 10$ and $L_Y = 2$ for FGQR, and $L_{X_1} = L_{X_2} = 6$ and $L_Y = 2$ for SDFGQR. The selected smoothing parameter for SDFGQR is $h_n = 0.1$.

To evaluate the predictive performance of each method, let $\{Y_i\}_{i=1}^{100}$ denote the density response curves in the test set. For any method, we use $\widehat{Y}_i$

to denote the density curve predicted from the corresponding pressure and temperature curves. We define the relative prediction error by comparing the predicted oil density curves with the observed density curves in the test set:

$$\text{RPE} = \frac{1}{100}\sum_{i=1}^{100}\frac{\left\|\widehat{Y}_i - Y_i\right\|}{\|Y_i\|}. \tag{S8.18}$$

Table 2 reports the relative prediction errors of all methods on the test set.

S9 Technical Lemmas

We begin by introducing the relevant notation and definitions used throughout the analysis. Let $\mathcal{F}$ be a class of measurable real-valued functionals, and define the empirical process by $\mathbb{G}_n(\mathfrak{f}) := \sqrt{n}(\mathbb{P}_n - P)(\mathfrak{f})$ for $\mathfrak{f} \in \mathcal{F}$, with its supremum norm given by $\|\mathbb{G}_n\|_{\mathcal{F}} := \sup_{\mathfrak{f}\in\mathcal{F}} |\mathbb{G}_n(\mathfrak{f})|$. $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ denotes the ϵ-bracketing number of the function class $\mathcal{F}$ with respect to the norm $\|\cdot\|$, for $\epsilon > 0$.

Let $m_{n,\boldsymbol{\theta},\eta}$ denote the loss function, where $\boldsymbol{\theta}$ is the parameter vector and η denotes the associated basis functions. For example, a specific form of $m_{n,\boldsymbol{\theta},\eta}$ is given by

$$m_{n,\boldsymbol{\theta},\eta} = \|Y - \sum_{i=1}^{L_Y(n)} f_i(X;\boldsymbol{\theta}_i)\phi_i^{(Y)}\| + \langle \tau, Y - \sum_{i=1}^{L_Y(n)} f_i(X;\boldsymbol{\theta}_i)\phi_i^{(Y)}\rangle.$$

The limiting version of this loss is denoted by $m_{\boldsymbol{\theta},\eta} = \lim_{n\to\infty} m_{n,\boldsymbol{\theta},\eta}$, which is assumed to be independent of the sample size n. Throughout the analysis, we let $\eta_0 = \{\phi_i^{(X)}, \phi_i^{(Y)}\}_{i=1}^{\infty}$ represent the true underlying FPCA basis.

Throughout this section, we work with negated empirical and population objectives to facilitate the use of standard empirical-process and M-estimation results. This is only a proof convention: maximizing these objectives is equivalent to minimizing the original losses used in the main text.

S9.1 Known Basis Functions

Given a fixed basis η, we define the empirical objective function as $\mathbb{M}_n(\boldsymbol{\theta}) := -\mathbb{P}_n m_{n,\boldsymbol{\theta},\eta}$ and the population objective function as $M(\boldsymbol{\theta}) := -P m_{\boldsymbol{\theta},\eta}$. The corresponding estimator and true parameter are then defined as:

$$\hat{\boldsymbol{\theta}}_n = \arg\max_{\boldsymbol{\theta}\in\Theta_n} \mathbb{M}_n(\boldsymbol{\theta}),$$

$$\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}\in\Theta} M(\boldsymbol{\theta}).$$

To establish Theorem 1, we present two auxiliary lemmas. Lemma 1 guarantees the consistency of $\hat{\boldsymbol{\theta}}_n$ under mild regularity conditions. Lemma 2 further characterizes the convergence rate of $\hat{\boldsymbol{\theta}}_n$ to $\boldsymbol{\theta}^*$.

Lemma 1. *Suppose the following Lipschitz condition holds:*

$$|m_{n,\boldsymbol{\theta}_1,\eta} - m_{n,\boldsymbol{\theta}_2,\eta}| \lesssim F(X)\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2,$$

for some envelope function $F(X)$ satisfying $\mathbb{E}[F(X)^2] < \infty$. Define the norm $\|F(X)\|_{P,2} := (\mathbb{E}[F(X)^2])^{1/2}$. Suppose $L(n)$ satisfies $\frac{L(n)}{\sqrt{n}} = o(1)$, and Assumptions A_1 and A_4–A_7 hold. Then we have $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$.

Proof. Fix any $\delta > 0$. By definition of $\boldsymbol{\theta}^*$ as the unique maximizer of $M(\boldsymbol{\theta})$, we have

$$M(\boldsymbol{\theta}^*) > \sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2>\delta} M(\boldsymbol{\theta}).$$

Now, if $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2 > \delta$, then by optimality of $\hat{\boldsymbol{\theta}}_n$, we have: $\mathbb{M}_n(\hat{\boldsymbol{\theta}}_n) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) > 0$, and

$$M(\hat{\boldsymbol{\theta}}_n) - M({}^{(n)}\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*)) < \sup_{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2>\delta} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) := -\psi(\delta).$$

Thus,

$$\begin{aligned}
&\mathbb{P}\left(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\| > \delta\right) \\
\leq &\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M({}^{(n)}\boldsymbol{\theta}^*))| + |M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)| > \psi(\delta)\right).
\end{aligned}$$

To prove this lemma, it is necessary to verify the following terms converge to zero in probability:

$$\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M({}^{(n)}\boldsymbol{\theta}^*))| + |M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)|$$

First, the approximation error $|M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)| = o(1)$ vanishes asymptotically. Also we have

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M({}^{(n)}\boldsymbol{\theta}^*))| \\
\leq &\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*))| \\
&\qquad + |\mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*))| \\
\leq &2 \sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*))|.
\end{aligned}$$

Define $\mathcal{M} = \{m_{n,\boldsymbol{\theta},\eta} - m_{n,\boldsymbol{\theta}^*,\eta} : \boldsymbol{\theta} \in \Theta_n, ||\boldsymbol{\theta} - \boldsymbol{\theta}^*|| < K\}$, where K is a given positive constant. Then, by Theorem 11.4 of Sen (2022), we have:

$$E||\mathbb{G}_n||_{\mathcal{M}} \leq \int_0^{K||F(X)||_{P,2}} \sqrt{\log N_{[]}(\epsilon, \mathcal{M}, L_2(P))} d\epsilon \lesssim \sqrt{L_X(n) L_Y(n)} \leq L(n).$$

Hence, for all $\delta > 0$, we have

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*))| > \delta \right)$$

$$\leq \frac{E||\mathbb{G}_n||_{\mathcal{M}}}{\sqrt{n}\delta} \to 0$$

as $n \to \infty$. We thus conclude that

$$\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M({}^{(n)}\boldsymbol{\theta}^*))| + |M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)| = o_p(1).$$

Therefore, it follows that $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$. □

Lemma 2. *Suppose the following Lipschitz condition holds:*

$$|m_{n,\boldsymbol{\theta}_1,\eta} - m_{n,\boldsymbol{\theta}_2,\eta}| \lesssim F(X)\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2,$$

for some envelope function $F(X)$ satisfying $\mathbb{E}(F(X)^2) < \infty$. Define the norm $\|F(X)\|_{P,2} = (\mathbb{E}(F(X)^2))^{\frac{1}{2}}$. Under Assumptions A_1–A_7, if $\frac{L(n)}{\sqrt{n}} = o(1)$, then the estimator $\hat{\boldsymbol{\theta}}_n$ satisfies

$$||\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*||_2 = O_p\left(\sqrt{\frac{L(n)^2}{n} + M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*)}\right).$$

Proof. By Lemma 1, we already have the consistency result $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$. Hence, it suffices to analyze the empirical process in a sufficiently small neighborhood of $\boldsymbol{\theta}$. Under Assumptions A_2 and A_3, and assuming that the approximation error satisfies $M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*) \leq \delta_n^2$, to complete

the proof of this lemma, it suffices to analyze the convergence rate of the following quantity:

$$\mathbb{E} \sup_{\theta \in \Theta_n : ||\boldsymbol{\theta} - \boldsymbol{\theta}^*|| < \delta} \sqrt{n} \left[\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}) + M(\boldsymbol{\theta}^*) \right], \tag{S9.19}$$

Define function class $\mathcal{M}_\delta = \{m_{n,\boldsymbol{\theta},\eta} - m_{n,\boldsymbol{\theta}^*,\eta} : ||\boldsymbol{\theta} - \boldsymbol{\theta}^*|| < \delta\}$. Then, according to Theorem 11.4 of Sen (2022), the expectation of the empirical process (S9.19) satisfies:

$$\begin{aligned}
E||\mathbb{G}_n||_{\mathcal{M}_\delta} &\leq \int_0^{\delta ||F(X)||_{P,2}} \sqrt{\log N_{[]}(\epsilon, \mathcal{M}_\delta, L_2(P))} d\epsilon \\
&\leq \int_0^{\frac{\delta}{2}} \sqrt{\log N(\epsilon, \{\theta \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < \delta\}, \|\cdot\|)} ||F(X)||_{P,2} \\
&\leq \int_0^{\frac{\delta}{2}} \sqrt{\log (\frac{C\delta}{\epsilon})^{L_X(n) L_Y(n)}} ||F(X)||_{P,2} \\
&\lesssim L(n)\delta.
\end{aligned}$$

Let $L(n)\delta_n \lesssim \sqrt{n}\delta_n^2$ and $M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*) \leq \delta_n^2$. Hence, $\delta_n^2 = \frac{L(n)^2}{n} + M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*)$. □

S9.2 Unknown Basis Functions

When the basis functions are unknown, they must be estimated. In this case, the loss function is denoted by $m_{n,\boldsymbol{\theta},\hat{\eta}_n}$, where $\hat{\eta}_n$ takes the form: $\hat{\eta}_n := \{\hat{\phi}_i^{(X)}\}_{i=1}^{L_X(n)} \cup \{\phi_i^{(X)}\}_{i=L_X(n)+1}^{\infty} \cup \{\hat{\phi}_j^{(Y)}\}_{j=1}^{L_Y(n)} \cup \{\phi_j^{(Y)}\}_{j=L_Y(n)+1}^{\infty}$, with

$\{\hat{\phi}_i^{(X)}\}_{i=1}^{L_X(n)} \cup \{\hat{\phi}_j^{(Y)}\}_{j=1}^{L_Y(n)}$ representing the basis functions estimated via FPCA. Define $\mathbb{M}_n(\boldsymbol{\theta}) = -\mathbb{P}_n m_{n,\boldsymbol{\theta},\hat{\eta}_n}$ and $M(\boldsymbol{\theta}) = -Pm_{\boldsymbol{\theta},\eta_0}$. Hence, the corresponding estimator is defined as:

$$\hat{\hat{\boldsymbol{\theta}}}_n = \arg\max_{\boldsymbol{\theta}\in\Theta_n} \mathbb{M}_n(\boldsymbol{\theta}),$$

To establish Theorem 2, we first introduce two auxiliary lemmas. The first lemma addresses the consistency of the estimator $\hat{\hat{\boldsymbol{\theta}}}_n$, while the second provides its convergence rate.

Lemma 3. *Assume that conditions A_1–A_9 hold. Let $\mathcal{V} = \max\left\{1, \frac{1+v}{2}\right\}$. If $\frac{L(n)^{\mathcal{V}}}{\sqrt{n}} = o(1)$ and*

$$|m_{n,\boldsymbol{\theta}_1,\eta_1} - m_{n,\boldsymbol{\theta}_2,\eta_2}| \lesssim \|\mathcal{O}_{\boldsymbol{\theta}_1,\eta_1} - \mathcal{O}_{\boldsymbol{\theta}_2,\eta_2}\|,$$

then the estimator $\hat{\hat{\boldsymbol{\theta}}}_n$ is consistent:

$$\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1).$$

Proof. Using the same reasoning as in Lemma 1, it suffices to verify that

$$\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*))| \tag{S9.20}$$

converges to zero in probability. To this end, decompose the above expression (S9.20) as:

$$
\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*))| \\
\leq &\sup_{\boldsymbol{\theta}\in\Theta_n} |(\mathbb{P}_n - P)\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - (m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0})\right)| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta_n} |(\mathbb{P}_n - P)\left(m_{n,\boldsymbol{\theta},\eta_0} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)| \\
&+ \sup_{\boldsymbol{\theta}\in\Theta_n} |P\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0}\right)| + |P\left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)| \\
:= &I_1 + I_2 + I_3.
\end{aligned}
$$

Denote $\bar{\hat{\phi}}_j^{(X)} = \hat{\phi}_j^{(X)}$ for $j \leq L_X(n)$, and $\bar{\hat{\phi}}_j^{(X)} = \phi_j^{(X)}$ for $j > L_X(n)$. Similarly, denote $\bar{\hat{\phi}}_i^{(Y)} = \hat{\phi}_i^{(Y)}$ for $i \leq L_Y(n)$, and $\bar{\hat{\phi}}_i^{(Y)} = \phi_i^{(Y)}$ for $i > L_Y(n)$. Without loss of generality, assume $f_i(0) = 0$ for all i.

For I_3 we have

$$
\begin{aligned}
|m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0}| \lesssim &\|\mathcal{O}_{\boldsymbol{\theta},\hat{\eta}_n} - \mathcal{O}_{\boldsymbol{\theta},\eta_0}\| \\
= &\|\sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j}\langle X, \bar{\hat{\phi}}_j^{(X)}\rangle)\bar{\hat{\phi}}_i^{(Y)} - \sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j}\langle X, \phi_j^{(X)}\rangle)\phi_i^{(Y)}\| \\
\leq &\|\sum_{i=1}^{\infty}\left(f_i(\sum_{j=1}^{\infty} \theta_{i,j}\langle X, \bar{\hat{\phi}}_j^{(X)}\rangle) - f_i(\sum_{j=1}^{\infty} \theta_{i,j}\langle X, \phi_j^{(X)}\rangle)\right)\bar{\hat{\phi}}_i^{(Y)}\| \\
&+ \|\sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j}\langle X, \phi_j^{(X)}\rangle)(\bar{\hat{\phi}}_i^{(Y)} - \phi_i^{(Y)})\|
\end{aligned}
$$

$$
\begin{aligned}
&\lesssim \sqrt{\sum_{i=1}^{\infty}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\hat{\phi}}_j^{(X)}\rangle)-f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\phi_j^{(X)}\rangle)\right)^2}\\
&\quad+\|\sum_{i=1}^{\infty}f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\phi_j^{(X)}\rangle)(\bar{\hat{\phi}}_i^{(Y)}-\phi_i^{(Y)})\|\\
&\lesssim \sqrt{\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\hat{\phi}}_j^{(X)}\rangle-\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\phi_j^{(X)}\rangle\right)^2}\\
&\quad+\sum_{i=1}^{L_Y(n)}|f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\phi_j^{(X)}\rangle)|\|\hat{\phi}_i^{(Y)}-\phi_i^{(Y)}\|\\
&\lesssim \sqrt{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\theta_{i,j}^2\sum_{j=1}^{L_X(n)}\langle X,\hat{\phi}_j^{(X)}-\phi_j^{(X)}\rangle^2}\\
&\quad+\sum_{i=1}^{L_Y(n)}|\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\phi_j^{(X)}\rangle|\|\hat{\phi}_i^{(Y)}-\phi_i^{(Y)}\|\\
&\lesssim \|X\|\sum_{j=1}^{L_X(n)}\|\hat{\phi}_j^{(X)}-\phi_j^{(X)}\|+\sum_{i=1}^{L_Y(n)}\sqrt{\sum_{j=1}^{\infty}\theta_{i,j}^2\sum_{j=1}^{\infty}\langle X,\phi_j^{(X)}\rangle^2}\|\hat{\phi}_i^{(Y)}-\phi_i^{(Y)}\|\\
&\lesssim \|X\|\left(\sum_{i=1}^{L_Y(n)}\|\hat{\phi}_i^{(Y)}-\phi_i^{(Y)}\|+\sum_{j=1}^{L_X(n)}\|\hat{\phi}_j^{(X)}-\phi_j^{(X)}\|\right).
\end{aligned}
$$

Under Assumptions A_7 and A_8, I_3 converges to zero in probability.

$I_2 = o_p(1)$ follows from Lemma 1.

To control I_1, we have

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2<K}|(\mathbb{P}_n-P)\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n}-m_{n,\boldsymbol{\theta},\eta_0}-(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n}-m_{n,\boldsymbol{\theta}^*,\eta_0})\right)|>\delta\right)\\
\leq&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2<K}|(\mathbb{P}_n-P)\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n}-m_{n,\boldsymbol{\theta},\eta_0}-(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n}-m_{n,\boldsymbol{\theta}^*,\eta_0})\right)|>\delta,\right.
\end{aligned}
$$

$$\left. \sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \leq K \right)$$

$$+ \mathbb{P}\left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \geq K \right)$$

$$\leq \mathbb{P}\left(\sup_{\substack{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2<K \\ \bar{\eta}_n \in H_{n,K}}} |(\mathbb{P}_n - P)\left(m_{n,\boldsymbol{\theta},\bar{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - (m_{n,\boldsymbol{\theta}^*,\bar{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0})\right)| > \delta \right)$$

$$+ \mathbb{P}\left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \geq K \right)$$

where $H_{n,K} := \{\bar{\phi}_j^{(X)}, \bar{\phi}_i^{(Y)} : 1 \leq i \leq L_Y(n); 1 \leq j \leq L_X(n); \sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \leq K\} \cup \{\phi_i^{(X)} \otimes \phi_j^{(Y)} : i > L_X(n); j > L_Y(n)\}$. Let $\mathcal{F}_K := \{m_{n,\boldsymbol{\theta},\bar{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - m_{n,\boldsymbol{\theta},\bar{\eta}_n} + m_{n,\boldsymbol{\theta}^*,\eta_0} | \boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 \leq K; \bar{\eta} \in H_{n,K}\}$. We now derive an envelope function for the function class $\mathcal{F}_K$. Specifically, for any $m_{n,\boldsymbol{\theta},\bar{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - m_{n,\boldsymbol{\theta}^*,\bar{\eta}_n} + m_{n,\boldsymbol{\theta}^*,\eta_0} \in \mathcal{F}_K$, we have

$$\begin{aligned} &|m_{n,\boldsymbol{\theta},\bar{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - m_{n,\boldsymbol{\theta}^*,\bar{\eta}_n} + m_{n,\boldsymbol{\theta}^*,\eta_0}| \\ &\lesssim \|\mathcal{O}_{\boldsymbol{\theta},\bar{\eta}_n}(X) - \mathcal{O}_{\boldsymbol{\theta}^*,\bar{\eta}_n}(X)\| + \|\mathcal{O}_{\boldsymbol{\theta},\eta_0}(X) - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}(X)\|. \end{aligned} \tag{S9.21}$$

Let $\{\bar{\phi}_i^{(Y)}, \bar{\phi}_j^{(X)}\}_{i,j=1}^{\infty} \in H_{n,K}$, we analyze the first term on the right-hand-

side (r.h.s.) of (S9.21).

$$
\begin{aligned}
&\|\mathcal{O}_{\boldsymbol{\theta},\bar{\eta}_n}(X) - \mathcal{O}_{\boldsymbol{\theta}^*,\bar{\eta}_n}(X)\| \\
=&\|\sum_{i=1}^{\infty}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\phi}_j^{(X)}\rangle) - f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle X,\bar{\phi}_j^{(X)}\rangle)\right)\bar{\phi}_i^{(Y)}\| \\
\le&\|\sum_{i=1}^{L_Y(n)}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\phi}_j^{(X)}\rangle) - f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle X,\bar{\phi}_j^{(X)}\rangle)\right)\bar{\phi}_i^{(Y)}\| \\
&+\|\sum_{i=L_Y(n)+1}^{\infty}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\phi}_j^{(X)}\rangle) - f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle X,\bar{\phi}_j^{(X)}\rangle)\right)\phi_i^{(Y)}\| \\
=&\sqrt{\sum_{i=1}^{L_Y(n)}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\phi}_j^{(X)}\rangle) - f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle X,\bar{\phi}_j^{(X)}\rangle)\right)^2} \\
&+\sqrt{\sum_{i=L_Y(n)+1}^{\infty}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\phi}_j^{(X)}\rangle) - f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle X,\bar{\phi}_j^{(X)}\rangle)\right)^2} \\
\le&\sqrt{2}\sqrt{\sum_{i=1}^{\infty}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}\langle X,\bar{\phi}_j^{(X)}\rangle) - f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle X,\bar{\phi}_j^{(X)}\rangle)\right)^2} \\
\le&C_1\sqrt{2}\sqrt{\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(\theta_{i,j}-\theta_{i,j}^*)^2\sum_{j=1}^{\infty}\langle X,\bar{\phi}_j^{(X)}\rangle^2} \\
\le&C_1\sqrt{2}\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2\left(\sqrt{\sum_{j=1}^{L_X(n)}\langle X,\bar{\phi}_j^{(X)}\rangle^2} + \sqrt{\sum_{j=L_X(n)+1}^{\infty}\langle X,\phi_j^{(X)}\rangle^2}\right) \\
\le&2\sqrt{2}C_1\|X\|\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|.
\end{aligned}
$$

The second term on the r.h.s. of (S9.21), $\|\mathcal{O}_{\boldsymbol{\theta},\eta_0}(X) - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}(X)\|$, admits an analogous bound. As the constant multiplier is asymptotically negli-

gible, we omit it and conclude that $\mathcal{F}_K$ has an envelope function $\|X\|K$. Consequently, we obtain

$$E\|\mathbb{G}_n\|_{\mathcal{F}_K} \lesssim \int_0^{\|X\|_{P,2}K} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_K, L_2(P))} d\epsilon. \tag{S9.22}$$

We now proceed to control the bracketing number $N_{[]}(\epsilon, \mathcal{F}_K, L_2(P))$ that appears on the r.h.s. of (S9.22). For any pair $f_1, f_2 \in \mathcal{F}_K$ corresponding to parameters $(\boldsymbol{\theta}_1, \eta_1)$ and $(\boldsymbol{\theta}_2, \eta_2)$, we have

$$\begin{aligned} |f_1 - f_2| =& |(m_{n,\boldsymbol{\theta}_1,\eta_1} - m_{n,\boldsymbol{\theta}_2,\eta_2}) - (m_{n,\boldsymbol{\theta}_1,\eta_0} - m_{n,\boldsymbol{\theta}_2,\eta_0}) + (m_{n,\boldsymbol{\theta}^*,\eta_1} - m_{n,\boldsymbol{\theta}^*,\eta_2})| \\ \lesssim& \|X\| \left(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2 + \sum_{i=1}^{L_Y(n)} \|\bar{\phi}_i^{(1)(Y)} - \bar{\phi}_i^{(2)(Y)}\| + \sum_{j=1}^{L_X(n)} \|\bar{\phi}_j^{(1)(X)} - \bar{\phi}_j^{(2)(X)}\| \right). \end{aligned} \tag{S9.23}$$

Define $\mathcal{F}_i^{(Y)} := \{\bar{\phi}_i^{(Y)} : \bar{\phi}_i^{(Y)} \in H_{n,K}\}$ and $\mathcal{F}_j^{(X)} := \{\bar{\phi}_j^{(X)} : \bar{\phi}_j^{(X)} \in H_{n,K}\}$. By Assumption A_9, the covering numbers of these classes satisfy

$$\begin{aligned} \log N\left(\epsilon, \mathcal{F}_i^{(Y)}, \|\cdot\|_{L_2(P)}\right) \lesssim \epsilon^{-v}; \\ \log N\left(\epsilon, \mathcal{F}_j^{(X)}, \|\cdot\|_{L_2(P)}\right) \lesssim \epsilon^{-v}. \end{aligned}$$

Recall that $L(n) = \max\{L_X(n), L_Y(n)\}$. The numbers of sieve coefficients and retained basis functions are $L_X(n)L_Y(n)$ and $L_X(n) + L_Y(n)$, respec-

tively, and are therefore bounded by $L(n)^2$ and $2L(n)$. Using these bounds and Assumption A_9 together with inequality (S9.23), it follows that the bracketing number for the class $\mathcal{F}_K$ satisfies:

$$
\begin{aligned}
&N_{[]}(2\epsilon E\|X\|^2, \mathcal{F}_K, L_2(P)) \\
&\leq N(\frac{\epsilon}{2}, \{\boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < K\}, L_2(P)) \cdot N^{2L(n)}(\frac{\epsilon}{4L(n)}, \mathcal{F}_1^{(Y)}, L_2(P)).
\end{aligned}
$$

In conclusion, for inequality (S9.22), we have

$$
\begin{aligned}
E||\mathbb{G}_n||_{\mathcal{F}_K} &\leq \int_0^{K||X||_{P,2}} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_K, L_2(P))} d\epsilon \\
&\leq 2K\|X\|_{P,2} \int_0^{\frac{1}{2}} \sqrt{\log N_{[]}(2K\|X\|_{P,2}\epsilon, \mathcal{F}_K, L_2(P))} d\epsilon \\
&\leq 2K\|X\|_{P,2} \int_0^{\frac{1}{2}} \sqrt{\log N(\frac{\epsilon}{2}, \{\boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < K\}, L_2(P)) \cdot N^{2L(n)}(\frac{\epsilon}{4L(n)}, \mathcal{F}_1^{(Y)}, L_2(P))} d\epsilon \\
&\leq 2K\|X\|_{P,2} \Bigg(\int_0^{1/2} \sqrt{\log N\left(\frac{\epsilon}{2}, \{\boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < K\}, L_2(P)\right)} d\epsilon \\
&\qquad + \int_0^{1/2} \sqrt{\log N^{2L(n)}\left(\frac{\epsilon}{4L(n)}, \mathcal{F}_1^{(Y)}, L_2(P)\right)} d\epsilon \Bigg),
\end{aligned}
$$

and

$$
\begin{aligned}
&2K\|X\|_{P,2} \int_0^{\frac{1}{2}} \sqrt{\log N^{2L(n)}(\frac{\epsilon}{4L(n)}, \mathcal{F}_1^{(Y)}, L_2(P))} d\epsilon \\
\lesssim &2K\|X\|_{P,2} L(n)^{\frac{1}{2}} \int_0^{\frac{1}{2}} \left(\frac{\epsilon}{4L(n)}\right)^{-\frac{v}{2}} d\epsilon
\end{aligned}
$$

$$\lesssim K L(n)^{\frac{1+v}{2}}.$$

Hence

$$E\|\mathbb{G}_n\|_{\mathcal{F}_K} \lesssim L(n)^{\frac{1+v}{2}}.$$

Since $\frac{L(n)^{\mathcal{V}}}{\sqrt{n}} = o(1)$, we can conclude that $I_1 = o_p(1)$. □

Lemma 4. *Assume that conditions A_1–A_9 hold. Let $\mathcal{V} = \max\left\{1, \frac{1+v}{2}\right\}$. Suppose $\frac{L(n)^{\mathcal{V}}}{\sqrt{n}} = o(1)$ and the loss function satisfies the Lipschitz condition:*

$$|m_{n,\boldsymbol{\theta}_1,\eta_1} - m_{n,\boldsymbol{\theta}_2,\eta_2}| \lesssim \|\mathcal{O}_{\boldsymbol{\theta}_1,\eta_1} - \mathcal{O}_{\boldsymbol{\theta}_2,\eta_2}\|.$$

In addition, assume the following condition holds:

$$\|D_\eta m_{n,\boldsymbol{\theta},\eta_0+t(\hat{\eta}_n-\eta_0)} - D_\eta m_{n,\boldsymbol{\theta}^*,\eta_0+t(\hat{\eta}_n-\eta_0)}\| \lesssim F(X)_{\boldsymbol{\theta},\eta}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|,$$

for some envelope function $F(X)_{\boldsymbol{\theta},\eta}$ satisfying $\mathbb{E}(|F(X)_{\boldsymbol{\theta},\eta}|) < \infty$ in a small neighborhood of $(\boldsymbol{\theta}^*, \eta_0)$.

Then, if $\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$, we have the following convergence rate:

$$\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 = O_p\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \alpha_n^2 + M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*)}\right).$$

Proof. Consider the probability $\mathbb{P}(\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 > 2^M\delta_n)$ for a large M. We have

$$\mathbb{P}\left(\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 > 2^M\delta_n\right) = \sum_{j>M} \mathbb{P}\left(2^{j-1}\delta_n < \|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 < 2^j\delta_n\right).$$

Define the "shells" related to n: $S_{n,j} = \{\boldsymbol{\theta} \in \Theta_n : 2^{j-1}\delta_n < \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < 2^j\delta_n\}$. Hence

$$\mathbb{P}\left(2^{j-1}\delta_n < \|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 < 2^j\delta_n\right) = \mathbb{P}\left(\hat{\hat{\boldsymbol{\theta}}}_n \in S_{n,j}\right).$$

As $\hat{\hat{\boldsymbol{\theta}}}_n$ maximizes $\mathbb{M}_n(\boldsymbol{\theta})$, it is obvious that

$$\mathbb{P}\left(\hat{\hat{\boldsymbol{\theta}}}_n \in S_{n,j}\right) \leq \mathbb{P}\left(\sup_{\theta \in S_{n,j}} (\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*)) \geq 0\right),$$

and

$$\begin{aligned} M(\boldsymbol{\theta}) - M({}^{(n)}\boldsymbol{\theta}^*) &= M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) + M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*) \\ &\lesssim -2^{2j-2}\delta_n^2 \end{aligned} \tag{S9.24}$$

where $\boldsymbol{\theta} \in S_{n,j}$ and let $M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*) \leq \delta_n^2$. Note that Assumption A_2 holds only in a small region around $\boldsymbol{\theta}^*$. Since we have $\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$, and without loss of generality, we omit the condition $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 < \epsilon$, where

ϵ is a small positive constant. Using inequality (S9.24) and the fact that $\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) \geq 0$, we obtain

$$\begin{aligned}
&\sqrt{n}\left(\mathbb{M}_n(\hat{\hat{\boldsymbol{\theta}}}_n) - M({}^{(n)}\boldsymbol{\theta}^*)\right) - \sqrt{n}\left(\mathbb{M}_n({}^{(n)}\boldsymbol{\theta}^*) - M(\hat{\hat{\boldsymbol{\theta}}}_n)\right) \\
\geq &\sqrt{n}\left(M({}^{(n)}\boldsymbol{\theta}^*) - M(\hat{\hat{\boldsymbol{\theta}}}_n)\right) \\
\gtrsim &\sqrt{n}2^{2j-2}\delta_n^2
\end{aligned}$$

if $\hat{\hat{\boldsymbol{\theta}}}_n \in S_{n,j}$. Since $D^2M(\boldsymbol{\theta}^*)$ is positive definite we have $\| {}^{(n)}\boldsymbol{\theta}^* - \boldsymbol{\theta}^*\|_2^2 \leq M(\boldsymbol{\theta}^*) - M({}^{(n)}\boldsymbol{\theta}^*) \leq \delta_n^2 < 2^j\delta_n^2$. Hence ${}^{(n)}\boldsymbol{\theta}^* \in S_{n,j}$ and we only need to consider $|\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*))|$. Define $U_n(\boldsymbol{\theta}) = \mathbb{M}_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})$ and consider $\sqrt{n}(U_n(\boldsymbol{\theta}) - U_n(\boldsymbol{\theta}^*))$,

$$\begin{aligned}
&|\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*))| \\
\leq &|(\mathbb{P}_n - P)\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - (m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0})\right)| \\
&+ |(\mathbb{P}_n - P)\left(m_{n,\boldsymbol{\theta},\eta_0} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)| \\
&+ |P\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0}\right) - P\left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)|.
\end{aligned}$$

Hence

$$\mathbb{P}(\hat{\hat{\boldsymbol{\theta}}}_n \in S_{n,j}) \leq \mathbb{P}\left(\sup_{\theta \in S_{n,j}} (U_n(\boldsymbol{\theta}) - U_n(\boldsymbol{\theta}^*)) \gtrsim \sqrt{n}2^{2j-2}\delta_n^2\right)$$

$$
\begin{aligned}
&\leq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in S_{n,j}} \left|\left(\mathbb{P}_n - P\right)\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - \left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right)\right| \gtrsim \sqrt{n} 2^{2j-2}\delta_n^2\right) \\
&+ \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in S_{n,j}} \left|\left(\mathbb{P}_n - P\right)\left(m_{n,\boldsymbol{\theta},\eta_0} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right| \gtrsim \sqrt{n} 2^{2j-2}\delta_n^2\right) \\
&+ \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in S_{n,j}} \left|P\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0}\right) - P\left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right| \gtrsim \sqrt{n} 2^{2j-2}\delta_n^2\right) \\
&:= I_1 + I_2 + I_3.
\end{aligned}
$$

For the term I_1, consider the following decomposition:

$$
\begin{aligned}
&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in S_{n,j}} \left|\left(\mathbb{P}_n - P\right)\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - \left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right)\right| \gtrsim 2^{2j-2}\delta_n^2\right) \\
&\leq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in S_{n,j}} \left|\left(\mathbb{P}_n - P\right)\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - \left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right)\right| \gtrsim 2^{2j-2}\delta_n^2,\right. \\
&\quad \left.\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \lesssim 2^j\delta_n\right) \\
&+ \mathbb{P}\left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \gtrsim 2^j\delta_n\right) \\
&\leq \mathbb{P}\left(\sup_{\substack{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2<2^j\delta_n \\ \bar{\eta}_n\in H_{n,2^j\delta_n}}} \left(\mathbb{P}_n - P\right)\left(m_{n,\boldsymbol{\theta},\bar{\eta}_n} - m_{n,\boldsymbol{\theta},\eta_0} - \left(m_{n,\boldsymbol{\theta}^*,\bar{\eta}_n} - m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right) \gtrsim 2^{2j-2}\delta_n^2\right) \\
&+\mathbb{P}\left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \gtrsim 2^j\delta_n\right),
\end{aligned}
$$

where $H_{n,2^j\delta_n} := \{\bar{\phi}_i^{(X)}, \bar{\phi}_j^{(Y)} : 1 \leq i \leq L_X(n); 1 \leq j \leq L_Y(n); \sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \lesssim 2^j\delta_n\} \cup \{\phi_i^{(X)} \otimes \phi_j^{(Y)} : i > L_X(n); j > L_Y(n)\}$.

Let $\mathcal{F}_\delta := \{m_{n,\boldsymbol{\theta},\bar{\eta}} - m_{n,\boldsymbol{\theta},\eta_0} - m_{n,\boldsymbol{\theta}^*,\bar{\eta}} + m_{n,\boldsymbol{\theta}^*,\eta_0} | \boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 \leq \delta; \bar{\eta} \in H_{n,\delta}\}$, where $\mathcal{F}_\delta$ has an envelope function $\|X\|\delta$. Hence

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}_\delta} \lesssim \int_0^{2\|X\|\delta} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_\delta, L_2(P))} d\epsilon.$$

By the same argument as in the proof of Lemma 3, we have

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}_\delta} \lesssim L(n)^{\frac{1+v}{2}}.$$

Letting $\gamma_n(u) = L(n)^{\frac{1+v}{2}} u$, one obtains:

$$\sqrt{n}\mathbb{E}\left(\sup_{\boldsymbol{\theta}\in\Theta_n : \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2 \leq 2^j\delta_n ; \bar{\eta}\in H_{n,2^j\delta_n}} (\mathbb{P}_n - P)\left(m_{n,\boldsymbol{\theta},\bar{\eta}} - m_{n,\boldsymbol{\theta},\eta_0} - (m_{n,\boldsymbol{\theta}^*,\bar{\eta}} - m_{n,\boldsymbol{\theta}^*,\eta_0})\right) \gtrsim 2^{2j-2}\delta_n^2\right)$$
$$\lesssim \gamma_n(2^j\delta_n).$$

Combining the bounds yields:

$$\mathbb{P}\left(\sup_{\theta\in S_{n,j}} |(\mathbb{P}_n - P)\left(m_{\theta,\hat{\eta}_n} - m_{\theta,\eta_0} - (m_{\theta_n,\hat{\eta}_n} - m_{\theta_n,\eta_0})\right)| \gtrsim 2^{2j-2}\delta_n^2\right)$$
$$\lesssim \frac{L(n)^{\frac{1+v}{2}}}{2^j\sqrt{n}\delta_n} + \frac{E\left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\|\right)}{2^j\delta_n}.$$

For the term I_2, a similar argument to that used in Lemma 1 implies:

$$\mathbb{P}\left(\sup_{\theta\in S_{n,j}}\left|\left(\mathbb{P}_n-P\right)\left(m_{\theta,\eta_0}-m_{\theta_n,\eta_0}\right)\right|\gtrsim\sqrt{n}2^{2j-2}\delta_n^2\right)\lesssim\frac{L(n)}{2^j\sqrt{n}\delta_n}$$

To bound the bias term I_3, we apply Markov's inequality:

$$\begin{aligned}&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in S_{n,j}}\left|P\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n}-m_{n,\boldsymbol{\theta},\eta_0}\right)-P\left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n}-m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right|\gtrsim 2^{2j-2}\delta_n^2\right)\\ \leq&\frac{\mathbb{E}\sup_{\boldsymbol{\theta}\in S_{n,j}}\left|P\left(m_{n,\boldsymbol{\theta},\hat{\eta}_n}-m_{n,\boldsymbol{\theta},\eta_0}\right)-P\left(m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n}-m_{n,\boldsymbol{\theta}^*,\eta_0}\right)\right|}{2^{2j}\delta_n^2}\end{aligned}\tag{S9.25}$$

To further analyze the difference terms on the r.h.s. of inequality (S9.25), we express them in the following form:

$$\begin{aligned}m_{n,\boldsymbol{\theta},\hat{\eta}_n}-m_{n,\boldsymbol{\theta},\eta_0}&=\int_0^1 D_\eta m_{n,\boldsymbol{\theta},\eta_0+t(\hat{\eta}_n-\eta_0)}\circ(\hat{\eta}_n-\eta_0)dt;\\ m_{n,\boldsymbol{\theta}^*,\hat{\eta}_n}-m_{n,\boldsymbol{\theta}^*,\eta_0}&=\int_0^1 D_\eta m_{n,\boldsymbol{\theta}^*,\eta_0+t(\hat{\eta}_n-\eta_0)}\circ(\hat{\eta}_n-\eta_0)dt.\end{aligned}$$

Since

$$\|D_\eta m_{n,\boldsymbol{\theta},\eta_0+t(\hat{\eta}_n-\eta_0)}-D_\eta m_{n,\boldsymbol{\theta}^*,\eta_0+t(\hat{\eta}_n-\eta_0)}\|\lesssim F(X)_{\boldsymbol{\theta},\eta}\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|,$$

where $\mathbb{E}(|F(X)_{\boldsymbol{\theta},\eta}|)<\infty$ in a small neighborhood of $(\boldsymbol{\theta}^*,\eta_0)$. Therefore, I_3

is bounded by

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\theta\in S_{n,j}}\left|P\left(m_{\theta,\hat{\eta}_n}-m_{\theta,\eta_0}\right)-P\left(m_{\theta_n,\hat{\eta}_n}-m_{\theta_n,\eta_0}\right)\right|\gtrsim 2^{2j-2}\delta_n^2\right)\\
\lesssim&\frac{E\|\hat{\eta}_n-\eta_0\|}{2^j\delta_n}+\mathbb{P}\left(\|\hat{\eta}_n-\eta_0\|>\epsilon\right).
\end{aligned}$$

Combining the above results, δ_n should satisfy:

$$\begin{aligned}
&\gamma_n(\delta_n)\lesssim\sqrt{n}\delta_n^2;\\
&\frac{L(n)}{\sqrt{n}}\lesssim\delta_n;\\
&\alpha_n\lesssim\delta_n;\\
&M(\boldsymbol{\theta}^*)-M({}^{(n)}\boldsymbol{\theta}^*)\leq\delta_n^2.
\end{aligned}$$

Consequently, we may take δ_n such that: $\delta_n^2=\frac{L(n)^{2\mathcal{V}}}{n}+\alpha_n^2+M(\boldsymbol{\theta}^*)-M({}^{(n)}\boldsymbol{\theta}^*)$. □

S9.3 Smoothed FoF Geometric Quantile Regression

In this subsection, the loss function is smoothed as follows:

$$m_{n,\boldsymbol{\theta},\eta}=\mathcal{K}_n\left(\|Y-\mathcal{O}_{\boldsymbol{\theta},\eta}(X)\|\right)+\langle\tau,Y-\mathcal{O}_{\boldsymbol{\theta},\eta}(X)\rangle.$$

Define the population objective function as $M_n(\boldsymbol{\theta}) = -Pm_{n,\boldsymbol{\theta},\eta_0}$ and the empirical objective function as $\mathbb{M}_n(\boldsymbol{\theta}) = -\mathbb{P}_n m_{n,\boldsymbol{\theta},\tilde{\eta}_n}$, where

$$\tilde{\eta}_n := \{\tilde{\phi}_j^{(X)}\}_{j=1}^{L_X(n)} \cup \{\phi_j^{(X)}\}_{j=L_X(n)+1}^{\infty} \cup \{\tilde{\phi}_i^{(Y)}\}_{i=1}^{L_Y(n)} \cup \{\phi_i^{(Y)}\}_{i=L_Y(n)+1}^{\infty}.$$

Denote the estimator and the oracle parameter by

$$\hat{\tilde{\boldsymbol{\theta}}}_n = \arg\max_{\boldsymbol{\theta}\in\Theta_n} \mathbb{M}_n(\boldsymbol{\theta});$$

$$\boldsymbol{\theta}_n^* = \arg\max_{\boldsymbol{\theta}\in\Theta} M_n(\boldsymbol{\theta}).$$

Due to the smoothing of the loss function, the parameter $\boldsymbol{\theta}_n^*$ differs from the parameter $\boldsymbol{\theta}^*$ associated with the non-smoothed loss $M(\boldsymbol{\theta})$. The first lemma in this section establishes that $\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}^*\|_2 = o(1)$. Under this setting, the consistency of the estimator $\hat{\tilde{\boldsymbol{\theta}}}_n$ with respect to $\boldsymbol{\theta}_n^*$ can be obtained by following the same reasoning used in the proofs of Lemma 3 and Lemma 4. Furthermore, the corresponding convergence rate can also be derived in a similar manner. These results are formally stated in Lemma 6. Let $\mathcal{O}_n^* := \mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0}$. At the end of this subsection, we provide a bound on the distance between $\mathcal{O}_n^*$ and the true operator $\mathcal{O}^*$, which we refer to as the smoothing bias.

Lemma 5. *Under Assumption A_1, we have*

$$\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}^*\|_2 = o(1).$$

Proof. We prove the claim by contradiction: suppose, to the contrary, that there exists a constant $\delta > 0$ such that $\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}^*\|_2 > \delta$ for all n, By the definition of $\boldsymbol{\theta}_n^*$, we have

$$M_n(\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}^*) > 0. \tag{S9.26}$$

Meanwhile, the definition of $\boldsymbol{\theta}^*$ further implies that

$$M(\boldsymbol{\theta}^*) > \sup_{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2>\delta} M(\boldsymbol{\theta}).$$

Then

$$M(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*) \leq \sup_{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2>\delta} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) := -\psi(\delta). \tag{S9.27}$$

Here, $\psi(\delta) > 0$ by Equation (S9.27) and $M(\boldsymbol{\theta}^*) > \sup_{\boldsymbol{\theta}\in\Theta:\,\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2>\delta} M(\boldsymbol{\theta})$. Combining inequalities (S9.26) and (S9.27), we obtain

$$M_n(\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}^*) - (M(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*)) > \psi(\delta). \tag{S9.28}$$

Moreover, for all $\boldsymbol{\theta} \in \Theta$ satisfying $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_2 > \delta$, we have

$$|M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})| \leq P\left(\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}\| + \mathcal{K}_n(\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}\|)\right) \mathbb{I}\{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}\| < h_n\}$$
$$\leq h_n \to 0.$$

This contradicts inequality (S9.28). Hence $\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}^*\| = o(1)$. □

Lemma 6. *Let* $\mathcal{V} = \max\left\{1, \frac{1+v}{2}\right\}$. *Under Assumptions* A_1–A_{11}, *if* $\frac{L(n)^{\mathcal{V}}}{\sqrt{n}} = o(1)$, *then we have*

$$\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\|_2 = O_p\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n(\boldsymbol{\theta}_n^*) - M_n({}^{(n)}\boldsymbol{\theta}_n^*)}\right).$$

Proof. For any $\delta > 0$, note that

$$M_n(\boldsymbol{\theta}_n^*) > \sup_{\boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}_n^*\|_2 > \delta} M_n(\boldsymbol{\theta}).$$

If $\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\|_2 > \delta$, we have $\mathbb{M}_n(\hat{\tilde{\boldsymbol{\theta}}}_n) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}_n^*) > 0$ and

$$M_n(\hat{\tilde{\boldsymbol{\theta}}}_n) - M_n({}^{(n)}\boldsymbol{\theta}_n^*) - (M_n(\boldsymbol{\theta}_n^*) - M_n({}^{(n)}\boldsymbol{\theta}_n^*)) < \sup_{\boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}_n^*\|_2 > \delta} M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_n^*).$$

Therefore

$$\mathbb{P}\left(\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\| > \delta\right)$$

$$\leq \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}_n^*) - (M_n(\boldsymbol{\theta}) - M_n({}^{(n)}\boldsymbol{\theta}_n^*))| + |M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)| \right.$$
$$\left. > - \sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M_n(\boldsymbol{\theta}) - M_n(\boldsymbol{\theta}_n^*) \right)$$
$$= \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}_n^*) - (M_n(\boldsymbol{\theta}) - M_n({}^{(n)}\boldsymbol{\theta}_n^*))| + |M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)| \right.$$
$$\left. + \sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta}) - (M_n(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*)) > - \sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) \right).$$

Since $\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}^*\|_2 = o(1)$, there exists N such that for all $n > N$, $\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}^*\|_2 < \frac{\delta}{2}$. Consequently,

$$\sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) \leq \sup_{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2>\frac{\delta}{2}} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) := -\psi(\delta).$$

Thus, for all $n > N$,

$$\sup_{n>N} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}_n^*) - (M_n(\boldsymbol{\theta}) - M_n({}^{(n)}\boldsymbol{\theta}_n^*))| + |M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)| \right.$$
$$\left. + \sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta}) - (M_n(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*)) > - \sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}^*) \right)$$
$$\leq \sup_{n>N} \mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}_n^*) - (M_n(\boldsymbol{\theta}) - M_n({}^{(n)}\boldsymbol{\theta}_n^*))| + |M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)| \right.$$
$$\left. + \sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta}) - (M_n(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*)) > \psi(\delta) \right). \tag{S9.29}$$

Hence, to complete the argument, it is necessary to show that the r.h.s of

the probability inequality tends to zero in probability (S9.29).

For the term $|M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)|$, we have

$$|M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)| \leq |M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}^*)| + |M_n(\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}^*)|.$$

To ensure convergence, we select a sequence ${}^{(n)}\boldsymbol{\theta}_n^*$ such that ${}^{(n)}\boldsymbol{\theta}_n^* \to \boldsymbol{\theta}^*$, This implies that $|M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}^*)| = o(1)$. Moreover, since $\boldsymbol{\theta}_n^* \to \boldsymbol{\theta}^*$, it follows that $|M_n(\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}^*)| \leq C\|X\|_{P,1}\|\boldsymbol{\theta}_n^* - \boldsymbol{\theta}^*\|$ which converges to zero as $n \to \infty$.

Next, we consider the term $\sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta}) - (M_n(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*))$ appearing on the r.h.s. of inequality (S9.29). Using Lemma 5, we have that $\sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|_2>\delta} M_n(\boldsymbol{\theta}) - M(\boldsymbol{\theta})$ converges to zero. Then

$$|M_n(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*)| \leq |M_n(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}_n^*)| + |M(\boldsymbol{\theta}_n^*) - M(\boldsymbol{\theta}^*)|$$

also converges to zero.

The remaining step is to show that the following supremum appearing on the r.h.s. of inequality (S9.29) converges to zero in probability: $\sup_{\boldsymbol{\theta}\in\Theta_n} |\mathbb{M}_n(\boldsymbol{\theta}) - \mathbb{M}_n({}^{(n)}\boldsymbol{\theta}_n^*) - (M_n(\boldsymbol{\theta}) - M_n({}^{(n)}\boldsymbol{\theta}_n^*))|$. To this end, observe

that

$$\begin{aligned}&\sup_{\boldsymbol{\theta}\in\Theta_n}|\mathbb{M}_n(\boldsymbol{\theta})-\mathbb{M}_n({}^{(n)}\boldsymbol{\theta}_n^*)-(M_n(\boldsymbol{\theta})-M_n({}^{(n)}\boldsymbol{\theta}_n^*))|\\ \leq&2\sup_{\boldsymbol{\theta}\in\Theta_n}|\mathbb{M}_n(\boldsymbol{\theta})-\mathbb{M}_n(\boldsymbol{\theta}_n^*)-(M_n(\boldsymbol{\theta})-M_n(\boldsymbol{\theta}_n^*))|\end{aligned}$$

Given that the loss satisfies

$$|m_{n,\boldsymbol{\theta}_1,\eta_1}-m_{n,\boldsymbol{\theta},\eta_2}|\lesssim\|\mathcal{O}_{\boldsymbol{\theta}_1,\eta_1}-\mathcal{O}_{\boldsymbol{\theta}_2,\eta_1}\|,$$

we can invoke Lemma 3 to conclude that $\|\hat{\tilde{\boldsymbol{\theta}}}_n-\boldsymbol{\theta}_n^*\|_2=o_p(1)$.

Moreover, the derivative of the smoothed loss function with respect to η takes the form

$$\begin{aligned}&D_\eta m_{n,\boldsymbol{\theta},\tilde{\eta}}(\eta)\\ =&\left[\frac{\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}-Y}{\|\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}-Y\|}\mathbb{I}\{\|\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}-Y\|>h_n\}+\left(\frac{2}{h_n}-\frac{\|\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}-Y\|}{h_n^2}\right)(\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}-Y)\mathbb{I}\{\|\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}-Y\|<h_n\}\right]\\ &\circ D_\eta\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\circ\eta\\ :=&D_{\mathcal{O}}\mathcal{K}_n(\|Y-\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\|)\circ D_\eta\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\circ\eta\end{aligned}$$

And we observe that

$$D_{\mathcal{O}}^2\mathcal{K}_n(\|Y-\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\|)$$

$$= \left(\frac{1}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\|} (Id - \frac{(Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}) \otimes (Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}})}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\|^2}) \right) \mathbb{I}\{\|Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\| > h_n\}$$
$$+ \left((\frac{2}{h_n} - \frac{\|\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}} - Y\|}{h_n^2}) Id - \frac{(Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}) \otimes (Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}})}{h_n^2 \|\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}} - Y\|} \right) \mathbb{I}\{\|Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\| < h_n\}$$

It can be shown that $\|D^2_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\|)\|$ satisfies the bound

$$\|D^2_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}}\|)\| \lesssim \frac{1}{\|\mathcal{O}_{\boldsymbol{\theta},\tilde{\eta}} - Y\|}.$$

By arguments similar to those in the proof of Theorem 2, one obtains the following convergence rate:

$$\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\|_2 = O_p \left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n(\boldsymbol{\theta}_n^*) - M_n({}^{(n)}\boldsymbol{\theta}_n^*)} \right).$$

□

Lemma 7. *Under Assumptions A_2, A_3, and A_{10}, the smoothing bias satisfies*

$$\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}^*\|_{\mathcal{X}} = O\left(h_n^p\right)$$

Proof. Define $f_{\mathcal{O}_{\boldsymbol{\theta}}} = \|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\| + \langle \tau, Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X) \rangle$ and $f_{n,\mathcal{O}_{\boldsymbol{\theta}}} = m_{n,\boldsymbol{\theta},\eta_0}$.

Let

$$Df_{\mathcal{O}_{\boldsymbol{\theta}}} = -\frac{Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|} - \tau$$
$$Df_{n,\mathcal{O}_{\boldsymbol{\theta}}} = -\mathcal{K}_n'(\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|)\frac{Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|} - \tau$$

and

$$D^2 f_{\mathcal{O}_{\boldsymbol{\theta}}} = \frac{1}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|}\left(Id - \frac{(Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)) \otimes (Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X))}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|^2}\right)$$
$$D^2 f_{n,\mathcal{O}_{\boldsymbol{\theta}}} = \mathcal{K}_n''(\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|)\frac{(Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)) \otimes (Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X))}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|^2}$$
$$+ \frac{\mathcal{K}_n'(\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|)}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|}\left(Id - \frac{(Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)) \otimes (Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X))}{\|Y - \mathcal{O}_{\boldsymbol{\theta},\eta_0}(X)\|^2}\right).$$

Consider the quantity

$$\left|\left(PDf_{n,\mathcal{O}_{\boldsymbol{\theta}_n^*}} - PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*}}\right) \circ (\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0}(X) - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}(X))\right|$$
$$\leq \|PDf_{\mathcal{O}_{\boldsymbol{\theta}^*},\eta_0} - PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*},\eta_0}\|\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}}$$
$$\leq h_n^p \|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}},$$

where the final inequality holds due to Assumption A_{10}, which ensures

$$\|PDf_{\mathcal{O}_{\boldsymbol{\theta}^*},\eta_0} - PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*},\eta_0}\| \leq 2P\mathbb{I}\{\|Y - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\| < h_n\} \leq h_n^p.$$

We further examine the expansion:

$$
\begin{aligned}
&\left(PDf_{n,\mathcal{O}_{\boldsymbol{\theta}_n^*}} - Pf_{n,\mathcal{O}_{\boldsymbol{\theta}^*}}\right) \circ (\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}) \\
&= PD^2 f_{n,\mathcal{O}_{\boldsymbol{\theta}^*}} \circ (\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}, \mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}) \\
&\quad + \int_0^1 PD^2 f_{n,\,\mathcal{O}_{\boldsymbol{\theta}^*}+(1-t)(\mathcal{O}_{\boldsymbol{\theta}_n^*}-\mathcal{O}_{\boldsymbol{\theta}^*})} \circ (\mathcal{O}_{\boldsymbol{\theta}_n^*} - \mathcal{O}_{\boldsymbol{\theta}^*}, \mathcal{O}_{\boldsymbol{\theta}_n^*} - \mathcal{O}_{\boldsymbol{\theta}^*})\, dt \\
&\quad - PD^2 f_{n,\mathcal{O}_{\boldsymbol{\theta}^*}} \circ (\mathcal{O}_{\boldsymbol{\theta}_n^*} - \mathcal{O}_{\boldsymbol{\theta}^*}, \mathcal{O}_{\boldsymbol{\theta}_n^*} - \mathcal{O}_{\boldsymbol{\theta}^*}) \\
&> \mu_- \left\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\right\|_{\mathcal{X}}^2 - \frac{1}{h_n^2}\int_0^1 (1-t)\, dt \left\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\right\|_{\mathcal{X}}^3.
\end{aligned}
$$

Thus, we obtain the inequality

$$
\frac{1}{2h_n^2}\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}}^2 - \mu_-\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}} + h_n^p \geq 0.
$$

Solving this quadratic inequality yields both upper and lower bounds:

$$
\begin{aligned}
\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}} &\leq \frac{h_n^p}{\mu_- + \sqrt{\mu_-^2 - 2h_n^{p-2}}} \\
\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}} &\geq h_n^2\left(\mu_- + \sqrt{\mu_-^2 - 2h_n^{p-2}}\right)
\end{aligned}
$$

Assume, for contradiction, that the second bound holds. Then $\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}} > \sqrt{2}h_n^{\frac{p}{2}+1} := r_0$. By the convexity of $Pf_{n,\mathcal{O}}$, and Lemma C.1 in the supplementary material of Sun et al. (2020), there exists some $\beta \in (0,1)$ such that $\tilde{\mathcal{O}} := (1-\beta)\mathcal{O}_{\boldsymbol{\theta}^*,\eta_0} + \beta\mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}$ satisfies $\|\tilde{\mathcal{O}} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}} = \eta\|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} -$

$\mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}\|_{\mathcal{X}} = r_0$ and

$$\begin{aligned}
&\left(PDf_{n,\tilde{\mathcal{O}}} - PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}}\right) \circ (\tilde{\mathcal{O}} - \mathcal{O}_{\boldsymbol{\theta}^*_n,\eta_0}) \\
\leq & c\left(PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*_n,\eta_0}} - PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}}\right) \circ (\mathcal{O}_{\boldsymbol{\theta}^*_n,\eta_0} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}) \\
= & \left(PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*_n,\eta_0}} - PDf_{n,\mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}}\right) \circ (\tilde{\mathcal{O}} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_0}).
\end{aligned}$$

Repeating the earlier analysis, we obtain

$$\begin{aligned}
\mu_- r_0^2 - \frac{1}{2h_n^2} r_0^3 &\leq h_n^p r_0 \\
\mu_- r_0^2 - h_n^{\frac{p}{2}-1} r_0^2 &\leq h_n^p r_0.
\end{aligned}$$

This implies $r_0 \leq \frac{h_n^p}{\mu_- - h_n^{\frac{p}{2}-1}} < h_n^{\frac{p}{2}+1} = r_0$, leading to a contradiction. □

S9.4 Supporting Lemma for Theorem 4

In this subsection, we define a family of "good events" and establish that, under a properly chosen sequence h_n, the probabilities of their complements converge to zero. Specifically, let M be a finite constant, and we define the events as follows:

$$\begin{aligned}
\mathcal{E}_0 &:= \left\{ \|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*_n\| \leq \min\{\frac{C_4\mu_-}{4} h_n^2, \frac{(1-C_4)\mu_-}{8\mu_+} h_n^2\} \right\}; \\
\mathcal{E}_j &:= \left\{ \frac{1}{n_j} \sum_{i=1}^{n_j} \|X_{i,j}\|^3 < M; \|\mathbb{P}_{n_j} D_{\boldsymbol{\theta}} m_{n,^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\| \leq \frac{(1-2C_4)}{4} h_n^2; \right.
\end{aligned}$$

$$\|\mathbb{P}_{n_j} D^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} - PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}\| \leq \frac{\mu_- C_4}{4} \Bigg\}.$$

Remark 2. Define the quantities

$$\mu_+ = \sup_{\|x\|=1} PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}(x,x), \quad \mu_- = \inf_{\|x\|=1} PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}(x,x),$$

which are assumed to exist for sufficiently large n. According to Assumptions A_2 and A_3, there exist strictly positive constants μ^*_+ and μ^*_- such that

$$\mu^*_+ = \sup_{\|x\|=1} PD^2_{\boldsymbol{\theta}} m_{\boldsymbol{\theta}^*,\eta_0}(x,x), \quad \mu^*_- = \inf_{\|x\|=1} PD^2_{\boldsymbol{\theta}} m_{\boldsymbol{\theta}^*,\eta_0}(x,x).$$

Furthermore, since $D^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0} \to D^2_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}^*_n,\eta_0}$ as $n \to \infty$, and by Assumption A_2 together with the continuity argument above, we also have $PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0} \to PD^2_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}^*_n,\eta_0}$. Therefore, for sufficiently large n, the quantities μ_+ and μ_- are well-defined. Since our subsequent analysis is asymptotic (i.e., as $n \to \infty$), it is without loss of generality to assume the existence of μ_+ and μ_- throughout.

Lemma 8. *Let* $\mathcal{V} = \max\left\{1, \frac{1+v}{2}\right\}$. *Suppose Assumptions* A_1–A_{11} *hold, and*

let h_n satisfy the following conditions:

$$h_n^2 > \max\left\{\frac{L(n)^{\mathcal{V}}\,\omega(1)}{\sqrt{N}},\ \sqrt{M_n(\boldsymbol{\theta}_n^*) - M_n({}^{(n)}\boldsymbol{\theta}_n^*)}\,\omega(1),\ \frac{L(n)}{\sqrt{n}}\,\omega(1)\right\};$$

$$h_n^2 > \frac{8}{1-2C_4}\left\|PD_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0} - PD_{\boldsymbol{\theta}}m_{n,\boldsymbol{\theta}_n^*,\eta_0}\right\|;$$

$$h_n^3 > \kappa_n\,\omega(1).$$

Then we have

$$\mathbb{P}\left(\bigcup_{j=0}^{K}\mathcal{E}_j^c\right) = o(1)$$

Proof. Since $n \to \infty$, the constant terms associated with the good events become asymptotically negligible and are thus omitted in the subsequent analysis.

To bound the probability of the event $\mathcal{E}_0^c$, we consider the following decomposition:

$$\mathbb{P}\left(\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\| > 2^M h_n^2\right) \le \sum_{j>M:2^j h_n^2<\epsilon} \mathbb{P}\left(2^{j-1}h_n^2 \le \|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\| \le 2^j h_n^2\right) + \mathbb{P}\left(\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\| > \eta\right),$$

By applying Lemma 4 with δ_n replaced by h_n^2, and invoking Lemma 6, we derive an upper bound for each term in the summation. In particular, if h_n

satisfies the following conditions:

$$
\begin{aligned}
h_n^2 &> \frac{L(n)^{\mathcal{V}}\omega(1)}{\sqrt{N}};\\
h_n^2 &> \sqrt{M_n(\boldsymbol{\theta}_n^*) - M_n({}^{(n)}\boldsymbol{\theta}_n^*)}\omega(1);\\
h_n^3 &> \kappa_n\omega(1),
\end{aligned}
$$

Then we have $P(\mathcal{E}_0^c) \to 0$ as $n \to \infty$.

For the Event $\mathcal{E}_j$, we first consider

$$
P\left(\|\mathbb{P}_{n_j} D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\| \leq \frac{(1-2C_4)}{4} h_n^2\right).
$$

To control this quantity, we decompose $\|\mathbb{P}_{n_j} D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|$ as follows:

$$
\begin{aligned}
&\|\mathbb{P}_{n_j} D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|\\
\leq &\|\mathbb{P}_{n_j} D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - P D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|\\
&+ \|P D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - P D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0}\|\\
&+ \|P D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0} - P D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\eta_0}\|. \qquad \text{(S9.30)}
\end{aligned}
$$

Note that

$$
D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} = D_{\mathcal{O}}\mathcal{K}_n(\|Y - \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|) \circ D_{\boldsymbol{\theta}}\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}.
$$

And

$$
\begin{aligned}
&D_{\boldsymbol{\theta}}\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}({}^{(n)}\boldsymbol{\theta}) \\
=&\lim_{t\to 0}\frac{1}{t}\sum_{i=1}^{L_Y(n)}\left(f_i(\sum_{j=1}^{L_X(n)}(\theta_{i,j}^*+t\theta_{i,j})\langle X,\tilde{\phi}_j^{(X)}\rangle)-f_i(\sum_{j=1}^{L_X(n)}\theta_{i,j}^*\langle X,\tilde{\phi}_j^{(X)}\rangle)\right)\tilde{\phi}_i^{(Y)} \\
=&\sum_{i=1}^{L_Y(n)}\left(f_i'(\sum_{j=1}^{L_X(n)}\theta_{i,j}^*\langle X,\tilde{\phi}_j^{(X)}\rangle)\right)\sum_{j=1}^{L_X(n)}\theta_{i,j}\langle X,\tilde{\phi}_j^{(X)}\rangle\tilde{\phi}_i^{(Y)}
\end{aligned}
$$

Hence, for the first term on the r.h.s. of inequality (S9.30), namely $\|\mathbb{P}_{n_j}D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}-PD_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|$, we obtain the bound

$$
\begin{aligned}
\|D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\| &\le 3\|D_{\boldsymbol{\theta}}\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\| \\
&\le 3C_1\sqrt{\sum_{i=1}^{L_Y(n)}\left(\sum_{j=1}^{L_X(n)}\theta_{i,j}\langle X,\tilde{\phi}_j^{(X)}\rangle\right)^2} \\
&\lesssim \|X\|.
\end{aligned}
$$

This implies that $\|X\|$ serves as an envelope function for the class $D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}$.

Hence we have

$$
\begin{aligned}
&\|D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_1)-D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_2)\| \\
&\lesssim \|D_{\boldsymbol{\theta}}\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\| \lesssim \|X\|\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|.
\end{aligned}
$$

Let us define the function class $\mathcal{F}_1 := \{D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}) : \boldsymbol{\theta}\in\Theta_n\}$. Then

the expected of the empirical process indexed by $\mathcal{F}_1$ can be bounded as

$$\begin{aligned}\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}_1} &\lesssim \int_0^{\|X\|_{P,2}} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}_1, L_2(P))} d\epsilon \\ &\lesssim \int_0^{\frac{1}{2}} \sqrt{\log N(\epsilon, \Theta_n, L_2(P))} d\epsilon \lesssim L(n)\end{aligned}$$

Therefore, we conclude

$$\begin{aligned}&\mathbb{P}\left(\|\mathbb{P}_{n_j} D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - P D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\| > h_n^2\right) \\ &\lesssim \frac{\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}_1}}{\sqrt{n} h_n^2} \lesssim \frac{L(n)}{\sqrt{n} h_n^2},\end{aligned}$$

which converges to zero provided that

$$h_n^2 > \frac{L(n)}{\sqrt{n}} \omega(1).$$

To prove that the second term on the r.h.s. of inequality (S9.30), $\|P D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - P D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0}\|$ converges to zero, we begin by bounding $\|D_{\boldsymbol{\theta}} \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - D_{\boldsymbol{\theta}} \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0}\|$:

$$\begin{aligned}&\|D_{\boldsymbol{\theta}} \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - D_{\boldsymbol{\theta}} \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0}\| \\ \leq &\| \sum_{i=1}^{L_Y(n)} f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)}\rangle\Big) \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)}\rangle \tilde{\phi}_i^{(Y)}\end{aligned}$$

$$
\begin{aligned}
&- \sum_{i=1}^{L_Y(n)} f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \phi_j^{(X)}\rangle\Big) \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \phi_j^{(X)}\rangle \phi_i^{(Y)} \| \\
\leq & \| \sum_{i=1}^{L_Y(n)} \left(f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)}\rangle\Big) - f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \phi_j^{(X)}\rangle\Big) \right) \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)}\rangle \tilde{\phi}_i^{(Y)} \| \\
&+ \| \sum_{i=1}^{L_Y(n)} f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)}\rangle\Big) \left(\sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)}\rangle \tilde{\phi}_i^{(Y)} - \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \phi_j^{(X)}\rangle \phi_i^{(Y)} \right) \|.
\end{aligned} \tag{S9.31}
$$

To handle the first term on the r.h.s. of (S9.31), we have

$$
\begin{aligned}
& \| \sum_{i=1}^{L_Y(n)} \left(f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)}\rangle\Big) - f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \phi_j^{(X)}\rangle\Big) \right) \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)}\rangle \tilde{\phi}_i^{(Y)} \| \\
= & \sqrt{\sum_{i=1}^{L_Y(n)} \left(f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)}\rangle\Big) - f_i'\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \phi_j^{(X)}\rangle\Big) \right)^2 \left(\sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)}\rangle \right)^2} \\
\lesssim & \sqrt{\sum_{i=1}^{L_Y(n)} \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)} - \phi_j^{(X)}\rangle \right)^2 \left(\sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)}\rangle \right)^2} \\
\leq & \sqrt{\sum_{i=1}^{L_Y(n)} \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{*2} \sum_{j=1}^{L_X(n)} \langle X, \tilde{\phi}_j^{(X)} - \phi_j^{(X)}\rangle^2 \right) \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^2 \sum_{j=1}^{L_X(n)} \langle X, \tilde{\phi}_j^{(X)}\rangle^2 \right)} \\
\lesssim & \|X\|^2 \left(\sum_{j=1}^{L_X(n)} \|\tilde{\phi}_j^{(X)} - \phi_j^{(X)}\| \right).
\end{aligned}
$$

For the second part on the r.h.s. of (S9.31), we have

$$
\begin{aligned}
&\| \sum_{i=1}^{L_Y(n)} f_i'(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)} \rangle) \left(\sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)} \rangle \tilde{\phi}_i^{(Y)} - \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \phi_j^{(X)} \rangle \phi_i^{(Y)} \right) \| \\
=&\| \sum_{i=1}^{L_Y(n)} f_i'(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)} \rangle) \left[\sum_{j=1}^{L_X(n)} \theta_{i,j} \left(\langle X, \tilde{\phi}_j^{(X)} - \phi_j^{(X)} \rangle \tilde{\phi}_i^{(Y)} + \langle X, \phi_j^{(X)} \rangle (\tilde{\phi}_i^{(Y)} - \phi_i^{(Y)}) \right) \right] \| \\
\lesssim&\| \sum_{i=1}^{L_Y(n)} \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}_j^{(X)} - \phi_j^{(X)} \rangle \tilde{\phi}_i^{(Y)} \| + \sum_{i=1}^{L_Y(n)} \sum_{j=1}^{L_X(n)} |\theta_{i,j} \langle X, \phi_j^{(X)} \rangle| \| \tilde{\phi}_i^{(Y)} - \phi_i^{(Y)} \| \\
\lesssim&\|X\| \left(\sum_{j=1}^{L_X(n)} \| \tilde{\phi}_j^{(X)} - \phi_j^{(X)} \| \right) + \sum_{i=1}^{L_Y(n)} \sqrt{\sum_{j=1}^{L_X(n)} \theta_{i,j}^2 \sum_{j=1}^{L_X(n)} \langle X, \phi_j^{(X)} \rangle^2} \| \tilde{\phi}_i^{(Y)} - \phi_i^{(Y)} \| \\
\lesssim&\|X\| \left(\sum_{i=1}^{L_Y(n)} \| \tilde{\phi}_i^{(Y)} - \phi_i^{(Y)} \| + \sum_{j=1}^{L_X(n)} \| \tilde{\phi}_j^{(X)} - \phi_j^{(X)} \| \right).
\end{aligned}
$$

In addition, we consider $\|D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \tilde{\eta}_n}\|) - D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \eta_0}\|)\|$:

$$
\begin{aligned}
&\|D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \tilde{\eta}_n}\|) - D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \eta_0}\|)\| \\
\leq&\frac{1}{h_n} \|\mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \tilde{\eta}_n} - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \eta_0}\| \lesssim \frac{\|X\|}{h_n} \left(\sum_{i=1}^{L_Y(n)} \| \tilde{\phi}_i^{(Y)} - \phi_i^{(Y)} \| + \sum_{j=1}^{L_X(n)} \| \tilde{\phi}_j^{(X)} - \phi_j^{(X)} \| \right).
\end{aligned}
$$

Combining the above bounds, we obtain:

$$
\begin{aligned}
&\|D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*, \tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*, \eta_0}\| \\
\leq&\|D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \tilde{\eta}_n}\|) - D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \eta_0}\|)\| \cdot \|D_{\boldsymbol{\theta}} \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \tilde{\eta}_n}\| \\
&+ \|D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \eta_0})\| \cdot \|D_{\boldsymbol{\theta}} \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \tilde{\eta}_n} - D_{\boldsymbol{\theta}} \mathcal{O}_{(n)\boldsymbol{\theta}_n^*, \eta_0}\|
\end{aligned}
$$

$$\lesssim(\frac{\|X\|^2}{h_n}+\|X\|)\left(\sum_{i=1}^{L_Y(n)}\|\tilde{\phi}_i^{(Y)}-\phi_i^{(Y)}\|+\sum_{j=1}^{L_X(n)}\|\tilde{\phi}_j^{(X)}-\phi_j^{(X)}\|\right)$$

Therefore, we have that

$$\mathbb{P}\left(\|D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}-D_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0}\|>h_n^2\right)$$

converges to zero provided that $h_n^3>\kappa_n\omega(1)$. Finally, we choose $\|PD_{\boldsymbol{\theta}}m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0}-PD_{\boldsymbol{\theta}}m_{n,\boldsymbol{\theta}_n^*,\eta_0}\|<\frac{(1-2C_4)}{8}h_n^2$.

Next we consider

$$\mathbb{P}\left(\|\mathbb{P}_{n_j}D_{\boldsymbol{\theta}}^2m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}-PD_{\boldsymbol{\theta}}^2m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0}\|>\frac{\mu_-C_4}{4}\right)$$

where

$$\begin{aligned}D_{\boldsymbol{\theta}}^2m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}=&D_{\boldsymbol{\theta}}\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\circ D_{\mathcal{O}}^2\mathcal{K}_n(\|Y-\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|)\circ D_{\boldsymbol{\theta}}\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\\&+D_{\mathcal{O}}\mathcal{K}_n(\|Y-\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|)\circ D_{\boldsymbol{\theta}}^2\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\end{aligned}$$

And

$D_{\boldsymbol{\theta}}^2\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$

$$= \sum_{i=1}^{L_Y(n)} f_i''\Big(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)} \rangle\Big) \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \tilde{\phi}_j^{(X)} \rangle\right) \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \tilde{\phi}_j^{(X)} \rangle\right) \tilde{\phi}_i^{(Y)}.$$

To establish the convergence of

$$\|\mathbb{P}_{n_j} D_{\boldsymbol{\theta}}^2 m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - P D_{\boldsymbol{\theta}}^2 m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|$$

we begin by verifying several key properties. First, we analyze $\|D_{\boldsymbol{\theta}}^2 \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\|$:

$$\begin{aligned}
&\|D_{\boldsymbol{\theta}}^2 \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\| \\
=&\sqrt{\sum_{i=1}^{L_Y(n)} (f_i''(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \tilde{\phi}_j^{(X)} \rangle))^2 \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \tilde{\phi}_j^{(X)} \rangle\right)^2 \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \tilde{\phi}_j^{(X)} \rangle\right)^2} \\
\lesssim&\sqrt{\sum_{i=1}^{L_Y(n)} \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)2} \sum_{j=1}^{L_X(n)} \langle X, \tilde{\phi}_j^{(X)} \rangle^2 \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)2} \sum_{j=1}^{L_X(n)} \langle X, \tilde{\phi}_j^{(X)} \rangle^2} \\
\lesssim&\|X\|^2.
\end{aligned}$$

Consequently, the second derivative of the smoothed loss is bounded as

$$\|D_{\boldsymbol{\theta}}^2 m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\| \lesssim (\frac{1}{h_n} + 1)\|X\|^2 \lesssim \frac{\|X\|^2}{h_n}.$$

Moreover, the second-order derivative satisfies

$$\|D_{\boldsymbol{\theta}}^2 m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_3) - D_{\boldsymbol{\theta}}^2 m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\theta}_2, \boldsymbol{\theta}_4)\| \lesssim \frac{\|X\|^2}{h_n} \left(\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2 + \|\boldsymbol{\theta}_3 - \boldsymbol{\theta}_4\|_2\right).$$

Define the function class $\mathcal{F}_2 := \{D^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n}(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) : \boldsymbol{\theta}_1,\boldsymbol{\theta}_2 \in \Theta_n\}$. Then

$$\begin{aligned}\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}_2} &\lesssim \int_0^{\frac{\|X\|^2_{P,2}}{h_n}} \sqrt{\log N_{[]}(\epsilon,\mathcal{F}_2,L_2(P))}d\epsilon \\ &\lesssim \frac{1}{h_n}\int_0^{\frac{1}{2}} \sqrt{\log N_{[]}(\frac{2\|X\|^2_{P,2}}{h_n}\epsilon,\mathcal{F}_2,L_2(P))}d\epsilon \\ &\lesssim \frac{1}{h_n}\int_0^{\frac{1}{2}} \sqrt{\log N(\frac{\epsilon}{2},\Theta_n,L_2(P))}d\epsilon \lesssim \frac{L(n)}{h_n}.\end{aligned}$$

Since $\frac{L(n)}{h_n\sqrt{n}} = o(1)$, it follows that

$$\mathbb{P}\left(\|\mathbb{P}_{n_j} D^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} - PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\| > \frac{\mu_- C_4}{8}\right)$$

converges to zero.

Next, we derive the conditions under which

$$\|PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} - PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}\| = o_p(1)$$

holds. We now analyze the discrepancy:

$$\begin{aligned}&\|D^2_{\mathcal{O}}\mathcal{K}_n(\|Y - \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\|) - D^2_{\mathcal{O}}\mathcal{K}_n(\|Y - \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}\|)\| \lesssim \frac{1}{h_n^2}\|\mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} - \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}\| \\ &\lesssim \frac{\|X\|}{h_n}\left(\sum_{i=1}^{L_Y(n)}\|\tilde{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)}\|\tilde{\phi}_j^{(X)} - \phi_j^{(X)}\|\right).\end{aligned}$$

And

$$
\begin{aligned}
&\|(D^2_{\boldsymbol{\theta}}\mathcal{O}_{(n)_{\boldsymbol{\theta}^*_n,\tilde{\eta}_n}} - D^2_{\boldsymbol{\theta}}\mathcal{O}_{(n)_{\boldsymbol{\theta}^*_n,\eta_0}})(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\| \\
\leq &\| \sum_{i=1}^{L_Y(n)} \left(f_i''(\sum_{j=1}^{L_X(n)} \theta^*_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle) - f_i''(\sum_{j=1}^{L_X(n)} \theta^*_{i,j}\langle X, \phi^{(X)}_j\rangle) \right) \qquad \text{(S9.32)} \\
&\cdot \sum_{j=1}^{L_X(n)} \theta^{(2)}_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle \sum_{j=1}^{L_X(n)} \theta^{(1)}_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle \tilde{\phi}^{(Y)}_i \| \\
&+ \| \sum_{i=1}^{L_Y(n)} f_i''(\sum_{j=1}^{L_X(n)} \theta^*_{i,j}\langle X, \phi^{(X)}_j\rangle) \left(\sum_{j=1}^{L_X(n)} \theta^{(2)}_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle \sum_{j=1}^{L_X(n)} \theta^{(1)}_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle \tilde{\phi}^{(Y)}_i \right. \\
&\qquad \left. - \sum_{j=1}^{L_X(n)} \theta^{(2)}_{i,j}\langle X, \phi^{(X)}_j\rangle \sum_{j=1}^{L_X(n)} \theta^{(1)}_{i,j}\langle X, \phi^{(X)}_j\rangle \phi^{(Y)}_i \| \right). \qquad \text{(S9.33)}
\end{aligned}
$$

We bound each term separately. For the first term on the r.h.s. of the inequality (S9.32), we have

$$
\begin{aligned}
&\| \sum_{i=1}^{L_Y(n)} \left(f_i''(\sum_{j=1}^{L_X(n)} \theta^*_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle) - f_i''(\sum_{j=1}^{L_X(n)} \theta^*_{i,j}\langle X, \phi^{(X)}_j\rangle) \right) \\
&\cdot \sum_{j=1}^{L_X(n)} \theta^{(2)}_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle \sum_{j=1}^{L_X(n)} \theta^{(1)}_{i,j}\langle X, \tilde{\phi}^{(X)}_j\rangle \tilde{\phi}^{(Y)}_i \| \\
\lesssim &\|X\|^3 \sum_{j=1}^{L_X(n)} \|\tilde{\phi}^{(X)}_j - \phi^{(X)}_j\|.
\end{aligned}
$$

For the second term on the r.h.s. of the inequality (S9.32), we have

$$\begin{aligned}
&\| \sum_{i=1}^{L_Y(n)} f_i''(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \phi_j^{(X)} \rangle) \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \tilde{\phi}_j^{(X)} \rangle \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \tilde{\phi}_j^{(X)} \rangle \tilde{\phi}_i^{(Y)} \right. \\
&\qquad \left. - \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \phi_j^{(X)} \rangle \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \phi_j^{(X)} \rangle \phi_i^{(Y)} \| \right) \\
&\lesssim \sqrt{\sum_{i=1}^{L_Y(n)} \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \tilde{\phi}_j^{(X)} - \phi_j^{(X)} \rangle \right)^2 \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \tilde{\phi}_j^{(X)} \rangle \right)^2} \\
&\lesssim \|X\|^2 \sum_{j=1}^{L_X(n)} \|\tilde{\phi}_j^{(X)} - \phi_j^{(X)}\|,
\end{aligned}$$

and

$$\begin{aligned}
&\| \sum_{i=1}^{L_Y(n)} f_i''(\sum_{j=1}^{L_X(n)} \theta_{i,j}^* \langle X, \phi_j^{(X)} \rangle) \left(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \phi_j^{(X)} \rangle \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \phi_j^{(X)} \rangle \tilde{\phi}_i^{(Y)} \right. \\
&\qquad \left. - \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \phi_j^{(X)} \rangle \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \phi_j^{(X)} \rangle \phi_i^{(Y)} \| \right) \\
&\lesssim \sum_{i=1}^{L_Y(n)} | \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \phi_j^{(X)} \rangle | | \sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \phi_j^{(X)} \rangle | \|\tilde{\phi}_i^{(Y)} - \phi_i^{(Y)}\| \lesssim \|X\|^2 \sum_{i=1}^{L_Y(n)} \|\tilde{\phi}_i^{(Y)} - \phi_i^{(Y)}\|.
\end{aligned}$$

Combining these yields

$$\begin{aligned}
&\|(D_{\boldsymbol{\theta}}^2 \mathcal{O}_{^{(n)}\boldsymbol{\theta}_n^*, \tilde{\eta}_n} - D_{\boldsymbol{\theta}}^2 \mathcal{O}_{^{(n)}\boldsymbol{\theta}_n^*, \eta_0})(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\| \\
&\lesssim (\|X\|^2 + \|X\|^3) \left(\sum_{i=1}^{L_Y(n)} \|\tilde{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\tilde{\phi}_j^{(X)} - \phi_j^{(X)}\| \right).
\end{aligned}$$

It then follows that

$$\|PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} - PD^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}\| = o_p(1)$$

provided that $\frac{\kappa_n}{h_n} = o(1)$.

Recall that

$$\begin{aligned} D^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} =& D_{\boldsymbol{\theta}} \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} \circ D^2_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\|) \circ D_{\boldsymbol{\theta}} \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n} \\ &+ D_{\mathcal{O}} \mathcal{K}_n(\|Y - \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\|) \circ D^2_{\boldsymbol{\theta}} \mathcal{O}_{{}^{(n)}\boldsymbol{\theta}^*_n,\tilde{\eta}_n}. \end{aligned}$$

For the first-order derivative, we have:

$$\begin{aligned} &\|D_{\boldsymbol{\theta}} \mathcal{O}_{\boldsymbol{\theta}_1,\tilde{\eta}_n} - D_{\boldsymbol{\theta}} \mathcal{O}_{\boldsymbol{\theta}_2,\tilde{\eta}_n}\| \\ =&\| \sum_{i=1}^{L_Y(n)} \left(f_i'(\sum_{j=1}^{L_X(n)} \theta^{(1)}_{i,j} \langle X, \tilde{\phi}^{(X)}_j \rangle) - f_i'(\sum_{j=1}^{L_X(n)} \theta^{(2)}_{i,j} \langle X, \tilde{\phi}^{(X)}_j \rangle) \right) \sum_{j=1}^{L_X(n)} \theta_{i,j} \langle X, \tilde{\phi}^{(X)}_j \rangle \tilde{\phi}^{(Y)}_i \| \\ \lesssim& \|X\|^2 \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|. \end{aligned}$$

Similarly, for the second-order derivative, we obtain:

$$\|D^2_{\boldsymbol{\theta}} \mathcal{O}_{\boldsymbol{\theta}_1,\tilde{\eta}_n} - D^2_{\boldsymbol{\theta}} \mathcal{O}_{\boldsymbol{\theta}_2,\tilde{\eta}_n}\| \lesssim \|X\|^3 \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Combining the above inequalities, we conclude that:

$$\|D^2_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_1,\tilde{\eta}_n} - D^2_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_2,\tilde{\eta}_n}\| \lesssim \frac{\|X\|^3}{h_n^2}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|$$

Now, since $\mathbb{E}(\|X\|^3) < \infty$, let $M = \mathbb{E}(\|X\|^3) + c$, where c is a positive number Then

$$\mathbb{P}\left(\frac{1}{n_j}\sum_{i=1}^{n_j}\|X_{i,j}\|^3 > M\right)$$

converges to zero by the law of large numbers. □

Lemma 9. *Suppose Assumptions A_1–A_{11} hold, and let h_n satisfy the following conditions:*

$$h_n^2 > \max\left\{\frac{L(n)^{\mathcal{V}}\,\omega(1)}{\sqrt{N}},\ \sqrt{M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\,\omega(1),\ \frac{L(n)}{\sqrt{n}}\,\omega(1)\right\};$$
$$h_n^2 > \frac{8}{1-2C_4}\left\|PD_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}_n^*,\eta_0} - PD_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\eta_0}\right\|;$$
$$h_n^3 > \kappa_n\,\omega(1).$$

We have

$$\|\tilde{\boldsymbol{\theta}}_n - \hat{\tilde{\boldsymbol{\theta}}}_n\|_2 = O_p\left(\frac{L(n)}{\sqrt{n}h_n}\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right).$$

Proof. Suppose that Assumptions A_1–A_{11} hold. By adapting arguments

from Jordan et al. (2019), under the events $\mathcal{E}_0 \cap \mathcal{E}_1$, we obtain

$$\|\tilde{\boldsymbol{\theta}}_n - \hat{\tilde{\boldsymbol{\theta}}}_n\|_2 \leq \frac{2\|\mathbb{P}_{n_1}(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n}) + \mathbb{P}_N(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n}) - \mathbb{P}_{n_1}(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n})\|}{(1-2C_4)\mu_-}$$

provided that n is sufficiently large so that $\|D^2_{\boldsymbol{\theta}} m_{n,{}^{(n)}\boldsymbol{\theta}^*_n,\eta_0}\| > \mu_-$. We now consider the decomposition:

$$\begin{aligned}
&\mathbb{P}_n(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n}) + \mathbb{P}_N(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n}) - \mathbb{P}_n(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n}) \\
=& (\mathbb{P}_n - P)\left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right) - (\mathbb{P}_n - P)\left(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right) \\
&- (\mathbb{P}_N - P)\left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right) + (\mathbb{P}_N - P)\left(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right).
\end{aligned}$$

Note that

$$\begin{aligned}
&\|\mathbb{P}_{n_1}(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n}) + \mathbb{P}_N(D m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n}) - \mathbb{P}_{n_1}(D m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n})\| \\
=& \sup_{\boldsymbol{\theta}\in\Theta_n} |\left(\mathbb{P}_{n_1}(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n}) + \mathbb{P}_N(D m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n}) - \mathbb{P}_{n_1}(D m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n})\right)(\boldsymbol{\theta})|.
\end{aligned}$$

Consequently, we define the following norm expression for analysis:

$$\begin{aligned}
&\|\left(\mathbb{P}_n - P\right)\left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right) - \left(\mathbb{P}_n - P\right)\left(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right) \\
&- \left(\mathbb{P}_N - P\right)\left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right) + \left(\mathbb{P}_N - P\right)\left(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}^*_n,\tilde{\eta}_n}\right)\|
\end{aligned} \tag{S9.34}$$

Let

$$\mathcal{F} := \left\{ \left(D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta},\tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n} \right)(\boldsymbol{\beta}) : \boldsymbol{\theta} \in \Theta_n,\ \boldsymbol{\beta} \in \Theta_n,\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_n^*\| < \delta_n,\ \|\boldsymbol{\beta}\| \leq 1 \right\}.$$

For all $f \in \mathcal{F}$, we have the uniform envelope bound $\|f\| \leq \|D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta},\tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\| \leq \frac{\delta_n}{h_n}\|X\|^2$. Moreover, for any f_1 and $f_2 \in \mathcal{F}$ corresponding to $(\boldsymbol{\theta}_1, \boldsymbol{\beta}_1)$ and $(\boldsymbol{\theta}_2, \boldsymbol{\beta}_2)$ respectively, we obtain:

$$\begin{aligned}
|f_1 - f_2| &= |D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_1,\tilde{\eta}_n}(\boldsymbol{\beta}_1) - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\beta}_1) - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_2,\tilde{\eta}_n}(\boldsymbol{\beta}_2) + D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}(\boldsymbol{\beta}_2)| \\
&\leq \|D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_1,\tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_2,\tilde{\eta}_n}\| + |\left(D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_2,\tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n} \right)(\boldsymbol{\beta}_1) \\
&\quad - \left(D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_2,\tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n} \right)(\boldsymbol{\beta}_2)| \\
&\leq \frac{1}{h_n}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|\|X\|_2 + \|D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_2,\tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\| \\
&\leq \frac{1}{h_n}\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|\|X\|^2 + \frac{\delta_n}{h_n}\|\boldsymbol{\beta}_1 - \boldsymbol{\beta}_2\|\|X\|^2.
\end{aligned}$$

Hence, the bracketing number of $\mathcal{F}$ satisfies

$$\begin{aligned}
&N_{[]}(2\epsilon(\mathbb{E}\|X\|^4)^{\frac{1}{2}}, \mathcal{F}, L_2(P)) \\
\leq &N(\frac{\epsilon h_n}{2}, \{\boldsymbol{\theta} \in \Theta_n : \|\boldsymbol{\theta} - \boldsymbol{\theta}_n^*\| < \delta_n\}, L_2(P)) \cdot N(\frac{h_n \epsilon}{2\delta_n}, \Theta_n, L_2(P)) \\
\lesssim &\left(\frac{\delta_n}{h_n \epsilon} \right)^{2L_X(n)L_Y(n)}
\end{aligned}$$

Therefore,

$$E\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \int_0^{\frac{\delta_n}{h_n}\|\|X\|^2\|_{P,2}} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))}d\epsilon$$
$$\leq \int_0^{\frac{\delta_n}{h_n}} \sqrt{\log\left(\frac{C_1\delta_n}{h_n\epsilon}\right)^{2L_X(n)L_Y(n)}}d\epsilon \lesssim \frac{\delta_n L(n)}{h_n}$$

For example, consider the first term of (S9.34), Let

$$\delta_{M,n} := 2^M\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right)$$

and $\|\left(\mathbb{P}_n - P\right)\left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - Dm_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right)\|$.

$$\begin{aligned}
&\mathbb{P}\left(\|\left(\mathbb{P}_n - P\right)\left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - Dm_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right)\| > 4^M\gamma_n\right)\\
=&\mathbb{P}\left(\|\left(\mathbb{P}_n - P\right)\left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - Dm_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right)\| > 4^M\gamma_n,\right.\\
&\left.\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\| < 2^M\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right)\right)\\
&+\mathbb{P}\left(\|\left(\mathbb{P}_n - P\right)D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - Dm_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\| > 4^M\gamma_n,\right.\\
&\left.\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\| > 2^M\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right)\right)\\
\leq&\mathbb{P}\left(\sup_{\boldsymbol{\theta}\in\Theta_n:\|\boldsymbol{\theta}-\boldsymbol{\theta}_n^*\|<\delta_{M,n};\boldsymbol{\beta}\in\Theta_n}|\left(\mathbb{P}_n - P\right)D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta},\tilde{\eta}_n} - Dm_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}| > 4^M\gamma_n\right)\\
&+\mathbb{P}\left(\|\hat{\tilde{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}_n^*\| > 2^M\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right)\right)\\
\lesssim&\frac{E\|\mathbb{G}_n\|_{\mathcal{F}_1}}{4^M\sqrt{n}\gamma_n} + \frac{1}{2^M}.
\end{aligned}$$

Let

$$\mathcal{F} := \Big\{ \left(D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta},\tilde{\eta}_n} - D_{\boldsymbol{\theta}} m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right)(\boldsymbol{\beta}) : \boldsymbol{\theta} \in \Theta_n,\ \boldsymbol{\beta} \in \Theta_n, \\ \|\boldsymbol{\theta} - \boldsymbol{\theta}_n^*\| < \sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)},\ \|\boldsymbol{\beta}\| \leq 1 \Big\}.$$

Hence, we choose γ_n satisfying

$$\gamma_n > \frac{L(n)}{\sqrt{n}h_n} \sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}$$

Therefore, we conclude

$$\begin{aligned} &\| \left(\mathbb{P}_n - P\right) \left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right) \| \\ =& O_p\left(\frac{L(n)}{\sqrt{n}h_n} \sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)} \right). \end{aligned}$$

Similarly, we obtain the following bounds:

$$\begin{aligned} &\| \left(\mathbb{P}_n - P\right) \left(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right) \| \\ =& O_p\left(\frac{L(n)}{\sqrt{n}h_n} \sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)} \right) \\ &\| \left(\mathbb{P}_N - P\right) \left(D_{\boldsymbol{\theta}} m_{n,\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right) \| \\ =& O_p\left(\frac{L(n)}{\sqrt{N}h_n} \sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)} \right) \end{aligned}$$

$$
\begin{aligned}
&\| \left(\mathbb{P}_N - P\right) \left(D_{\boldsymbol{\theta}} m_{n,\bar{\boldsymbol{\theta}},\tilde{\eta}_n} - D m_{n,\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\right) \| \\
=&O_p\left(\frac{L(n)}{\sqrt{N}h_n}\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right)
\end{aligned}
$$

Combining these results, we obtain the final bound:

$$
\begin{aligned}
&\|\tilde{\boldsymbol{\theta}}_n - \hat{\tilde{\boldsymbol{\theta}}}_n\| \\
=&O_p\left(\frac{L(n)}{\sqrt{n}h_n}\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right)
\end{aligned}
$$

□

S10 Proofs of the Main Theorems

For technical convenience, Section S9 reformulates the original estimation problem by negating the loss function, thereby converting the minimization problem in the main text into an equivalent maximization problem. In the subsequent proofs, we leverage these results in Section S9 and restate the conclusions in the original loss-minimization notation used throughout the main text.

S10.1 Proof of Theorem 1

Proof. By Lemma 2, it suffices to verify the conditions required therein. First, observe that by the continuity of the population loss function $P\left[\rho_\tau\left(Y-\sum_{i=1}^{\infty} f_i(X;\boldsymbol{\theta}_i^*)\phi_i^{(Y)}\right)\right]$, we have the approximation error $|M({}^{(n)}\boldsymbol{\theta}^*)-M(\boldsymbol{\theta}^*)| = o(1)$. Next, we verify the Lipschitz condition in Lemma 1. Note that

$$
\begin{aligned}
|m_{n,\boldsymbol{\theta}_1,\eta} - m_{\boldsymbol{\theta}_2,\eta}| &= |\|Y - {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_1,\eta}(X)\| - \|Y - {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_2,\eta}(X)\| \\
&\quad + < \tau, {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_1,\eta}(X) - {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_2,\eta}(X) >| \\
&\le || {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_1,\eta}(X) - {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_2,\eta}(X)|| + | < \tau, {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_1,\eta}(X) - {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_2,\eta}(X) > | \\
&\le 2|| {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_1,\eta}(X) - {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_2,\eta}(X)||
\end{aligned}
$$

Now, consider the following bound:

$$
\begin{aligned}
&|| {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_1,\eta}(X) - {}^{(n)}\mathcal{O}_{\boldsymbol{\theta}_2,\eta}(X)|| \\
=&\| \sum_{i=1}^{L_Y(n)} \left(f_i(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \phi_j^{(X)} \rangle) - f_i(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \phi_j^{(X)} \rangle) \right) \phi_i^{(Y)} \| \\
=&\sqrt{\sum_{i=1}^{L_Y(n)} \left(f_i(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(1)} \langle X, \phi_j^{(X)} \rangle) - f_i(\sum_{j=1}^{L_X(n)} \theta_{i,j}^{(2)} \langle X, \phi_j^{(X)} \rangle) \right)^2} \\
\lesssim&\sqrt{\sum_{i=1}^{L_Y(n)} \left(\sum_{j=1}^{L_X(n)} (\theta_{i,j}^{(1)} - \theta_{i,j}^{(2)}) \langle X, \phi_j^{(X)} \rangle \right)^2}
\end{aligned}
$$

$$\leq \sqrt{\sum_{i=1}^{L_Y(n)} \sum_{j=1}^{L_X(n)} (\theta_{i,j}^{(1)} - \theta_{i,j}^{(2)})^2 \sum_{j=1}^{L_X(n)} \langle X, \phi_j^{(X)} \rangle^2}$$

$$\leq \|X\| \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2,$$

Combining the inequalities above, we conclude that

$$|m_{n,\boldsymbol{\theta}_1,\eta} - m_{\boldsymbol{\theta}_2,\eta}| \lesssim \|X\| \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|_2,$$

where $E\|X\|^2 < \infty$ by Assumption A_7. Applying Lemma 2 yields the desired result. □

S10.2 Proof of Theorem 2

Proof. In Theorem 2, the loss function is defined as

$$m_{n,\boldsymbol{\theta},\eta} = \|Y - \mathcal{O}_{\boldsymbol{\theta},\eta}\| + < \tau, Y - \mathcal{O}_{\boldsymbol{\theta},\eta} >,$$

and satisfies the Lipschitz-type condition:

$$|m_{n,\boldsymbol{\theta}_1,\eta_1} - m_{n,\boldsymbol{\theta}_2,\eta_2}| \leq 2\|\mathcal{O}_{\boldsymbol{\theta}_1,\eta_1} - \mathcal{O}_{\boldsymbol{\theta}_2,\eta_2}\|.$$

Under Assumptions A_1–A_9, Lemma 3 implies the consistency of the estimator, i.e., $\|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 = o_p(1)$.

Let $\eta_t = \eta_0 + t(\hat{\eta}_n - \eta_0)$, and write $\eta_t = \{\phi_{t,j}^{(X)}, \phi_{t,i}^{(Y)}\}_{i,j=1}^{\infty}$, where $\phi_{t,j}^{(X)} = \phi_j^{(X)} + t(\hat{\phi}_j^{(X)} - \phi_j^{(X)})$ and $\phi_{t,i}^{(Y)} = \phi_i^{(Y)} + t(\hat{\phi}_i^{(Y)} - \phi_i^{(Y)})$. By the definition of $\hat{\eta}_n$, we have $\hat{\phi}_j^{(X)} = \phi_j^{(X)}$ and $\hat{\phi}_i^{(Y)} = \phi_i^{(Y)}$ for $j > L_X(n)$ and $i > L_Y(n)$, respectively. For any direction $h = \{h_j^{(X)}, h_i^{(Y)}\}_{i,j=1}^{\infty}$, To use the Lemma 4, we proceed to compute the Fréchet derivative $D_\eta m_{n,\boldsymbol{\theta},\eta}$:

$$D_\eta m_{n,\boldsymbol{\theta},\eta_t}(h) = \left(\frac{\mathcal{O}_{\boldsymbol{\theta},\eta_t} - Y}{\|\mathcal{O}_{\boldsymbol{\theta},\eta_t} - Y\|} - \tau \right) \circ D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} \circ h.$$

The Fréchet derivative $D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} \circ h$ is computed as follows:

$$\begin{aligned}
D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} \circ h &= \lim_{s\to 0} \frac{\mathcal{O}_{\boldsymbol{\theta},\eta_t + sh} - \mathcal{O}_{\boldsymbol{\theta},\eta_t}}{s} \\
&= \lim_{s\to 0} \frac{1}{s} \left(\sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} + sh_j^{(X)} >)(\phi_{t,i}^{(Y)} + sh_i^{(Y)}) \right. \\
&\qquad \left. - \sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} >)\phi_{t,i}^{(Y)} \right) \\
&= \lim_{s\to 0} \frac{1}{s} \left[\sum_{i=1}^{\infty} \left(f_i(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} + sh_j^{(X)} >) \right. \right. \\
&\quad \left. - f_i(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} >)(\phi_{t,i}^{(Y)} + sh_i^{(Y)}) \right) \\
&\qquad \left. + s \sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} >) h_i^{(Y)} \right] \\
&= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \theta_{i,j} < X, h_j^{(X)} > f_i'(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} >)\phi_{t,i}^{(Y)}
\end{aligned}$$

$$+\sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty}\theta_{i,j} < X, \phi_{t,j}^{(X)} >)h_i^{(Y)}.$$

Then we have

$$\begin{aligned}
&\|D_\eta m_{n,\boldsymbol{\theta},\eta_t} - D_\eta m_{n,\boldsymbol{\theta}^*,\eta_t}\| \\
\leq &\| \left(\frac{\mathcal{O}_{\boldsymbol{\theta},\eta_t} - Y}{\|\mathcal{O}_{\boldsymbol{\theta},\eta_t} - Y\|} - \frac{\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y}{\|\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y\|} \right) \circ D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} \| \\
+ &\| \left(\frac{\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y}{\|\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y\|} - \tau \right) \circ (D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} - D_\eta \mathcal{O}_{\boldsymbol{\theta}^*,\eta_t}) \|
\end{aligned}$$

The quantity $\|D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} - D_\eta \mathcal{O}_{\boldsymbol{\theta}^*,\eta_t}\|$ can be controlled via the following inequality:

$$\begin{aligned}
&\| \frac{\mathcal{O}_{\boldsymbol{\theta},\eta_t} - Y}{\|\mathcal{O}_{\boldsymbol{\theta},\eta_t} - Y\|} - \frac{\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y}{\|\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y\|} \| \\
\leq &\| \int_0^1 \frac{1}{\|\mathcal{O}_{\boldsymbol{\theta}^*+s(\boldsymbol{\theta}-\boldsymbol{\theta}^*),\eta_t} - Y\|} \circ (\mathcal{O}_{\boldsymbol{\theta},\eta_t} - \mathcal{O}_{\boldsymbol{\theta}^*,\eta_t}) ds \| \\
\lesssim &\int_0^1 \frac{1}{\|\mathcal{O}_{\boldsymbol{\theta}^*+s(\boldsymbol{\theta}-\boldsymbol{\theta}^*),\eta_t} - Y\|} ds \|X\| \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|;
\end{aligned}$$

$$\begin{aligned}
&\|D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} \circ (\hat{\eta}_n - \eta_0)\| \\
&= \| \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \theta_{i,j} < X, \hat{\phi}_j^{(X)} - \phi_j^{(X)} > f_i'(\sum_{j=1}^{\infty}\theta_{i,j} < X, \phi_{t,j}^{(X)} >)\phi_{t,i}^{(Y)} \\
&\quad + \sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty}\theta_{i,j} < X, \phi_{t,j}^{(X)} >)(\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}) \|
\end{aligned}$$

$$\lesssim \|X\| \left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \right),$$

where the inequality follows from the uniform boundedness of the derivatives $f_i'(\cdot)$. This bound is structurally analogous to the argument used in Lemma 3. It is straightforward to observe that

$$\|\frac{\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y}{\|\mathcal{O}_{\boldsymbol{\theta}^*,\eta_t} - Y\|} - \tau\| \leq 2.$$

And

$$\begin{aligned}
&\| \left(D_\eta \mathcal{O}_{\boldsymbol{\theta},\eta_t} - D_\eta \mathcal{O}_{\boldsymbol{\theta}^*,\eta_t}\right) \circ (\hat{\eta}_n - \eta_0)\| \\
=&\| \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \theta_{i,j} < X, \hat{\phi}_j^{(X)} - \phi_j^{(X)} > f_i'(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} >)\phi_{t,i}^{(Y)} \\
&+ \sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j} < X, \phi_{t,j}^{(X)} >)(\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}) \\
&- \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \theta_{i,j}^* < X, \hat{\phi}_j^{(X)} - \phi_j^{(X)} > f_i'(\sum_{j=1}^{\infty} \theta_{i,j}^* < X, \phi_{t,j}^{(X)} >)\phi_{t,i}^{(Y)} \\
&+ \sum_{i=1}^{\infty} f_i(\sum_{j=1}^{\infty} \theta_{i,j}^* < X, \phi_{t,j}^{(X)} >)(\hat{\phi}_i^{(Y)} - \phi_i^{(Y)})\|. \qquad \text{(S10.35)}
\end{aligned}$$

This expression can be decomposed into three terms. The first term on the

r.h.s. of (S10.35) is bounded as follows:

$$
\begin{aligned}
&\|\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(\theta_{i,j}-\theta_{i,j}^*)<X,\hat{\phi}_j^{(X)}-\phi_j^{(X)}>f_i'(\sum_{j=1}^{\infty}\theta_{i,j}<X,\phi_{t,j}^{(X)}>)\phi_{t,i}^{(Y)}\| \\
\lesssim&\sqrt{\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty}(\theta_{i,j}-\theta_{i,j}^*)<X,\hat{\phi}_j^{(X)}-\phi_j^{(X)}>\right)^2} \\
\leq&\|X\|\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|_2\left(\sum_{j=1}^{L_X(n)}\|\hat{\phi}_j^{(X)}-\phi_j^{(X)}\|\right).
\end{aligned}
$$

The second term on the r.h.s. of Equation (S10.35) satisfies

$$
\begin{aligned}
&\|\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\theta_{i,j}^*<X,\hat{\phi}_j^{(X)}-\phi_j^{(X)}>\left(f_i'(\sum_{j=1}^{\infty}\theta_{i,j}<X,\phi_{t,j}^{(X)}>)-f_i'(\sum_{j=1}^{\infty}\theta_{i,j}^*<X,\phi_{t,j}^{(X)}>)\right)\phi_{t,i}^{(Y)}\| \\
\lesssim&\sqrt{\sum_{i=1}^{\infty}\left(f_i'(\sum_{j=1}^{\infty}\theta_{i,j}<X,\phi_{t,j}^{(X)}>)-f_i'(\sum_{j=1}^{\infty}\theta_{i,j}^*<X,\phi_{t,j}^{(X)}>)\right)^2\left(\sum_{j=1}^{\infty}\theta_{i,j}^*<X,\hat{\phi}_j^{(X)}-\phi_j^{(X)}>\right)^2} \\
\lesssim&\sqrt{\sum_{i=1}^{\infty}\left(\sum_{j=1}^{\infty}(\theta_{i,j}-\theta_{i,j}^*)<X,\phi_{t,j}^{(X)}>\right)^2\left(\sum_{j=1}^{\infty}\theta_{i,j}^*<X,\hat{\phi}_j^{(X)}-\phi_j^{(X)}>\right)^2} \\
\lesssim&\|X\|^2\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\left(\sum_{j=1}^{L_X(n)}\|\hat{\phi}_j^{(X)}-\phi_j^{(X)}\|\right).
\end{aligned}
$$

The third term on the r.h.s. of Equation (S10.35) can be bounded by

$$
\begin{aligned}
&\|\sum_{i=1}^{\infty}\left(f_i(\sum_{j=1}^{\infty}\theta_{i,j}<X,\phi_{t,j}^{(X)}>)-f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*<X,\phi_{t,j}^{(X)}>)\right)(\hat{\phi}_i^{(Y)}-\phi_i^{(Y)})\| \\
\leq&\sum_{i=1}^{L_Y(n)}|\sum_{j=1}^{\infty}(\theta_{i,j}-\theta_{i,j}^*)<X,\phi_{t,j}^{(X)}>|\|\hat{\phi}_i^{(Y)}-\phi_i^{(Y)}\|
\end{aligned}
$$

$$\lesssim \|X\| \sum_{i=1}^{L_Y(n)} \sqrt{\sum_{j=1}^{\infty} (\theta_{i,j} - \theta_{i,j}^*)^2} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\|$$
$$\leq \|X\| \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| \right).$$

Combining the three components above, it follows that

$$\begin{aligned} &\| \left(D_\eta m_{n,\boldsymbol{\theta},\eta_0+t(\hat{\eta}_n-\eta_0)} - D_\eta m_{n,\boldsymbol{\theta}^*,\eta_0+t(\hat{\eta}_n-\eta_0)} \right) \circ (\hat{\eta}_n - \eta_0) \| \\ \lesssim & F(X)_{\boldsymbol{\theta},\eta} \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \left(\sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\| \right), \end{aligned}$$

Therefore, by Assumptions A_3 and A_7, it holds that $\mathbb{E}(|F(X)_{\boldsymbol{\theta},\eta}|) < \infty$ in a small neighborhood of $\boldsymbol{\theta}^*$ and η_0. Then Lemma 4 is applied to complete the proof. □

S10.3 Proof of Theorem 4

Proof. Recall that $\tilde{\mathcal{O}}_n = \mathcal{O}_{\tilde{\boldsymbol{\theta}}_n,\tilde{\eta}_n}$ and $\hat{\tilde{\mathcal{O}}} = \mathcal{O}_{\hat{\tilde{\boldsymbol{\theta}}}_n,\tilde{\eta}_n}$. We begin by decomposing the total estimation error into four parts:

$$\begin{aligned} &\|\tilde{\mathcal{O}}_n - \mathcal{O}^*\|_{\mathcal{X}} \\ \leq & \|\tilde{\mathcal{O}}_n - \hat{\tilde{\mathcal{O}}}\|_{\mathcal{X}} + \|\hat{\tilde{\mathcal{O}}} - \mathcal{O}_{\boldsymbol{\theta}_n^*,\tilde{\eta}_n}\|_{\mathcal{X}} + \|\mathcal{O}_{\boldsymbol{\theta}_n^*,\tilde{\eta}_n} - \mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0}\|_{\mathcal{X}} + \|\mathcal{O}_{\boldsymbol{\theta}_n^*,\eta_0} - \mathcal{O}^*\|_{\mathcal{X}} \end{aligned}$$

By applying Lemmas 6, 7, and 9, we obtain:

$$\begin{aligned}
&\|\tilde{\mathcal{O}}_n - \mathcal{O}^*\|_{\mathcal{X}} \\
=&O_p\left(\frac{L(n)}{\sqrt{n}h_n}\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)}\right. \\
&\left. + \sqrt{\frac{L(n)^{2\mathcal{V}}}{N} + \kappa_n^2 + M_n({}^{(n)}\boldsymbol{\theta}_n^*) - M_n(\boldsymbol{\theta}_n^*)} + h_n^p\right).
\end{aligned}$$

□

S10.4 Proof of Theorem 3

Proof. We begin by establishing the estimation error bound for the first eigenfunction ϕ_1. Because eigenfunctions are only determined up to sign, we choose the sign of ϕ_1^* such that $\langle \phi_1^*, \phi_1 \rangle \geq 0$, and then choose the sign of $\widetilde{\phi}_1$ such that $\langle \widetilde{\phi}_1, \phi_1^* \rangle \geq 0$. Define $\tilde{C}_1 = \frac{1}{K}\sum_{k=1}^K \hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)}$, and $C_1^* = \mathbb{E}\left(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)}\right)$. Let $\tilde{\phi}_1$ and ϕ_1^* be the leading eigenfunctions of $\tilde{C}_1$, and C_1^*, respectively. Then we have

$$\|\tilde{\phi}_1 - \phi_1^*\| \leq 2\sqrt{2}\left(\lambda_{1,1}^* - \lambda_{1,2}^*\right)^{-1}\|\tilde{C}_1 - C_1^*\|,$$

where

$$\|\tilde{C}_1 - C_1^*\| = \|\frac{1}{K}\sum_{k=1}^K \left(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)})\right)\|$$

To analyze the convergence of $\|\tilde{C}_1 - C_1^*\|$, we first bound each term $\|\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)})\|$:

$$\begin{aligned}
&\|\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}\left(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)}\right)\| \leq 2\|\hat{\phi}_1^{(k)} - \phi_1\| + 2\mathbb{E}\left(\|\hat{\phi}_1^{(k)} - \phi_1\|\right) \\
\leq & 4\sqrt{2}(\lambda_1 - \lambda_2)^{-1}\left(\|\hat{C} - C\| + \mathbb{E}(\|\hat{C} - C\|)\right),
\end{aligned}$$

where $\hat{C} = \frac{1}{n}\sum_{i=1}^n X_i \otimes X_i$, $C = \mathbb{E}\,(X \otimes X)$. And λ_1, λ_2 denote the first two eigenvalues of C.

Next, observe that

$$\|\hat{C} - C\| = \sup_{\substack{x,y \in \mathcal{H}_X \\ \|x\|=1,\ \|y\|=1}} |\frac{1}{n}\sum_{i=1}^n \left(\langle X_i, x\rangle\langle X_i, y\rangle - \mathbb{E}\langle X_i, x\rangle\langle X_i, y\rangle\right)\|$$

Define the function class: $\mathcal{F} := \{f_{x,y}(\cdot) = \langle \cdot, x\rangle\langle \cdot, y\rangle : x, y \in \mathcal{H}_X; \|x\| = 1,\ \|y\| = 1\}$. And $|f_{x,y}(\cdot)| \leq \|X\|^2$. Thus,

$$\sqrt{\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}}^2} \lesssim \mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} + \sqrt{\mathbb{E}\|X\|^4}.$$

Additionally,

$$\mathbb{E}\|\mathbb{G}_n\|_{\mathcal{F}} \lesssim \int_0^{\sqrt{\mathbb{E}\|X\|^4}} \sqrt{\log N_{[]}\left\{\epsilon, \mathcal{F}, L_2(P)\right\}} d\epsilon$$

For $f_{x_1,y_1}, f_{x_2,y_2} \in \mathcal{F}$, we have

$$|f_{x_1,y_1} - f_{x_2,y_2}| \leq \|X\|^2 (\|x_1 - x_2\| + \|y_1 - y_2\|)$$

Hence

$$\int_0^{\sqrt{\mathbb{E}\|X\|^4}} \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon \lesssim \int_0^1 \sqrt{\log N(\epsilon, \mathcal{H}_X, L_2(P))} d\epsilon \lesssim \frac{2}{2-v}$$

Therefore, we have the following bound:

$$\sqrt{\mathbb{E}\|\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}\left(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)}\right)\|^2} \lesssim \frac{1}{\lambda_1 - \lambda_2} \frac{1}{\sqrt{n}}.$$

Moreover, it is clear that

$$\begin{aligned}
&\|\frac{1}{K}\sum_{k=1}^{K}\left(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)})\right)\| \\
= &\sup_{\substack{x,y \in \mathcal{H}_X \\ \|x\|=1,\ \|y\|=1}} |\frac{1}{K}\sum_{k=1}^{K}\left(\langle\hat{\phi}_1^{(k)}, x\rangle\langle\hat{\phi}_1^{(k)}, y\rangle - \mathbb{E}(\langle\hat{\phi}_1^{(k)}, x\rangle\langle\hat{\phi}_1^{(k)}, y\rangle)\right)|,
\end{aligned}$$

and

$$\begin{aligned}
&|\langle\hat{\phi}_1^{(k)}, x\rangle\langle\hat{\phi}_1^{(k)}, y\rangle - \mathbb{E}(\langle\hat{\phi}_1^{(k)}, x\rangle\langle\hat{\phi}_1^{(k)}, y\rangle)| \\
\leq &\sup_{\substack{x,y \in \mathcal{H}_X \\ \|x\|=1,\ \|y\|=1}} |\langle\hat{\phi}_1^{(k)}, x\rangle\langle\hat{\phi}_1^{(k)}, y\rangle - \mathbb{E}(\langle\hat{\phi}_1^{(k)}, x\rangle\langle\hat{\phi}_1^{(k)}, y\rangle)|
\end{aligned}$$

$$=\|\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)})\|.$$

Define $\mathcal{G} := \{g_{x,y}(\cdot) = \langle \cdot, x\rangle\langle \cdot, y\rangle - \mathbb{E}\left(\langle \cdot, x\rangle\langle \cdot, y\rangle\right) : x, y \in \mathcal{H}_X; \|x\| = 1,\ \|y\| = 1\}$. Then, it is straightforward to see that $g_{x,y} \leq \|\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)})\|$ and

$$\sqrt{\mathbb{E}\|\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}\left(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)}\right)\|^2} \lesssim \frac{1}{\lambda_1 - \lambda_2}\frac{1}{\sqrt{n}}.$$

And thus we have the bound

$$|g_{x_1,y_1} - g_{x_2,y_2}| \leq \|\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)} - \mathbb{E}(\hat{\phi}_1^{(k)} \otimes \hat{\phi}_1^{(k)})\| \left(\|x_1 - x_2\| + \|y_1 - y_2\|\right).$$

Consequently, similar arguments lead us to:

$$\mathbb{E}\|\tilde{C}_1 - C_1^*\| \lesssim \frac{1}{\lambda_1 - \lambda_2}\frac{1}{\sqrt{N}},$$

which implies that

$$\mathbb{E}\|\tilde{\phi}_1 - \phi_1^*\| \lesssim \frac{1}{(\lambda_{1,1}^* - \lambda_{1,2}^*)(\lambda_1 - \lambda_2)}\frac{1}{\sqrt{N}}.$$

We further obtain

$$\|\phi_1^* - \phi_1\| \leq 2\sqrt{2}\|\mathbb{E}(\hat{\phi}_1 \otimes \hat{\phi}_1) - \phi_1 \otimes \phi_1\| \lesssim \frac{1}{(\lambda_1 - \lambda_2)^2}\mathbb{E}\|\hat{C} - C\|^2,$$

where the second inequality follows by applying the second-order bias bound in the proof of Theorem 3 in Koltchinskii and Lounici (2017) to $\|\mathbb{E}(\hat{\phi}_1 \otimes \hat{\phi}_1) - \phi_1 \otimes \phi_1\|$. Using the known result $\mathbb{E}\|\hat{C} - C\|^2 \lesssim \frac{1}{n}$ and combining the results above, we thus derive the overall bound:

$$\mathbb{E}\|\tilde{\phi}_1 - \phi_1\| \lesssim \frac{1}{(\lambda_{1,1}^* - \lambda_{1,2}^*)(\lambda_1 - \lambda_2)\sqrt{N}} + \frac{1}{(\lambda_1 - \lambda_2)^2}\frac{1}{n}.$$

For $i \geq 2$, define $\widetilde{C}_i$ and C_i^* in the same way as $\widetilde{C}_1$ and C_1^*. Let $V^{(i)}$ and ϕ_i^* be their leading eigenfunctions, respectively. We choose the sign of ϕ_i^* such that $\langle \phi_i^*, \phi_i \rangle \geq 0$, and then choose the sign of $V^{(i)}$ such that $\langle V^{(i)}, \phi_i^* \rangle \geq 0$. If $\left(Id - \sum_{j=1}^{i-1} \widetilde{\phi}_j \otimes \widetilde{\phi}_j\right) V^{(i)} = 0$, $\widetilde{\phi}_i$ is chosen to be a unit element in $\text{span}\{\widetilde{\phi}_1, \ldots, \widetilde{\phi}_{i-1}\}^\perp$. Thus, we have the following decomposition for $\widetilde{\phi}_i$:

$$\begin{aligned}\|\tilde{\phi}_i - \phi_i\| \leq & 2\|Id - \sum_{j=1}^{i-1} \tilde{\phi}_j \otimes \tilde{\phi}_j\|\|V^{(i)} - \phi_i^*\| + 2\|\sum_{j=1}^{i-1}\left(\tilde{\phi}_j \otimes \tilde{\phi}_j - \phi_j \otimes \phi_j\right)\| \\ & + 2\|Id - \sum_{j=1}^{i-1} \phi_j \otimes \phi_j\|\|\phi_i^* - \phi_i\|.\end{aligned}$$

Through an analogous analysis, the following bounds are obtained:

$$\mathbb{E}\|V^{(i)} - \phi_i^*\| \lesssim \frac{1}{(\lambda_{i,1}^* - \lambda_{i,2}^*)\min\{\lambda_{i-1} - \lambda_i,\, \lambda_i - \lambda_{i+1}\}} \frac{1}{\sqrt{N}},$$

$$\|\phi_i^* - \phi_i\| \lesssim \frac{1}{\left[\min\{\lambda_{i-1} - \lambda_i,\, \lambda_i - \lambda_{i+1}\}\right]^2} \frac{1}{n}.$$

For notational convenience, define $\Delta_i := \min\{\lambda_{i-1} - \lambda_i,\, \lambda_i - \lambda_{i+1}\}$ for $2 \leq i \leq L$. Hence, combining these results, we conclude that:

$$\mathbb{E}\|\tilde{\phi}_i - \phi_i\| \lesssim \frac{1}{(\lambda_{i,1}^* - \lambda_{i,2}^*)\Delta_i} \frac{1}{\sqrt{N}} + \frac{1}{\Delta_i^2} \frac{1}{n} + \sum_{j=1}^{i-1} \mathbb{E}\|\tilde{\phi}_j - \phi_j\|.$$

□

S11 Proof of Corollaries

S11.1 Proof of Corollary 1

Proof. According to the argument used in the proof of Theorem 1, for any $x \in \mathcal{H}_X$,

$$\|\hat{\mathcal{O}}_n(x) - \mathcal{O}^*(x)\| \lesssim \|x\| \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2.$$

Taking the supremum over all $x \in \mathcal{H}_X$ satisfying $\|x\| \leq 1$, we obtain

$$\|\hat{\mathcal{O}}_n(\cdot) - \mathcal{O}^*(\cdot)\|_{\mathcal{X}} \lesssim \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*\|_2.$$

The result then follows directly from Theorem 1. □

S11.2 Proof of Corollary 2

Proof. We begin by decomposing the estimation error into two terms:

$$\|\hat{\hat{\mathcal{O}}}_n(\cdot) - \mathcal{O}^*(\cdot)\|_{\mathcal{X}} \lesssim \|\hat{\hat{\mathcal{O}}}_n(\cdot) - \mathcal{O}_{\boldsymbol{\theta}^*,\hat{\eta}_n}(\cdot)\|_{\mathcal{X}} + \|\mathcal{O}_{\boldsymbol{\theta}^*,\hat{\eta}_n}(\cdot) - \mathcal{O}^*(\cdot)\|_{\mathcal{X}}.$$

Applying Theorem 2, the first term on the right-hand side satisfies:

$$\|\hat{\hat{\mathcal{O}}}_n(\cdot) - \mathcal{O}_{\boldsymbol{\theta}^*,\hat{\eta}_n}(\cdot)\|_{\mathcal{X}} \lesssim \|\hat{\hat{\boldsymbol{\theta}}}_n - \boldsymbol{\theta}^*\|_2 = O_p\left(\sqrt{\frac{L(n)^{2\mathcal{V}}}{n} + \alpha_n^2 + M({}^{(n)}\boldsymbol{\theta}^*) - M(\boldsymbol{\theta}^*)}\right).$$

Next, we focus on bounding the second term. Denote by $\{\bar{\phi}_j^{(X)}, \bar{\phi}_i^{(Y)}\}$ the basis elements in $\hat{\eta}_n$. Then, for any $x \in \mathcal{H}_X$ satisfying $\|x\| \leq 1$, we have

$$\begin{aligned}
\|\mathcal{O}_{\boldsymbol{\theta}^*,\hat{\eta}_n}(x) - \mathcal{O}^*(x)\| \lesssim & \sqrt{\sum_{i=1}^{\infty}\left[f_i\left(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle x, \bar{\phi}_j^{(X)}\rangle\right) - f_i\left(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle x, \phi_j^{(X)}\rangle\right)\right]^2} \\
& + \sum_{i=1}^{L_Y(n)} |f_i(\sum_{j=1}^{\infty}\theta_{i,j}^*\langle x, \phi_j^{(X)}\rangle)| \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| \\
\lesssim & \sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\|.
\end{aligned}$$

Taking the supremum over all $x \in \mathcal{H}_X$ satisfying $\|x\| \leq 1$ yields

$$\|\mathcal{O}_{\boldsymbol{\theta}^*,\hat{\eta}_n}(\cdot) - \mathcal{O}^*(\cdot)\|_{\mathcal{X}} \lesssim \sum_{i=1}^{L_Y(n)} \|\hat{\phi}_i^{(Y)} - \phi_i^{(Y)}\| + \sum_{j=1}^{L_X(n)} \|\hat{\phi}_j^{(X)} - \phi_j^{(X)}\|.$$

Combining the bounds above proves both displays in Corollary 2. □

Bibliography

Beyaztas, U., H. L. Shang, and S. Saricam (2025). Penalized function-on-function linear quantile regression. *Computational Statistics 40*, 1–29.

Boyd, S. P. and L. Vandenberghe (2004). *Convex optimization*. Cambridge University Press.

Chakraborty, A. and P. Chaudhuri (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *Annals of Statistics 42*, 1203–1231.

Chaouch, M. and C. Goga (2010). Design-based estimation for geometric quantiles with application to outlier detection. *Computational Statistics & Data Analysis 54*(10), 2214–2229.

Chaudhuri, P. (1996). On a geometric notion of quantiles for multivariate data. *Journal of the American Statistical Association 91*(434), 862–872.

Chen, X., J. D. Lee, H. Li, and Y. Yang (2022). Distributed estimation for principal component analysis: An enlarged eigenspace analysis. *Journal of the American Statistical Association 117*(540), 1775–1786.

Chowdhury, J. and P. Chaudhuri (2019). Nonparametric depth and quantile regression for functional data. *Bernoulli 25*, 395–423.

Fan, J., D. Wang, K. Wang, and Z. Zhu (2019). Distributed estimation of principal eigenspaces. *The Annals of Statistics 47*(6), 3009 – 3031.

Jordan, M. I., J. D. Lee, and Y. Yang (2019). Communication-efficient distributed statistical inference. *Journal of the American Statistical Association 114*(526), 668–681.

Koltchinskii, V. and K. Lounici (2017). Normal approximation and concentration of spectral projectors of sample covariance. *The Annals of Statistics 45*(1), 121–157.

Mutis, M., U. Beyaztas, F. Karaman, and H. L. Shang (2025). On function-on-function linear quantile regression. *Journal of Applied Statistics 52*(4), 814–840.

Sen, B. (2022). *A Gentle Introduction to Empirical Process Theory and Applications*. Columbia University. Lecture notes.

Shen, S., J. Lu, and X. Lin (2026). Dimension reduction for large-scale federated data: Statistical rate and asymptotic inference. *Journal of the American Statistical Association 121*(553), 585–597.

Sun, Q., W.-X. Zhou, and J. Fan (2020). Adaptive Huber regression. *Journal of the American Statistical Association 115*(529), 254–265.

Yan, X. and S. He (2026). Poisson subsampling for large-scale functional data analysis. *Computational Statistics 41*, 101.