SUPPLEMENTARY MATERIAL FOR: A PARTIALLY LINEAR FRAMEWORK FOR MASSIVE HETEROGENOUS DATA

By Tianqi Zhao[‡], Guang Cheng[§] and Han Liu[‡]

In this supplemental material, we provide the detailed proofs of results presented in the main text. Appendix A contains theoretical justification of RKHS extension to the partially linear function space, as defined in Section 2. Appendix B, C and D present proofs of results in Section 3, 4 and 5 respectively. Appendix E contains proof of lemmas used in proving main theorems. Appendix F proves an exponential inequality for empirical processes in a Banach space, and Appendix G provides proofs of auxiliary lemmas which are used in Appendix E.

APPENDIX A: RKHS EXTENSION TO PARTIALLY LINEAR FUNCTION SPACE

In this section, we provide detailed theoretical justifications for the RKHS extension to the partially linear space. We first study the properties of the inner product $\langle\cdot,\cdot\rangle_{\mathcal{C}}$ and its induced kernel $\widetilde{K}$, and then prove Proposition 2.3. In the end we provide technical lemmas for the properties of R_u and P_λ .

A.1. A Collection of Lemmas. The following lemmas are direct consequences of defining the new inner product $\langle\cdot,\cdot\rangle_{\mathcal{C}}$. Lemma A.1 proves the existence of kernel $\widetilde{K}$ under the new inner product $\langle\cdot,\cdot\rangle_{\mathcal{C}}$, and derives its closed form. Lemma A.2 justifies the existence of the linear operator W_λ and derives its closed form. Lemma A.3 studies the asymptotic limit of the $\boldsymbol{B}-\boldsymbol{A}$, where recall $\boldsymbol{A}$ is the Reisz representer of $\boldsymbol{B}=\mathbb{E}[\boldsymbol{X}\mid Z]$ under the inner product $\langle\cdot,\cdot\rangle_{\mathcal{C}}$.

Lemma A.1. The linear evaluation functional E_z of $\mathcal{H}$ under the inner product $\langle\cdot,\cdot\rangle_{\mathcal{C}}$ is bounded. Hence $\langle\cdot,\cdot\rangle_{\mathcal{C}}$ induces a new kernel $\widetilde{K}(z,z)$ with the form

$$\widetilde{K}_z(\cdot)=\sum_{\ell=1}^{\infty}\frac{\phi_\ell(z)}{1+\lambda/\mu_\ell}\phi_\ell(\cdot). \tag{A.1}$$

Moreover, we have that $\|\widetilde{K}_z\|_{\mathcal{C}}\le c_\phi h^{-1/2}$, where c_ϕ is the constant specified in Assumption 3.2. This implies that $\|f\|_{\sup}\le c_\phi h^{-1/2}\|f\|_{\mathcal{C}}$ for all $f\in\mathcal{H}$.

PROOF. (i) Boundedness of E_z: We have for any $f \in \mathcal{H}$,

$$|E_z f| = |f(z)| = |\langle f, K_z\rangle_{\mathcal{H}}| \le \|K_z\|_{\mathcal{H}}\|f\|_{\mathcal{H}} \le \lambda^{-1/2} c_k \|f\|_{\mathcal{C}},$$

where the last inequality follows from the relationship $\lambda\|f\|_{\mathcal{H}}^2 \le \|f\|_{\mathcal{C}}$ implied by the definition of $\langle\cdot,\cdot\rangle_{\mathcal{C}}$. It follows that E_z is bounded.

(ii) Existence and exact form of $\widetilde{K}$: By Definition 2.1, we have that $\langle\cdot,\cdot\rangle_{\mathcal{C}}$ induces a new kernel $\widetilde{K}(z,z)$. As $\widetilde{K}_z \in \mathcal{H}$ for all $z \in \mathcal{Z}$, by Fourier expansion, $\widetilde{K}_z = \sum_{\ell=1}^{\infty} \kappa_\ell \phi_\ell$. Then we have

$$\begin{aligned}\kappa_\ell = \langle \widetilde{K}_z, \phi_\ell\rangle_{L_2(\mathbb{P}_Z)} &= \langle \widetilde{K}_z, \phi_\ell\rangle_{\mathcal{C}} - \lambda\langle \widetilde{K}_z, \phi_\ell\rangle_{\mathcal{H}} \\ &= \phi_\ell(z) - \lambda\kappa_\ell/\mu_\ell.\end{aligned}$$

Solving for κ_ℓ, we have $\kappa_\ell = \phi_\ell(z)/(1+\lambda/\mu_\ell)$. Hence we get the formula for $\widetilde{K}_z(\cdot)$ in (A.1).

(iii) Uniform bound of $\widetilde{K}_z$: By (A.1) and reproducing property, we have that

$$\begin{aligned}\|\widetilde{K}_z\|_{\mathcal{C}}^2 = \langle \widetilde{K}_z, \widetilde{K}_z\rangle_{\mathcal{C}} &= \widetilde{K}(z,z) \\ &= \sum_{\ell=1}^{\infty} \frac{\phi_\ell^2(z)}{1+\lambda/\mu_\ell} \le c_\phi^2 h^{-1}.\end{aligned}$$

as desired. Hence for all $z \in \mathcal{Z}$,

$$|f(z)| \le \|f\|_{\mathcal{C}}\|\widetilde{K}_z\|_{\mathcal{C}} \le c_\phi h^{-1/2}\|f\|_{\mathcal{C}},$$

by Cauchy-Schwarz. This implies that $\|f\|_{\sup} \le c_\phi h^{-1/2}\|f\|_{\mathcal{C}}$ for all $f \in \mathcal{H}$. □

Lemma A.2. There exists a bounded linear operator $W_\lambda : \mathcal{H} \to \mathcal{H}$ such that for any $f, \widetilde{f} \in \mathcal{H}$, we have

$$\langle W_\lambda f, \widetilde{f}\rangle_{\mathcal{C}} = \lambda\langle f, \widetilde{f}\rangle_{\mathcal{H}}. \tag{A.2}$$

Moreover, we have for all eigenfunctions ϕ_ℓ, $\ell = 1, 2, \ldots$

$$W_\lambda \phi_\ell(\cdot) = \frac{\lambda}{\lambda + \mu_\ell}\phi_\ell(\cdot). \tag{A.3}$$

PROOF. The proof for the existence of W_λ uses Riesz representation theorem. Define the bilinear form $V(f, \widetilde{f}) := \lambda\langle f, \widetilde{f}\rangle_{\mathcal{H}}$, for any $f, \widetilde{f} \in \mathcal{H}$. For any fixed f, this defines a functional $V_f(\cdot) = V(f, \cdot)$. It is easy to verify that V_f is linear. Moreover, V_f is bounded under the inner product $\langle\cdot,\cdot\rangle_{\mathcal{C}}$, as

$$|V_f(\widetilde{f})| = |\lambda\langle f, \widetilde{f}\rangle_{\mathcal{H}}| \le \lambda\|f\|_{\mathcal{H}}\|\widetilde{f}\|_{\mathcal{H}} \le \lambda^{1/2}\|f\|_{\mathcal{H}}\|\widetilde{f}\|_{\mathcal{C}},$$

for all $\widetilde{f} \in \mathcal{H}$. Hence by Riesz representation theorem, there exists an unique element $f_1 \in \mathcal{H}$ such that $V_f(\widetilde{f}) = \langle f_1, \widetilde{f}\rangle_{\mathcal{C}}$ for all $\widetilde{f} \in \mathcal{H}$. We let $W_\lambda f = f_1$, and it follows that $\langle W_\lambda f, \widetilde{f}\rangle_{\mathcal{C}} = \lambda\langle f, \widetilde{f}\rangle_{\mathcal{H}}$ for all $f, \widetilde{f} \in \mathcal{H}$.

We next prove linearity of W_λ. By definition, for any $f_1, f_2, \widetilde{f} \in \mathcal{H}$ and $a, b \in \mathbb{R}$, we have

$$\begin{aligned}\langle W_\lambda(af_1 + bf_2), \widetilde{f}\rangle_{\mathcal{C}} &= \lambda\langle af_1 + bf_2, \widetilde{f}\rangle_{\mathcal{H}} \\ &= a\lambda\langle f_1, \widetilde{f}\rangle_{\mathcal{H}} + b\lambda\langle f_2, \widetilde{f}\rangle_{\mathcal{H}} \\ &= \langle aW_\lambda f_1 + bW_\lambda f_2, \widetilde{f}\rangle_{\mathcal{C}},\end{aligned}$$

which implies $W_\lambda(af_1 + bf_2) = aW_\lambda f_1 + bW_\lambda f_2$.

Furthermore,W_λ is a bounded operator under $\|\cdot\|_{\mathcal{C}}$, as for any $f \in \mathcal{H}$

$$\begin{aligned}\|W_\lambda f\|_{\mathcal{C}} &= \sup_{\|\widetilde{f}\|_{\mathcal{C}}\le 1} \langle W_\lambda f, \widetilde{f}\rangle_{\mathcal{C}} \\ &= \sup_{\|\widetilde{f}\|_{\mathcal{C}}\le 1} \lambda\langle \lambda f, \widetilde{f}\rangle_{\mathcal{H}} \\ &\le \lambda\|f\|_{\mathcal{H}} \sup_{\|\widetilde{f}\|_{\mathcal{C}}\le 1} \|\widetilde{f}\|_{\mathcal{H}} \\ &\le \|f\|_{\mathcal{C}},\end{aligned}$$

where the last inequality follows from the fact that $\lambda^{1/2}\|f\|_{\mathcal{H}} \le \|f\|_{\mathcal{C}}$ implied by the definition of $\langle\cdot,\cdot\rangle_{\mathcal{C}}$. This shows that W_λ has operator norm bounded by 1.

To prove the second half of the lemma, we have that $\langle W_\lambda f, \widetilde{f}\rangle_{\mathcal{C}} = \lambda\langle f, \widetilde{f}\rangle_{\mathcal{H}}$. Also, by the definition of $\langle\cdot,\cdot\rangle_{\mathcal{C}}$ we have $\langle W_\lambda f, \widetilde{f}\rangle_{\mathcal{C}} = \langle W_\lambda f, \widetilde{f}\rangle_{L_2(P_Z)} + \lambda\langle W_\lambda f, \widetilde{f}\rangle_{\mathcal{H}}$. It follows from the two equations that

$$\langle W_\lambda f, \widetilde{f}\rangle_{L_2(\mathbb{P}_Z)} = \lambda\langle (id - W_\lambda)f, \widetilde{f}\rangle_{\mathcal{H}}, \tag{A.4}$$

for any $f, \widetilde{f} \in \mathcal{H}$. $W_\lambda\phi_\ell$ has a Fourier expansion: $W_\lambda\phi_\ell = \sum_{k=1}^{\infty} w_k\phi_k$. Letting $f = \widetilde{f} = \phi_\ell$ in (A.4) yields $w_\ell = \lambda/(\lambda + \mu_\ell)$, and letting $f = \phi_\ell$ and $\widetilde{f} = \phi_r$ in (A.4) yields $w_r = 0$ for $r \ne \ell$. Hence the conclusion follows. $\square$

Lemma A.3. We have the following equations hold:

$$\lim_{\lambda\to 0}\mathbb{E}\Big[\boldsymbol{X}\big(\boldsymbol{B}(Z) - \boldsymbol{A}(Z)\big)^T\Big] = 0 \tag{A.5}$$

$$\lim_{\lambda\to 0}\mathbb{E}\Big[\boldsymbol{B}(Z)\big(\boldsymbol{B}(Z) - \boldsymbol{A}(Z)\big)^T\Big] = 0 \tag{A.6}$$

$$\lim_{\lambda\to 0}\mathbb{E}\Big[\big(\boldsymbol{B}(Z) - \boldsymbol{A}(Z)(\boldsymbol{B}(Z) - \boldsymbol{A}(Z))^T\big)\Big] = 0. \tag{A.7}$$

The lemma shows that the difference $\boldsymbol{B} - \boldsymbol{A}$ goes to zero as $\lambda \to 0$. Intuititively, as $\lambda \to 0$, the inner product $\langle\cdot,\cdot\rangle_{\mathcal{C}}$ converges to $\langle\cdot,\cdot\rangle_{L_2(\mathbb{P}_Z)}$ by its definition, hence the the representer $\boldsymbol{A}$ of $\boldsymbol{B}$ converges to $\boldsymbol{B}$ itself. The following is the formal proof of this lemma.

PROOF. By reproducing property of $\widetilde{K}_z$, the definition of A_k and (A.1), we have

$$A_k(z) = \langle A_k, \widetilde{K}_z\rangle_{\mathcal{C}} = \langle B_k, \widetilde{K}_z\rangle_{L_2(\mathbb{P}_Z)} = \sum_{i=1}^{\infty} \frac{\langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)}}{1+\lambda/\mu_i}\phi_i(z). \tag{A.8}$$

By Fourier expansion of B_k, it follows from the above equation that

$$B_k(z) - A_k(z) = \sum_{i=1}^{\infty} \frac{\langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)}\lambda/\mu_i}{1+\lambda/\mu_i}\phi_i(z) \to 0, \tag{A.9}$$

As ϕ_i are uniformly bounded and $\sum_{i=1}^{\infty}\langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)} \leq \infty$, it follows from dominated convergence theorem that $B_k(z) - A_k(z) \to 0$ for all $z \in \mathcal{Z}$. Therefore (A.7) holds. Moreover,

$$\begin{aligned}\mathbb{E}\big[B_j(Z)(B_k(Z) - A_k(Z))\big] &= \langle B_k, B_k - A_k\rangle_{L_2(\mathbb{P}_Z)} \\ &= \sum_{i=1}^{\infty} \frac{\lambda/\mu_i}{1+\lambda/\mu_i}\langle B_j, \phi_i\rangle_{L_2(\mathbb{P}_Z)}\langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)} \\ &\to 0,\end{aligned} \tag{A.10}$$

where the second equality is by (A.9) and the limit is by dominated convergence theorem. Hence (A.6) holds. Lastly,

$$\mathbb{E}\big[X_j(B_k(Z) - A_k(Z))\big] = \sum_{i=1}^{\infty} \frac{\lambda/\mu_i}{1+\lambda/\mu_i}\langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)}\mathbb{E}[X_j\phi_i(Z)] \to 0,$$

again by dominated convergence and the fact that X_j and ϕ_i are uniformly bounded. Hence (A.5) also holds. □

A.2. Proof of Proposition 2.3. With the theoretical foundations laid up in the previous section, we are now ready to construct R_u and P_λ, whose exact forms are presented in Proposition 2.3.

PROOF. The proof follows similarly as Proposition 2.1 in Cheng and Shang (2013). We first want to construct $R_u \in \mathcal{A}$ such that it possess the following reproducing property:

$$\langle R_u, m\rangle_{\mathcal{A}} = \boldsymbol{\beta}^T\boldsymbol{x} + f(z), \tag{A.11}$$

for any $u = (\boldsymbol{x}, z)$ and $m = (\boldsymbol{\beta}, f) \in \mathcal{A}$. As $R_u \in \mathcal{A}$, it has two components:

$R_u = (L_u, N_u)$. Hence the L.H.S. of (A.11) can be written as

$$
\begin{aligned}
\langle R_u, m\rangle_{\mathcal{A}} &= \mathbb{E}\big[(\boldsymbol{X}^T L_u + N_u(Z))(\boldsymbol{X}^T\boldsymbol{\beta} + f(Z))\big] + \lambda\langle N_u, f\rangle_{\mathcal{H}}\\
&= \boldsymbol{\beta}^T\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]L_u + \boldsymbol{\beta}^T\mathbb{E}[\boldsymbol{X}N_u(Z)] + L_u^T\mathbb{E}[\boldsymbol{X}f(Z)]\\
&\qquad\qquad + \mathbb{E}[N_u(Z)f(Z)] + \lambda\langle N_u, f\rangle_{\mathcal{H}}\\
&= \boldsymbol{\beta}^T\big(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]L_u + \mathbb{E}[\boldsymbol{B}(Z)N_u(Z)]\big) + L_u^T\mathbb{E}[\boldsymbol{B}(Z)f(Z)] + \langle N_u, f\rangle_{\mathcal{C}}\\
&= \boldsymbol{\beta}^T\big(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]L_u + \mathbb{E}[\boldsymbol{B}(Z)N_u(Z)]\big) + \langle \boldsymbol{A}^T L_u + N_u, f\rangle_{\mathcal{C}} \qquad \text{(A.12)}
\end{aligned}
$$

where in the second last inequality we used the definition of $\langle\cdot,\cdot\rangle_{\mathcal{C}}$ and in the last equality we used $\mathbb{E}[\boldsymbol{B}(Z)f(Z)] = \langle \boldsymbol{B}, f\rangle_{L_2(\mathbb{P}_Z)} = \langle \boldsymbol{A}, f\rangle_{\mathcal{C}}$. On the other hand, the R.H.S of (A.11) is

$$
\boldsymbol{\beta}^T\boldsymbol{x} + f(z) = \boldsymbol{\beta}^T\boldsymbol{x} + \langle \widetilde{K}_z, f\rangle. \qquad \text{(A.13)}
$$

Comparing (A.12) and (A.13), we have the following set of equations:

$$
\begin{aligned}
\boldsymbol{x} &= \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]L_u + \mathbb{E}[\boldsymbol{B}(Z)N_u(Z)]\\
\widetilde{K}_z &= \boldsymbol{A}^T L_u + N_u.
\end{aligned}
$$

From the second equation we get $N_u = \widetilde{K}_z - \boldsymbol{A}^T L_u$. Substitute it into the first equation, we get

$$
\begin{aligned}
\boldsymbol{x} &= \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]L_u + \mathbb{E}[\boldsymbol{B}(Z)(\widetilde{K}_z(Z) - \boldsymbol{A}(Z)^T L_u)]\\
&= \big(\Omega + \mathbb{E}[\boldsymbol{B}(Z)\boldsymbol{B}^T(Z)]\big)L_u + \mathbb{E}[\boldsymbol{B}(Z)\widetilde{K}_z(Z)] - \mathbb{E}[\boldsymbol{B}(Z)\boldsymbol{A}(Z)^T L_u]\\
&= (\Omega + \Sigma_\lambda)L_u + \langle \boldsymbol{B}, \widetilde{K}_z\rangle_{L_2(\mathbb{P}_Z)}\\
&= (\Omega + \Sigma_\lambda)L_u + \langle \boldsymbol{A}, \widetilde{K}_z\rangle_{\mathcal{C}} = (\Omega + \Sigma_\lambda)L_u + \boldsymbol{A}(z),
\end{aligned}
$$

where in the second inequality we used the fact that $\boldsymbol{B}(Z)$ and $\boldsymbol{X} - \boldsymbol{B}(Z)$ are orthogonal. Therefore it follows that

$$
L_u = (\Omega + \Sigma_\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{A}(z)).
$$

This finishes the proof for the construction of R_u.

We next construct P_λ such that

$$
\langle P_\lambda m, \widetilde{m}\rangle_{\mathcal{A}} = \lambda\langle f, \widetilde{f}\rangle_{\mathcal{H}}, \qquad \text{(A.14)}
$$

for any $m = (\boldsymbol{\beta}, f), \widetilde{m} = (\widetilde{\boldsymbol{\beta}}, \widetilde{f}) \in \mathcal{A}$. As $P_\lambda m \in \mathcal{A}$, it has two components: $P_\lambda m = (L_\lambda f, N_\lambda f)$. Similar to the derivation of (A.12), the L.H.S. of (A.14) can be written as

$$
\begin{aligned}
\langle P_\lambda m, \widetilde{m}\rangle_{\mathcal{A}} &= \mathbb{E}\big[(\boldsymbol{X}^T L_\lambda f + N_\lambda f(Z))(\boldsymbol{X}^T\widetilde{\boldsymbol{\beta}} + \widetilde{f})\big] + \lambda\langle N_\lambda f, \widetilde{f}\rangle_{\mathcal{H}}\\
&= \widetilde{\boldsymbol{\beta}}^T\big(\mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]L_\lambda f + \mathbb{E}[\boldsymbol{B}(Z)N_\lambda f(Z)]\big) + \langle \boldsymbol{A}^T L_\lambda f + N_\lambda f, \widetilde{f}\rangle_{\mathcal{C}} \qquad \text{(A.15)}
\end{aligned}
$$

The R.H.S. of (A.14) is

$$\lambda\langle f, \widetilde{f}\rangle_{\mathcal{H}} = \langle W_\lambda f, \widetilde{f}\rangle_{\mathcal{C}}. \tag{A.16}$$

Comparing (A.15) and (A.16), we obtain the following set of equations:

$$\begin{aligned} \mathbf{0} &= \mathbb{E}[\boldsymbol{X}\boldsymbol{X}^T]L_\lambda f + \mathbb{E}[\boldsymbol{B}(Z)N_\lambda f(Z)] \\ W_\lambda f &= \boldsymbol{A}^T L_\lambda f + N_\lambda f. \end{aligned}$$

Solving the above two equations, we get

$$L_\lambda f = -(\Omega + \Sigma_\lambda)^{-1}\langle \boldsymbol{B}, W_\lambda f\rangle_{L_2(\mathbb{P}_Z)} \text{ and } N_\lambda f = W_\lambda f - \boldsymbol{A}^T L_\lambda f,$$

as desired. □

A.3. Properties of R_u and P_λ.

We first present a lemma that bounds the $\mathcal{A}$-norm of R_u.

Lemma A.4. There exists a constant $c_r > 0$ independent to u such that $\|R_u\|_{\mathcal{A}} \leq c_r h^{-1/2}$. It follows that $\|m\|_{\sup} \leq c_r h^{-1/2}\|m\|_{\mathcal{A}}$ for all m.

Proof. By Proposition 2.3, we have that for $u = (\boldsymbol{x}, z)$,

$$\begin{aligned} \langle R_u, R_u\rangle_{\mathcal{A}} &= \boldsymbol{x}^T L_u + N_u(z) \\ &= \widetilde{K}_z(z) + (\boldsymbol{x} - \boldsymbol{A}(z))^T L_u \\ &= \widetilde{K}_z(z) + (\boldsymbol{x} - \boldsymbol{A}(z))^T(\Omega + \Sigma_\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{A}(z)). \end{aligned} \tag{A.17}$$

From Lemma A.1, we have $\widetilde{K}_z(z) = \|\widetilde{K}\|_{\mathcal{C}}^2 \leq c_\phi^2 h^{-1}$. For the second term in (A.17), we first show that Σ_λ is positive definite. Recall the definition of $\Sigma_\lambda = \mathbb{E}\big[\boldsymbol{B}(Z)(\boldsymbol{B}(Z) - \boldsymbol{A}(Z))\big]$. (A.10) shows that

$$[\Sigma_\lambda]_{jk} = \mathbb{E}\big[B_j(Z)(B_k(Z) - A_k(Z))\big] = \sum_{\ell=1}^{\infty} \frac{\lambda/\mu_\ell}{1 + \lambda/\mu_\ell}\langle B_j, \phi_\ell\rangle_{L_2}\langle B_k, \phi_\ell\rangle_{L_2},$$

which implies that Σ_λ is positive definite. Indeed, for any $\boldsymbol{x} \in \mathbb{R}^p$ and $\boldsymbol{x} \neq 0$,

$$\boldsymbol{x}^T\Sigma_\lambda\boldsymbol{x} = \sum_{j=1}^{p}\sum_{k=1}^{p} x_j x_k[\Sigma_\lambda]_{jk} = \sum_{\ell=1}^{\infty}\frac{\lambda/\mu_\ell}{1 + \lambda/\mu_\ell}\Big(\sum_{j=1}^{p} x_j\langle B_j, \phi_\ell\rangle_{L_2}\Big)^2 > 0.$$

Therefore, it follows that the second term in (A.17) is bounded by

$$\begin{aligned} (\boldsymbol{x} - \boldsymbol{A}(z))^T(\Omega + \Sigma_\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{A}(z)) &\leq \|\Omega + \Sigma_\lambda\|^{-1}\|\boldsymbol{x} - \boldsymbol{A}(z)\|_2^2 \\ &\leq \tau_{\min}(\Omega)^{-1}\|\boldsymbol{x} - \boldsymbol{A}(z)\|_2^2. \end{aligned} \tag{A.18}$$

where recall $\tau_{\min}(\Omega)$ is the minimum eigenvalue of Ω. As $\boldsymbol{x}$ is uniformly bounded, we are left to bound $\boldsymbol{A}(z)$. By (A.8), we have

$$A_k(z) = \sum_{\ell=1}^{\infty} \frac{\langle B_k, \phi_\ell \rangle_{L_2(\mathbb{P}_Z)}}{1+\lambda/\mu_\ell} \phi_\ell(z), \tag{A.19}$$

Hence by the assumption that $B_j \in L_2(\mathbb{P}_Z)$ for $j = 1, \ldots, p$ and uniform boundedness of ϕ_ℓ, we have

$$A_k^2(z) \le c_\phi \sum_{\ell=1}^{\infty} \langle B_k, \phi_\ell \rangle^2_{L_2(\mathbb{P}_Z)} \sum_{\ell=1}^{\infty} (1+\lambda/\mu_\ell)^{-2} \le c_\phi \|B_k\|^2_{L_2(\mathbb{P}_Z)} h^{-1}. \tag{A.20}$$

Therefore, by (A.17) and (A.18), we have $\|R_u\|_{\mathcal{A}} \le c_r h^{-1/2}$, where c_r is determined by $p, c_\phi, c_x, \tau_{\min}$ and $\|B_k\|_{L_2(\mathbb{P}_Z)}$. Therefore, for any $u = (\boldsymbol{x}, z)$

$$|m(u)| = |\langle m, R_u \rangle_{\mathcal{A}}| \le \|m\|_{\mathcal{A}} \|R_u\|_{\mathcal{A}} \le c_r h^{-1/2} \|m\|_{\mathcal{A}},$$

which implies that $\|m\|_{\sup} \le c_r h^{-1/2} \|m\|_{\mathcal{A}}$. $\square$

Based on the above lemma, if we have the extra condition that $B_k(z)$ are smooth functions, then we can bound the parametric and nonparametric components of R_u and $P_\lambda m_0$ more precisely.

Lemma A.5. Suppose Assumptions 3.1 - 3.3 hold. Then we have

(i) $\|L_u\|_2^2 \le C_1'$ where $C_1' = 2\tau_{\min}^{-2}\Big(c_x^2 p + c_\phi^2 \operatorname{Tr}(K) \sum_{k=1}^p \|B_k\|^2_{\mathcal{H}}\Big)$, and $\|N_u\|^2_{\mathcal{C}} \le C_1 h^{-1}$ where $C_1 = 2c_\phi^2\Big(1 + C_1' \sum_{k=1}^p \|B_k\|^2_{L_2(\mathbb{P}_Z)}\Big)$;

(ii) Moreover, $\|L_\lambda f_0\|_2^2 \le C_2' \lambda^2$ and $\|N_\lambda f_0\|^2_{\mathcal{C}} \le 2\|f_0\|^2_{\mathcal{H}} \lambda + C_2 \lambda^2$, where $C_2' = \tau_{\min}^{-2} \|f_0\|^2_{\mathcal{H}} \sum_{k=1}^p \|B_k\|^2_{\mathcal{H}}$ and $C_2 = 2C_2' \sum_{k=1}^p \|B_k\|^2_{L_2(\mathbb{P}_Z)}$.

Proof. (i) By Proposition 2.3,

$$L_u = (\Omega + \Sigma_\lambda)^{-1}(\boldsymbol{x} - \boldsymbol{A}(z)) \text{ and } N_u = \widetilde{K}_z - \boldsymbol{A}^T L_u,$$

By the first equation we have

$$\|L_u\|_2^2 \le \|(\Omega + \Sigma_\lambda)^{-1}\|_2^2 \|\boldsymbol{x} - \boldsymbol{A}(z)\|_2^2.$$

Recall

$$A_k(z) = \sum_{\ell=1}^{\infty} \frac{\langle B_k, \phi_\ell \rangle_{L_2}}{1+\lambda/\mu_\ell} \phi_\ell(z).$$

Hence by Assumption 3.3 it follows that

$$A_k(z)^2 \le \sum_{\ell=1}^{\infty} \frac{\langle B_k, \phi_\ell \rangle^2_{L_2}}{\mu_\ell} \sum_{\ell=1}^{\infty} \mu_\ell \frac{\phi_\ell(z)^2}{(1+\lambda/\mu_\ell)^2} \le c_\phi^2 \|B_k\|^2_{\mathcal{H}} \operatorname{Tr}(K),$$

where the first inequality is by Cauchy-Schwarz and the second is by Assumption 3.2 that ϕ_ℓ are uniformly bounded. Hence for all $z \in \mathcal{Z}$,

$$\|\boldsymbol{A}(z)\|_2^2 \le c_\phi^2 \operatorname{Tr}(K) \sum_{k=1}^{p} \|B_k\|_{\mathcal{H}}^2. \tag{A.21}$$

Also, we showed in the proof of Lemma A.4 that $\|(\Omega + \Sigma_\lambda)^{-1}\|$ is bounded by $\tau_{\min}^{-1}$. Finally, by the boundedness of the support of $\mathcal{X}$, we have

$$\|L_U\|_2^2 \le 2\tau_{\min}^{-2}\Big(c_x^2 p + c_\phi^2 \operatorname{Tr}(K) \sum_{k=1}^{p} \|B_k\|_{\mathcal{H}}^2\Big) = C_1'.$$

To control N_u, we have

$$\|N_u\|_{\mathcal{C}}^2 \le 2\big(\|\widetilde{K}_z\|_{\mathcal{C}}^2 + \|L_u^T \boldsymbol{A}\|_{\mathcal{C}}^2\big) \tag{A.22}$$

For the first term in (A.22), by Lemma A.1, we have $\|\widetilde{K}_z\|_{\mathcal{C}}^2 \le c_\phi^2 h^{-1}$. For the second term, by (A.19) we have

$$\begin{aligned} \|A_k\|_{\mathcal{C}}^2 = \langle A_k, A_k\rangle_{\mathcal{C}} = \langle B_k, A_k\rangle_{L_2(\mathbb{P}_Z)} &= \sum_{\ell=1}^{\infty} \frac{\langle B_k, \phi_\ell\rangle_{L_2(\mathbb{P}_Z)}^2}{1+\lambda/\mu_\ell} \\ &\le c_\phi^2 \|B_k\|_{L_2(\mathbb{P}_Z)}^2 \sum_{\ell=1}^{\infty} \frac{1}{1+\lambda/\mu_\ell} = c_\phi^2 \|B_k\|_{L_2(\mathbb{P}_Z)}^2 h^{-1}. \end{aligned} \tag{A.23}$$

where the inequality is by Cauchy-Schwartz and uniform boundedness of ϕ_ℓ. Hence it follows that

$$\begin{aligned} \|L_u^T \boldsymbol{A}\|_{\mathcal{C}} &= \|\sum_{k=1}^{p} (L_u)_k A_k\|_{\mathcal{C}} \le \sum_{k=1}^{p} |(L_u)_k| \|A_k\|_{\mathcal{C}} \\ &\le \|L_u\|_2 \Big(\sum_{k=1}^{p} \|A_k\|_{\mathcal{C}}^2\Big)^{1/2} \le (C_1')^{1/2} c_\phi h^{-1/2} \Big(\sum_{k=1}^{p} \|B_k\|_{L_2(\mathbb{P}_Z)}^2\Big)^{1/2} \end{aligned} \tag{A.24}$$

Therefore, by (A.22) we obtain

$$\|N_u\|_{\mathcal{C}}^2 \le 2c_\phi^2\Big(1 + C_1' \sum_{k=1}^{p} \|B_k\|_{L_2(\mathbb{P}_Z)}^2\Big) h^{-1} = C_1 h^{-1}$$

(ii) By Proposition 2.3, we have

$$\begin{aligned} L_\lambda f_0 &= -(\Omega + \Sigma_\lambda)^{-1} \langle W_\lambda f_0, \boldsymbol{B}\rangle_{L_2(\mathbb{P}_Z)} \\ N_\lambda f_0 &= (L_\lambda f_0)^T \boldsymbol{A} + W_\lambda f_0 \end{aligned}$$

To control $L_\lambda f_0$, we have by Fourier expansion of f_0 and (A.3)

$$\begin{aligned}\langle B_j, W_\lambda f_0\rangle^2_{L_2(\mathbb{P}_Z)} &= \Big(\sum_{\ell=1}^{\infty}\langle B_k,\phi_\ell\rangle_{L_2}\frac{\lambda\theta_\ell}{\mu_\ell+\lambda}\Big)^2\\ &\leq \lambda^2\Big(\sum_{i=1}^{\infty}\frac{\langle B_k,\phi_\ell\rangle^2_{L_2}}{\mu_\ell+\lambda}\Big)\Big(\sum_{\ell=1}^{\infty}\frac{\theta_\ell^2}{\mu_\ell+\lambda}\Big)\\ &\leq \lambda^2\|B_k\|^2_{\mathcal{H}}\|f_0\|^2_{\mathcal{H}}.\end{aligned}$$

Hence by the positive definiteness of Σ_λ, we have

$$\begin{aligned}\|L_{f_0}\|_2^2 &\leq \|(\Omega+\Sigma_\lambda)^{-1}\|\|\langle W_\lambda f_0, \boldsymbol{B}\rangle_{L_2(\mathbb{P}_Z)}\|_2^2\\ &\leq \tau_{\min}^{-2}\|f_0\|^2_{\mathcal{H}}\sum_{k=1}^{p}\|B_k\|^2_{\mathcal{H}}\lambda^2 = C_2'\lambda^2.\end{aligned}$$

For $N_\lambda f_0$, we have

$$\|N_\lambda f_0\|^2_{\mathcal{C}} = 2\|(L_\lambda f_0)^T\boldsymbol{A}\|^2_{\mathcal{C}} + 2\|W_\lambda f_0\|^2_{\mathcal{C}} \tag{A.25}$$

For the first term (A.25), by the fact that $\sum_{\ell=1}^{\infty}\langle B_k,\phi_\ell\rangle^2_{L_2(\mathbb{P}_Z)} = \|B_k\|^2_{L_2(\mathbb{P}_Z)}$, we first get an inequality for $\|A_k\|^2_{\mathcal{C}}$ that is different than (A.23):

$$\|A_k\|^2_{\mathcal{C}} = \langle A_k, A_k\rangle_{\mathcal{C}} = \langle B_k, A_k\rangle_{L_2(\mathbb{P}_Z)} = \sum_{\ell=1}^{\infty}\frac{\langle B_k,\phi_\ell\rangle^2_{L_2(\mathbb{P}_Z)}}{1+\lambda/\mu_\ell} \leq \|B_k\|^2_{L_2(\mathbb{P}_Z)}.$$

Therefore, following the same derivation as (A.24), we have

$$\|(L_\lambda f_0)^T\boldsymbol{A}\|^2_{\mathcal{C}} \leq \|L_\lambda f_0\|_2^2\sum_{k=1}^{p}\|A_k\|^2_{\mathcal{C}} \leq C_2'\lambda^2\sum_{k=1}^{p}\|B_k\|^2_{L_2(\mathbb{P}_Z)}.$$

For the second term in (A.25) we have

$$\begin{aligned}\|W_\lambda f_0\|_{\mathcal{C}} &= \sup_{\|f\|_{\mathcal{C}}=1}|\langle W_\lambda f_0, f\rangle_{\mathcal{C}}| = \sup_{\|f\|_{\mathcal{C}}=1}\lambda|\langle f_0, f\rangle_{\mathcal{H}}|\\ &\leq \sup_{\|f\|_{\mathcal{C}}=1}\sqrt{\lambda\|f_0\|^2_{\mathcal{H}}}\sqrt{\lambda\|f\|^2_{\mathcal{H}}} \leq \lambda^{1/2}\|f_0\|_{\mathcal{H}},\end{aligned}$$

where the last equality follows from the fact that $\lambda\|f\|^2_{\mathcal{H}} \leq \|f\|^2_{\mathcal{C}}$. Therefore, by (A.25), we obtain

$$\|N_\lambda f_0\|^2_{\mathcal{C}} \leq 2\lambda\|f_0\|^2_{\mathcal{H}} + 2C_2'\lambda^2\sum_{k=1}^{p}\|B_k\|^2_{L_2(\mathbb{P}_Z)} = 2\|f_0\|^2_{\mathcal{H}}\lambda + C_2\lambda^2,$$

as desired. □

APPENDIX B: PROOFS IN SECTION 3

B.1. Proof of Theorem 3.5.

PROOF. By first order optimality condition, we have

$$\sum_{i \in L_j} \boldsymbol{X}_i\big(Y_i - \boldsymbol{X}_i^T \check{\boldsymbol{\beta}}^{(j)} - \bar{f}(Z_i)\big) = 0.$$

Hence we have

$$\check{\boldsymbol{\beta}}^{(j)} = \Big(\sum_{i \in L_j} \boldsymbol{X}_i \boldsymbol{X}_i^T\Big)^{-1} \sum_{i \in L_j} \boldsymbol{X}_i\big(Y_i - \bar{f}(Z_i)\big). \tag{B.1}$$

As for $i \in L_j$ we have $Y_i = \boldsymbol{X}_i^T \boldsymbol{\beta}_0^{(j)} + f_0(Z_i) + \varepsilon_i$, hence

$$\begin{aligned} \sqrt{n}(\check{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0) = {} & n^{-1/2} \sum_{i \in L_j} (\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} \boldsymbol{X}_i \varepsilon_i \\ & + n^{-1/2} \sum_{i \in L_j} (\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} \boldsymbol{X}_i\big(f_0(Z_i) - \bar{f}(Z_i)\big), \end{aligned} \tag{B.2}$$

where the $\widehat{\boldsymbol{\Sigma}}^{(j)} = \frac{1}{n}\sum_{i \in L_j} \boldsymbol{X}_i \boldsymbol{X}_i^T$ is sample covariance of $\boldsymbol{X}$ based on data from the j-th subpopulation. For the first term on the R.H.S. of (B.2), it is the same as the one for ordinary least squares, and so

$$n^{-1/2} \sum_{i \in L_j} (\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} \boldsymbol{X}_i \varepsilon_i \rightsquigarrow N(0, \sigma^2 \Sigma^{-1}). \tag{B.3}$$

For the second term on the R.H.S. of (B.2), by triangular inequality, we have for all $1 \le j \le s$ that

$$\begin{aligned} \Big\| \frac{1}{\sqrt{n}} \sum_{i \in L_j} \big(\widehat{\boldsymbol{\Sigma}}^{(j)}\big)^{-1} \boldsymbol{X}_i\big(f_0(Z_i) - \bar{f}(Z_i)\big)\Big\|_2 &\le \|\bar{f} - f_0\|_{\sup} n^{-1/2} \sum_{i \in L_j} \|\widehat{\boldsymbol{\Sigma}}^{(j)} \boldsymbol{X}_i\|_2 \\ &\le C n^{1/2} \|\bar{f} - f_0\|_{\sup}, \end{aligned}$$

for some constant C that is related to the dimension p and boundedness of the support of $\mathcal{X}$. As $\|\bar{f} - f_0\|_{\sup} \le \|\bar{m} - m_0\|_{\sup} \le c_r h^{-1/2} \|\bar{m} - m_0\|_{\mathcal{A}}$ by Lemma A.4, it suffices to bound $\|\bar{m} - m_0\|_{\mathcal{A}}$. Taking average of equations (7.2) for all j over s, we get

$$\bar{m} - m_0 = \frac{1}{N} \sum_{i=1}^{N} R_{U_i} \varepsilon_i - P_\lambda m_0 - \frac{1}{s} \sum_{j=1}^{s} Rem^{(j)}.$$

By triangular inequality,

$$\|\bar{m} - m_0\|_{\mathcal{A}} \le \Big\| \frac{1}{N} \sum_{i=1}^{N} R_{U_i} \varepsilon_i \Big\|_{\mathcal{A}} + \|P_\lambda m_0\|_{\mathcal{A}} + \Big\| \frac{1}{s} \sum_{j=1}^{s} Rem^{(j)} \Big\|_{\mathcal{A}}. \tag{B.4}$$

For the first term on the R.H.S., define $\mathcal{Q}_i = \{|\varepsilon_i| \le \log N\}$. Since $\|R_U\|_{\mathcal{A}} \le c_r h^{-1/2}$, we have that $\big\{\varepsilon_i R_{U_i} I_{\mathcal{Q}_i}\big\}_{i=1}^N$ is a sequence of random variables in

Hilbert space $\mathcal{F}$ that are i.i.d. with mean zero and bounded by $c_r h^{-1/2}\log N$. Therefore by Lemma G.1 we have

$$\mathbb{P}\Big(\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i R_{U_i}\|_{\mathcal{A}} > c_r \log^2 N (Nh)^{-1/2}\Big)$$

$$\leq \mathbb{P}\Big(\big\{\cap_i \mathcal{Q}_i\big\}\cap\big\{\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i R_{U_i}\|_{\mathcal{A}} > c_r \log^2 N (Nh)^{-1/2}\big\}\Big) + \mathbb{P}\big((\cap_i \mathcal{Q}_i)^c\big)$$

$$\leq 2\exp(-\log^2 N) + 2N\exp(-\log^2 N) \to 0. \tag{B.5}$$

Therefore, we have $||\frac{1}{N}\sum_{i=1}^{N}\varepsilon_i R_{U_i}|| = o_P\big(\log^2 N (Nh)^{-1/2}\big)$. Moreover, by (A.26), we have $\|P_\lambda m_0\|_{\mathcal{A}} = O(\lambda^{-1/2})$. Furthermore, by Lemma 7.3, when s satisfies Condition (3.17) , we have $\|\frac{1}{s}\sum_{j=1}^{s} Rem^{(j)}\|_{\mathcal{A}} = o_P(s^{-1/2}b_{n,s}\log N)$. By the definition of $b_{n,s}$, we have if s satisfies (3.18), then $\|\frac{1}{s}\sum_{j=1}^{s} Rem^{(j)}\|_{\mathcal{A}} = o_P((Nh)^{-1/2})$. Therefore, by (B.4), we have

$$\|\bar{m} - m_0\|_{\mathcal{A}} = o_P\big(\log^2 N (Nh)^{-1/2} + \lambda^{1/2}\big).$$

Hence we have

$$\max_{1\leq j\leq s}\|\frac{1}{\sqrt{n}}\sum_{i\in L_j}\big(\widehat{\boldsymbol{\Sigma}}^{(j)}\big)^{-1}\boldsymbol{X}_i\big(f_0(Z_i) - \bar{f}(Z_i)\big)\|_2$$

$$= o_P\big(n^{1/2}h^{-1/2}(\log^2 N (Nh)^{-1/2} + \lambda^{1/2})\big). \tag{B.6}$$

Plugging the relationship $n = N/s$, we have when $s \gtrsim h^{-2}\log^4 N$ and $\lambda = O((Nh)^{-1})$,

$$\|\frac{1}{\sqrt{n}}\sum_{i\in L_j}\big(\widehat{\boldsymbol{\Sigma}}^{(j)}\big)^{-1}\boldsymbol{X}_i\big(f_0(Z_i) - \bar{f}(Z_i)\big)\|_2 = o_P(1).$$

Hence by (B.2) and (B.3),

$$\sqrt{n}(\check{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \rightsquigarrow N(0, \sigma^2\Sigma^{-1}),$$

as desired. □

B.2. Proof of Theorem 3.6.

PROOF. (i) By (7.3), we have for the jth sub-population

$$\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)} = \frac{1}{n}\sum_{i\in L_j} L_{U_i}\varepsilon_i - L_\lambda f_0 - Rem_\beta^{(j)}, \tag{B.7}$$

where $Rem_\beta^{(j)} = 1/n\sum_{i\in L_j}\Big(L_{U_i}\Delta m^{(j)}(U_i) - \mathbb{E}_U[L_U\Delta m^{(j)}(U)]\Big)$. Equation

(B.7) also holds for the k-th sub-population. Hence under $H_0 : \boldsymbol{\beta}_0^{(j)} = \boldsymbol{\beta}_0^{(k)}$, we have

$$\widehat{\boldsymbol{\beta}}^{(j)} - \widehat{\boldsymbol{\beta}}^{(k)} = \frac{1}{n}\sum_{i\in L_j} L_{U_i}\varepsilon_i - \frac{1}{n}\sum_{i\in L_k} L_{U_i}\varepsilon_i - (Rem_\beta^{(j)} - Rem_\beta^{(k)}), \tag{B.8}$$

By independence between two sub-populations, we have $\frac{1}{n}\sum_{i\in L_j} L_{U_i}\varepsilon_i - \frac{1}{n}\sum_{i\in L_k} L_{U_i}\varepsilon_i \rightsquigarrow N(\mathbf{0}, 2\sigma^2\Omega^{-1})$. Moreover, when the conditions in Theorem 3.4 are satisfied, we have $\sqrt{n}\|Q\|\|Rem_\beta^{(j)} - Rem_\beta^{(j)}\|_2 = o_P(1)$ by triangular inequality. Therefore the result follows.

(ii) By (B.2), we have

$$\begin{aligned}\sqrt{n}(\check{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)}) = {} & \frac{1}{\sqrt{n}}\big(\widehat{\Sigma}^{(j)}\big)^{-1}\sum_{i\in L_j}\boldsymbol{X}_i\varepsilon_i \\ & + \frac{1}{\sqrt{n}}\big(\widehat{\Sigma}^{(j)}\big)^{-1}\sum_{i\in L_j}\boldsymbol{X}_i\big(f_0(Z_i) - \bar{f}(Z_i)\big),\end{aligned} \tag{B.9}$$

where $\widehat{\Sigma}^{(j)} = \frac{1}{n}\sum_{i\in L_j}\boldsymbol{X}_i\boldsymbol{X}_i^T$. The above equation is also true for k-th sub-population. So if s satisfies Condition (3.16), (3.17) and (3.18), we have

$$\Big\|\frac{1}{\sqrt{n}}\big(\widehat{\Sigma}^{(j)}\big)^{-1}\sum_{i\in L_j}\boldsymbol{X}_i\big(f_0(Z_i) - \bar{f}(Z_i)\big)\Big\|_2 = o_P(1).$$

We have another equation that is same as (B.9) with j replaced by k. Hence subtracting the two equations, we have under $H_0 : \boldsymbol{\beta}_0^{(j)} = \boldsymbol{\beta}_0^{(k)}$,

$$\begin{aligned}\sqrt{n}(\check{\boldsymbol{\beta}}^{(j)} - \check{\boldsymbol{\beta}}^{(k)}) = {} & \frac{1}{\sqrt{n}}\big(\widehat{\Sigma}^{(j)}\big)^{-1}\sum_{i\in L_j}\boldsymbol{X}_i\varepsilon_i - \frac{1}{\sqrt{n}}\big(\widehat{\Sigma}^{(k)}\big)^{-1}\sum_{i\in L_k}\boldsymbol{X}_i\varepsilon_i \\ & + o_P(1).\end{aligned}$$

Hence the conclusion follows from CLT and independence of sub-populations j and k. □

B.3. Proof of Theorem 3.7. Before presenting the proof, we define the following preliminaries: for any $\mathcal{G} \subset \{1, 2, \ldots, s\}$ with $|\mathcal{G}| = d$, let

$$T_{0,\mathcal{G}} := \max_{j\in\mathcal{G}, 1\le k\le p}\frac{1}{\sqrt{n}}\sum_{i\in L_j}\big(\boldsymbol{\Sigma}^{-1}\big)_k\boldsymbol{X}_i\varepsilon_i,$$

where $\big(\boldsymbol{\Sigma}^{-1}\big)_k$ denotes the k-th row of the precision matrix $\boldsymbol{\Sigma}^{-1}$ of $\boldsymbol{X}$. Furthermore, let

$$W_{0,\mathcal{G}} := \max_{j\in\mathcal{G}, 1\le k\le p} n^{-1/2}\sum_{i\in L_j}\Gamma_{i,k},$$

where $\{\boldsymbol{\Gamma}_i = (\Gamma_{i,1}, \dots \Gamma_{i,p})\}$ for each $i \in L_j$, $j \in \mathcal{G}$ is a sequence of mean zero independent Gaussian vector with $\mathbb{E}[\boldsymbol{\Gamma}_i \boldsymbol{\Gamma}_i^T] = (\boldsymbol{\Sigma})^{-1}\sigma^2$. Lastly, it is useful to recall

$$W_{\mathcal{G}} := \max_{j\in\mathcal{G}, 1\le k\le p} \frac{1}{\sqrt{n}} \sum_{i\in L_j} \big(\widehat{\boldsymbol{\Sigma}}^{(j)}\big)^{-1}_k \boldsymbol{X}_i e_i,$$

and $c_{\mathcal{G}}(\alpha) = \inf\{t \in \mathbb{R} : \mathbb{P}(W_{\mathcal{G}} \le t \,|\, \mathbb{X}) \ge 1 - \alpha\}$.

The proof strategy is similar to that of Theorem 3.2 in Chernozhukov et al. (2013). Specifically, we first approximate $T_{\mathcal{G}}$ by $T_{0,\mathcal{G}}$, and then apply Gaussian approximation to $T_{0,\mathcal{G}}$ and $W_{0,\mathcal{G}}$. Then, we argue that $W_{0,\mathcal{G}}$ and $W_{\mathcal{G}}$ are close. Hence we can approximate the quantiles of $T_{\mathcal{G}}$ by those of $W_{\mathcal{G}}$. The detailed proof is presented as follows.

Proof. By (B.2), we have

$$\sqrt{n}(\check{\boldsymbol{\beta}}^{(j)} - \widetilde{\boldsymbol{\beta}}^{(j)}) = n^{-1/2} \sum_{i\in L_j} (\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} \boldsymbol{X}_i \varepsilon_i + \boldsymbol{\Delta}^{(j)}, \tag{B.10}$$

where

$$\boldsymbol{\Delta}^{(j)} = n^{-1/2} \sum_{i\in L_j} (\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} \boldsymbol{X}_i \big(f_0(Z_i) - \bar{f}(Z_i)\big).$$

By (B.6) in the proof of Theorem 3.5, we have

$$\max_{j\in\mathcal{G}} \|\boldsymbol{\Delta}^{(j)}\|_\infty = o_P\big(n^{1/2} h^{-1/2} (\log^2 N (Nh)^{-1/2} + \lambda^{1/2})\big).$$

and when $s \gtrsim h^{-2} \log(pd) \log^4 N$ and $\lambda = O((Nh)^{-1})$, we have

$$\max_{j\in\mathcal{G}} \|\boldsymbol{\Delta}^{(j)}\|_\infty = o_P(\log^{-1/2}(pd)).$$

By the definitions of $T_{\mathcal{G}}$ and $T_{0,\mathcal{G}}$ and (B.10), we have

$$|T_{\mathcal{G}} - T_{0,\mathcal{G}}| \le \max_{j\in\mathcal{G}} \frac{1}{\sqrt{n}} \Big\| \sum_{i\in L_j} (\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} \boldsymbol{X}_i \varepsilon_i - \boldsymbol{\Sigma}^{-1} \boldsymbol{X}_i \varepsilon_i \Big\|_\infty + \max_{j\in\mathcal{G}} \|\boldsymbol{\Delta}^{(j)}\|_\infty,$$

where we used the fact that $\max_j a_j - \max_j b_j \le \max_j |a_j - b_j|$ for any two finite sequences $\{a_j\}, \{b_j\}$. By the above inequality and Lemma B.2, there exist ζ_1 and ζ_2 such that

$$\mathbb{P}\big(|T_{\mathcal{G}} - T_{0,\mathcal{G}}| \ge \zeta_1\big) \le \zeta_2, \tag{B.11}$$

where $\zeta_1 \sqrt{1 \vee \log(pd/\zeta_1)} = o(1)$ and $\zeta_2 = o(1)$.

We next turn to bound the distance between quantiles of $W_{\mathcal{G}}$ and $W_{0,\mathcal{G}}$. Let $c_{0,\mathcal{G}}(\alpha) := \inf\{t \in \mathbb{R} : \mathbb{P}(W_{0,\mathcal{G}} \le t) \ge 1 - \alpha\}$, and let $\pi(\nu) := C_2 \nu^{1/3} (1 \vee \log(pd/\nu))^{2/3}$ with $C_2 > 0$, and

$$\Psi := \max_{\substack{1\le k,\ell\le p \\ j\in\mathcal{G}}} \sigma^2 |(\widehat{\boldsymbol{\Sigma}}^{(j)} - \boldsymbol{\Sigma})_{k\ell}|.$$

As the data size in each L_j is the same, we can relabel $\{\boldsymbol{X}_i \in \mathbb{R}^p\}_{i \in L_j, j \in \mathcal{G}}$ as $\{\boldsymbol{X}_i^{(j)} \in \mathbb{R}^p\}_{1 \le i \le n, j \in \mathcal{G}}$ and $\{\Gamma_i \in \mathbb{R}^p\}_{1 \le i \le n, j \in \mathcal{G}}$ as $\{\Gamma_i^{(j)} \in \mathbb{R}^p\}_{1 \le i \le n, j \in \mathcal{G}}$. Then we can re-write $W_{0,\mathcal{G}} = \max_{j \in \mathcal{G}, 1 \le k \le p} U_k^{(j)}$, and $W_{\mathcal{G}} = \max_{j \in \mathcal{G}, 1 \le k \le p} V_k^{(j)}$, where

$$U_k^{(j)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Gamma_{ik}^{(j)} \quad \text{and} \quad V_k^{(j)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\widehat{\boldsymbol{\Sigma}}^{(j)})_k^{-1} \boldsymbol{X}_i^{(j)} e_i^{(j)}.$$

($e_i^{(j)}$ is defined in the similar way). Notice that $\boldsymbol{U} = \{U_k^{(j)}\}_{1 \le k \le p, j \in \mathcal{G}}$ can be viewed as an $(p \cdot d)$-dimensional Gaussian random vector with mean zero and covariance

$$\begin{pmatrix} \sigma^2 \boldsymbol{\Sigma}^{-1} & 0 & \dots & 0 \\ 0 & \sigma^2 \boldsymbol{\Sigma}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \boldsymbol{\Sigma}^{-1} \end{pmatrix} \in \mathbb{R}^{(p \cdot d) \times (p \cdot d)}.$$

Conditioned on $\mathbb{X}$, $\boldsymbol{V} = \{V_k^{(j)}\}_{1 \le k \le p, j \in \mathcal{G}}$ can be viewed as an $(p{\cdot}d)$-dimensional Gaussian random vector with mean zero and covariance

$$\begin{pmatrix} \sigma^2 \big(\widehat{\boldsymbol{\Sigma}}^{(1)}\big)^{-1} & 0 & \dots & 0 \\ 0 & \sigma^2 \big(\widehat{\boldsymbol{\Sigma}}^{(2)}\big)^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \big(\widehat{\boldsymbol{\Sigma}}^{(d)}\big)^{-1} \end{pmatrix} \in \mathbb{R}^{(p \cdot d) \times (p \cdot d)}.$$

Using Gaussian comparison (Lemma 3.1 in Chernozhukov et al. (2013)) and applying the same argument as in the proof of Lemma 3.2 in Chernozhukov et al. (2013), we obtain for any $\nu > 0$

$$\mathbb{P}\big(c_{0,\mathcal{G}}(\alpha) \le c_{\mathcal{G}}(\alpha + \pi(\nu))\big) \ge 1 - \mathbb{P}(\Psi > \nu), \tag{B.12}$$

$$\mathbb{P}\big(c_{\mathcal{G}}(\alpha) \le c_{0,\mathcal{G}}(\alpha + \pi(\nu))\big) \ge 1 - \mathbb{P}(\Psi > \nu). \tag{B.13}$$

By Lemma B.1, we have

$$\sup_{\alpha \in (0,1)} \big|\mathbb{P}\big(T_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha)\big) - \alpha\big| \le \sup_{\alpha \in (0,1)} \big|\mathbb{P}\big(W_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha)\big) - \alpha\big| + n^{-c}. \tag{B.14}$$

To further control $\mathbb{P}\big(W_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha)\big)$, we define $\mathcal{E}_1 = \{c_{0,\mathcal{G}}(\alpha - \pi(\nu)) \le c_{\mathcal{G}}(\alpha)\}$, $\mathcal{E}_2 = \{c_{\mathcal{G}}(\alpha) \le c_{0,\mathcal{G}}(\alpha + \pi(\nu))\}$. We have

$$\begin{aligned} \mathbb{P}\big(W_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha)\big) &= \mathbb{P}\big(W_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha), \mathcal{E}_1\big) + \mathbb{P}\big(W_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha), \mathcal{E}_1^c\big) \\ &\le \mathbb{P}\big(W_{0,\mathcal{G}} > c_{0,\mathcal{G}}(\alpha - \pi(\nu))\big) + \mathbb{P}(\mathcal{E}_1^c) \\ &\le \alpha - \pi(\nu) + \mathbb{P}(\Psi > \nu), \end{aligned}$$

where the last inequality is by the definition of $c_{0,\mathcal{G}}(\alpha)$ and (B.12). Similarly, we have

$$\begin{aligned}\mathbb{P}\big(W_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha)\big) &= 1 - \mathbb{P}\big(W_{0,\mathcal{G}} \le c_{\mathcal{G}}(\alpha)\big) \\ &= 1 - \mathbb{P}\big(W_{0,\mathcal{G}} \le c_{\mathcal{G}}(\alpha), \mathcal{E}_2\big) - \mathbb{P}\big(W_{0,\mathcal{G}} \le c_{\mathcal{G}}(\alpha), \mathcal{E}_2^c\big) \\ &\ge 1 - \mathbb{P}\big(W_{0,\mathcal{G}} \le c_{0,\mathcal{G}}(\alpha + \pi(\nu))\big) - \mathbb{P}(\mathcal{E}_2^c) \\ &\ge \alpha + \pi(\nu) - \mathbb{P}(\Psi > \nu),\end{aligned}$$

where the last inequality is by the definition of $c_{0,\mathcal{G}}(\alpha)$ and (B.13). Hence it follows from (B.14) that

$$\sup_{\alpha\in(0,1)} \big|\mathbb{P}\big(T_{0,\mathcal{G}} > c_{\mathcal{G}}(\alpha)\big) - \alpha\big| \le \pi(\nu) + \mathbb{P}(\Psi > \nu) + n^{-c}. \tag{B.15}$$

Define the event $\mathcal{E}_3 = \{|T_{0,\mathcal{G}} - T_{\mathcal{G}}| \le \zeta_1\}$. By (B.11), we have $\mathbb{P}(\mathcal{E}_3^c) \le \zeta_2$. Hence, we deduce that for any α

$$\begin{aligned}\mathbb{P}(T_{\mathcal{G}} \ge c_G(\alpha)) - \alpha &\le \mathbb{P}(T_{\mathcal{G}} \ge c_G(\alpha), \mathcal{E}_3) + \mathbb{P}(\mathcal{E}_3^c) - \alpha \\ &\le \mathbb{P}(T_{0,\mathcal{G}} \ge c_G(\alpha) - \zeta_1) + \zeta_2 - \alpha, \\ &\le \mathbb{P}(T_{0,\mathcal{G}} \ge c_G(\alpha)) + C\zeta_1\sqrt{1 \vee \log(ps/\zeta_1)} + \zeta_2 - \alpha \\ &\le \pi(\nu) + \mathbb{P}(\Psi > \nu) + n^{-c} + C\zeta_1\sqrt{1 \vee \log(ps/\zeta_1)} + \zeta_2,\end{aligned}$$

where the second last inequality is by Corollary 16 of Wasserman (2014) (Gaussian anti-concentration). By similar arguments, we get the same bound for $\alpha - \mathbb{P}(T_{\mathcal{G}} \ge c_G(\alpha))$, so we have

$$\sup_{\alpha} \big|\mathbb{P}(T_{\mathcal{G}} \ge c_G(\alpha)) - \alpha\big| \le \pi(\nu) + \mathbb{P}(\Psi > \nu) + n^{-c} + C\zeta_1\sqrt{1 \vee \log(pd/\zeta_1)} + \zeta_2.$$

Lastly, we bound Ψ. By (B.18) in the proof of Lemma B.2 and the fact that elementwise infinity norm is bounded by spectral norm, we obtain

$$\Psi \le \max_{j\in\mathcal{G}} \|\widehat{\mathbf{\Sigma}}^{(j)} - \mathbf{\Sigma}\|_\infty = o_P\Big(p\sqrt{(\log d)/n}\Big).$$

Hence, choosing $\nu = p\sqrt{(\log d)/n}$, we get

$$\sup_{\alpha} \big|\mathbb{P}(T_{\mathcal{G}} \ge c_G(\alpha)) - \alpha\big| = o(1),$$

which concludes the proof. □

Lemma B.1. Suppose Assumption 3.1 holds. For any $\mathcal{G} \subset \{1, 2, \dots, s\}$ with $d = d$, if $(\log(pdn))^7/n \le C_1 n^{-c_1}$ for some constants $c_1, C_1 > 0$, then we have

$$\sup_{x\in\mathbb{R}} \Big|\mathbb{P}\big(T_{0,\mathcal{G}} \le x\big) - \mathbb{P}\big(W_{0,\mathcal{G}} \le x\big)\Big| \le n^{-c},$$

for some constant $c > 0$.

PROOF. As the data size in each L_j is the same, we can relabel $\{\boldsymbol{X}_i \in$

$\mathbb{R}^p\}_{i\in L_j, j\in\mathcal{G}}$ as $\{\boldsymbol{X}_i^{(j)} \in \mathbb{R}^p\}_{1\le i\le n, j\in\mathcal{G}}$. Then $T_{0,\mathcal{G}} = \max_{j\in\mathcal{G}, 1\le k\le p} n^{-1/2}\sum_{i=1}^n \xi_{ik}^{(j)}$, where $\xi_{ik}^{(j)} = (\boldsymbol{\Sigma}^{-1})_k \boldsymbol{X}_i^{(j)}\varepsilon_i$. For each i, $\{\xi_{ik}^{(j)}\}_{1\le k\le p, j\in\mathcal{G}}$ can be viewed as a $(p\cdot d)$-dimensional vector with covariance matrix

$$\begin{pmatrix} \sigma^2\boldsymbol{\Sigma}^{-1} & 0 & \dots & 0 \\ 0 & \sigma^2\boldsymbol{\Sigma}^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2\boldsymbol{\Sigma}^{-1} \end{pmatrix} \in \mathbb{R}^{(p\cdot d)\times(p\cdot d)}.$$

The same thing can be done for $\boldsymbol{\Gamma}$ which results in the same covariance matrix. Then we apply Corollary 2.1 of Chernozhukov et al. (2013) to prove the Gaussian approximation result stated in the lemma. It suffices to verify Condition (E1) therein. We have $\mathbb{E}\big[(\xi_{ik}^{(j)})^2\big] = (\boldsymbol{\Sigma}^{-1})_{kk}$ is a constant, and $\max_{\ell=1,2}\mathbb{E}\big[|\xi_{ik}^{(j)}|^{2+\ell}/B^\ell\big] + \mathbb{E}\big[\exp(|\xi_{ik}^{(j)}|/B)\big] \le 4$ for some large enough constant B, by the sub-Gaussianity of $\varepsilon_i^{(j)}$ and the boundedness of $\boldsymbol{X}_i^{(j)}$. Hence Condition (E1) is verified, and by the assumption that $(\log(pdn))^7/n \le C_1 n^{-c_1}$, we get the desired result. $\square$

Lemma B.2. Suppose Assumption 3.1 holds. For any $\mathcal{G} \subset \{1, 2, \dots, s\}$ with $d = d$, suppose $p^2\log(pd)/\sqrt{n} = o(1)$. Then there exist ζ_1 and ζ_2 such that

$$\mathbb{P}\bigg(\max_{j\in\mathcal{G}}\frac{1}{\sqrt{n}}\Big\|\sum_{i\in L_j}(\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1}\boldsymbol{X}_i\varepsilon_i - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i\varepsilon_i\Big\|_\infty > \zeta_1\bigg) \le \zeta_2,$$

where $\zeta_1\sqrt{1\vee\log(pd/\zeta_1)} = o(1)$ and $\zeta_2 = o(1)$.

Proof. We have

$$\begin{aligned}
&\max_{j\in\mathcal{G}}\frac{1}{\sqrt{n}}\Big\|\sum_{i\in L_j}(\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1}\boldsymbol{X}_i\varepsilon_i - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i\varepsilon_i\Big\|_\infty \\
&\qquad\le \max_{j\in\mathcal{G}}\big\|(\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} - \boldsymbol{\Sigma}^{-1}\big\|_1 \max_{j\in\mathcal{G}}\Big\|\frac{1}{\sqrt{n}}\sum_{i\in L_j}\boldsymbol{X}_i\varepsilon_i\Big\|_\infty \\
&\qquad\le \max_{j\in\mathcal{G}} p\big\|(\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} - \boldsymbol{\Sigma}^{-1}\big\| \max_{j\in\mathcal{G}}\Big\|\frac{1}{\sqrt{n}}\sum_{i\in L_j}\boldsymbol{X}_i\varepsilon_i\Big\|_\infty, \qquad \text{(B.16)}
\end{aligned}$$

where $\|\cdot\|_1$ denotes the elementwise L_1 norm of matrices. As ε_i are i.i.d. sub-Gaussian random variables, we have by Hoeffding's inequality that for any $j \in \mathcal{G}$ and $1 \le k \le p$

$$\begin{aligned}\mathbb{P}\Big(\frac{1}{\sqrt{n}}\sum_{i\in L_j} X_{ik}\varepsilon_i > t \,|\, \mathbb{X}\Big) &\le \exp\Big(-\frac{nt^2}{\sum_{i\in L_j} X_{ik}^2\sigma^2}\Big)\\ &\le \exp\Big(-\frac{t^2}{c_x^2\sigma^2}\Big),\end{aligned}$$

where the second inequality is by the boundedness of X_{ik}. By law of iterated expectation and union bound we have

$$\mathbb{P}\Big(\max_{j\in\mathcal{G}}\Big\|\frac{1}{\sqrt{n}}\sum_{i\in L_j}\boldsymbol{X}_i\varepsilon_i\Big\|_\infty > t\Big) \le pd\exp\Big(-\frac{t^2}{c_x^2\sigma^2}\Big).$$

Letting $t = 2c_x\sigma\sqrt{\log(pd)}$, we get with probability at least $1-(pd)^{-1}$ that

$$\max_{j\in\mathcal{G}}\Big\|\frac{1}{\sqrt{n}}\sum_{i\in L_j}\boldsymbol{X}_i\varepsilon_i\Big\|_\infty \le 2c_x\sigma\sqrt{\log(pd)}. \tag{B.17}$$

By the boundedness of $\boldsymbol{X}$, we have $\|\boldsymbol{X}_i\boldsymbol{X}_i^T - \mathbb{E}[\boldsymbol{X}_i\boldsymbol{X}_i^T]\| \le 2\|\boldsymbol{X}_i\boldsymbol{X}_i^T\| \le 2\|\boldsymbol{X}_i\|_2^2 \le 2pc_x^2$. Therefore, by Lemma G.3, we have for all $j \in \mathcal{G}$ that

$$\begin{aligned}\mathbb{P}\Big(\|\widehat{\boldsymbol{\Sigma}}^{(j)} - \boldsymbol{\Sigma}\| \ge t\Big) &\le \mathbb{P}\Big(\Big\|\frac{1}{n}\sum_{i\in L_j}\boldsymbol{X}_i\boldsymbol{X}_i^T - \mathbb{E}[\boldsymbol{X}_i\boldsymbol{X}_i^T]\Big\| \ge t\Big)\\ &\le p\exp\Big(-\frac{nt^2}{32p^2c_x^4}\Big).\end{aligned}$$

and so it follows from union bound that

$$\mathbb{P}\Big(\max_{j\in\mathcal{G}}\|\widehat{\boldsymbol{\Sigma}}^{(j)} - \boldsymbol{\Sigma}\| \ge t\Big) \le pd\exp\Big(-\frac{nt^2}{32p^2c_x^4}\Big).$$

Choosing $t = 64p\sqrt{(\log d)/n}$, we obtain

$$\max_{j\in\mathcal{G}}\|\widehat{\boldsymbol{\Sigma}}^{(j)} - \boldsymbol{\Sigma}\| = o_P\Big(p\sqrt{\frac{\log d}{n}}\Big). \tag{B.18}$$

Thus, by Lemma G.4, we get

$$\max_{j\in\mathcal{G}}\|(\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1} - \boldsymbol{\Sigma}^{-1}\| = o_P\Big(p\sqrt{\frac{\log d}{n}}\Big). \tag{B.19}$$

Combining (B.16), (B.17) and (B.19), we have

$$\max_{j\in\mathcal{G}}\frac{1}{\sqrt{n}}\Big\|\sum_{i\in L_j}(\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1}\boldsymbol{X}_i\varepsilon_i - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i\varepsilon_i\Big\|_\infty = o_P\Big(p^2\frac{\log(pd)}{\sqrt{n}}\Big)$$

We choose ζ_1 such that $p^2\log(pd)/(\sqrt{n}\zeta_1) = o(1)$ and $\zeta_1\sqrt{1\vee\log(pd/\zeta_1)} =$

$o(1)$, e.g., $\zeta_1^2 = p^2 \log(pd)/\sqrt{n}$. Then by the above equation we have

$$\mathbb{P}\left(\max_{j\in\mathcal{G}} \frac{1}{\sqrt{n}} \Big\| \sum_{i\in L_j} (\widehat{\boldsymbol{\Sigma}}^{(j)})^{-1}\boldsymbol{X}_i\varepsilon_i - \boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i\varepsilon_i \Big\|_\infty \geq \zeta_1 \right) < \zeta_2,$$

where $\zeta_2 = o(1)$. □

APPENDIX C: PROOFS IN SECTION 4

C.1. Proof of Corollary 4.1.

Proof. We begin by computing h. As $\mu_i = 0$ for $i > r$, we have that $h^{-1} = \sum_{i=1}^r \frac{1}{1+\lambda/\mu_i} \asymp r$. Hence $h \asymp r^{-1}$.

Therefore by Theorem 3.4, $\lambda = o\big((Nh)^{-1/2} \wedge n^{-1/2}\big) = o(N^{-1/2})$. We next calculate the asymptotic covariance.

$$\begin{aligned} A_k(z_0) &= \langle A_k, \widetilde{K}_{z_0}\rangle_{\mathcal{C}} = \langle B_k, \widetilde{K}_{z_0}\rangle_{L_2(\mathbb{P}_Z)} \\ &= \sum_{i=1}^r \frac{\langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)}}{1+\frac{\lambda}{\mu_i}}\phi_i(z_0) \to \sum_{i=1}^r \langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)}\phi_i(z_0). \end{aligned}$$

Hence $\gamma_{z_0} = h^{1/2}\sum_{i=1}^r \langle B_k, \phi_i\rangle_{L_2(\mathbb{P}_Z)}\phi_i(z_0)$. The formula for Σ_{12}^* and Σ_{22}^* then follows from Theorem 3.4.

We next calculate the entropy integral $\omega(\mathcal{F}, \delta)$ for finite rank RKHS and the upper bound for s. Define $\widetilde{\mathcal{F}}_2 = \{f \in \mathcal{H} : \|f\|_{\sup} \leq 1, \|f\|_{\mathcal{H}} \leq 1\}$. By Carl and Triebel (1980), for finite rank RKHS,

$$\log \mathcal{N}(\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, \delta) \asymp r \log \delta^{-1}.$$

We have that $\mathcal{N}(\mathcal{F}, \|\cdot\|_{\sup}, \delta) \leq \mathcal{N}(\mathcal{F}_1, \|\cdot\|_{\sup}, \delta)\mathcal{N}(\mathcal{F}_2, \|\cdot\|_{\sup}, \delta)$. As $\mathcal{N}(\mathcal{F}_1, \|\cdot\|_{\sup}, \delta)$ is dominated by $\mathcal{N}(\mathcal{F}_2, \|\cdot\|_{\sup}, \delta)$, it suffices to bound $\mathcal{N}(\mathcal{F}_2, \|\cdot\|_{\sup}, \delta)$. Now by Van Der Vaart and Wellner (1996), we have that

$$\begin{aligned} \mathcal{N}(\mathcal{F}_2, \|\cdot\|_{\sup}, \delta) &\leq \mathcal{N}(h^{1/2}\lambda^{-1/2}\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, \delta) \\ &= \mathcal{N}(\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, h^{-1/2}\lambda^{1/2}\delta). \end{aligned}$$

Hence

$$\begin{aligned} \omega(\mathcal{F}, \delta) &\leq \int_0^\delta \sqrt{\log \mathcal{N}(\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, h^{-1/2}\lambda^{1/2}\varepsilon)}\,d\varepsilon \\ &\asymp \int_0^\delta \sqrt{r \log(h^{1/2}\lambda^{-1/2}\varepsilon^{-1})}\,d\varepsilon \\ &\asymp \sqrt{r}\delta\sqrt{\log(h^{1/2}\lambda^{-1/2}\delta^{-1})} \end{aligned}$$

Now we are ready to calculate the upper bound for s. We plug in $n = N/s$ and $h \asymp r^{-1}$ into (3.6) and (3.7), and by the condition $\lambda = o\Big(\frac{1}{\sqrt{N}}\Big)$, we get

$s = o\Big(\frac{N}{\sqrt{\log\lambda^{-1}}\log^6 N}\Big)$. This upper bound needs to allow the case that $s = 1$, which yields the lower bound for λ: $\sqrt{\log(\lambda^{-1})} = o\big(N\log^{-6}N\big)$. $\square$

C.2. Proof of Corollary 4.2.

PROOF. Recall that we have $h \asymp r^{-1}$. To optimize the rate, we choose λ such that $\frac{1}{Nh} \asymp \lambda$, which yields $\lambda = \frac{r}{N}$. By Theorem 3.1 we have

$$\mathbb{E}\big[\|\bar{f}_{N,\lambda} - f_0\|^2_{L_2(\mathbb{P}_Z)}\big] \le Cr/N + s^{-1}a(n,s,h,\lambda,\omega).$$

For the remainder term to be small, we need $s^{-1}a(n,s,h,\lambda,\omega) \lesssim N^{-1}$. Plugging in $a(n,s,h,\lambda,\omega), h$ and λ, we get the upper bound for s. $\square$

C.3. A Lemma for Exponentially Decaying RKHS.

Lemma C.1. Let $h = (-\log\lambda)^{-1/p}$. For all $t > 0$, $p \ge 1$ and some positive constants c, α, we have

$$\lim_{\lambda\to 0}\sum_{\ell=1}^{\infty}\frac{1}{(1+\lambda c\exp(\alpha\ell^p))^t} = \alpha^{-1/p}.$$

PROOF. We have by convexity that

$$\sum_{\ell=1}^{\infty}\frac{1}{(1+\lambda c\exp(\alpha\ell^p))^t} \le \int_0^{\infty}\frac{1}{(1+\lambda c\exp(\alpha x^p))^t}dx.$$

We then approximate the integral by

$$\begin{aligned}
&\int_0^{\infty}\frac{dx}{(1+\lambda c\exp(\alpha x^p))^t}\\
&= \int_0^{(\alpha^{-1}\log(1/\lambda))^{1/p}}\frac{dx}{(1+\lambda c\exp(\alpha x^p))^t}\\
&\qquad\qquad + \int_{(\alpha^{-1}\log(1/\lambda))^{1/p}}^{\infty}\frac{dx}{(1+\lambda c\exp(\alpha x^p))^t}\\
&\le (\alpha^{-1}\log(1/\lambda))^{1/p} + \int_{(\alpha^{-1}\log(1/\lambda))^{1/p}}^{\infty}(c\lambda)^{-t}\exp(-t\alpha x^p)dx\\
&= (\alpha^{-1}\log(1/\lambda))^{1/p} + o(1), \qquad \text{(C.1)}
\end{aligned}$$

where the last equality is by L'Hospital's Rule for $\lambda \to 0$.

Moreover, we have for any $\epsilon \in (0,1)$ that

$$\begin{aligned}\sum_{\ell=1}^{\infty} \frac{1}{(1+\lambda c\exp(\alpha \ell^p))^t} &\geq \int_1^{\infty} \frac{1}{(1+\lambda c\exp(\alpha x^p))^t} dx \\ &\geq \int_1^{(\epsilon\alpha^{-1}\log(1/\lambda))^{1/p}} \frac{1}{(1+\lambda c\exp(\alpha x^p))^t} dx \\ &\geq \frac{1}{(1+c\lambda^{1-\epsilon})^t}\Big((\epsilon\alpha^{-1}\log(1/\lambda))^{1/p} - 1\Big) \\ &= \frac{1}{(1+c\lambda^{1-\epsilon})^t}\Big((\epsilon\alpha^{-1}\log(1/\lambda))^{1/p}\Big) + O(1). \qquad \text{(C.2)}\end{aligned}$$

Combining (C.1) and (C.2), we get

$$\Big(\frac{\epsilon}{\alpha}\Big)^{1/p} \leq \lim_{\lambda\to 0}\sum_{\ell=1}^{\infty} \frac{h}{(1+\lambda c\exp(\alpha \ell^p))^t} \leq \Big(\frac{1}{\alpha}\Big)^{1/p}.$$

for any $\epsilon \in (0,1)$. Lastly, letting $t \to 1$, we get the desired result. □

C.4. Proof the Corollary 4.3.

PROOF. As before, we start by calculating h. By Lemma C.1 with $t = 1$, we have $h \asymp (-\log\lambda/c)^{-1/p}$.

As $h \to 0$, Theorem 3.4 shows that $\alpha_{z_0} = \gamma_{z_0} = 0$. Moreover,

$$\begin{aligned}|W_\lambda f_0(z_0)| &= \lambda\Big|\sum_{\ell=1}^{\infty} \frac{\theta_\ell}{\lambda+\mu_\ell}\phi_\ell(z_0)\Big| \\ &\leq \lambda\sum_{\ell=0}^{\infty} |\phi_\ell(z_0)\langle f_0, \phi_\ell\rangle_{\mathcal{H}}| = O(\lambda).\end{aligned}$$

Therefore, by Theorem 3.4, we can completely remove the asymptotic bias by choosing $\lambda = o((Nh)^{-1/2} \wedge n^{-1/2}) = o\big(N^{-1/2}\log^{1/(2p)} N \wedge n^{-1/2}\big)$. We next calculate the entropy integral. We have that for RKHS with exponentially decaying eigenvalues, by Proposition 17 in Williamson et al. (2001) with $p = 2$,

$$\log\mathcal{N}(\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, \delta) \asymp \Big(\log\frac{1}{\delta}\Big)^{\frac{p+1}{p}}.$$

Then following the deduction in the proof of Corollary 4.1, we have

$$\begin{aligned}\omega(\delta) &\le \int_0^\delta \sqrt{\log(1+\mathcal{N}(\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, (c_r^{-2}h\lambda^{-1})^{-1/2}\varepsilon))}d\varepsilon \\ &\asymp \int_0^\delta \sqrt{\Big(\log\frac{1}{(c_r^{-2}h\lambda^{-1})^{-1/2}\varepsilon}\Big)^{\frac{p+1}{p}}}d\varepsilon \\ &\asymp \delta\log^{\frac{p+1}{2p}}(h^{1/2}\lambda^{-1/2}\delta^{-1}).\end{aligned}$$

For the range on s, we plug in $n = N/s$ and $h \asymp (-\log\lambda/c)^{-1/p}$ into (3.6) and (3.7), and we get that it suffices to take

$$s = o\Big(\frac{N}{\log^6 N \log^{(p+4)/p}\lambda^{-1}}\Big).$$

Again the upper bound must allow the case that $s = 1$, which yields the lower bound for the choice of λ. □

C.5. Proof of Corollary 4.4.

PROOF. Recall that we have $h \asymp (-\log\lambda/c)^{-1/p}$. To balance variance and bias, we choose $\lambda = \frac{(\log N)^{1/p}}{N}$. By Theorem 3.1 we have

$$\mathbb{E}\big[\|\bar{f}_{N,\lambda} - f_0\|^2_{L_2(\mathbb{P}_Z)}\big] \le C(\log N)^{1/p}/N + s^{-1}a(n,s,h,\lambda,\omega).$$

For the remainder term to be small, we need $s^{-1}a(n,s,h,\lambda,\omega) \lesssim (\log N)^{1/p}/N$. Plugging in h, λ and $\omega(\mathcal{F},1)$, we get the upper bound for s. □

C.6. Proof of Corollary 4.5.

PROOF. Again, we begin by calculating h. As $\mu_j \le cj^{-2\nu}$, we approximate h using integration. For simplicity, let $c = 1$ here. We have

$$\begin{aligned} h^{-1} \le \int_0^\infty \frac{1}{1+\lambda x^{2\nu}}dx &= \int_0^{\lambda^{-\frac{1}{2\nu}}} \frac{1}{1+\lambda x^{2\nu}}dx + \int_{\lambda^{-\frac{1}{2\nu}}}^\infty \frac{1}{1+\lambda x^{2\nu}}dx \\ &\le \Big(1+\frac{1}{1-2\nu}\Big)\lambda^{-\frac{1}{2\nu}}.\end{aligned}$$

On the other hand, we also have

$$\begin{aligned} h^{-1} \ge \int_1^\infty \frac{1}{1+\lambda x^{2\nu}}dx &= \int_1^{\lambda^{-\frac{1}{2\nu}}} \frac{1}{1+\lambda x^{2\nu}}dx + \int_{\lambda^{-\frac{1}{2\nu}}}^\infty \frac{1}{1+\lambda x^{2\nu}}dx \\ &\ge \Big(2+\frac{1}{2-4\nu}\Big)\lambda^{-\frac{1}{2\nu}}.\end{aligned}$$

Hence we conclude that $h^{-1} \asymp \lambda^{-\frac{1}{2\nu}}$ and thus $h \asymp \lambda^{\frac{1}{2\nu}}$.

As $h \to 0$, Theorem 3.4 shows that $\alpha_{z_0} = \gamma_{z_0} = 0$. Similar to proof of Corollary 4.3, we get $|W_\lambda f_0(z_0)| = o(\lambda)$, and by Theorem 3.4, we can remove the asymptotic bias by choosing $\lambda = o\big((Nh)^{-1/2} \wedge n^{-1/2}\big) = o\big(N^{-\frac{2\nu}{4\nu+1}} \wedge n^{-1/2}\big)$. We next calculate the entropy integral. We have that for RKHS with polynomially decaying eigenvalues, by Proposition 16 in Williamson et al. (2001),

$$\log \mathcal{N}(\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, \delta) \asymp \Big(\frac{1}{\delta}\Big)^{\frac{1}{\nu}}.$$

Then following the deduction in the proof of Corollary 4.1, we have

$$\begin{aligned}
\omega(\mathcal{F}, \delta) &\leq \int_0^\delta \sqrt{\log \mathcal{N}(\widetilde{\mathcal{F}}_2, \|\cdot\|_{\sup}, h^{-1/2}\lambda^{1/2}\varepsilon)}d\varepsilon \\
&\asymp \int_0^\delta \sqrt{\Big(\frac{1}{h\lambda^{-1})^{-1/2}\varepsilon}\Big)^{\frac{1}{\nu}}}d\varepsilon \\
&\asymp (h\lambda^{-1})^{\frac{1}{4\nu}}\delta^{1-\frac{1}{2\nu}}.
\end{aligned}$$

For the range on s, we plug in $n = N/s$ and $h \asymp \lambda^{\frac{1}{2\nu}}$ into (3.6) and (3.7), and it follows that s needs to satisfy

$$s = o\Big(\lambda^{\frac{10\nu-1}{4\nu^2}} N \log^{-6} N\Big).$$

Again the upper bound must allow the case that $s = 1$, which yields the lower bound for the choice of λ: $\lambda^{-1} = o\big(N^{\frac{4\nu^2}{10\nu-1}}\big)$. $\square$

C.7. Proof of Corollary 4.6 .

PROOF. Recall that we have $h \asymp \lambda^{1/2\nu}$. To optimize the rate, we choose λ such that $\frac{1}{Nh} \asymp \lambda$, which yields $\lambda = N^{-\frac{2\nu}{2\nu+1}}$. By Theorem 3.1 we have

$$\mathbb{E}\big[\|\bar{f}_{N,\lambda} - f_0\|^2_{L_2(\mathbb{P}_Z)}\big] \leq CN^{-\frac{2\nu}{2\nu+1}} + s^{-1}a(n, s, h, \lambda, \omega)$$

For the remainder term to be small, we need $s^{-1}a(n, s, h, \lambda, \omega) \lesssim N^{-\frac{2\nu}{2\nu+1}}$. Plugging in a_n, h and λ, we get the upper bound for s. $\square$

C.8. Proof of Lemma 4.1. Recall that $\lambda = h^{2\nu}$. By Theorem 3.3 for the asymptotic variance, we compute that

$$\begin{aligned} h\|\widetilde{K}_{z_0}\|^2_{L_2(\mathbb{P}_Z)} &= h\sum_{\ell=1}^{\infty}\big(\frac{\phi_\ell(z_0)}{1+\lambda/\mu_\ell}\big)^2 \\ &= h\Big(1+\sum_{\ell=1}^{\infty}\frac{2cos^2(2\ell\pi z_0)+2sin^2(2\ell\pi z_0)}{(1+\lambda(2\ell\pi)^{2\nu})^2}\Big) \\ &= h\Big(1+\sum_{\ell=1}^{\infty}\frac{2}{(1+(2\ell\pi h)^{2\nu})^2}\Big) \end{aligned}$$

And we have that

$$\begin{aligned} \sum_{\ell=0}^{\infty}\frac{2\pi h}{(1+(2\ell\pi h)^{2\nu})^2} &\le \sum_{\ell=1}^{\infty}\int_{2\pi h(\ell-1)}^{2\pi h\ell}\frac{1}{(1+x^{2\nu})^2}dx \\ &\to \int_0^{\infty}\frac{1}{(1+x^{2\nu})^2}dx \end{aligned}$$

and similarly

$$\begin{aligned} \sum_{\ell=1}^{\infty}\frac{2\pi h}{(1+(2\ell\pi h)^{2\nu})^2} &\ge \sum_{\ell=1}^{\infty}\int_{2\pi h(\ell-1)}^{2\pi h\ell}\frac{1}{(1+x^{2\nu})^2}dx \\ &\to \int_0^{\infty}\frac{1}{(1+x^{2\nu})^2}dx \end{aligned}$$

The two inequalities yield

$$h\|\widetilde{K}_{z_0}\|^2_{L_2(\mathbb{P}_Z)} \to \int_0^{\infty}\frac{1}{\pi(1+x^{2\nu})^2}dx$$

and so

$$\sigma^2_{z_0} = \int_0^{\infty}\frac{1}{\pi(1+x^{2\nu})^2}dx. \tag{C.3}$$

APPENDIX D: PROOF OF RESULTS IN SECTION 5

D.1. Proof of Proposition 5.1.

PROOF. The proof follows similarly as the proof for Theorem 3.1. □

D.2. Proof of Proposition 5.2.

PROOF. Recall (7.2) from proof of Theorem 3.1 in Section 7.2

$$\widehat{m}^{(j)} - m_0 = \frac{1}{n}\sum_{i\in L_j} R_{U_i}\varepsilon_i - P_\lambda m_0 - Rem^{(j)}.$$

Also recall $m_0^* = m_0 - P_\lambda m_0$. Taking average of the above equation for all j over s, we have

$$\bar{m} - m_0^* = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_i R_{U_i} + \frac{1}{s}\sum_{j=1}^{s} Rem^{(j)}, \tag{D.1}$$

which decomposes into

$$\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0^* = \frac{1}{N}\sum_{i=1}^{N} L_{U_i}\varepsilon_i - \frac{1}{s}\sum_{j=1}^{s} Rem_\beta^{(j)}, \tag{D.2}$$

and

$$\bar{f} - f_0^* = \frac{1}{N}\sum_{i=1}^{N} N_{U_i}\varepsilon_i - \frac{1}{s}\sum_{j=1}^{s} Rem_f^{(j)}. \tag{D.3}$$

Similar to proof in Section 7.2, we can show that the first term weakly converges to a normal distribution, and the remainder term is asymptotically ignorable. Recall the definition of $m_0^* = (id - P_\lambda)f_0$.

Therefore, we deduct that

$$\begin{aligned}
&(\boldsymbol{x}^T, 1)\begin{pmatrix} \sqrt{N}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0^*) \\ \sqrt{Nh}\big(\bar{f}(z_0) - f_0^*(z_0)\big) \end{pmatrix} \\
&= \sqrt{N}\boldsymbol{x}^T(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0^*) + (Nh)^{\frac{1}{2}}(\bar{f}(z_0) - f_0^*(z_0)) \\
&\le \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\varepsilon_i \boldsymbol{x}^T L_{U_i} + h^{1/2}\varepsilon_i N_{U_i}) \\
&\qquad \frac{1}{s}\sum_{j=1}^{s}\sqrt{N}\boldsymbol{x}^T Rem_\beta^{(j)} - \frac{1}{s}\sum_{j=1}^{s}\sqrt{Nh}Rem_f^{(j)}(z_0).
\end{aligned} \tag{D.4}$$

We can show that the first term is asymptotic normal by central limit theorem: first note that the summands are i.i.d. and with mean zero. Moreover, by Proposition 2.3,

$$\begin{aligned}
\boldsymbol{x}^T L_U + h^{1/2}N_U(z_0) &= \boldsymbol{x}^T L_U + h^{1/2}(\widetilde{K}_Z(z_0) - \boldsymbol{A}(z_0)^T L_U) \\
&= (\boldsymbol{x} - h^{1/2}\boldsymbol{A}(z_0))^T L_U + h^{1/2}\widetilde{K}_Z(z_0)
\end{aligned}$$

We compute

$$
\begin{aligned}
&\mathbb{E}\big[(\boldsymbol{x}^T L_{U_i} + h^{1/2} N_{U_i}(z_0))^2\big] \\
&= \underbrace{\mathbb{E}\big[(\boldsymbol{x} - h^{1/2}\boldsymbol{A}(z_0))^T L_U L_U^T (\boldsymbol{x} - h^{1/2}\boldsymbol{A}(z_0))\big]}_{(I)} \\
&\quad + \underbrace{\mathbb{E}\big[h\widetilde{K}_Z(z_0)^2\big]}_{(II)} + \underbrace{\mathbb{E}\big[2h^{1/2}\widetilde{K}_Z(z_0)(\boldsymbol{x} - h^{1/2}\boldsymbol{A}(z_0))^T L_U\big]}_{(III)}.
\end{aligned} \tag{D.5}
$$

For the second term in (D.5), we have

$$
(II) = \mathbb{E}\big[h\widetilde{K}_Z(z_0)^2\big] = h\|\widetilde{K}_{z_0}\|^2_{L_2(\mathbb{P}_Z)} \to \sigma^2_{z_0}. \tag{D.6}
$$

For the first term in (D.5), recall from Section 7.2 that $\mathbb{E}[L_U L_U^T] \to \Omega^{-1}$, $h^{1/2}\boldsymbol{A}(z_0) \to -\gamma_{z_0}$. Hence we have

$$
(I) \to (\boldsymbol{x} + \gamma_{z_0})^T \Omega^{-1} (\boldsymbol{x} + \gamma_{z_0}). \tag{D.7}
$$

Moreover, recall from Section 7.2 that

$$
h^{1/2}\mathbb{E}\big[\widetilde{K}_Z(z_0) L_U\big] \to \Omega^{-1}\alpha_{z_0}.
$$

Therefore it follows that

$$
(III) \to 2(\boldsymbol{x} + \gamma_{z_0})^T \Omega^{-1}\alpha_{z_0}. \tag{D.8}
$$

Hence combining (D.7), (D.6) and (D.8), the limit of (D.5) is

$$
\mathbb{E}\big[(\boldsymbol{x}^T L_{U_i} + h^{1/2} N_{U_i}(z_0))^2\big] \to \boldsymbol{x}^T \Omega^{-1}\boldsymbol{x} + 2\boldsymbol{x}^T \Sigma_{12} + \Sigma_{22},
$$

for any $\boldsymbol{x} \in \mathbb{R}^p$. Therefore the limit distribution follows by central limit theorem. Now for the remainder terms, by Lemma 7.3, if Condition (3.6) is satisfied, we have

$$
\begin{aligned}
\Big|\frac{1}{s}\sum_{j=1}^{s} \sqrt{N}\boldsymbol{x}^T Rem_\beta^{(j)}\Big| &\le C\sqrt{N}\Big\|\frac{1}{s}\sum_{j=1}^{s} Rem_\beta^{(j)}\Big\|_2 \\
&= o_P(N^{1/2}s^{-1/2}b_{n,s}\log n).
\end{aligned}
$$

and

$$
\begin{aligned}
\Big|\frac{1}{s}\sum_{j=1}^{s} \sqrt{Nh}\boldsymbol{x}^T Rem_f^{(j)}(z_0)\Big| &\le C\sqrt{Nh}\Big\|\frac{1}{s}\sum_{j=1}^{s} Rem_f^{(j)}\Big\|_{\sup} \\
&\le C' N^{1/2}\Big\|\frac{1}{s}\sum_{j=1}^{s} Rem_f^{(j)}\Big\|_{\mathcal{C}} \\
&= o_P(N^{1/2}s^{-1/2}b_{n,s}\log n).
\end{aligned}
$$

where in the second inequality we used Lemma A.1. Then if Condition (3.7) is satisfied, we have $N^{1/2}s^{-1/2}b_{n,s}\log n \to 0$. Hence by (D.4), it follows that

$$\big(\boldsymbol{x}^T, 1\big)\begin{pmatrix} \sqrt{N}(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_0^*) \\ \sqrt{Nh}\big(\bar{f}(z_0) - f_0^*(z_0)\big) \end{pmatrix} \to N\big(0, \sigma^2(\boldsymbol{x}^T\Omega^{-1}\boldsymbol{x} + 2\boldsymbol{x}^T\Sigma_{12} + \Sigma_{22})\big).$$

Hence the conclusion follows by the arbitrariness of $\boldsymbol{x}$ using Wold device. □

APPENDIX E: PROOFS OF LEMMAS IN SECTION 7

E.1. Proof of Lemma 7.1.

PROOF. Recall from Section 7.3 that

$$Z_n(\widetilde{m}) = \frac{1}{2}h^{1/2}n^{1/2}d_{n,s}^{-1}Rem^{(j)} = \frac{1}{2}c_r^{-1}hn^{1/2}r_{n,s}^{-1}Rem^{(j)}$$

We showed in Section 7.3 that $Z_n(m)$ is a sub-Gaussian process. Letting $\mathbb{U}^{(j)} = (\mathbb{X}^{(j)}, \mathbb{Z}^{(j)})$, where $\mathbb{X}^{(j)}$ and $\mathbb{Z}^{(j)}$ are designs on j-th sub-population. Without causing any confusion, we can remove the the superscript (j). We have

$$\begin{aligned} \mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2] &= \mathbb{E}\big[\mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]\big] \\ &= \mathbb{E}\big[\mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}}\big] + \mathbb{E}\big[\mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}^c}\big], \end{aligned} \tag{E.1}$$

where $\mathcal{E}$ is the event defined in Section 7.3. For the first term in (E.1), we have

$$\begin{aligned} &\mathbb{E}\big[\mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}}\big] \\ &= 4c_r^2h^{-2}n^{-1}r_{n,s}^2\mathbb{E}\big[\mathbb{E}[\|Z_n(\widetilde{m})\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}}\big] \\ &= 4c_r^2h^{-2}n^{-1}r_{n,s}^2\int_0^\infty \mathbb{P}\big(\mathbb{E}[\|Z_n(\widetilde{m})\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}} \geq x\big)dx \\ &= 4c_r^2h^{-2}n^{-1}r_{n,s}^2\Big\{\int_0^{\omega(\mathcal{F},1)^2} \mathbb{P}\big(\mathbb{E}[\|Z_n(\widetilde{m})\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}} \geq x\big)dx \\ &\qquad + \int_{\omega(\mathcal{F},1)^2}^\infty \mathbb{P}\big(\mathbb{E}[\|Z_n(\widetilde{m})\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}} \geq x\big)dx\Big\} \\ &\leq 4c_r^2h^{-2}n^{-1}r_{n,s}^2\Big\{\omega(\mathcal{F},1)^2 + \int_0^\infty \mathbb{P}\big(\mathbb{E}[\|Z_n(\widetilde{m})\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}} \geq x + \omega(\mathcal{F},1)^2\big)dx\Big\}. \end{aligned}$$

In Section 7.3 we proved that $\mathcal{E} \subset \{\widetilde{m} \in \mathcal{F}\}$. Therefore we have

$$\mathbb{E}[\|Z_n(\widetilde{m})\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}} \leq \sup_{m\in\mathcal{F}} \|Z_n(m)\|_{\mathcal{A}}^2.$$

and by Lemma F.1 and the fact that $\mathrm{diam}(\mathcal{F}) \leq 1$, we have

$$\begin{aligned} \mathbb{P}\Big(\sup_{m\in\mathcal{F}} \|Z_n(m)\|_{\mathcal{A}}^2 \geq x + \omega(\mathcal{F},1)^2\Big) &\leq \mathbb{P}\Big(\sup_{m\in\mathcal{F}} \|Z_n(m)\|_{\mathcal{A}} \geq (\sqrt{x} + \omega(\mathcal{F},1))/2\Big) \\ &\leq C\exp\big(-x/C\big). \end{aligned}$$

Hence it follows that

$$\begin{aligned}\mathbb{E}\big[\mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2 \,|\, \mathbb{U}]I_{\mathcal{E}}\big] &\le 2c_r^2h^{-2}n^{-1}r_{n,s}^2\Big(\omega(\mathcal{F},1)^2 + \int_0^\infty C\exp(-x/C)dx\Big)\\ &= 2c_r^2h^{-2}n^{-1}r_{n,s}^2\big(\omega(\mathcal{F},1)^2 + C^2\big). \qquad \text{(E.2)}\end{aligned}$$

We now turn to control the second term in (E.1). By Lemma G.2, we get that $\mathbb{E}\big[\|\Delta f^{(j)}\|_{\mathcal{H}}^2 \,|\, \mathbb{U}\big] \le 2\sigma^2/\lambda + 4\|f_0\|_{\mathcal{H}}^2$. Also by first order optimality condition with respect to $\widehat{\boldsymbol{\beta}}$,

$$\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(f_0(\mathbb{Z}) - \widehat{f}^{(j)}(\mathbb{Z}) + \boldsymbol{\varepsilon}^{(j)}),$$

where we omitted the superscript of (j) for the designs $\mathbb{X}^{(j)}$ and $\mathbb{Z}^{(j)}$. Hence

$$\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_2^2 \le 2\|(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\boldsymbol{\varepsilon}^{(j)}\|_2^2 + 2\|(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(f_0(\mathbb{Z}) - \widehat{f}^{(j)}(\mathbb{Z}))\|_2^2,$$

Taking conditional expectation yields

$$\begin{aligned}\mathbb{E}[\|\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)}\|_2^2 \,|\, \mathbb{U}] \le{}& 2\mathbb{E}\big[\|(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\boldsymbol{\varepsilon}^{(j)}\|_2^2 \,|\, \mathbb{U}\big]\\ &+ 2\mathbb{E}\big[\|(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(f_0(\mathbb{Z}) - \widehat{f}^{(j)}(\mathbb{Z}))\|_2^2 \,|\, \mathbb{U}\big]. \qquad \text{(E.3)}\end{aligned}$$

Denote $\widehat{\Sigma}^{(j)} = \mathbb{X}^T\mathbb{X}$, we first control the first term (E.3). Note that

$$(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\boldsymbol{\varepsilon}^{(j)} = \frac{1}{n}\sum_{i\in L_j}(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\varepsilon_i.$$

Taking conditional expectation and by independence, we have

$$\begin{aligned}\mathbb{E}\big[\|(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T\boldsymbol{\varepsilon}^{(j)}\|_2^2 \,|\, \mathbb{U}\big] &= \mathbb{E}\big[\|\frac{1}{n}\sum_{i\in L_j}(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\varepsilon_i\|_2^2 \,|\, \mathbb{U}\big]\\ &= \frac{1}{n^2}\sum_{i\in L_j}\mathbb{E}\Big[(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\varepsilon_i\|_2^2 \,|\, \mathbb{U}\Big]\\ &= \frac{1}{n^2}\sum_{i\in L_j}\sigma^2\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2. \qquad \text{(E.4)}\end{aligned}$$

For the second term in (E.3), we have similar to above that

$$(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(f_0(\mathbb{Z}) - \widehat{f}^{(j)}(\mathbb{Z})) = \frac{1}{n}\sum_{i\in L_j}(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i(f_0(Z_i) - \widehat{f}^{(j)}(Z_i)).$$

Taking conditional expectation, we have

$$\begin{aligned}\mathbb{E}\big[\|(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(f_0(\mathbb{Z}) - \widehat{f}^{(j)}(\mathbb{Z}))\|_2^2 \,|\, \mathbb{U}\big] \\ &= \mathbb{E}\big[\|\frac{1}{n}\sum_{i\in L_j}(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i(f_0(Z_i) - \widehat{f}^{(j)}(Z_i))\|_2^2 \,|\, \mathbb{U}\big] \\ &\le \frac{1}{n}\sum_{i\in L_j}\mathbb{E}\big[\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i(f_0(Z_i) - \widehat{f}^{(j)}(Z_i))\|_2^2 \,|\, \mathbb{U}\big],\end{aligned}$$

where the inequality is by $(\sum_{i=1}^n a_i)^2 \le n\sum_{i=1}^n a_i^2$. Hence we have

$$\begin{aligned}\mathbb{E}\big[\|(\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T(f_0(\mathbb{Z}) - \widehat{f}^{(j)}(\mathbb{Z}))\|_2^2 \,|\, \mathbb{U}\big] \\ &\le \frac{1}{n}\sum_{i\in L_j}\mathbb{E}\big[(f_0(Z_i) - \widehat{f}^{(j)}(Z_i))^2 \,|\, \mathbb{U}\big]\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2 \\ &\le \frac{1}{n}\sum_{i\in L_j}\mathbb{E}\big[\|f_0 - \widehat{f}^{(j)}\|_{\sup}^2 \,|\, \mathbb{U}\big]\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2 \\ &\le \frac{1}{n}\sum_{i\in L_j}c_k\mathbb{E}\big[\|f_0 - \widehat{f}^{(j)}\|_{\mathcal{H}}^2 \,|\, \mathbb{U}\big]\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2 \\ &\le \frac{1}{n}\sum_{i\in L_j}c_k(2\sigma^2/\lambda + 4\|f_0\|_{\mathcal{H}}^2)\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2, \qquad \text{(E.5)}\end{aligned}$$

where $c_k = \sup_z K(z,z)$. The second last inequality follows from the fact that $\|f\|_{\sup} \le \sup_z \|K_z\|_{\mathcal{H}}\|f\|_{\mathcal{H}} = c_k^{1/2}\|f\|_{\mathcal{H}}$ and the last inequality is by Lemma G.2. Combing (E.4) and (E.5), we have by (E.3) that

$$\mathbb{E}[\|\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)}\|_2^2 \,|\, \mathbb{U}] \le C\lambda^{-1}\frac{1}{n}\sum_{i\in L_j}\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2$$

holds almost surely for some constant C As we have

$\|\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)}\|_{L_2(\mathbb{P}_X)}^2 = (\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)})^T\Sigma(\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)}) \le \|\Sigma^{-1/2}\|\|\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)}\|_2^2$, it follows that

$$\mathbb{E}[\|\widehat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}_0^{(j)}\|_{L_2(\mathbb{P}_X)}^2 \,|\, \mathbb{U}] \le C\lambda^{-1}\frac{1}{n}\sum_{i\in L_j}\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2, \qquad \text{(E.6)}$$

for a constant C that is different from above. Lastly, we have $\|\widehat{f}^{(j)} - f_0\|_{L_2(\mathbb{P}_Z)} \le \|\widehat{f}^{(j)} - f_0\|_{\sup} \le c_k^{1/2}\|\widehat{f}^{(j)} - f_0\|_{\mathcal{H}}$, and it follows that

$$\mathbb{E}\big[\|\widehat{f}^{(j)} - f_0\|_{L_2(\mathbb{P}_Z)}^2 \,|\, \mathbb{U}\big] \lesssim \lambda^{-1}. \qquad \text{(E.7)}$$

Note that for any $m = (\boldsymbol{\beta}, f)$, $\|m\|_{\mathcal{A}}^2 = \|\boldsymbol{X}^T\boldsymbol{\beta} + f(Z)\|_{L_2(\mathbb{P}_U)}^2 + \lambda\|f\|_{\mathcal{H}}^2 \le$

$2\|\boldsymbol{\beta}\|_{L_2(\mathbb{P}_X)}^2+2\|f\|_{L_2(\mathbb{P}_Z)}^2+\lambda\|f\|_{\mathcal{H}}^2$. Hence by (E.6), (E.7) and (G.1), we have

$$\mathbb{E}\big[\|\Delta m^{(j)}\|_{\mathcal{A}}^2\,|\,\mathbb{U}\big]\le C\lambda^{-1}n^{-1}\sum_{i\in L_j}\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2.$$

Moreover,

$$\begin{aligned}\|Rem^{(j)}\|_{\mathcal{A}}&\le\|\frac{1}{n}\sum_{i\in L_j}\Delta m^{(j)}(U_i)R_{U_i}\|_{\mathcal{A}}+\|\mathbb{E}_U[\Delta m^{(j)}(U)R_U]\|\\&=\frac{1}{n}\sum_{i\in L_j}\|\Delta m^{(j)}(U_i)R_{U_i}\|_{\mathcal{A}}+\mathbb{E}_U\big[\|\Delta m^{(j)}(U)R_U\|\big]\\&\le 2h^{-1}\|\Delta m^{(j)}\|_{\mathcal{A}},\end{aligned}$$

where in the last inequality we used $|\Delta m^{(j)}(U)|\le h^{-1/2}\|\Delta m^{(j)}\|_{\mathcal{A}}$ and $\|R_U\|_{\mathcal{A}}\le h^{-1/2}$. Hence

$$\mathbb{E}\big[\|Rem^{(j)}\|_{\mathcal{A}}^2\,|\,\mathbb{U}\big]\le 4h^{-2}\mathbb{E}\big[\|\Delta m^{(j)}\|_{\mathcal{A}}^2\,|\,\mathbb{U}\big]\le Ch^{-2}\lambda^{-1}n^{-1}\sum_{i\in L_j}\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2.$$

Hence, we have that the second term in (E.1)

$$\begin{aligned}\mathbb{E}\big[\mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2\,|\,\mathbb{U}]I_{\mathcal{E}^c}\big]&\le Ch^{-2}\lambda^{-1}n^{-1}\sum_{i\in L_j}\mathbb{E}\big[\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^2I_{\mathcal{E}^c}\big]\\&\le Ch^{-2}\lambda^{-1}n^{-1}\mathbb{P}(\mathcal{E}^c)\sum_{i\in L_j}\mathbb{E}\big[\|(\widehat{\Sigma}^{(j)})^{-1}\boldsymbol{X}_i\|_2^4\big]\\&\le C'h^{-2}\lambda^{-1}\mathbb{P}(\mathcal{E}^c),\end{aligned}\tag{E.8}$$

where the second last inequality is by Holder's inequality and the last one by assumption on the design. By (E.8) and Lemma 7.4, we obtain

$$\mathbb{E}\big[\mathbb{E}[\|Rem^{(j)}\|_{\mathcal{A}}^2\,|\,\mathbb{U}]I_{\mathcal{E}^c}\big]\lesssim h^{-2}\lambda^{-1}n\exp(-c\log^2N).\tag{E.9}$$

Finally, plugging (E.2) and (E.9) into (E.1), we have for sufficiently large n,

$$\mathbb{E}\big[\|Rem^{(j)}\|_{\mathcal{A}}^2\big]\le 2c_r^2h^{-2}n^{-1}r_{n,s}^2\big(\omega(\mathcal{F},1)^2+C\big)+C'h^{-2}\lambda^{-1}n\exp(-c\log^2N),\tag{E.10}$$

as desired.

We can apply similar arguments as above to bound $\|Rem_f^{(j)}\|_{\mathcal{C}}$ and $\|1/s\sum_{j=1}^s Rem_f^{(j)}\|_{\mathcal{C}}$, by changing $\omega(\mathcal{F},1)$ to $\omega(\mathcal{F}_2,1)$, which is dominated by $\omega(\mathcal{F},1)$. The bounds of $\|Rem_\beta^{(j)}\|_2$ and $\|1/s\sum_{j=1}^s Rem_\beta^{(j)}\|_2$ then follow from triangular inequality. □

E.2. Proof of Lemma 7.2.

PROOF. The main term (I) can be rearranged as follows:

$$(I) = \underbrace{\frac{1}{\sqrt{n}}\sum_{i\in L_j}\varepsilon_i\Big(\boldsymbol{x}^T L_{U_i} + s^{-1/2}h^{1/2}N_{U_i}(z_0)\Big)}_{(III)} + \underbrace{\frac{1}{\sqrt{N}}\sum_{i\notin L_j} h^{1/2}N_{U_i}(z_0)\varepsilon_i}_{(IV)}$$

When analyzing (I), we consider two cases: (1) $s\to\infty$ and (2) s is fixed.

Case 1: $s\to\infty$. We first apply CLT to the first component of term (III), i.e., $\frac{1}{\sqrt{n}}\sum_{i\in L_j}\varepsilon_i\boldsymbol{x}^T L_{U_i}$. The summands are i.i.d. with mean zero. Moreover,

$$\mathbb{E}\big[(\varepsilon\boldsymbol{x}^T L_U)^2\big] = \sigma^2\boldsymbol{x}^T\mathbb{E}[L_U L_U^T]\boldsymbol{x}.$$

By Proposition 2.3,

$$\mathbb{E}[L_U L_U^T] \;=\; (\Omega+\Sigma_\lambda)^{-1}\mathbb{E}[(\boldsymbol{X}-\boldsymbol{A}(Z))(\boldsymbol{X}-\boldsymbol{A}(Z))^T](\Omega+\Sigma_\lambda)^{-1}$$

By Lemma A.3 in Section A.1, we have that $\Sigma_\lambda = \mathbb{E}_Z\big[\boldsymbol{B}(Z)\big(\boldsymbol{B}(Z)-\boldsymbol{A}(Z)\big)^T\big]\to 0$, and also

$$\begin{aligned}
&\mathbb{E}\big[(\boldsymbol{X}-\boldsymbol{A}(Z))(\boldsymbol{X}-\boldsymbol{A}(Z))^T\big]\\
&\;=\; \mathbb{E}[(\boldsymbol{X}-\boldsymbol{B}(Z))(\boldsymbol{X}-\boldsymbol{B}(Z))^T] + \mathbb{E}[(\boldsymbol{B}(Z)-\boldsymbol{A}(Z))(\boldsymbol{B}(Z)-\boldsymbol{A}(Z))^T]\\
&\qquad +2\mathbb{E}[(\boldsymbol{X}-\boldsymbol{A}(Z))(\boldsymbol{B}(Z)-\boldsymbol{A}(Z))^T]\to\Omega^{-1}
\end{aligned}$$

This implies $\mathbb{E}[L_U L_U^T]\to\Omega^{-1}$. Therefore by CLT, we have

$$\frac{1}{\sqrt{n}}\sum_{i\in L_j}\varepsilon_i\boldsymbol{x}^T L_{U_i} \rightsquigarrow N(0,\sigma^2\boldsymbol{x}^T\Omega^{-1}\boldsymbol{x}). \tag{E.11}$$

We next consider $\frac{1}{\sqrt{n}}\sum_{i\in L_j}s^{-1/2}h^{1/2}N_{U_i}(z_0)\varepsilon_i$ which is the second component in (III). Again the summands are i.i.d. with mean zero. By Proposition 2.3 we have

$$\begin{aligned}
&\mathbb{E}\big[(h^{1/2}\varepsilon N_U(z_0))^2\big] \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (\text{E.12})\\
&\;=\; \sigma^2 h\mathbb{E}\big[(\widetilde{K}_Z(z_0)-L_U^T\boldsymbol{A}(z_0))^2\big]\\
&\;=\; \sigma^2 h\mathbb{E}\big[\widetilde{K}_Z(z_0)^2\big] + \sigma^2 h\boldsymbol{A}(z_0)^T\mathbb{E}[L_U L_U^T]\boldsymbol{A}(z_0)\\
&\qquad -2\sigma^2 h\mathbb{E}[\widetilde{K}_Z(z_0)L_U^T\boldsymbol{A}(z_0)].
\end{aligned}$$

For the first term in (E.12), by condition in the lemma, we have

$$\sigma^2 h\mathbb{E}\big[\widetilde{K}_Z(z_0)^2\big] = h\|\widetilde{K}_{z_0}\|^2_{L_2(\mathbb{P}_Z)}\to\sigma^2\sigma^2_{z_0}$$

For the second term in (E.12), as $h^{1/2}\boldsymbol{A}(z_0)\to-\gamma_{z_0}$, and $\mathbb{E}[L_U L_U^T]\to\Omega^{-1}$, we have

$$\sigma^2 h\boldsymbol{A}(z_0)^T\mathbb{E}[L_U L_U^T]\boldsymbol{A}(z_0)\to\sigma^2\gamma_{z_0}^T\Omega^{-1}\gamma_{z_0}.$$

For the last term in (E.12), we consider

$$h^{1/2}\mathbb{E}\big[\widetilde{K}_Z(z_0)L_U\big] = h^{1/2}(\Omega+\Sigma_\lambda)^{-1}\mathbb{E}[\widetilde{K}_Z(z_0)(\boldsymbol{X}-\boldsymbol{A}(Z))].$$

We have $(\Omega+\Sigma_\lambda)^{-1}\to\Omega^{-1}$ and

$$\begin{aligned}
\mathbb{E}[\widetilde{K}_Z(z_0)(\boldsymbol{X}-\boldsymbol{A}(Z))] &= h^{1/2}(\langle \boldsymbol{B}, \widetilde{K}_{z_0}\rangle_{L_2(\mathbb{P}_Z)} - \langle \boldsymbol{A}, \widetilde{K}_{z_0}\rangle_{L_2(\mathbb{P}_Z)})\\
&= h^{1/2}(\langle \boldsymbol{A}, \widetilde{K}_{z_0}\rangle_{\mathcal{C}} - \langle \boldsymbol{A}, \widetilde{K}_{z_0}\rangle_{L_2(\mathbb{P}_Z)})\\
&= h^{1/2}\lambda\langle \boldsymbol{A}, \widetilde{K}_{z_0}\rangle_{\mathcal{H}}\\
&= h^{1/2}\langle W_\lambda\boldsymbol{A}, \widetilde{K}_{z_0}\rangle_{\mathcal{C}}\\
&= h^{1/2}W_\lambda\boldsymbol{A}(z_0)\to\alpha_{z_0}
\end{aligned}$$

Hence $h^{1/2}\mathbb{E}\big[\widetilde{K}_Z(z_0)L_U\big]\to\Omega^{-1}\alpha_{z_0}$ and so $h\mathbb{E}[\widetilde{K}_Z(z_0)L_U^T\boldsymbol{A}(z_0)]\to\gamma_{z_0}^T\Omega^{-1}\alpha_{z_0}$. In summary, we have

$$\mathbb{E}\big[(h^{1/2}\varepsilon N_U(z_0))^2\big]\to\sigma^2\big(\sigma_{z_0}^2+\gamma_{z_0}^T\Omega^{-1}\gamma_{z_0}+2\gamma_{z_0}^T\Omega^{-1}\alpha_{z_0}\big)=\Sigma_{22}.$$

By central limit theorem, it follows that

$$\frac{1}{\sqrt{n}}\sum_{i\in L_j}\varepsilon_i h^{1/2}N_{U_i}(z_0)\rightsquigarrow N(0,\sigma^2\Sigma_{22}). \tag{E.13}$$

As $s\to\infty$, we have $\frac{1}{\sqrt{n}}\sum_{i=1}^n\varepsilon_i s^{-1/2}h^{1/2}N_{U_i}(z_0)\to 0$. So the second component in (III) is asymptotically ignorable. Therefore by (E.11), we obtain

$$(III)\rightsquigarrow N\left(0,\sigma^2\boldsymbol{x}^T\Omega^{-1}\boldsymbol{x}\right).$$

As for (IV), we apply similar arguments as in the previous paragraph and consider $s\to\infty$. It follows that

$$(IV)=\sqrt{1-s^{-1}}\Big\{\frac{1}{\sqrt{N-n}}\sum_{i\notin L_j}h^{1/2}N_{U_i}(z_0)\varepsilon_i\Big\}\rightsquigarrow N(0,\sigma^2\Sigma_{22}).$$

Lastly, note that (III) and (IV) are independent, so are their limits. Therefore, it follows that

$$(I)\rightsquigarrow N(0,\sigma^2(\boldsymbol{x}^T\Omega^{-1}\boldsymbol{x}+\Sigma_{22})).$$

Case 2: s fixed. Instead of decomposing (III) into two components as in previous case, we apply CLT to term (III) as a whole. Note that the summands in (III) are i.i.d. with mean zero. Moreover,

$$\begin{aligned}
&\mathbb{E}\Big[\varepsilon^2\big(\boldsymbol{x}^T L_U+s^{-1/2}h^{1/2}N_U(z_0)\big)^2\Big]\\
&=\sigma^2\mathbb{E}\big[(\boldsymbol{x}^T L_U)^2\big]+s^{-1}\sigma^2\mathbb{E}\big[(h^{1/2}N_U(z_0))^2\big]\\
&\qquad\qquad+2s^{-1/2}\sigma^2\mathbb{E}\big[h^{1/2}\boldsymbol{x}^T L_U N_U(z_0)\big]
\end{aligned}$$

The first two terms are considered in Case 1. For the third term, we have

$$\mathbb{E}\big[\boldsymbol{x}^T L_U h^{1/2} N_U(z_0)\big] = \mathbb{E}\big[h^{1/2}\boldsymbol{x}^T L_U\big(\widetilde{K}_Z(z_0) - L_U^T \boldsymbol{A}(z_0)\big)\big]$$

From Case 1, we have $h^{1/2}\mathbb{E}\big[\widetilde{K}_Z(z_0)L_U\big] \to \alpha_{z_0}$, $\mathbb{E}[L_U L_U^T] \to \Omega^{-1}$, and $h^{1/2}\boldsymbol{A}(z_0) \to -\gamma_{z_0}$. It follows that

$$\mathbb{E}\big[h^{1/2}\boldsymbol{x}^T L_U N_U(z_0)\big] \to \boldsymbol{x}^T\Omega^{-1}(\alpha_{z_0} + \gamma_{z_0}). \tag{E.14}$$

Therefore, we have

$$\mathbb{E}\Big[\varepsilon^2\big(\boldsymbol{x}^T L_U + s^{-1/2}h^{1/2}N_U(z_0)\big)^2\Big] \to \sigma^2(\boldsymbol{x}^T\Omega^{-1}\boldsymbol{x} + s^{-1}\Sigma_{22} + 2s^{-1/2}\boldsymbol{x}^T\Sigma_{12}).$$

Hence by central limit theorem, we have

$$(III) \rightsquigarrow N(0, \sigma^2(\boldsymbol{x}^T\Omega^{-1}\boldsymbol{x} + s^{-1}\Sigma_{22} + 2s^{-1/2}\boldsymbol{x}^T\Sigma_{12})).$$

Similarly, we have

$$(IV) \rightsquigarrow N\big(0, (1 - s^{-1})\sigma^2\Sigma_{22}\big).$$

As (III) and (IV) are independent, so are their limits. Therefore in the case that s is fixed, we have

$$(I) \rightsquigarrow N(0, \sigma^2(\boldsymbol{x}^T\Omega^{-1}\boldsymbol{x} + \Sigma_{22} + 2s^{-1/2}\boldsymbol{x}^T\Sigma_{12})).$$

This finishes the proof. □

E.3. Proof of Lemma 7.4.

PROOF. Recall that $\Delta m^{(j)} = \widehat{m}^{(j)} - m_0^{(j)}$. As $\Delta m^{(j)}$ minimizes the objective function (3.2), we have

$$\frac{1}{n}\sum_{i\in L_j}(\widehat{m}^{(j)}(U_i) - Y_i)^2 + \lambda\|\widehat{f}\|_{\mathcal{H}}^2 \le \frac{1}{n}\sum_{i\in L_j}(m_0(U_i) - Y_i)^2 + \lambda\|f_0\|_{\mathcal{H}}^2,$$

On the j-th sub-population, we have $Y_i = m_0^{(j)}(U_i) + \varepsilon_i$, hence it follows that

$$\frac{1}{n}\sum_{i\in L_j}(\widehat{m}^{(j)}(U_i) - m_0^{(j)}(U_i))^2 + \frac{2}{n}\sum_{i\in L_j}\varepsilon_i(\widehat{m}^{(j)}(U_i) - m_0^{(j)}(U_i)) + \lambda\|\widehat{f}^{(j)}\|_{\mathcal{H}}^2 \le \lambda\|f_0\|_{\mathcal{H}}^2.$$

Adding and subtracting $\mathbb{E}_U[\Delta m^{(j)}(U)^2]$, we transform the above inequality to

$$\begin{aligned}\frac{1}{n}\sum_{i\in L_j}\Delta m^{(j)}(U_i)^2 - \mathbb{E}_U[\Delta m^{(j)}(U)^2] + \mathbb{E}_U[\Delta m^{(j)}(U)^2] + \lambda\|\Delta f^{(j)}\|_{\mathcal{H}}^2 \\ + \frac{2}{n}\sum_{i\in L_j}\varepsilon_i\Delta m^{(j)}(U_i) - 2\lambda\|f_0\|_{\mathcal{H}}^2 + 2\lambda\langle\widehat{f}^{(j)}, f_0\rangle_{\mathcal{H}} \le 0.\end{aligned}$$

As we have $\mathbb{E}_U[\Delta m^{(j)}(U)^2] + \lambda\|\Delta f^{(j)}\|_{\mathcal{H}}^2 = \|\Delta m^{(j)}\|_{L_2(\mathbb{P}_U)}^2 + \lambda\|\Delta f^{(j)}\|_{\mathcal{H}}^2 =$

$\|\Delta m^{(j)}\|_{\mathcal{A}}^2$. It follows that

$$\begin{aligned}
\|\Delta m^{(j)}\|_{\mathcal{A}}^2 &\leq -2\Big(\frac{1}{n}\sum_{i\in L_j}\varepsilon_i \Delta m^{(j)}(U_i) - \lambda\langle \Delta f^{(j)}, f_0\rangle_{\mathcal{H}}\Big) \\
&\quad -\frac{1}{n}\sum_{i\in L_j}\langle \Delta m^{(j)}(U_i)R_{U_i} - \mathbb{E}_U[\Delta m^{(j)}(U)R_U], \Delta m^{(j)}\rangle_{\mathcal{A}} \\
&= -2\langle \frac{1}{n}\sum_{i\in L_j}\varepsilon_i R_{U_i} - P_\lambda m_0^{(j)}, \Delta m^{(j)}\rangle_{\mathcal{A}} - \langle Rem^{(j)}, \Delta m^{(j)}\rangle_{\mathcal{A}}. \qquad \text{(E.15)}
\end{aligned}$$

Define the following two events:

$$\mathcal{B}_1 := \big\{\|\frac{1}{n}\sum_{i\in L_j}\varepsilon_i R_{U_i}\| \leq C\log^2 N (nh)^{-1/2}\big\},$$

$$\mathcal{B}_2 := \big\{\|Rem^{(j)}\|_{\mathcal{A}} \leq 2c_r h^{-1}n^{-1/2}\big(C\omega(1) + \log N\big)\|\Delta m^{(j)}\|_{\mathcal{A}}\big\}.$$

We bound the two terms in (E.15) respectively. First, note that

$$\begin{aligned}
\|P_\lambda m_0^{(j)}\| &= \sup_{\|m\|_{\mathcal{A}}=1}|\langle P_\lambda m_0^{(j)}, m\rangle_{\mathcal{A}}| = \sup_{\|m\|_{\mathcal{A}}=1}\lambda|\langle f_0, f\rangle_{\mathcal{H}}| \\
&\leq \sup_{||m||=1}\sqrt{\lambda\|f_0\|_{\mathcal{H}}^2}\sqrt{\lambda\|f\|_{\mathcal{H}}^2} \leq \lambda^{1/2}\|f_0\|_{\mathcal{H}},
\end{aligned}$$

where the last inequality follows from the fact that $\lambda\|f\|_{\mathcal{H}}^2 \leq \|m\|_{\mathcal{A}}^2 = 1$. Therefore on event $\mathcal{B}_1$, the first term in (E.15) can be bounded by

$$\begin{aligned}
\Big|\langle\frac{1}{n}\sum_{i\in L_j}\varepsilon_i R_{U_i} - P_\lambda f_0, \Delta m^{(j)}\rangle_{\mathcal{A}}\Big| &\leq \big\|\frac{1}{n}\sum_{i\in L_j}\varepsilon_i R_{U_i} - P_\lambda f_0\big\|_{\mathcal{A}}\|\Delta m^{(j)}\|_{\mathcal{A}} \\
&\leq C\big(\log^2 N (nh)^{-1/2} + \lambda^{1/2}\big)\|\Delta m^{(j)}\|_{\mathcal{A}}. \qquad \text{(E.16)}
\end{aligned}$$

Furthermore, on the event $\mathcal{B}_2$, the second term in (E.15) can be bounded by

$$\begin{aligned}
\Big|\frac{1}{n}\sum_{i\in L_j}\langle Rem^{(j)}, \Delta m^{(j)}\rangle_{\mathcal{A}}\Big| &\leq \|\Delta m^{(j)}\|_{\mathcal{A}}\|Rem^{(j)}\|_{\mathcal{A}} \\
&\leq 2c_r h^{-1}n^{-1/2}\big(C\omega(\mathcal{F},1) + \log N\big)\|\Delta m^{(j)}\|_{\mathcal{A}}^2. \qquad \text{(E.17)}
\end{aligned}$$

Therefore, by (E.15), (E.16) and (E.17), it yields that on the event $\mathcal{B}_1 \cap \mathcal{B}_2$ there exists a constant C,

$$\begin{aligned}
\|\Delta m^{(j)}\|_{\mathcal{A}}^2 \leq C'\big((nh)^{-1/2}\log^2 N + \lambda^{1/2}\big)\|\Delta m^{(j)}\|_{\mathcal{A}} \\
+ 2c_r h^{-1}n^{-1/2}\big(C\omega(\mathcal{F},1) + \log N\big)\|\Delta m^{(j)}\|_{\mathcal{A}}^2.
\end{aligned}$$

If Condition (3.6) is satisfied, it implies that

$$h^{-1}n^{-1/2}\big(C\omega(1) + \log N\big) \leq Cs^{1/2}N^{-1/2}h^{-1}\big(\omega(1) + \log N\big) = o(1).$$

Therefore it follows that

$$\|\Delta m^{(j)}\|_{\mathcal{A}} \le C_1\big((nh)^{-1/2}\log^2 N + \lambda^{1/2}\big) + o(1)\|\Delta m^{(j)}\|_{\mathcal{A}}$$

which implies that for sufficiently large n,

$$\|\Delta m^{(j)}\|_{\mathcal{A}} \le C(\log^2 N(nh)^{-1/2} + \lambda^{1/2}).$$

Now we are left to bound the probability of $\mathcal{B}_1^c \cup \mathcal{B}_2^c$. For $\mathcal{B}_1$, define $\mathcal{Q}_i = \{|\varepsilon_i| \le \log N\}$. Since $\|R_U\| \le c_r h^{-1/2}$, we have that on the event of $\cap_{i\in L_j}\mathcal{Q}_i$, $\big\{\varepsilon_i R_{U_i}\big\}_{i\in L_j}$ is a sequence of random variables in Hilbert space $\mathcal{A}$ that are independent with mean zero and bounded by $ch^{-1/2}\log N$. Therefore we have

$$\begin{aligned}
\mathbb{P}(\mathcal{B}_1^c) &= \mathbb{P}\Big(\|\frac{1}{n}\sum_{i\in L_j}\varepsilon_i R_{U_i}\|_{\mathcal{A}} > C\log^2 N(nh)^{-1/2}\Big)\\
&\le \mathbb{P}\Big(\cap_{i\in L_j}\mathcal{Q}_i, \|\frac{1}{n}\sum_{i\in L_j}\varepsilon_i R_{U_i}\|_{\mathcal{A}} > C\log^2 N(nh)^{-1/2}\Big) + \mathbb{P}\big((\cap_i\mathcal{Q}_i)^c\big)\\
&\le 2\exp(-\log^2 N) + 2n\exp(-\log^2 N), \qquad \text{(E.18)}
\end{aligned}$$

where the first term in the last inequality is by Lemma G.1, and the second term is by union bound and the fact that ε_i are i.i.d. sub-Gaussian.

Now we turn to $\mathcal{B}_2$. Define $\widetilde{m} := (2c_r)^{-1}h^{1/2}\frac{\Delta m^{(j)}}{\|\Delta m^{(j)}\|_{\mathcal{A}}}$. Then it follows that

$$\|\widetilde{m}\|_{\sup} \le c_r h^{-1/2}\|\widetilde{m}\|_{\mathcal{A}} \le 1/2.$$

By the same argument as in Section 7.3, it follows that $\|\Delta f^{(j)}\|_{\sup} \le 1/2$ and $|\boldsymbol{x}^T\Delta\boldsymbol{\beta}| \le 1$ for all $\boldsymbol{x}$. Moreover, we have

$$\|\widetilde{f}\|_{\mathcal{H}} \le \lambda^{-1/2}\|\widetilde{m}\|_{\mathcal{A}} \le (2c_r)^{-1}h^{1/2}\lambda^{-1/2}.$$

Hence we proved that $\widetilde{m} \in \mathcal{F}$. By Lemma F.1, it follows that

$$\mathbb{P}\bigg(\|Z_n(\widetilde{m})\|_{\mathcal{A}} \ge C\omega(1, \mathrm{diam}(\mathcal{F})) + \log N\bigg) \le C\exp\Big(-\log^2 N/C\Big).$$

By definition of $Z_n(m)$ and $\widetilde{m}$, we have $Z_n(\widetilde{m}) = (2c_r)^{-1}hn^{1/2}\|\Delta m^{(j)}\|_{\mathcal{A}}^{-1}Rem^{(j)}$. Hence it follows that

$$\begin{aligned}
\mathbb{P}(\mathcal{B}_2^c) &= \mathbb{P}\Big(\|Rem^{(j)}\|_{\mathcal{A}} \ge 2c_r h^{-1}n^{-1/2}\big(C\omega(1) + \log N\big)\|\Delta m^{(j)}\|_{\mathcal{A}}\Big)\\
&\le C\exp\Big(-\log^2 n/C\Big). \qquad \text{(E.19)}
\end{aligned}$$

Combining (E.18) and (E.19), we have that for some universal constants c, C

and sufficiently large n,

$$\begin{aligned}\mathbb{P}\Big(\|\Delta m^{(j)}\|_{\mathcal{A}} \geq C\big((nh)^{-1/2}\log^2 N + \lambda^{1/2}\big)\Big) &\leq \mathbb{P}(\mathcal{B}_1^c) + \mathbb{P}(\mathcal{B}_2^c) \\ &\lesssim n\exp(-c\log^2 N). \qquad \text{(E.20)}\end{aligned}$$

This finishes the proof. □

APPENDIX F: EMPIRICAL PROCESS LEMMA

Lemma F.1 (Local chaining inequality)**.** If $Z_n(m) \in \mathcal{A}$ is a separable process on the metric space $(\mathcal{F}, \|\cdot\|_{\sup})$ and satisfies (7.15):

$$\mathbb{P}\Big(\|Z_n(m_1) - Z_n(m_2)\|_{\mathcal{A}} \geq t\Big) \leq 2\exp\Big(-\frac{t^2}{8\|m_1 - m_2\|_{\sup}^2}\Big)$$

Then for all $m_0 \in \mathcal{F}$ and $x \geq 0$, we have

$$\mathbb{P}\Big(\sup_{m\in\mathcal{F}}\|Z_n(m) - Z_n(m_0)\|_{\mathcal{A}} \geq C\omega(\mathcal{F}, \mathrm{diam}(\mathcal{F})) + x\Big) \leq C\exp\big(-x^2/C\mathrm{diam}(\mathcal{F})^2\big), \qquad \text{(F.1)}$$

where C is a generic constant.

Proof. The proof follows by modifying the proof of Theorem 5.28 in van Handel (2014). Let k_0 be the largest integer such that $2^{-k_0} \geq \mathrm{diam}(\mathcal{F})$. Then $N(\mathcal{F}, d, 2^{-k}) = 1$ for all $k \leq k_0$. We employ a chaining argument, and start at the scale 2^{-k_0}. For every $k > k_0$, let N_k be a 2^{-k} net such that $|N_k| = N(T, d, 2^{-k})$. We define the singleton $N_{k_0} = \{m_0\}$. We claim that

$$\begin{aligned}Z_n(m) &= \lim_{k\to\infty} Z_n(\pi_k(m)) - Z_n(m_0) \\ &= \sum_{k>k_0}\big\{Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\big\} \quad \text{a.s.}\end{aligned}$$

where $\pi_k(m)$ is the closes point in N_k to m. To prove this identity, note that the sub-Gaussian property of $\{Z_n(m)\}_{m\in\mathcal{F}}$ implies that $Z_n(m) - Z_n(\pi_k(m))$ is $d(m, \pi_k(m))$-sub-Gaussian. Thus

$$\sum_{k=k_0}^{\infty}\mathbb{E}\big[\|Z_n(m) - Z_n(\pi_k(m))\|_{\mathcal{A}}^2\big] \leq \sum_{k=k_0}^{\infty} d(m, \pi_k(m))^2 \leq \sum_{k=k_0}^{\infty} 2^{-k} < \infty.$$

It follows that $\|Z_n(m) - Z_n(\pi_k(m))\|_{\mathcal{A}} \to 0$ a.s. as $k \to \infty$, and the chaining identity follows readily using the telescoping property of the sum. By the chaining identity and separability of $\mathcal{F}$, we obtain

$$\sup_{m\in\mathcal{F}}\|Z_n(m) - Z_n(m_0)\|_{\mathcal{A}} \leq \sum_{k>k_0}\sup_{m\in\mathcal{F}}\|Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\|_{\mathcal{A}}.$$

By union bound and sub-Gaussian property, it follows that

$$\mathbb{P}\Big(\sup_{m\in\mathcal{F}} \|Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\|_{\mathcal{A}} > t\Big) \le 2|N_k| \exp\Big(-\frac{t^2}{8\cdot 2^{-2k}}\Big)$$
$$= 2\exp\Big(\log|N_k| - \frac{t^2}{8\cdot 2^{-2k}}\Big).$$

For large enough t, let $u = \frac{t^2}{8\cdot 2^{-2k}} - \log|N_k|$, and we have

$$\mathbb{P}\Big(\sup_{m\in\mathcal{F}} \|Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\|_{\mathcal{A}} > 2\sqrt{2}\cdot 2^{-k}(\sqrt{\log|N_k|} + u)\Big) \le 2\exp^{u^2/2},$$

which implies the link $\|Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\|_{\mathcal{A}}$ at scale k is small. To show that the links at all scale of k are small simultaneously, we again use the union bound. Define the event $\mathcal{D} := \big\{\exists k > k_0 \text{ s.t.} \sup_{m\in\mathcal{F}} \|Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\|_{\mathcal{A}} > 2\sqrt{2}\cdot 2^{-k}(\sqrt{\log|N_k|} + u_k)\big\}$, where $u_k = x + \sqrt{k-k_0}$. Then

$$\begin{aligned}
&\mathbb{P}(\mathcal{D})\\
&\le \sum_{k>k_0} \mathbb{P}\Big(\sup_{m\in\mathcal{F}} \|Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\|_{\mathcal{A}} > 2\sqrt{2}\cdot 2^{-k}(\sqrt{\log|N_k|} + u_k)\Big)\\
&\le \sum_{k>k_0} \exp(-u_k^2/2) \le \exp(-x^2/2)\sum_{k>0}\exp(-k/2) \le C\exp(-x^2/2).
\end{aligned}$$

Moreover, by the fact that $2^{-k_0} \le 2\mathrm{diam}(\mathcal{F})$ and

$$2^{-k_0} \le C 2^{-k_0-1}\sqrt{\log N(\mathcal{F}, d, 2^{-k_0-1})} \le C\sum_{k>k_0}\sqrt{\log|N_k|},$$

we have on the event $\mathcal{D}^c$,

$$\begin{aligned}
&\sup_{m\in\mathcal{F}} \|Z_n(m) - Z_n(m_0)\|_{\mathcal{A}}\\
&\le \sum_{k>k_0}\sup_{m\in\mathcal{F}} \|Z_n(\pi_k(m)) - Z_n(\pi_{k-1}(m))\|_{\mathcal{A}}\\
&\le 2\sqrt{2}\sum_{k>k_0} 2^{-k}(\sqrt{\log|N_k|} + u_k)\\
&\le 2\sqrt{2}\sum_{k>k_0} 2^{-k}\sqrt{\log|N_k|} + 2\sqrt{2}\cdot 2^{-k_0}\sum_{k>k_0} 2^{-k}\sqrt{k} + 2\sqrt{2}\sum_{k>k_0} 2^{-k}x\\
&\le C\int_0^{\mathrm{diam}(\mathcal{F})}\sqrt{\log N(\mathcal{F}, d, \epsilon)}d\epsilon + C\mathrm{diam}(\mathcal{F})x\\
&= C\omega(\mathcal{F}, \mathrm{diam}(\mathcal{F})) + C\mathrm{diam}(\mathcal{F})x.
\end{aligned}$$

Therefore

$$\mathbb{P}\Big(\sup_{m\in\mathcal{F}}\|Z_n(m)-Z_n(m_0)\|_{\mathcal{A}}\geq C\omega(\mathcal{F},\mathrm{diam}(\mathcal{F}))+C\mathrm{diam}(\mathcal{F})x\Big)$$
$$\leq\mathbb{P}(\mathcal{D})\leq C\exp(-x^2/2).$$

Replacing $C\mathrm{diam}(\mathcal{F})x$ with a new variable x, we reach the conclusion of the lemma. □

APPENDIX G: AUXILIARY LEMMAS

Lemma G.1. (Pinelis, 1994) If $\Xi_1,\ldots,\Xi_s$ are zero mean independent random variables in a separable Hilbert space and $\|\Xi_i\|\leq M$ for $i=1,\ldots,n$, then

$$\mathbb{P}\Big(\Big\|\frac{1}{n}\sum_{i=1}^n\Xi_i\Big\|>t\Big)<2\exp\big(-\frac{nt^2}{2M^2}\big).$$

Lemma G.2. We have for all $j=1,\ldots,s$,

$$\mathbb{E}\big[\|\Delta f^{(j)}\|_{\mathcal{H}}^2\,|\,\mathbb{U}\big]\leq 2\sigma^2/\lambda+4\|f_0\|_{\mathcal{H}}^2. \tag{G.1}$$

Proof. By the zero order optimality condition, we have

$$\begin{aligned}\lambda\|\widehat{f}^{(j)}\|_{\mathcal{H}}^2&\leq\frac{1}{n}\sum_{i\in L_j}(Y_i-\widehat{m}^{(j)}(U_i))^2+\lambda\|\widehat{f}^{(j)}\|_{\mathcal{H}}^2\\&\leq\frac{1}{n}\sum_{i\in L_j}(Y_i-m_0^{(j)}(U_i))^2+\lambda\|f_0\|_{\mathcal{H}}^2\\&=\frac{1}{n}\sum_{i\in L_j}\varepsilon_i^2+\lambda\|f_0\|_{\mathcal{H}}^2.\end{aligned}$$

Hence taking expectation conditioned on $\mathbb{U}$, we get

$$\lambda\mathbb{E}[\|\widehat{f}^{(j)}\|_{\mathcal{H}}^2\,|\,\mathbb{U}]\leq\sigma^2+\lambda\|f_0\|_{\mathcal{H}}^2.$$

Then, applying triangular inequality along with the inequality $(a+b)^2\leq 2a^2+2b^2$, we have

$$\begin{aligned}\mathbb{E}[\|\Delta f^{(j)}\|_{\mathcal{H}}^2\,|\,\mathbb{U}]&\leq 2\|f_0\|_{\mathcal{H}}^2+2\mathbb{E}[\|\widehat{f}^{(j)}\|_{\mathcal{H}}^2\,|\,\mathbb{U}]\\&\leq\frac{2\sigma^2}{\lambda}+4\|f_0\|_{\mathcal{H}}^2,\end{aligned}$$

as desired. □

Lemma G.3 (Matrix Heoffding in Tropp (2012))**.** Consider a finite sequence $\{\mathbf{A}_i\}_{i=1}^n$ of independent, random, symmetric matrices with dimension p. Assume that each random matrix satisfies

$$\mathbb{E}[\mathbf{A}_i]=\mathbf{0}\quad\text{and}\quad\|\mathbf{A}_i^2\|\leq M\text{ almost surely.}$$

Then, for all $t > 0$,

$$\mathbb{P}\Big(\Big\|\frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_i\Big\| \geq t\Big) \leq p \exp\Big(-\frac{nt^2}{8M}\Big).$$

Lemma G.4. Let $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{k\times k}$ be given. If $\mathbf{A}$ is invertible, and $\|\mathbf{A}^{-1}\mathbf{E}\| < 1$, then $\widetilde{\mathbf{A}} := \mathbf{A} + \mathbf{E}$ is invertible, and

$$\|\widetilde{\mathbf{A}}^{-1} - \mathbf{A}^{-1}\| \leq \frac{\|\mathbf{E}\|\|\mathbf{A}^{-1}\|^2}{1 - \|\mathbf{A}^{-1}\mathbf{E}\|}$$

PROOF. See Theorem 2.5, p. 118 in Stewart and Sun (1990). □

DEPARTMENT OF OPERATIONS RESEARCH
AND FINANCIAL ENGINEERING
PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08544
USA
E-MAIL: tianqi@princeton.edu
hanliu@princeton.edu

DEPARTMENT OF STATISTICS
PURDUE UNIVERSITY
WEST LAFAYETTE, IN 47906
USA
E-MAIL: chengg@purdue.edu