Supplementary material to

Kernel Meets Sieve: Post-Regularization Confidence Bands for Sparse Additive Model

Junwei Lu* Mladen Kolar† Han Liu‡

Abstract

This document contains the supplementary material to the paper "Kernel Meets Sieve: Post-Regularization Confidence Bands for Sparse Additive Model". All the proofs in the supplementary material assume that true nonparametric function $f(x_1, \ldots, x_d)$ belongs to the ATLAS model $\mathcal{A}_d(s)$. In Appendix A, we outlines the proof of Theorem 3.2. In Appendix B, we introduce an accelerated method to derive our estimator. Appendix C proves the validity of bootstrap confidence bands. In Appendix D, we prove Propositions 3.1.Appendix E collects the technical lemmas on the estimation rate. Appendix F states some auxiliary results on the bootstrap confidence bands. In Appendix H, we lists several useful results on empirical processes.

A Proof of the Statistical Rate of Kernel-Sieve Hybrid Estimator

For all the proofs in the following of the paper (including the supplementary material), we consider the most general case that true nonparametric function $f(x_1, \ldots, x_d)$ belongs to the ATLAS model $\mathcal{A}_d(s)$. Since SpAM is a strictly smaller family of $\mathcal{A}_d(s)$, all the proofs apply to $\mathcal{K}_d(s)$ as well.

This section outlines the proof of Theorem 3.2 on the statistical estimation rate of the kernel-sieve hybrid estimator in (2.7). Before presenting the main proof, we list several technical lemmas whose proofs are deferred to Appendix E in the supplementary material.

The following lemma provides the restricted eigenvalue condition on the empirical Hessian matrix of the kernel-sieve hybrid loss in (2.6), which is $\widehat{\boldsymbol{\Sigma}}_z = n^{-1}\boldsymbol{\Psi}\mathbf{W}_z\boldsymbol{\Psi}^T$.

*Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115; Email: junweilu@hsph.harvard.edu, †Booth School of Business, The University of Chicago, Chicago, IL 60637, USA; Email: mkolar@chicagobooth.edu, ‡Department of Computer Science, Northwestern University, Evanston, IL 60208; Email: hanliu@northwestern.edu

Lemma A.1. Under Assumptions (**A1**)-(**A5**), suppose $\boldsymbol{\beta} \in \mathbb{R}^{(d-1)m}$ and $\alpha \in \mathbb{R}$ satisfy the cone restriction

$$\sum_{j \in S^c} \|\boldsymbol{\beta}_j\|_2 \le 3 \sum_{j \in S} \|\boldsymbol{\beta}_j\|_2 + 3\sqrt{m}|\alpha|$$

for some index set $S \subset [d]$ with cardinality s. Denote $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^T)^T$. If $s\sqrt{m^3 \log(dm)/(nh)} + sm^2/(nh) = o(1)$, there exists a constant $\rho_{\min}$ such that with high probability,

$$\inf_{z \in \mathcal{X}} \boldsymbol{\theta}^T \widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta} \ge \frac{\rho_{\min}}{2m} \|\boldsymbol{\beta}\|_2^2 + \frac{\rho_{\min}}{2} |\alpha|_2^2.$$

The estimation error for the kernel-sieve hybrid estimator comes from three sources: (1) noise ε, (2) approximation error by finite B-spline bases, and (3) approximation error by s local additive functions to the true function. The following lemma provides the rate for the B-spline approximation error, which further illustrates how the number of B-spline basis functions m influences the rate.

Lemma A.2. Recall that $\{f_{jz}\}_{j=1}^d$ are defined in Definition 4.1. Let $\boldsymbol{\delta}_z = (\delta_1(z), \ldots, \delta_n(z))^T$ where $\delta_i(z) = \sum_{j=2}^d f_{jz}(X_{ji}) - f_{mj}(X_{ji})$ for $i = 1, \ldots n$, where $f_{mj}(\cdot)$ is defined in (2.4). Under Assumptions (**A1**)-(**A5**) there exists a constant $C > 0$ such that the following three inequalities hold with probability at least $1 - 1/n$,

$$\sup_{z \in \mathcal{X}} \max_{j \ge 2} \frac{1}{n} \|\boldsymbol{\Psi}_{\bullet j}^T \mathbf{W}_z \boldsymbol{\delta}_z\|_2 \le C\sqrt{s} \cdot m^{-5/2}, \tag{A.1}$$

$$\sup_{z \in \mathcal{X}} \frac{1}{n} |\boldsymbol{\Psi}_{\bullet 1}^T \mathbf{W}_z \boldsymbol{\delta}_z| \le C\sqrt{s} \cdot m^{-2}, \tag{A.2}$$

$$\sup_{z \in \mathcal{X}} \frac{1}{n} \|\mathbf{W}_z^{1/2} \boldsymbol{\delta}_z\|_2^2 \le Csm^{-4}. \tag{A.3}$$

Our next lemma bounds the approximation error of charts under the ATLAS model (4.1). We can see that both the number of bases m and the bandwidth h play a role in the estimation.

Lemma A.3. Let $\boldsymbol{\xi}_z = (\xi_1(z), \ldots, \xi_n(z))^T$ and $\boldsymbol{\zeta}_z = (\zeta_1(z), \ldots, \zeta_n(z))^T$, where $\xi_i(z) = f_1(X_{1i}) - f_1(z)$ and $\zeta_i(z) = f(X_{1i}, \ldots, X_{di}) - \sum_{j=1}^d f_{jz}(X_{ji})$ for $i \in [n]$. Under Assumptions (**A1**)-(**A5**) there exists a constant $C > 0$ such that the following three inequalities hold with probability at least

$1 - 1/n$,

$$\sup_{z\in\mathcal{X}} \max_{j\in[d]} \frac{1}{n}\|\mathbf{\Psi}_{\bullet j}^T \mathbf{W}_z(\boldsymbol{\xi}_z + \boldsymbol{\zeta}_z)\|_2 \le C\left(\sqrt{\frac{h\log(dh^{-1})}{n}} + \frac{m^{3/2}\log(dh^{-1})}{n} + \frac{h^2}{\sqrt{m}}\right),$$
$$\sup_{z\in\mathcal{X}} \frac{1}{n}|\mathbf{\Psi}_{\bullet 1}^T \mathbf{W}_z(\boldsymbol{\xi}_z + \boldsymbol{\zeta}_z)| \le C\left(h^2 + \sqrt{h/n}\right),$$
$$\sup_{z\in\mathcal{X}} \frac{1}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{\xi}_z + \boldsymbol{\zeta}_z)\|_2^2 \le Ch^2.$$

The following lemma quantifies the statistical error arising from the noise ε.

Lemma A.4. Let $T_n = \sup_{z\in\mathcal{X}} \max_{j\in[d]} n^{-1}\|\mathbf{\Psi}_{\bullet j}^T \mathbf{W}_z \boldsymbol{\varepsilon}\|_2$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$. Under Assumptions (**A1**)-(**A5**) and if $m(nh)^{-1} = o(1)$, there exists a constant $C > 0$ such that with probability at least $1 - 1/n$,

$$T_n \le C\sqrt{\log(dm^2h^{-2})/(nh)}.$$

We are now ready to present the main proof of Theorem 3.2.

Proof of Theorem 3.2. We denote $\boldsymbol{\eta}_z = \boldsymbol{\varepsilon} + \boldsymbol{\delta}_z + \boldsymbol{\xi}_z + \boldsymbol{\zeta}_z$ and define the event

$$\mathcal{E} = \left\{\sup_{z\in\mathcal{X}} \max_{j\ge 2} \frac{4}{n}\|\mathbf{\Psi}_{\bullet j}^T \mathbf{W}_z \boldsymbol{\eta}_z\|_2 \le \lambda\right\} \bigcup \left\{\sup_{z\in\mathcal{X}} \frac{4}{n}|\mathbf{\Psi}_{\bullet 1}^T \mathbf{W}_z \boldsymbol{\eta}_z|_2 \le \lambda\sqrt{m}\right\}.$$

Using Lemma A.2, Lemma A.3 and Lemma A.4, there exist constants c, C such that $\mathbb{P}(\mathcal{E}) \ge 1 - c/n$ if the tuning parameter satisfies the following inequality

$$\lambda \ge C\left(\sqrt{\frac{\log(dm^2h^{-2})}{nh}} + \sqrt{s}\cdot m^{-5/2} + \sqrt{\frac{h\log(dh^{-1})}{n}} + \frac{m^{3/2}\log(dh^{-1})}{n} + \frac{h^2}{\sqrt{m}}\right). \tag{A.4}$$

In the rest of this proof, we are always conditioning on the event $\mathcal{E}$.

Denote $S_z := \{j \in \{2, \ldots, d\} \,|\, f_{jz} \not\equiv 0\}$ and $\mathbf{\Delta} = \widehat{\boldsymbol{\beta}}_+ - \boldsymbol{\beta}_+$, where $\mathbf{\Delta}_1 = \widehat{\alpha}_z - f_1(z)$ and $\mathbf{\Delta}_j = \widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j$ for $j \ge 2$. We start by showing that $\mathbf{\Delta}$ falls into the cone

$$\mathcal{A}_z := \left\{\mathbf{\Delta} : \textstyle\sum_{j\in S_z^c} \|\mathbf{\Delta}_j\|_2 \le 3\sum_{j\in S_z} \|\mathbf{\Delta}_j\|_2 + 3\sqrt{m}|\mathbf{\Delta}_1|\right\}.$$

Since $\widehat{\boldsymbol{\beta}}_+$ is a minimizer of the objective function,

$$\frac{1}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{Y}-\boldsymbol{\Psi}\widehat{\boldsymbol{\beta}}_+)\|_2^2-\frac{1}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{Y}-\boldsymbol{\Psi}\boldsymbol{\beta}_+)\|_2^2+\lambda\|\widehat{\boldsymbol{\beta}}\|_{1,2}-\lambda\|\boldsymbol{\beta}\|_{1,2}+\lambda\sqrt{m}(|\widehat{\alpha}_z|-|f_1(z)|)\leq 0.$$

On the event $\mathcal{E}$, we have the following inequality

$$\begin{aligned}\sup_{z\in\mathcal{X}}\frac{4}{n}\boldsymbol{\eta}_z^T\mathbf{W}_z\boldsymbol{\Psi}\boldsymbol{\Delta}&\leq\sup_{z\in\mathcal{X}}\max_{j\geq 2}\frac{4}{n}\|\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z\boldsymbol{\eta}_z\|_2\|\boldsymbol{\Delta}_{2:d}\|_{1,2}+\sup_{z\in\mathcal{X}}\frac{4}{n}\|\boldsymbol{\Psi}_{\bullet 1}^T\mathbf{W}_z\boldsymbol{\eta}_z\|_2|\boldsymbol{\Delta}_1|\\&\leq\lambda\|\boldsymbol{\Delta}_{2:d}\|_{1,2}+\lambda\sqrt{m}|\boldsymbol{\Delta}_1|.\end{aligned}$$

The first inequality is due to the Hölder's inequality and the second one is by the definition of $\mathcal{E}$. Furthermore, we derive the following inequality

$$\begin{aligned}\frac{1}{n}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}\boldsymbol{\Delta}\|_2^2&\leq\frac{2}{n}\boldsymbol{\eta}_z^T\mathbf{W}_z\boldsymbol{\Psi}\boldsymbol{\Delta}-\lambda\sum_{j=2}^{d}(\|\widehat{\boldsymbol{\beta}}_j\|-\|\boldsymbol{\beta}_j\|)-\lambda\sqrt{m}(|\widehat{\alpha}_z|-|f_1(z)|)\\&\leq\frac{\lambda}{2}\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\|_{1,2}+\lambda\sqrt{m}|\widehat{\alpha}_z-f_1(z)|-\lambda\sum_{j=2}^{d}(\|\widehat{\boldsymbol{\beta}}_j\|-\|\boldsymbol{\beta}_j\|)\\&\leq\frac{3\lambda}{2}\sum_{j\in S_z}\|\boldsymbol{\Delta}_j\|+\frac{3\lambda}{2}\sqrt{m}|\boldsymbol{\Delta}_1|-\frac{\lambda}{2}\sum_{j\in S_z^c}\|\boldsymbol{\Delta}_j\|.\end{aligned}$$

The last inequality shows that $\boldsymbol{\Delta}\in\mathcal{A}_z$.

Next, we prove the rate of convergence by contradiction. Suppose that for some fixed t, which will be specified later, we have

$$\exists z\in\mathcal{X},\quad\frac{1}{\sqrt{n}}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}\boldsymbol{\Delta}\|>t. \tag{A.5}$$

Equation (A.5) implies that there exists some $z\in\mathcal{X}$ such that

$$0>\min_{\boldsymbol{\Delta}\in\mathcal{A}_z,\|\widehat{\boldsymbol{\Sigma}}_z^{1/2}\boldsymbol{\Delta}\|\geq t}\frac{1}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{Y}-\boldsymbol{\Psi}\widehat{\boldsymbol{\beta}}_+)\|_2^2-\frac{1}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{Y}-\boldsymbol{\Psi}\boldsymbol{\beta}_+)\|_2^2+\lambda\|\widehat{\boldsymbol{\beta}}_+\|_{1,2}-\lambda\|\boldsymbol{\beta}_+\|_{1,2}.$$

Using the fact that $\mathcal{A}_z$ is a cone, we can replace the constraint $\|\widehat{\boldsymbol{\Sigma}}_z^{1/2}\boldsymbol{\Delta}\|\geq t$ by $\|\widehat{\boldsymbol{\Sigma}}_z^{1/2}\boldsymbol{\Delta}\|=t$ and the above inequality still preserves. Combining the event $\mathcal{E}$, we have

$$
\begin{aligned}
0 &> \min_{\boldsymbol{\Delta}\in\mathcal{A}_z, \|\widehat{\boldsymbol{\Sigma}}_z^{1/2}\boldsymbol{\Delta}\|=t} \frac{1}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{Y}-\boldsymbol{\Psi}\widehat{\boldsymbol{\beta}}_+)\|_2^2 - \frac{1}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{Y}-\boldsymbol{\Psi}\boldsymbol{\beta}_+)\|_2^2 + \lambda\mathcal{R}(\widehat{\boldsymbol{\beta}}_+) - \lambda\mathcal{R}(\boldsymbol{\beta}_+) \\
&\geq \min_{\boldsymbol{\Delta}\in\mathcal{A}_z, \|\widehat{\boldsymbol{\Sigma}}_z^{1/2}\boldsymbol{\Delta}\|=t} \frac{1}{n}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}\boldsymbol{\Delta}\|_2^2 - 2\lambda\mathcal{R}(\boldsymbol{\Delta}) + \lambda\mathcal{R}(\widehat{\boldsymbol{\beta}}_+) - \lambda\mathcal{R}(\boldsymbol{\beta}_+).
\end{aligned}
\tag{A.6}
$$

From Lemma A.1, we can bound the R.H.S. by

$$
\begin{aligned}
2\mathcal{R}(\boldsymbol{\Delta}) - \mathcal{R}(\widehat{\boldsymbol{\beta}}_+) + \mathcal{R}(\boldsymbol{\beta}_+) &\leq 3\sum_{j\in S}\|\boldsymbol{\Delta}_j\| + 3\sqrt{m}|\boldsymbol{\Delta}_1| \\
&\leq 3\sqrt{s}\,\|\boldsymbol{\Delta}_{S\cup\widehat{S}}\|_2 + 3\sqrt{m}|\boldsymbol{\Delta}_1| \\
&\leq 6\sqrt{2sm/\rho_{\min}}\,\|\widehat{\boldsymbol{\Sigma}}_z^{1/2}\boldsymbol{\Delta}\|_2.
\end{aligned}
\tag{A.7}
$$

Combining (A.6) and (A.7), we get a quadratic inequality

$$
0 > t^2 - \frac{2\lambda\sqrt{sm}}{\rho_{\min}}t. \tag{A.8}
$$

Setting $t = 2\sqrt{sm/\rho_{\min}}\cdot\lambda$, we obtain from (A.8) that $0 > t^2 - \left[2\lambda\sqrt{sm/\rho_{\min}}\right]t = 0$, which is a contradiction. Therefore, $\sup_{z\in\mathcal{X}} n^{-1/2}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}\boldsymbol{\Delta}\|_2 \leq 2\lambda\sqrt{sm}/\rho_{\min}$. Using the rate for λ in (A.4) and $h = o(1)$, we have

$$
\sup_{z\in\mathcal{X}} \frac{1}{\sqrt{n}}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}\boldsymbol{\Delta}\|_2 \leq C\sqrt{sm}\left(\sqrt{\frac{\log(dmh^{-1})}{nh}} + \frac{\sqrt{s}}{m^{5/2}} + \frac{m^{3/2}\log(dh^{-1})}{n} + \frac{h^2}{\sqrt{m}}\right). \tag{A.9}
$$

Now, using Lemma A.1, since $\boldsymbol{\Delta}\in\mathcal{A}_z$, we have that

$$
\|\boldsymbol{\Delta}_{2:d}\|_{1,2} + \sqrt{m}|\boldsymbol{\Delta}_1| \leq \sqrt{s}\|\boldsymbol{\Delta}_{2:d}\|_2 + \sqrt{m}|\boldsymbol{\Delta}_1| \leq \sqrt{sm/(\rho_{\min}n)}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}\boldsymbol{\Delta}\|_2
$$

for any $z\in\mathcal{X}$, which leads to the following inequality

$$
\sup_{z\in\mathcal{X}} \sqrt{m}|\widehat{\alpha}_z - f_1(z)| + \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\|_{1,2} \leq Csm\left(\sqrt{\frac{\log(dmh^{-1})}{nh}} + \frac{\sqrt{s}}{m^{5/2}} + \frac{m^{3/2}\log(dh^{-1})}{n} + \frac{h^2}{\sqrt{m}}\right),
$$

with probability at least $1 - c/n$. To obtain the best rate on the right hand side of the equation, we choose $h \asymp n^{-1/6}$ and $m \asymp n^{1/6}$ to obtain

$$\sup_{z\in\mathcal{X}} \left\{ \sqrt{m}|\widehat{a}_z - f_1(z)| + \sum_{i=2}^{d} \|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_2 \right\} = O_P\left(\log(dn)n^{-1/4}\right).$$

According to Corollary 15 in Chapter XI of de Boor (2001), given a function $g(x) = \sum_{k=1}^{m} \beta_k \phi_k(x)$, we have

$$\|g\|_2^2 \asymp m^{-1} \sum_{k=1}^{m} \beta_k^2. \tag{A.10}$$

Therefore, we have $\|\widehat{f} - f\|_2 \le \rho_{\min}^{-1} s\sqrt{m}\lambda$ and, when $h \asymp n^{-1/6}$ and $m \asymp n^{1/6}$, the rate becomes

$$\|\widehat{f} - f\|_2^2 = O_P\left(n^{-2/3}\log(dn)\right).$$

This completes the proof. □

B Accelerated Algorithm

This section presents details of our method to accelerate Algorithm 1. To estimate f_1, we need to compute the estimator $\widehat{\alpha}_z$ for a number of z values $z \in \{z_1, \ldots, z_M\}$. A naïve approach is to run Algorithm 1 M times, once for each value of z's. We provide a more efficient algorithm which significantly reduces the computational cost. From Algorithm 1 and (2.15), the most expensive operation is evaluation of the gradient

$$\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)}) = -\frac{1}{n}\boldsymbol{\Psi}_{\bullet j}^T \mathbf{W}_z \left(\boldsymbol{Y} - \boldsymbol{\Psi}\boldsymbol{\beta}_+^{(t)}\right). \tag{B.1}$$

Computing $\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})$ for a single z requires $O(dm^2n)$ flops. If we trivially repeat the computation for M different z's, the computational complexity is $O(dm^2nM)$ which is challenging when M is large. However, we can exploit the structure of $\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})$ to reduce the computational complexity. According to (B.1) and the fact that $\psi_{jk}(X_{i1}) = \phi_k(X_{i1}) - \bar{\phi}_{jk}(z)$, the k-th coordinate of $\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})$

has a formulation

$$
\begin{aligned}
\Big(\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})\Big)_k = & -\frac{1}{n}\sum_{i\in[n]} K_h(X_{i1}-z)\phi_k(X_{i1})Y_i + \bar{\phi}_{jk}(z)\cdot\frac{1}{n}\sum_{i\in[n]} K_h(X_{i1}-z)Y_i \\
& + \sum_{\ell\in[d],s\in[m]} \boldsymbol{\beta}_{\ell s}^{(t)}\bigg\{\frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1}-z)\big[\phi_k(X_{i1})-\bar{\phi}_{jk}(z)\big]\big[\phi_s(X_{i1})-\bar{\phi}_{\ell s}(z)\big]\bigg\}.
\end{aligned} \tag{B.2}
$$

The computation of $\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})$ is mostly spent on evaluating the formulation

$$
q(z) = \sum_{i=1}^{n} K_h(X_{i1}-z)u_i \tag{B.3}
$$

for $z \in \{z_1,\ldots,z_M\}$ where $u_1,\ldots,u_n$ are fixed quantities (e.g., u_i could be Y_i, $\phi_k(X_{i1})$ or $Y_i\phi_k(X_{i1})$ when evaluating (B.2)) independent of z. We introduce a fast method to calculate the general form $q(z)$ and apply it to the computation of (B.2). Without loss of generality, we assume that $z_1 < \ldots < z_M$. The naïve method to evaluate $\{q(z_\ell)\}_{\ell\in[M]}$ separately for different z has the computational complexity $O(nM)$. However, if the kernel function has some special structure, we can reduce the complexity to $O(n+M)$. For example, for the uniform kernel $K(u) = \frac{1}{2}\mathbb{1}\{|u|\le 1\}$, when we vary the value of z from z_ℓ to $z_{\ell+1}$, we just need to subtract u_i for those $i \in \{v : X_v \in [z_\ell - h, z_{\ell+1}-h)\}$ and add u_i for those $i \in \{v : X_v \in (z_\ell + h, z_{\ell+1}+h]\}$. For $M \gg h^{-1}$, the cardinality of $\{i : X_{i1} \in (z_\ell - h, z_{\ell+1}-h] \cup (z_\ell + h, z_{\ell+1}+h]\}$ does not increase with n or d. Therefore, the complexity to evaluate $\{q(z_\ell)\}_{\ell\in[M]}$ is reduced to $O(n+K)$. For the Epanechnikov kernel $K(u) = (3/4)\cdot(1-u^2)\mathbb{1}\{|u|\le 1\}$, suppose $q(z_\ell)$ is known and define $I_\ell = \{i : X_{i1} \in (z_\ell - h, z_{\ell+1}-h] \cup (z_\ell + h, z_{\ell+1}+h]\}$. We have $q(z_{\ell+1}) = q(z_\ell) + \Delta q(z_\ell)$, where

$$
\Delta q(z_\ell) = q(z_{\ell+1}) - q(z_\ell) = \frac{3}{4}\sum_{i\in I_\ell}\big(1-(X_{i1}/h)^2\big)u_i + \frac{3z}{2h^2}\sum_{i\in I_\ell} X_{i1} + \frac{3z^2}{4}\sum_{i\in I_\ell} u_i.
$$

Similar to the argument for the case of uniform kernel, we also have $|I_\ell| = O(1)$ if $K \gg h^{-1}$. The computational complexity of $\sum_{i\in I_z}(1-X_{i1}^2)u_i$ and the other two summations above for $z = 1,\ldots,z_K$ is $O(n+K)$ and hence the computational complexity of $\{q(z_\ell)\}_{k\in[K]}$ for Epanechnikov kernel is also $O(n+K)$. We can also apply a similar trick to many other kernels. Now we turn back to the calculation of the gradient $\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})$. Let $\widehat{p}_1(z) = n^{-1}\sum_{i=1}^n K_h(X_{i1} - z)$,

$$Y_k^{(1)}(z) = \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)Y_i, \quad Y_k^{(2)}(z) = \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\phi_k(X_{ij})Y_i,$$
$$Y_k^{(3)}(z) = \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\phi_k(X_{ij}) \text{ and } R_{ks}(z) = \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\phi_k(X_{ij})\phi_s(X_{iu}).$$

For different values of z, we denote the components of $\boldsymbol{\beta}_+$ corresponding to the k-th B-spline basis for the j-th covariate as $\boldsymbol{\beta}_{jk;z}$. According to the expansion in (B.2), we can write the k-th coordinate of $\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})$ as

$$\begin{aligned}\Big(\nabla_j \mathcal{L}_z(\boldsymbol{\beta}_+^{(t)})\Big)_k &= -Y_k^{(1)}(z) + \bar{\phi}_{jk}(z)Y_k^{(2)}(z) + \frac{1}{n}\sum_{\ell\in[d],s\in[m]} \boldsymbol{\beta}_{\ell s;z}\Big(R_{kv}(z) - \bar{\phi}_{jk}(z)Y_v^{(2)}(z)\Big) \\ &\quad - \frac{1}{n}\sum_{\ell\in[d],s\in[m]} \boldsymbol{\beta}_{\ell s;z}\Big(\bar{\phi}_{\ell s}(z)Y_k^{(3)}(z) - \bar{\phi}_{jk}(z)\bar{\phi}_{\ell s}(z)\widehat{p}_1(z)\Big).\end{aligned}$$

Based on the previous discussion on the calculation of $q(z)$ in (B.3), we note that it takes $O(n + M)$ operations to evaluate $\widehat{p}_1(z), Y_k^{(1)}(z), Y_k^{(2)}(z), Y_k^{(3)}(z)$ and $R_{ks}(z)$ for M different values of z. Therefore, the computational complexity of each iteration in Algorithm 1 can be reduced from $O(dm^2nM)$ to $O(dm^2(n+M))$. Therefore under the case $M = O(n)$, we can estimate f_1 using the introduced procedure with the same computational complexity as (2.10). Since most of existing algorithms for the group Lasso involve evaluating the gradient (Yuan and Lin, 2007; Friedman et al., 2010; Farrell, 2013; Qin et al., 2013), the above argument is applicable to other solvers as well.

C Covering Properties of the Bootstrap Confidence Bands

In this section, we prove the theorems on the coverage probabilities for the Gaussian multiplier bootstrap confidence bands $\mathcal{C}_{n,\alpha}^b$ in (2.19). We will first prove Theorem 3.7 and Theorem 3.5 can be proven by following the same steps.

C.1 Proof of Theorem 3.7

We first prove that $\mathcal{C}_{n,\alpha}^b$ in (2.19) is honest. To simplify the notation, we will use $\mathcal{X}$ to represent the interval $[-D_n, D_n]$ when there is no confusion. The strategy to prove the result is to establish a sequence of processes from $\widehat{\mathbb{H}}_n(z)$ that approximate $\widetilde{Z}_n(z)$. We consider the following four stochastic processes

$$\widehat{\mathbb{H}}_n(z) = \frac{1}{\sqrt{nh^{-1}}}\sum_{i=1}^{n}\xi_i \cdot \frac{\widehat{\sigma} K_h(X_{i1}-z)\boldsymbol{\Psi}_{i\bullet}^T\widehat{\boldsymbol{\theta}}_z}{\widehat{\sigma}_n(z)}; \tag{C.1}$$

$$\widehat{\mathbb{H}}_n^{(1)}(z) = \frac{1}{\sqrt{nh^{-1}}}\sum_{i=1}^{n}\xi_i \cdot \frac{\sigma K_h(X_{i1}-z)\boldsymbol{\Psi}_{i\bullet}^T\widehat{\boldsymbol{\theta}}_z}{\widehat{\sigma}_n(z)}, \tag{C.2}$$

$$\widetilde{\mathbb{H}}_n(z) = \frac{1}{\sqrt{nh^{-1}}}\sum_{i=1}^{n}\varepsilon_i \frac{K_h(X_{i1}-z)\boldsymbol{\Psi}_{i\bullet}^T\widehat{\boldsymbol{\theta}}_z}{\widehat{\sigma}_n(z)}, \tag{C.3}$$

$$\widetilde{Z}_n(z) = \sqrt{nh}\cdot\widehat{\sigma}_n^{-1}(z)\left(\widehat{f}_1^u(z) - f_1(z)\right). \tag{C.4}$$

Corollary 3.1 of Chernozhukov et al. (2014a) provides sufficient conditions for the confidence band to be asymptotically honest. Specifically, we need to verify the following high-level conditions:

H1 There exists a Gaussian process $\mathbb{G}_n(z)$ and a sequence of random variables W_n^0 such that $W_n^0 \overset{d}{=} \sup_{z\in\mathcal{X}}\mathbb{G}_n(z)$. Furthermore, $\mathbb{E}[\sup_{z\in\mathcal{X}}\mathbb{G}_n] \le C\sqrt{\log n}$ and

$$\mathbb{P}(|W_n^Z - W_n^0| > \varepsilon_{1n}) < \delta_{1n}$$

for some ε_{1n} and δ_{1n}.

H2 For any $\epsilon > 0$, the anti-concentration inequality

$$\sup_{x\in\mathbb{R}}\mathbb{P}\left(\left|\sup_{z\in\mathcal{X}}|\mathbb{G}_n(z)| - x\right| \le \epsilon\right) \le C\epsilon\sqrt{\log n}.$$

holds.

H3 Let $c_n(\alpha)$ be the $(1-\alpha)$-quantile of W_n^Z and $\widehat{c}_n(\alpha)$ be the $1-\alpha$ quantile of $\widehat{W}_n$. There exists τ_n, ε_{2n} and δ_{2n} such that

$$\mathbb{P}\left(\widehat{c}_n(\alpha) < c_n(\alpha+\tau_n) - \varepsilon_{2n}\right) \le \delta_{2n} \quad \text{and} \quad \mathbb{P}\left(\widehat{c}_n(\alpha) > c_n(\alpha-\tau_n) + \varepsilon_{2n}\right) \le \delta_{2n}.$$

H4 There exists ε_{3n} and δ_{3n} such that

$$\mathbb{P}\left(\sup_{z\in\mathcal{X}}\left|\frac{\widehat{\sigma}\sqrt{\widehat{p}_1(z)}}{\sigma\sqrt{p_1(z)}}-1\right|>\varepsilon_{3n}\right)\leq\delta_{3n}.$$

If the high-level conditions **H1** - **H4** are verified, Corollary 3.1 in Chernozhukov et al. (2014a) implies that

$$\mathbb{P}(f_1\in\mathcal{C}^b_{n,\alpha})\geq 1-\alpha-(\varepsilon_{1n}+\varepsilon_{2n}+\varepsilon_{3n}+\delta_{1n}+\delta_{2n}+\delta_{3n}).$$

In the remaining part of the proof, we show that the conditions are satisfied.

The roadmap is to establish that the process in (C.4) is close to the process in (C.1) following the chain $\widetilde{Z}_n\to\widetilde{\mathbb{H}}_n\to\widehat{\mathbb{H}}^{(1)}_n\to\widehat{\mathbb{H}}_n$. After that, we can check conditions **H1** – **H3**. Since we do not use the population $\sigma_n(z)=\mathbb{E}[\widehat{\sigma}_n(z)]$ in the intermediate processes, we do not need to check the condition **H4**.

In order to verify the condition **H1**, we first bound the difference between $\sup_{z\in\mathcal{X}}\widetilde{\mathbb{H}}_n(z)$ and $\sup_{z\in\mathcal{X}}Z_n(z)$. We begin by considering two auxiliary processes

$$\widetilde{\mathbb{H}}'_n(z)=\frac{1}{\sqrt{nh^{-1}}}\sum_{i=1}^n\varepsilon_iK_h(X_{i1}-z)\mathbf{\Psi}^T_{i\bullet}\widehat{\boldsymbol{\theta}}_z\quad\text{ and }\quad\widetilde{Z}'_n(z)=\sqrt{nh}\left(\widehat{f}^u_1(z)-f_1(z)\right).$$

Notice that the above processes are un-normalized version of (C.2) and (C.4), that is, $\widetilde{\mathbb{H}}'_n(z)=\widehat{\sigma}_n(z)\widetilde{\mathbb{H}}_n(z)$ and $\widetilde{Z}'_n(z)=\widehat{\sigma}_n(z)\widetilde{Z}_n(z)$. The following lemma provides a direct bound for the difference between $\widetilde{\mathbb{H}}'_n(z)$ and $\widetilde{Z}'_n(z)$.

Lemma C.1. Under the same conditions of Theorem 3.7, there exists a constant $c_0>\delta/2$ such that with probability $1-c/n$,

$$\sup_{z\in\mathcal{X}}\left|\widetilde{\mathbb{H}}'_n(z)-\widetilde{Z}'_n(z)\right|\leq Cn^{-c_0}.$$

We defer the proof of the lemma to Section F.5 and proceed to prove Theorem 3.7. We also need to study $\widehat{\sigma}$ and $\widehat{\sigma}_n(z)$ in the following lemmas.

Lemma C.2. Let the estimator for $\mathrm{Var}(\varepsilon) = \sigma^2$ be $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \widehat{\varepsilon}_i^2$. Let

$$r_n := \sqrt{\frac{s^2 \log(dmh^{-1})}{nm^{-2}h}} + \frac{s^{3/2}}{m^{3/2}} + \frac{s\log(dh^{-1})}{nm^{-5/2}} + s\sqrt{m}h^2. \tag{C.5}$$

Under the same conditions of Theorem 3.7,, there exists constants c, C such that $\mathbb{P}\big(|\widehat{\sigma}^2 - \sigma^2| \geq Cn^{-c}\big) \leq 6/n$.

Lemma C.3. Let $\boldsymbol{\Sigma}'_z = n^{-1}\boldsymbol{\Psi}\mathbf{W}_z^2\boldsymbol{\Psi}^T$. Under the same conditions of Theorem 3.7, there exist constants c, C such that for sufficiently large n, with probability $1 - c/n$, for any $z \in \mathcal{X}$,

$$ch^{-1}\mathbf{e}_1^T\boldsymbol{\theta}_z \leq \widehat{\boldsymbol{\theta}}_z^T\boldsymbol{\Sigma}'_z\widehat{\boldsymbol{\theta}}_z \leq Ch^{-1}\mathbf{e}_1^T\boldsymbol{\theta}_z.$$

We defer the proof of this lemma to Section F.6. Notice that we can no longer choose $h \asymp n^{-1/6}$, $m \asymp n^{1/6}$ used for the estimation rate in Theorem 3.2. This is because we need to under-regularize our estimator to make the bias terms ignorable.

From Lemmas C.3 and F.3, we have an upper bound of the inverse of $\widehat{\sigma}_n^2(z) = \widehat{\boldsymbol{\theta}}_z^T\boldsymbol{\Sigma}'_z\widehat{\boldsymbol{\theta}}_z$ as

$$\sup_{z\in\mathcal{X}} \sqrt{h} \cdot \widehat{\sigma}_n^{-1}(z) \leq C. \tag{C.6}$$

With Lemma C.1 and Lemma C.3, we are ready to bound the difference between $\sup_{z\in\mathcal{X}} \widetilde{\mathbb{H}}_n(z)$ and $\sup_{z\in\mathcal{X}} Z_n(z)$. Let c_0 be the constant in Lemma C.1. We choose h, m satisfying the scaling condition of Theorem 3.7. We denote $c = c_0 - \delta/2$ and observe that $c > 0$ by Lemma C.1. From Lemma C.1 and (C.6), we have

$$\begin{aligned}
\mathbb{P}\left(\sup_{z\in\mathcal{X}} \left|\widetilde{\mathbb{H}}_n(z) - \widetilde{Z}_n(z)\right| \geq Cn^{-c}\right) &\leq \mathbb{P}\left(\sup_{z\in\mathcal{X}} \left|\widetilde{\mathbb{H}}'_n(z) - \widetilde{Z}'_n(z)\right| \geq C\widehat{\sigma}_n(z)n^{-c_0}/\sqrt{h}\right) \\
&\leq \mathbb{P}\left(\sup_{z\in\mathcal{X}} \left|\widetilde{\mathbb{H}}'_n(z) - \widetilde{Z}'_n(z)\right| \geq C^2 n^{-c_0}\right) \leq 1/n.
\end{aligned}$$

Define $V_n^0 = \sup_{z\in\mathcal{X}} \widetilde{\mathbb{H}}_n(z)$ and $\widetilde{V}^Z = \sup_{z\in\mathcal{X}} \widetilde{Z}_n(z)$. Since $\sup_{z\in\mathcal{X}} \widetilde{\mathbb{H}}_n(z)$ is a Gaussian process conditional on $\{\boldsymbol{X}_{i1}\}_{i\in[n]}$, we verify **H1** by

$$\mathbb{P}\left(|V_n^0 - \widetilde{V}^Z| \geq Cn^{-c}\right) \leq \mathbb{P}\left(\sup_{z\in\mathcal{X}} |\widetilde{\mathbb{H}}_n(z) - \widetilde{Z}_n(z)| \geq Cn^{-c}\right) \leq \frac{1}{n}. \tag{C.7}$$

The condition **H2** follows from **H1** and the anti-concentration inequality in Corollary 2.1 of Chernozhukov et al. (2014a).

Next, we check **H3** by bounding the difference between (C.1) and (C.2). We first approximate $\widehat{\mathbb{H}}_n(z)$ by $\widehat{\mathbb{H}}_n^{(1)}(z)$. By Lemma C.2, if we choose h, m satisfying the scaling condition of Theorem 3.7., with probability $1-6/n$, $|\widehat{\sigma}-\sigma| < C\sqrt{r_n}m^{1/4} = o(n^{-c})$, where r_n is defined in (C.5).

We denote $\widehat{V}_n = \sup_{z\in\mathcal{X}} \widehat{\mathbb{H}}_n(z)$, $\widehat{V}_n^{(1)} = \sup_{z\in\mathcal{X}} \widehat{\mathbb{H}}_n^{(1)}(z)$ and the difference between $\widehat{V}_n - \widehat{V}_n^{(1)}$. Let $\Delta\mathbb{H}^{(1)}(z) = \widehat{\mathbb{H}}_n^{(1)}(z) - \widehat{\mathbb{H}}_n(z)$. We have

$$\sup_{z\in\mathcal{X}} \left|\Delta\mathbb{H}^{(1)}(z)\right| \le |\widehat{\sigma}-\sigma| \sup_{z\in\mathcal{X}} \sqrt{h}\cdot\widehat{\sigma}_n^{-1}(z)\left(\sup_{z\in\mathcal{X}} I_1(z) + \sup_{z\in\mathcal{X}} I_2(z)\right),$$

where $I_1(z) = n^{-1}\sum_{i=1}^n K_h(X_{i1}-z)\left|\mathbf{\Psi}_{i\bullet}^T(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)\right|$ and $I_2(z) = n^{-1}\sum_{i=1}^n K_h(X_{i1}-z)\left|\mathbf{\Psi}_{i\bullet}^T\boldsymbol{\theta}_z\right|$.

In order to bound $I_1(z)$, we first state a technical lemma that characterizes the estimation error between $\widehat{\boldsymbol{\theta}}_z$ and $\boldsymbol{\theta}_z$.

Lemma C.4. Let $\widehat{\boldsymbol{\theta}}_z$ be a minimizer of (2.16). Suppose that Assumptions (**A1**), (**A2**), (**A4**)-(**A6**) hold and $\bar{\beta}_{2,\kappa}$ in (3.5) is finite. If the parameter γ in the optimization program (2.16) is chosen as in (3.11), then with probability $1-c/d$,

$$\sup_{z\in\mathcal{X}}(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T\widehat{\mathbf{\Sigma}}_z(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z) \le C\inf_{z\in\mathcal{X}}\frac{\log(D_n/p_1(z))}{p_1^2(z)}\cdot m\left(\sqrt{\frac{m\log(dm)}{nh}} + \frac{m}{nh} + \sqrt{\frac{\log(1/h)}{nh}}\right). \quad \text{(C.8)}$$

We defer the proof of this lemma to Section F.4. Using Lemma C.4 we bound $I_1(z)$. Applying Cauchy-Schwarz inequality, we have

$$\begin{aligned}\sup_{z\in\mathcal{X}}|I_1(z)| &\le \sup_{z\in\mathcal{X}}\left(\frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\left(\mathbf{\Psi}_{i\bullet}^T(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)\right)^2\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\right)^{1/2} \\ &\le C\inf_{z\in\mathcal{X}}\left(\frac{\log(D_n/p_1(z))}{p_1^2(z)}\right)^{1/2}\cdot\sqrt{m}\left(\sqrt{\frac{m\log(dm)}{nh}} + \frac{m}{nh} + \sqrt{\frac{\log(1/h)}{nh}}\right)^{1/2},\end{aligned} \quad \text{(C.9)}$$

where the last inequality is due to Proposition 3.1, Lemma C.4 and

$$\sup_{z\in\mathcal{X}} n^{-1}\sum_{i=1}^n K_h(X_{i1}-z) = \sup_{z\in\mathcal{X}} p_1(z) + o(1).$$

For $I_2(z)$, we have the following inequality

$$
\begin{aligned}
\sup_{z\in\mathcal{X}} |I_2(z)| &\le \sup_{z\in\mathcal{X}} \frac{1}{n}\|\boldsymbol{\Psi}^T\mathbf{W}_z\mathbf{1}\|_{2,\infty}\|\boldsymbol{\theta}_z\|_1 \\
&\le \sup_{z\in\mathcal{X}} \frac{1}{n}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}_{\bullet j}\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z^{1/2}\|_2\|\mathbf{W}_z^{1/2}\mathbf{1}\|_2\|\boldsymbol{\theta}_z\|_1 \qquad \text{(C.10)}\\
&\le \frac{C}{m}\cdot \sup_{z\in\mathcal{X}} \sqrt{p(z)}\cdot\sqrt{m} = O(1/\sqrt{m}).
\end{aligned}
$$

Therefore, combining (C.9) and (C.10), we have

$$
\mathbb{P}\left(\left|\widehat{V}_n - \widehat{V}_n^{(1)}\right| > C\sqrt{r_n}m^{1/4}\right) \le n^{-1}.
$$

Under the scaling conditions of Theorem 3.7, there exists a constant c such that $\sqrt{r_n}m^{1/4} = O(n^{-c})$. Since $\sigma\xi_i \overset{d}{=} \varepsilon_i$, we also have $\sup_{z\in\mathcal{X}} \widehat{\mathbb{H}}_n^{(1)}(z) \overset{d}{=} \sup_{z\in\mathcal{X}} \widetilde{\mathbb{H}}_n(z)$. Combining with (C.7), we have

$$
\mathbb{P}\left(\left|\widehat{V}_n - \widetilde{V}_n^Z\right| > 2Cn^{-c}\right) \le 2n^{-1}.
$$

Therefore, we can bound the probability

$$
\begin{aligned}
\mathbb{P}(\widetilde{V}_n^Z \le \widehat{c}_n(\alpha) + 2Cn^{-c}) &\ge \mathbb{P}(\widetilde{V}_n^Z \le \widehat{c}_n(\alpha)) - \mathbb{P}(|\widehat{V}_n - \widetilde{V}_n^Z| > 2Cn^{-c}) \\
&\ge 1-\alpha - 2c/n^c, \qquad \text{(C.11)}
\end{aligned}
$$

which implies that the estimated quantile has the following lower bound

$$
\widehat{c}_n(\alpha) \ge c_n(\alpha + 2Cn^{-c}) - 2cn^{-c}. \qquad \text{(C.12)}
$$

Similarly, we also have $\widehat{c}_n(\alpha) \le c_n(\alpha - 2Cn^{-c}) + 2cn^{-c}$. By setting $\tau_n = 2Cn^{-c}$, $\varepsilon_{2n} = 2cn^{-c}$ and $\delta_{2n} = 2cn^{-c}$, we have

$$
\mathbb{P}\left(\widehat{c}_n(\alpha) \ge c_n(\alpha + 2Cn^{-c}) - 2cn^{-c} \text{ and } \ \widehat{c}_n(\alpha) \le c_n(\alpha - 2Cn^{-c}) + 2cn^{-c}\right) \le 2c/n^c,
$$

which verifies the condition **H3**.

Now, since we have checked the high-level conditions **H1** – **H3**, since the high-level conditions

H1 – **H4** are verified, the result follows from Corollary 3.1 in Chernozhukov et al. (2014a) such that

$$\mathbb{P}\big(f_1(z) \in \mathcal{C}^b_{n,\alpha}(z), \forall z \in \mathcal{X}\big) \geq 1 - \alpha - Cn^{-c},$$

which completes the proof of the theorem.

D Proof of Propositions 3.1

Let J be arbitrary subset of $[d]$ and for any $(\alpha, \boldsymbol{\beta}) = (\alpha, \boldsymbol{\beta}_2^T, \ldots, \boldsymbol{\beta}_d^T)^T \in \mathbb{C}_\beta^{(\kappa)}(J)$, where $\mathbb{C}_\beta^{(\kappa)}(J)$ is defined (3.1), we consider the functions $h_1(x_1) \equiv \alpha$ and $h_j(x_j) = \sum_{k=1}^m \boldsymbol{\beta}_{jk}\psi_{jk}(x_j)$, for $j = 2, \ldots, d$. From the cone restriction in (3.1), we have

$$\sum_{j \notin J, j \neq 1} \|\boldsymbol{\beta}_j\|_2 \leq \kappa \sum_{j \in J, j \neq 1} \|\boldsymbol{\beta}_j\|_2 + \kappa\sqrt{m}|\alpha|.$$

By the B-spline property de Boor (2001), there exists a constant $c_1 > 0$ such that

$$\begin{aligned}
\sqrt{\frac{m}{c_1}} \sum_{j \notin J} \|h_j\|_{L^2(\mu_z)} &\leq \sum_{j \notin J \setminus \{1\}} \|\boldsymbol{\beta}_j\|_2 + \sqrt{\frac{m}{c_1}}|\alpha| \\
&\leq \kappa \sum_{j \in J \setminus \{1\}} \|\boldsymbol{\beta}_j\|_2 + \left(\kappa + c_1^{-1/2}\right)\sqrt{m}|\alpha| \\
&\leq \left((\sqrt{c_1} + 1)\kappa + c_1^{-1/2}\right)\sqrt{m} \cdot \kappa \sum_{j \in J} \|h_j\|_{L^2(\mu_z)}. \qquad \text{(D.1)}
\end{aligned}$$

Let c be the smallest constant satisfying $c\kappa \geq \left(\sqrt{c_1} + 1\right)\kappa + c_1^{-1/2}$. Combining the above with (D.1), we have $(h_1, \ldots, h_d) \in \mathbb{C}_h^{(c\kappa)}(J)$.

We first aim to bound $\boldsymbol{\beta}_+^T \boldsymbol{\Sigma}_z \boldsymbol{\beta}_+ = \mathbb{E}\Big[K_h(X_1 - z)\big(\sum_{j=2}^d h_j(X_j)\big)^2\Big]$ as

$$\begin{aligned}
\mathbb{E}\Big[K_h(X_1 - z)\big(\textstyle\sum_{j=1}^d h_j(X_j)\big)^2\Big] &= \int K(u) \textstyle\sum_{j,k} h_j(x_j)h_k(x_k)p_{1,j,k}(z + uh, x_j, x_k)du dx_j dx_k \\
&= \int \textstyle\sum_{j,k} h_j(x_j)h_k(x_k)p_{1,j,k}(z, x_j, x_k)dx_j dx_k + O(h^2) \\
&\geq \frac{p_1(z)}{2} \cdot \mathbb{E}\Big[\big(\textstyle\sum_{j=2}^d h_j(X_j)\big)^2 \Big| X_1 = z\Big] = \frac{p_1(z)}{2}\big\|\textstyle\sum_{j=1}^d h_j\big\|^2_{L^2(\mu_z)}.
\end{aligned}$$

Therefore, for any $z \in \mathcal{X}$, by the definition in (3.4),

$$\boldsymbol{\beta}_+^T \boldsymbol{\Sigma}_z \boldsymbol{\beta}_+ \geq \frac{p_1(z)}{2} \left\| \sum_{j=1}^d h_j \right\|_{L^2(\mu_z)}^2 \geq \frac{p_1(z)}{2\beta_{2,c\kappa}^2(J)} \sum_{j \in J} \|h_j\|_{L^2(\mu_z)}^2 \tag{D.2}$$

Applying the cone restriction on $(h_1, \ldots, h_d)$, we further have

$$\begin{aligned}\left(\sum_{j \in J} \|h_j\|_{L^2(\mu_z)} \right)^2 &\geq \frac{1}{(c\kappa+1)^2} \left(\sum_{j=1}^d \|h_j\|_{L^2(\mu_z)} \right)^2 \\ &\geq \frac{1}{(c\kappa+1)^2} \sum_{j=1}^d \|h_j\|_{L^2(\mu_z)}^2 \geq \frac{Cm^{-1}}{(c\kappa+1)^2} \left(\sum_{j=2}^d \|\boldsymbol{\beta}_j\|_2^2 + m\alpha^2 \right).\end{aligned}$$

Combining the above inequality with (D.2), we obtain that with probability $q(\delta)$ for any $z \in \mathcal{X}$,

$$\boldsymbol{\beta}_+^T \boldsymbol{\Sigma}_z \boldsymbol{\beta}_+ \geq p_1(z) \cdot \frac{Cm^{-1} \bar{\beta}_{2,c\kappa}^{-2}}{s(c\kappa+1)^2} \left(\|\boldsymbol{\beta}\|_2^2 + m\alpha^2 \right), \quad \text{for any } \boldsymbol{\beta}_+ \in \mathbb{C}_\beta^{(\kappa)}(J).$$

This completes the proof.

E Auxiliary Lemmas for Estimation Rate

In this section, we give detailed proofs of technical lemmas stated in Section A. The principal technique used in the proofs of this section is the control of the suprema of empirical processes. In the entire section, we will abuse the notation σ_P^2 as the variance for certain empirical process.

E.1 Restricted eigenvalue condition

We provide a proof of Lemma A.1 in this section. Before stating the main part of the proof, we begin with a technical lemma.

Lemma E.1 (Resctricted eigenvalue condition)**.** With probability larger than $1 - c/(dm)$, for any $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^T)^T$, with $\alpha \in \mathbb{R}$ and $\boldsymbol{\beta} \in \mathbb{R}^{(d-1)m}$, and any $z \in \mathcal{X}$ it holds that

$$\boldsymbol{\theta}^T \widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta} \geq \boldsymbol{\theta}^T \boldsymbol{\Sigma}_z \boldsymbol{\theta} - C\left(\sqrt{\frac{m \log(dm)}{nh}} + \frac{m}{nh} + \frac{1}{\sqrt{nh}} + h^2 \right) \|\boldsymbol{\theta}\|_{1,2}^2,$$

where $\|\boldsymbol{\theta}\|_{1,2} = |\alpha| + \|\boldsymbol{\beta}\|_{1,2}$. Moreover, for any $j \in [d]$ there exists a constant C such that

$$\sup_{x \in \mathcal{X}} \frac{1}{n} \left\| \boldsymbol{\Psi}_{\bullet j} \mathbf{W}_z \boldsymbol{\Psi}_{\bullet j}^T \right\|_2^2 \leq C m^{-1}.$$

Proof. The proof strategy is to study suprema of the entries of $\widehat{\boldsymbol{\Sigma}}_z - \boldsymbol{\Sigma}_z$. We denote the (u, v) entry of $\boldsymbol{\Sigma}_z$ as $\boldsymbol{\Sigma}_z(u, v)$ and similarly for $\widehat{\boldsymbol{\Sigma}}_z$. Let $\mathbb{E}_n$ denote the empirical expectation. We first study the random variable

$$Z_{kk'jj'} = \sup_{z \in \mathcal{X}} (\mathbb{E}_n - \mathbb{E})[K_h(X_{i1} - z)\psi_{jk}(X_{ij})\psi_{j'k'}(X_{ij'})].$$

Notice that $Z_{kk'jj'} = \sup_{z \in \mathcal{X}} [\widehat{\boldsymbol{\Sigma}}_z - \boldsymbol{\Sigma}_z](1 + (j-2)m + k, 1 + (j'-2)m + k')$ for $j, j' \geq 2$ and $k \in [m]$. In order to bound $Z_{kk'jj'}$, we turn to study the covering number of the space

$$\mathcal{G}_h = \left\{ g_z(x_1, x_2, x_3) = h^{-1}K(h^{-1}(x_1 - z))\psi_{jk}(x_2)\psi_{j'k'}(x_3) \,\middle|\, z \in \mathcal{X}, x_1, x_2, x_3 \in \mathcal{X} \right\}.$$

Let $\mathcal{F}_h = \left\{ h^{-1}K(h^{-1}(\cdot - z)) \,|\, z \in \mathcal{X} \right\}$ and let $\|K\|_{\mathrm{TV}}$ be the total variation of $K(\cdot)$. From Lemma H.3, we have

$$\sup_Q N\left(\mathcal{F}_h, L^2(Q), \epsilon\right) \leq \left(\frac{2\|K\|_{\mathrm{TV}} A}{h\epsilon} \right)^4, \quad 0 < \epsilon < 1,$$

where the supremum is taken over all probability measures Q on $\mathbb{R}$. Let $\widetilde{\mathcal{F}}_h$ be an ϵ/L-cover of $\mathcal{F}_h$ with respect to Q, where $L \geq \|\psi_{jk}\|_\infty$ for any k. We construct an ϵ-cover for $\mathcal{G}_h$ with respect to $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{X_{i1}, \ldots, X_{id}}$ as

$$\widetilde{\mathcal{G}}_h = \left\{ f_1(x_1)\psi_{jk}(x_2)\psi_{j'k'}(x_3) \,|\, f_1 \in \widetilde{\mathcal{F}}_h \right\}.$$

For a function $g_z = h^{-1}K(h^{-1}(x_1 - z))\psi_{jk}(x_2)\psi_{j'k'}(x_3) \in \mathcal{G}_h$, let

$$\widetilde{g}_z = h^{-1}K(h^{-1}(x_1 - \widetilde{z}))\psi_{jk}(x_2)\psi_{j'k'}(x_3) \in \widetilde{\mathcal{G}}_h$$

be the corresponding element in the cover. Here $h^{-1}K(h^{-1}(x_1 - \widetilde{z})) \in \widetilde{\mathcal{F}}_h$ is the corresponding element in the cover for $h^{-1}K(h^{-1}(x_1 - z)) \in \mathcal{F}_h$. If we are able to show that $\|g_z - \widetilde{g}_z\|_{L^2(Q)}^2 \leq \epsilon^2$, then $\widetilde{\mathcal{G}}_h$ is the ϵ-cover for $\mathcal{G}_h$. Combining with Lemma 3, Giné and Nickl (2009) (also see Lemma

H.3), the covering number of $\mathcal{G}_h$ can be bounded as

$$N\left(\mathcal{G}_h, L^2(\mathbb{P}_n), \epsilon\right) \leq \left(\frac{2||K||_{\mathrm{TV}} AL}{h\epsilon}\right)^4. \tag{E.1}$$

Now we show $\|g_z - \widetilde{g}_z\|^2_{L^2(Q)} \leq \epsilon^2$:

$$\begin{aligned}\|g_z - \widetilde{g}_z\|^2_{L^2(Q)} &= \mathbb{E}_Q\left[\left(\left(K_h(X_1 - z) - K_h(X_1 - \widetilde{z})\right)\psi_{jk}(X_{ji})\psi_{j'k}(X_{j'i})\right]^2 \\ &\leq \mathbb{E}_Q\left[\left(\left(K_h(X_1 - z) - K_h(X_1 - \widetilde{z})\right)\right]^2 \leq \epsilon^2.\end{aligned}$$

Observe that all functions in $\mathcal{G}_h$ are bounded by $U = 4h^{-1}\|K\|_\infty$ and

$$\begin{aligned}\sigma_P^2 &:= \mathbb{E}\left[\left(K_h\left(X_1 - z\right)\left(\psi_{jk}(X_j)\psi_{j'k'}(X_{j'})\right)^2\right]\right. \\ &= h^{-2}\mathbb{E}\left[K^2\left(h^{-1}(X_1 - z)\right)\mathbb{E}\left[\left(\psi^2_{jk}(X_j)\psi^2_{j'k'}(X_{j'}) \,|\, X_1\right]\right]\right. \\ &\leq Bm^{-2}h^{-2}\mathbb{E}\left[K^2\left(h^{-1}(X_1 - z)\right)\right] \\ &= Bm^{-2}h^{-1}\int K^2(u)\, p_1(z + uh)du \leq B^2m^{-2}h^{-1},\end{aligned}$$

where the first and last inequalities are due to Assumption (**A1**). The bound above does not depend on the particular choice of z. If $m(nh)^{-1} = o(1)$, we have $n\sigma_P^2 \geq CU^2\log\left(U\sigma^{-1}\right)$, and from Lemma H.2, we have

$$\mathbb{E}[Z_{kk'jj'}] \leq C_1\sqrt{\frac{\log(C_2 m)}{nm^2h}}, \tag{E.2}$$

where the constants C_1, C_2 are independent of k, k', j, j'. As $|Z_{kk'jj'}| \leq 4h^{-1}$ and $\sigma_P^2 \leq Cm^{-2}h^{-2}$, we can apply Lemma H.4 to obtain

$$\mathbb{P}\Big(Z_{kk'jj'} \geq \mathbb{E}[Z_{kk'jj'}] + t\sqrt{Cm^{-2}h^{-1} + 4h^{-1}\mathbb{E}[Z_{kk'jj'}]} + 4t^2h^{-1}/3\Big) \leq \exp(-nt^2). \tag{E.3}$$

For $t = \log d/\sqrt{n}$, there exists a constant C such that

$$Z_{kk'jj'} \leq C\log dm/\sqrt{nm^2h} + C/(nh)$$

with probability $1 - 1/d$. Combining (E.2) with (E.3), there exists a constant C such that

$$\begin{aligned}
&\mathbb{P}\left(\max_{j,j'\geq 2,k,k'\in[m]} |Z_{kk'jj'}| > 2\mathbb{E}[Z_{kk'jj'}] + t\sqrt{Cm^{-2}h^{-1} + 4h^{-1}\mathbb{E}[Z_{kk'jj'}]} + 4t^2h^{-1}/3\right) \\
&\leq \mathbb{P}\Big(\max_{k,k'\in[m],j,j'\geq 2} |Z_{kk'jj'} - \mathbb{E}[Z_{kk'jj'}]| > t\sqrt{Cm^{-2}h^{-1} + 4h^{-1}\mathbb{E}[Z_{kk'jj'}]} + 4t^2h^{-1}/3 + \mathbb{E}[Z_{kk'jj'}]\Big) \\
&\leq (dm)^2 \exp\left(-nt^2\right).
\end{aligned}$$

Let $t = 3\sqrt{\log(dm)/n}$, $n_{jk} = 1 + (j-2)m + k$ and $n_{j'k'} = 1 + (j'-2)m + k')$ and we obtain that

$$\sup_{z\in\mathcal{X}} \max_{j,j'\geq 2,k,k'\in[m]} \left|\left[\widehat{\mathbf{\Sigma}}_z - \mathbf{\Sigma}_z\right](n_{jk}, n_{j'k'})\right| = O_P\left(\frac{1}{nh} + \sqrt{\frac{\log(dm)}{nm^2h}}\right). \tag{E.4}$$

Similarly, we define $\bar{Z}_{kj} = \sup_{z\in\mathcal{X}}(\mathbb{E}_n - \mathbb{E})[K_h(X_{i1} - z)\psi_{jk}(X_{ij})]$. Following the similar procedure as above, we apply Lemma H.2 to obtain that for some constant C,

$$\sigma_P^2 := \mathbb{E}\left[\left(K_h\left(X_1 - z\right)\psi_{jk}(X_j)\right)^2\right] \leq Cm^{-1}h^{-1},$$

and $U \leq h^{-1}$, which implies the following inequality

$$\mathbb{E}[\bar{Z}_{kj}] \leq C_1\sqrt{\frac{\log(C_2 m)}{nmh}}. \tag{E.5}$$

We now turn to study the remaining entries of $\widehat{\mathbf{\Sigma}}_z - \mathbf{\Sigma}_z$. Using the same arguments as in (E.4) and (E.5), we can derive an upper bound on the difference

$$\sup_{z\in\mathcal{X}} \max_{j\geq 2} \left|\widehat{\mathbf{\Sigma}}_z(1 + (j-2)m + k, 1) - \mathbf{\Sigma}_z(1 + (j-2)m + k, 1)\right| = O_P\left(\frac{1}{nh} + \sqrt{\frac{\log(dm)}{nmh}}\right). \tag{E.6}$$

From Assumption **(A1)**, the density function of X_1, $p_1(x)$, is smooth. Recall that $\widehat{p}_1(z) = n^{-1}\sum_{i=1}^n K_h(X_{i1} - z)$. Applying the supreme norm rate for a kernel density estimator established in Theorem 2.3 of Giné and Guillou (2002), we have $\|\widehat{p}_1 - \mathbb{E}[\widehat{p}_1]\|_\infty = O_P\big(\sqrt{\log(1/h)/(nh)}\big)$ and therefore we can get the rate

$$\sup_{z\in\mathcal{X}} \left|\widehat{\mathbf{\Sigma}}_z(1,1) - \mathbf{\Sigma}_z(1,1)\right| = \sup_{z\in\mathcal{X}} \left|\widehat{p}_1(z) - \mathbb{E}[\widehat{p}_1(z)]\right| = O_P\left(\sqrt{\frac{\log(1/h)}{nh}}\right). \tag{E.7}$$

Combining (E.4), (E.6) and (E.7), according to Hölder inequality, we have for any $z \in \mathcal{X}$

$$\begin{aligned}\left|\boldsymbol{\theta}^T(\widehat{\boldsymbol{\Sigma}}_z - \boldsymbol{\Sigma}_z)\boldsymbol{\theta}\right| &\le \|\boldsymbol{\theta}\|_1^2 \|\widehat{\boldsymbol{\Sigma}}_z - \boldsymbol{\Sigma}_z\|_{\max} \\ &\le \|\boldsymbol{\theta}\|_{1,2}^2 \left\{ \sup_{z\in\mathcal{X}} \max_{t,t'\neq 1} m \left|\widehat{\boldsymbol{\Sigma}}_z(t,t') - \boldsymbol{\Sigma}_z(t,t')\right| + \sup_{z\in\mathcal{X}} \left|\widehat{\boldsymbol{\Sigma}}_z(1,1) - \boldsymbol{\Sigma}_z(1,1)\right| \right\} \\ &\le C\left(\sqrt{\frac{m\log(dm)}{nh}} + \frac{m}{nh} + \sqrt{\frac{\log(1/h)}{nh}}\right)\|\boldsymbol{\theta}\|_{1,2}^2, \end{aligned} \tag{E.8}$$

which completes the proof of the first part of the lemma.

An upper bound on $\sup_{x\in\mathcal{X}} n^{-1}\|\boldsymbol{\Psi}_{\bullet j}\mathbf{W}_z\boldsymbol{\Psi}_{\bullet j}^T\|_2^2$ can be obtained in a way similar to the proof of Lemma 6.2 in Zhou et al. (1998). For any $\boldsymbol{\beta}_j = (\beta_1, \ldots, \beta_m)^T$, let $u(x_j) = \sum_{k=1}^m \beta_k \psi_{jk}(x_j)$. Let the joint density function between X_1 and X_j be $p_{1,j}(x_1, x_j)$. From Assumption (**A1**), we have for any $z \in \mathcal{X}$,

$$\begin{aligned}\frac{1}{n}\boldsymbol{\beta}_j^T\mathbb{E}[\boldsymbol{\Psi}_{\bullet j}\mathbf{W}_z\boldsymbol{\Psi}_{\bullet j}^T]\boldsymbol{\beta}_j &= \int \frac{1}{h}K\left(\frac{x_1 - z}{h}\right)u^2(x_j)p_{1,j}(x_1,x_j)dx_1dx_j \\ &\le C\int K(u)du \int u^2(x_j)dx_j \le Cm^{-1}\sum_{k=1}^m \beta_k^2. \end{aligned} \tag{E.9}$$

Furthermore, we also have

$$\begin{aligned}\sup_{x\in\mathcal{X}} \frac{1}{n}\boldsymbol{\beta}_j^T\mathbb{E}[\boldsymbol{\Psi}_{\bullet j}\boldsymbol{\Psi}_{\bullet j}^T \mid X_1 = x]\boldsymbol{\beta}_j &= \int u^2(x_j)\frac{p_{1,j}(x_1,x_j)}{p_1(x)}dx_1dx_j \\ &\le B\int K(u)du\int u^2(x_j)dx_j \le Cm^{-1}\sum_{k=1}^m \beta_k^2. \end{aligned} \tag{E.10}$$

Let $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \boldsymbol{\delta}_{X_{i1},X_{ij}}$. We write the integration as

$$\sup_{z\in\mathcal{X}} \int K_h(x_1 - z)u^2(x_j)d\mathbb{P}_n = I_1 + I_2, \text{ where}$$

$$I_1 = \sup_{z\in\mathcal{X}} \int K_h(x_1 - z)u^2(x_j)d\mathbb{P}_{X_1,X_j} \text{ and } I_2 = \sup_{z\in\mathcal{X}} \left|\int K_h(x_1 - z)u^2(x_j)d(\mathbb{P}_n - \mathbb{P}_{X_1,X_j})\right|.$$

Due to (E.9), we have $I_1 = O_P(m^{-1})\|\boldsymbol{\beta}_j\|_2^2$ and a similar argument to one in Lemma 6.2 of Zhou et al. (1998) will derive $I_2 = o(h)\|\boldsymbol{\beta}_j\|_2^2$. This completes the proof. □

Based on Lemma E.1, the remaining step is to prove Lemma A.1.

Proof of Lemma A.1. We can derive the restricted eigenvalue condition on the cone from Lemma E.1.

We apply Lemma E.1 in the last step. If the cone condition

$$\sum_{j\in S^c}\|\boldsymbol{\beta}_j\|_2 \le 3\sum_{j\in S}\|\boldsymbol{\beta}_j\|_2 + 3\sqrt{m}|\alpha|$$

is satisfied, by Hölder inequality, we have the upper bound

$$\|\boldsymbol{\beta}\|_{1,2} \le 4\sum_{j\in S}\|\boldsymbol{\beta}_j\|_2 + 3\sqrt{m}|\alpha| \le 4\sqrt{s}\|\boldsymbol{\beta}\|_2 + 3\sqrt{m}|\alpha|.$$

With large probability, we have the following inequality

$$\begin{aligned}
\boldsymbol{\theta}^T\widehat{\boldsymbol{\Sigma}}_z\boldsymbol{\theta} &\ge \boldsymbol{\theta}^T\boldsymbol{\Sigma}_z\boldsymbol{\theta} - C\left(\sqrt{\frac{m\log(dm)}{nh}} + \frac{m}{nh} + \frac{1}{\sqrt{nh}} + h^2\right)\|\boldsymbol{\theta}\|_{1,2}^2 \\
&\ge \rho_{\min}|\alpha|^2 + \rho_{\min}\|\boldsymbol{\beta}\|_2/m - C\left(\sqrt{\frac{m\log(dm)}{nh}} + \frac{m}{nh} + \frac{1}{\sqrt{nh}} + h^2\right)\left(4m|\alpha|^2 + 4s\|\boldsymbol{\beta}\|_2^2\right) \\
&\ge \rho_{\min}|\alpha|^2/2 + \rho_{\min}m^{-1}\|\boldsymbol{\beta}\|_2^2/2
\end{aligned}$$

for any $z \in \mathcal{X}$ and sufficiently large n if $s\sqrt{m^3\log(dm)/(nh)} + sm^2/(nh) = o(1)$. $\square$

E.2 Proof of Lemma A.2

The proof can be separated into two cases: $j = 1$ and $j \ge 2$. For the simplicity of notation, we write $\delta_i(z)$ as δ_i in this proof. We first consider the situation when $j \ge 2$ and prove (A.1) and (A.3). From Lemma A.1,

$$\sup_{z\in\mathcal{X}}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}_{\bullet j}\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z^{1/2}\|_2/\sqrt{n} = \sup_{z\in\mathcal{X}}\|\boldsymbol{\Psi}_{\bullet j}\mathbf{W}_z\boldsymbol{\Psi}_{\bullet j}^T\|_2/\sqrt{n} \le \rho_{\max}m^{-1/2}$$

with high probability. Therefore

$$\sup_{z\in\mathcal{X}}\frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z\boldsymbol{\delta}\|_2 \le \sup_{z\in\mathcal{X}}\frac{1}{n}\|\mathbf{W}_z^{1/2}\boldsymbol{\Psi}_{\bullet j}\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z^{1/2}\|_2\|\mathbf{W}_z^{1/2}\boldsymbol{\delta}\|_2 \le \frac{C}{\sqrt{m}}\cdot\sup_{z\in\mathcal{X}}\frac{1}{\sqrt{n}}\|\mathbf{W}_z^{1/2}\boldsymbol{\delta}\|_2. \quad \text{(E.11)}$$

To complete the proof, we need a bound on

$$\sup_{z\in\mathcal{X}} \frac{1}{n}\|\mathbf{W}_z^{1/2}\boldsymbol{\delta}\|_2^2 = \sup_{z\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1}-z)\delta_i^2. \tag{E.12}$$

Using Equation (20) in Zhou et al. (1998) on B-spline, we have

$$\delta_i^2 = \Big(\sum_{j=2}^{d} f_{jz}(X_{ji}) - f_{nj;z}(X_{ji})\Big)^2 \le s^2 m^{-4}.$$

Define the following empirical process

$$U_n(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1}-z)\delta_i^2 - \mathbb{E}\big[K_h(X_{11}-z)\delta_1^2\big].$$

Applying Hoeffding's inequality (Hoeffding, 1963), we have

$$\mathbb{P}\left(\sup_{z\in\mathcal{X}} U_n(z) - \mathbb{E}\left[\sup_{z\in\mathcal{X}} U_n(z)\right] > t\right) \le \exp\left(-C\frac{nh^2t^2}{(sm^{-4})^2}\right). \tag{E.13}$$

Let

$$\mathcal{G}_h'' = \left\{g_z(x_1,x_2) = h^{-1}K(h^{-1}(x_1-z))\delta^2(x_2)\,|\, z\in\mathcal{X}, x_1\in\mathcal{X}, x_2\in\mathcal{X}^{d-1}\right\},$$

where $\delta(x_2) = \sum_{j=2}^{d} f_j(x_{2j}) - f_{nj}(x_{2j})$. Similar to the covering number of $\mathcal{G}_h$ in (E.1), since $\delta^2(x_2) \le sm^{-4}$ for any x_2, we have for any measure Q,

$$\sup_{Q} N\left(\mathcal{G}_h'', L^2(Q), \epsilon\right) \le \left(\frac{2\sqrt{s}||K||_{\mathrm{TV}}A}{m^2h\epsilon}\right)^4.$$

Furthermore, $\sigma_P^2 := \mathbb{E}[K_h(X_{i1}-z)\delta_i^2]^2 = O((sm^{-4})^2h^{-1})$. Since $g \le U := Ch^{-1}(sm^{-4})$ for any $g\in\mathcal{G}_h''$ and $m^4(sn)^{-1} = o(1)$, we have $n\sigma_P^2 \ge C_1U^2\log(C_2\sqrt{sm^{-4}}U/\sigma)$. By Lemma H.2, we have

$$\mathbb{E}\left[\sup_{z\in\mathcal{X}} U_n(z)\right] \le C\frac{sm^{-4}}{\sqrt{nh}}\sqrt{\log(m^2/\sqrt{s}h)}. \tag{E.14}$$

We set $t = Cs(m^4h)^{-1}\sqrt{\log n/n}$ in (E.13) and combine it with (E.14) to obtain that, with probability at least $1-1/n$,

$$\sup_{z\in\mathcal{X}} U_n(z) \le C\frac{sm^{-4}}{\sqrt{nh}}\sqrt{\log\left(m^2/\sqrt{s}h\right)} + C\frac{s\sqrt{\log n/n}}{m^4h}. \tag{E.15}$$

Finally, we bound the maximal of the expectation by

$$\begin{aligned}\sup_{z\in\mathcal{X}} \mathbb{E}[K_h(X_{11}-z)\delta_1^2] &= \sup_{z\in\mathcal{X}} \int K_h(t-z)dP_{X_1}(t)\delta^2(u)dP_{X_{>2}|X_1=t}(u) \\ &\le Csm^{-4} \sup_{z\in\mathcal{X}} \int K_h(t-z)dP_{X_1}(t) \le Csm^{-4}.\end{aligned} \tag{E.16}$$

Combining (E.15) and (E.16), with probability at least $1-1/n$, we have

$$\begin{aligned}\sup_{z\in\mathcal{X}} \frac{1}{n}\|\mathbf{W}_z^{1/2}\boldsymbol{\delta}\|_2^2 &= \sup_{z\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\delta_i^2 \\ &\le \sup_{z\in\mathcal{X}} U_n(z) + \sup_{z\in\mathcal{X}} \mathbb{E}[K_h(X_{11}-z)\delta_1^2] \\ &\le C\frac{sm^{-4}}{\sqrt{nh}}\sqrt{\log(m^2/\sqrt{sh})} + C\frac{s\sqrt{\log n/n}}{m^4 h} + Csm^{-4} \\ &= O(sm^{-4}),\end{aligned} \tag{E.17}$$

where the last equality is due to $2/\sqrt{nh^2} = o(1)$. Therefore, we prove the upper bound in (A.3). Combing (E.17) with (E.11), we have we can conclude that

$$\sup_{z\in\mathcal{X}} \max_{j\ge 2} \frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T \mathbf{W}_z\boldsymbol{\delta}\|_2 \le \frac{C}{\sqrt{m}}\cdot \sup_{z\in\mathcal{X}} \frac{1}{\sqrt{n}}\|\mathbf{W}_z^{1/2}\boldsymbol{\delta}\|_2 \le C\sqrt{\frac{s}{m^5}}.$$

This gives us the rate in (A.1).

The final step is to prove (A.2). Recall that $\boldsymbol{\Psi}_{\bullet 1} = (1,\dots,1)^T$. For the case $j=1$, following the proof for (A.3). According to (E.12), we have $|\delta_i| \le sm^{-2}$ for any $i\in[n]$. Let

$$U_n'(z) = \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\delta_i - \mathbb{E}[K_h(X_{11}-z)\delta_1].$$

We use Hoeffding's inequality (Hoeffding, 1963) again and obtain

$$\mathbb{P}\left(\sup_{z\in\mathcal{X}} U_n'(z) - \mathbb{E}\left[\sup_{z\in\mathcal{X}} U_n'(z)\right] > t\right) \le \exp\left(-C\frac{nh^2t^2}{sm^{-4}}\right). \tag{E.18}$$

Applying symmetrization inequality again, we have

$$\mathbb{E}\left[\sup_{z\in\mathcal{X}} U_n'(z)\right] \le 2\mathbb{E}\bigg[\sup_{z\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^n \xi_i K_h(X_{i1}-z)|\delta_i|\bigg],$$

where $\{\xi_i\}_{i=1}^n$ are i.i.d. Rademacher variables independent of data. Let

$$\widetilde{\mathcal{G}}_h'' = \left\{ g_z(x_1, x_2) = h^{-1}K(h^{-1}(x_1 - z))\delta(x_2) \,|\, z \in \mathcal{X}, x_1 \in \mathcal{X}, x_2 \in \mathcal{X}^{d-1} \right\},$$

where $\delta(x_2) = \left|\sum_{j=2}^d f_j(x_{2j}) - f_{nj}(x_{2j})\right|$. Just as the covering number of $\mathcal{G}_h''$, we also have $\delta(x_2) \leq sm^{-2}$ for any x_2, we have for any measure Q,

$$\sup_Q N\left(\widetilde{\mathcal{G}}_h'', L^2(Q), \epsilon\right) \leq \left(\frac{2s^{1/2}||K||_{\mathrm{TV}}A}{m^2h\epsilon}\right)^4.$$

The variance of the process $\sigma_P^2 := \mathbb{E}[K_h(X_{i1}-z)\delta_i]^2 = O(sm^{-4}h^{-1})$. Since $g \leq U := Ch^{-1}(sm^{-4})^{1/2}$ for any $g \in \mathcal{G}_h''$ and $m^4(sn)^{-1} = o(1)$, we have $n\sigma_P^2 \geq C_1U^2\log(C_2s^{1/4}m^{-1}U/\sigma)$. Applying Lemma H.2 again, we have

$$\mathbb{E}\left[\sup_{z\in\mathcal{X}} U_n'(z)\right] \leq C\frac{\sqrt{s}m^{-2}}{\sqrt{nh}}\sqrt{\log(m/\sqrt{sh})}. \tag{E.19}$$

We let $t = C\sqrt{s}(m^2h)^{-1}\sqrt{\log n/n}$ in (E.18) and use it with (E.19). Therefore, we achieve with probability at least $1 - 1/n$,

$$\sup_{z\in\mathcal{X}} U_n'(z) \leq C\frac{\sqrt{s}m^{-2}}{\sqrt{nh}}\sqrt{\log\left(m^2/\sqrt{sh}\right)} + C\frac{\sqrt{s\log n/n}}{m^2h}. \tag{E.20}$$

We again bound the supreme of the expectation

$$\begin{aligned}\sup_{z\in\mathcal{X}} \mathbb{E}[K_h(X_{11} - z)\delta_1] &= \sup_{z\in\mathcal{X}} \int K_h(t-z)dP_{X_1}(t)\delta(u)dP_{X_{>2}|X_1=t}(u) \\ &\leq C\sqrt{s}m^{-2}\sup_{z\in\mathcal{X}} \int K_h(t-z)dP_{X_1}(t) \leq C\sqrt{s}m^{-2}.\end{aligned} \tag{E.21}$$

Combining (E.20) and (E.21), with probability at least $1 - 1/n$, we have

$$\begin{aligned}\sup_{z\in\mathcal{X}} \frac{1}{n}|\mathbf{\Psi}_{\bullet 1}^T\mathbf{W}_z\boldsymbol{\delta}_z| &\leq \sup_{z\in\mathcal{X}} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1} - z)|\delta_i| \\ &\leq \sup_{z\in\mathcal{X}} U_n'(z) + \sup_{z\in\mathcal{X}} \mathbb{E}[K_h(X_{11} - z)\delta_1] = O(\sqrt{s}m^{-2}).\end{aligned}$$

Therefore, we prove the upper bound in (A.2) which completes the proof of the lemma.

E.3 Proof of Lemma A.3

For $j \geq 2$, we bound the two terms $\sup_{z\in\mathcal{X}} \max_{j\geq 2} \frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T \mathbf{W}_z \boldsymbol{\xi}_z\|_2$ and $\sup_{z\in\mathcal{X}} \max_{j\geq 2} \frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T \mathbf{W}_z \boldsymbol{\zeta}_z\|_2$ separately. To bound the first term, let $\Delta f_z(x) = f_1(x) - f_1(z)$ and $\boldsymbol{\Psi}_{ij}$ be the ith row of $\boldsymbol{\Psi}_{\bullet j}$. We can rewrite the suprema as

$$\sup_{z\in\mathcal{X}} \max_{j\geq 2} \frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T \mathbf{W}_z \boldsymbol{\xi}_z\|_2 = \max_{j\geq 2} \sup_{z\in\mathcal{X}} \sup_{\mathbf{v}\in\mathbb{B}^m} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}^T\boldsymbol{\Psi}_{ij}. \tag{E.22}$$

Let $N_v = \{\mathbf{v}_1, \ldots, \mathbf{v}_M\}$ be a 1/2-covering of the sphere $\mathbb{B}^m = \{\mathbf{v} \in \mathbb{R}^m \mid \|\mathbf{v}\|_2 \leq 1\}$. Observe that for any $\mathbf{v} \in \mathbb{B}^m$, there exists $\pi(\mathbf{v}) \in N_v$ such that $\|\mathbf{v} - \pi(\mathbf{v})\|_2 \leq 1/2$. Therefore we have

$$\begin{aligned}
&\sup_{\mathbf{v}\in\mathbb{B}^m} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}^T\boldsymbol{\Psi}_{ij} \\
&\leq \sup_{k\in[M]} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}_k^T\boldsymbol{\Psi}_{ij} + \sup_{\mathbf{v}\in\frac{1}{2}\mathbb{B}^m} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}^T\boldsymbol{\Psi}_{ij} \\
&\leq \sup_{k\in[M]} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}_k^T\boldsymbol{\Psi}_{ij} + \frac{1}{2}\sup_{\mathbf{v}\in\mathbb{B}^m} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}^T\boldsymbol{\Psi}_{ij}.
\end{aligned}$$

If we move the second term on the right hand side of the last inequality to the left hand side, we obtain the inequality that

$$\sup_{\mathbf{v}\in\mathbb{B}^m} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}^T\boldsymbol{\Psi}_{ij} \leq 2\sup_{k\in[M]} \frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\Delta f_z(X_{1i})\mathbf{v}_k^T\boldsymbol{\Psi}_{ij}.$$

Therefore, in order to bound (E.22), we need to study the following empirical process

$$V_n(z) = \max_{j\geq 2, k\in[M]} \sup_{z\in\mathcal{X}} \left\{ \frac{1}{n}\sum_{i=1}^n K_h(X_{1i}-z)\Delta f_z(X_{1i})\mathbf{v}_k^T\boldsymbol{\Psi}_{ij} - \mathbb{E}[K_h(X_{11}-z)\Delta f_z(X_{11})\mathbf{v}_k^T\boldsymbol{\Psi}_{ij}] \right\}.$$

We define the following function class

$$\mathcal{G}_h''' = \left\{ g_z(x_1, x_2) = h^{-1}K((x_1-z)/h)\Delta f_z(x_1)\sum_{t=1}^m \mathbf{v}_{kt}\psi_t(x_j) \,\Big|\, j \geq 2, k \in [M], z \in \mathcal{X} \right\}$$

and, similarly to argument in the covering number of $\mathcal{G}_h$ in (E.1), we have

$$\sup_Q N\left(\mathcal{G}_h''', L^2(Q), \epsilon\right) \leq dM\left(\frac{2\sqrt{m}||K||_{\mathrm{TV}}A}{h\epsilon}\right)^4.$$

From (E.10), we bound the maximal of the expectation by

$$\sup_{x\in\mathcal{X}} \mathbb{E}\big[(\mathbf{v}_k^T\mathbf{\Psi}_{ij})^2 \,|\, X_1 = x\big] \leq C\|\mathbf{v}_k\|^2 m^{-1}.$$

Furthermore, we can bound the variance by expanding the expectation as the integration and applying the Taylor expansion as follows

$$\begin{aligned}
\sigma_P^2 &:= \mathbb{E}[K_h(X_1 - z)\Delta f_z(X_1)\mathbf{v}_k^T\mathbf{\Psi}_{ij}]^2 \\
&= h^{-2}\int K^2\left(\frac{x-z}{h}\right)(f_1(x) - f_1(z))^2 p_{X_1}(x)dx \cdot \mathbb{E}\left[(\mathbf{v}_k^T\mathbf{\Psi}_{ij})^2 \,|\, X_1 = x\right] \\
&\leq C(mh)^{-1}\int K^2(u)(f_1(z+hu) - f_1(z))^2 p_{X_1}(z+hu)du \\
&= C(mh)^{-1}\int K^2(u)(f_1'(z)uh + o(uh))^2(p_{X_1}(z) + p_{X_1}'(z)uh + o(uh))du \\
&= Cm^{-1}[f'(z)]^2 p_{X_1}(z)\int u^2K^2(u)du \cdot h + o(m^{-1}h) = Chm^{-1}.
\end{aligned}$$

The uniform upper bound of $K_h(x-z)\Delta f_z(x)$ can be studied under two cases: (1) x is out of the support and (2) x is in the support. In particular, we have

- if $x \notin [z-h, z+h]$, then $K_h(x-z)\Delta f_z^2(x) = 0$;
- if $x \in [z-h, z+h]$, then, by mean value theorem,

$$K_h(x-z)\Delta f_z(x) \leq h^{-1}K(h^{-1}(x-z))|f(x) - f(z)| \leq h^{-1}\|K\|_\infty\|f'\|_\infty^2\cdot(2h) = 4\|f_1'\|_\infty^2\|K\|_\infty.$$

Combining with the fact that $|\mathbf{v}_k^T\mathbf{\Psi}_{ij}| \leq \sqrt{m}$ for any i, j, k, we conclude that $g \leq U := 4\|f_1'\|_\infty^2\|K\|_\infty\sqrt{m}$ for any $g \in \mathcal{G}_h'''$. Therefore by Lemma H.2 and $M = 6^m$, we have

$$\begin{aligned}
\mathbb{E}V_n(z) &\leq C\sqrt{\frac{h\log(dMh^{-1})}{mn}} + C\frac{\sqrt{m}\cdot\log(dMh^{-1})}{n} \\
&= C\sqrt{\frac{h\log(dh^{-1})}{n}} + C\frac{m^{3/2}\log(dh^{-1})}{n}.
\end{aligned} \tag{E.23}$$

Similar to the analysis of σ_P^2, we also expand the expectation of the process as the integration and use the Taylor expansion to bound it as follows

$$\begin{aligned}
&\mathbb{E}\big[K_h(X_1-z)\Delta f_z(X_1)\mathbf{v}_k^T\boldsymbol{\Psi}_{ij}\big] \\
&\quad= h^{-1}\int K\left(\frac{x-z}{h}\right)(f_1(x)-f_1(z))p_{X_1}(x)dx\cdot\mathbb{E}\left[\mathbf{v}_k^T\boldsymbol{\Psi}_{1j}\,|\,X_1=x\right]dx \\
&\quad= \int K(u)\big[f_1'(z)uh+f_1''(z)(uh)^2/2+o(uh)^2\big]\big[p_{X_1}(z)+p_{X_1}'(z)uh+o(uh)\big] \\
&\qquad\quad\cdot\left(\mathbb{E}\left[\mathbf{v}_k^T\boldsymbol{\Psi}_{1j}\,|\,X_1=z\right]+uh\frac{d}{dz}\mathbb{E}\left[\mathbf{v}_k^T\boldsymbol{\Psi}_{1j}\,|\,X_1=z\right]+o(uh)\right)du\le Ch^2/\sqrt{m}.
\end{aligned}\tag{E.24}$$

The last inequality is due to the fact that $K(\cdot)$ is an even function, $\|\psi_{jk}\|_\infty\le 1$ and from (E.10). Moreover, the constant is independent to j,k and z. Using Lemma H.4, we have

$$\mathbb{P}\left(V_n(z)-\mathbb{E}V_n(z)>t\sqrt{2(\sigma_P^2+2U\mathbb{E}V_n(z))}+\frac{2Ut^2}{3}\right)\le\exp\left(-nt^2\right).\tag{E.25}$$

Combining (E.23) and (E.24) with (E.25) for $t=\sqrt{\log n/n}$, with probability at least $1-1/n$, we have

$$\begin{aligned}
\sup_{z\in\mathcal{X}}\max_{j\ge 2}\frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z\boldsymbol{\xi}_z\|_2 &\le 2\max_{j\ge 2,k\in[M]}\sup_{z\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^n K_h(X_{1i}-z)\Delta f_z(X_{1i})\mathbf{v}_k^T\boldsymbol{\Psi}_{ij} \\
&\le V_n(z)+\max_{j,k}\sup_{z\in\mathcal{X}}\mathbb{E}[K_h(X_1-z)\Delta f_z(X_1)\mathbf{v}_k^T\boldsymbol{\Psi}_{ij}] \\
&\le C\sqrt{\frac{h\log(dh^{-1})}{n}}+C\frac{m^{3/2}\log(dh^{-1})}{n}+C\frac{h^2}{\sqrt{m}},
\end{aligned}\tag{E.26}$$

where the last equality is because of $n^{-1}h=o(1)$.

Now we bound $\sup_{z\in\mathcal{X}}\max_{j\ge 2}\frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z\boldsymbol{\zeta}_z\|_2$. The procedure is similar to the first part of the proof. We again apply the $1/2$-covering of $\mathbb{B}^d$ so that

$$\sup_{z\in\mathcal{X}}\max_{j\ge 2}\frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z\boldsymbol{\zeta}_z\|_2\le 2\max_{j\ge 2,k\in[M]}\sup_{z\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^n K_h(X_{i1}-z)\zeta_i(z)\mathbf{v}_k^T\boldsymbol{\Psi}_{ij}.$$

Motived by the above argument, we now turn to study the following empirical process

$$V_n'(z)=\max_{j\ge 2,k\in[M]}\sup_{z\in\mathcal{X}}\left\{\frac{1}{n}\sum_{i=1}^n K_h(X_{1i}-z)\zeta_i(z)\mathbf{v}_k^T\boldsymbol{\Psi}_{ij}-\mathbb{E}[K_h(X_{11}-z)\zeta_i(z)\mathbf{v}_k^T\boldsymbol{\Psi}_{ij}]\right\}$$

and the function class inspired from the above empirical process

$$\mathcal{G}_h'''' = \left\{ g_z(x_1, x_2) = h^{-1} K((x_1 - z)/h) \zeta_i(z) \sum_{t=1}^{m} \mathbf{v}_{kt} \psi_t(x_j) \,\middle|\, j \geq 2, k \in [M], z, x_1, x_j \in \mathcal{X} \right\}.$$

Our method of bound the supreme of the process is same as the proof of previous lemmas. We need to study the covering number of the function space. Assembling the concentration inequality of the suprema with the upper bound of the expectation of suprema and suprema of the expectation, we will arrive at the final bound. Therefore, we first bound the covering number

$$\sup_Q N\left(\mathcal{G}_h'''', L^2(Q), \epsilon\right) \leq dM \left(\frac{2\sqrt{m} ||K||_{\mathrm{TV}} A}{h\epsilon} \right)^4.$$

Using Definition 4.1, there exists $L_1(z, x_{\backslash 1})$ such that the approximation error is bounded by

$$|\zeta_i(z)| = \left| f(X_{i1}, \ldots . X_{id}) - \sum_{j=1}^{d} f_{jz}(X_{ij}) \right| \leq |L_1(z, x_{\backslash 1})| \cdot |X_{i1} - z| + |U_j(z)|(X_{i1} - z)^2.$$

Therefore, the variance of the process V_n' can be bounded by computing the expectation

$$\sigma_P^2 := \mathbb{E}[K_h(X_1 - z)\zeta_i(z)\mathbf{v}_k^T \mathbf{\Psi}_{ij}]^2 \leq Chm^{-1} \tag{E.27}$$

and $g \leq U := 4\|L_1(z, x_{\backslash 1})\|_\infty^2 \|K\|_\infty \sqrt{m}$ for all $g \in \mathcal{G}_h''''$. Lemma H.2 gives us

$$\mathbb{E}V_n'(z) \leq C\sqrt{\frac{h\log(dh^{-1})}{n}} + C\frac{m^{3/2}\log(dh^{-1})}{n}. \tag{E.28}$$

Denote $\kappa(x) = \mathbb{E}[L_1(z, X_{\backslash 1}) \,|\, X_1 = x]$. Using Definition 4.1,

$$\begin{aligned}
&\mathbb{E}[K_h(X_1 - z)\zeta_i(z)\mathbf{v}_k^T \mathbf{\Psi}_{ij}] \\
&\quad = h^{-1} \int K\left(\frac{x - z}{h}\right) (\mathbb{E}[L_1(z, X_{\backslash 1}) \,|\, X_1 = x](x_1 - z) + U_j(z)(x - z)^2) \\
&\qquad\qquad \cdot \mathbb{E}\left[\mathbf{v}_k^T \mathbf{\Psi}_{1j} \,|\, X_1 = x\right] p_{X_1}(x) dx \\
&\quad = \int K(u)(\kappa(z)uh + (\kappa''(z) + \|U_j\|_\infty)(uh)^2/2 + o(uh)^2)(p_{X_1}(z) + p_{X_1}'(z)uh + o(uh)) \\
&\qquad \cdot \left(\mathbb{E}\left[\mathbf{v}_k^T \mathbf{\Psi}_{1j} \,|\, X_1 = z\right] + uh\frac{d}{dz}\mathbb{E}\left[\mathbf{v}_k^T \mathbf{\Psi}_{1j} \,|\, X_1 = z\right] + o(uh) \right) du \leq Ch^2/\sqrt{m}.
\end{aligned} \tag{E.29}$$

The last inequality is due to Assumption (**A1**). Since $\mathcal{X}$ is compact, $\mathbb{E}\left[\mathbf{v}_k^T\boldsymbol{\Psi}_{1j} \,|\, X_1 = z\right]$ and $\kappa(z)$ are uniformly bounded on $z \in \mathcal{X}$. According to Lemma H.2 and Lemma H.4, similar to the first part of the proof, (E.27), (E.28) and (E.29) can yield that for some constant C, with probability at least $1 - 1/d$, we can bound the suprema

$$\sup_{z\in\mathcal{X}}\max_{j\geq 2}\frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z\boldsymbol{\zeta}_z\|_2 \leq C\sqrt{\frac{h\log(dh^{-1})}{n}} + C\frac{m^{3/2}\log(dh^{-1})}{n} + C\frac{h^2}{\sqrt{m}}. \tag{E.30}$$

Combining (E.26) and (E.30), we have the rate of $\sup_{z\in\mathcal{X}}\max_{j\geq 2}\frac{1}{n}\|\boldsymbol{\Psi}_{\bullet j}^T\mathbf{W}_z(\boldsymbol{\xi}_z + \boldsymbol{\zeta}_z)\|_2$.

For the case when $j = 1$, $\boldsymbol{\Psi}_{\bullet 1} = (1, \ldots, 1)^T \in \mathbb{R}^n$ and we can follow similar procedure to derive

$$\sup_{z\in\mathcal{X}}\frac{1}{n}\left\|\boldsymbol{\Psi}_{\bullet 1}^T\mathbf{W}_z(\boldsymbol{\xi}_z + \boldsymbol{\zeta}_z)\right\|_2 = \sup_{z\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^n K_h(z - X_{i1})\big(\xi_i(z) + \zeta_i(z)\big) = O_P\left(h^2 + \sqrt{h/n}\right).$$

The final step is to bound $\sup_{z\in\mathcal{X}}\frac{1}{n}\|W_z^{1/2}\boldsymbol{\xi}_z^2\|_2^2$ and $\sup_{z\in\mathcal{X}}\frac{1}{n}\|W_z^{1/2}\boldsymbol{\zeta}_z^2\|_2^2$. We just repeat the procedure again and consider $V_n'''(z) = \sup_{z\in\mathcal{X}} n^{-1}\sum_{i=1}^n K_h(X_{1i} - z)\xi_i(z) - \mathbb{E}[K_h(X_{11} - z)\xi_i(z)]$. First, we find that

$$\mathbb{E}V_n'''(z) \leq C\sqrt{\frac{h^3\log(h^{-1/2})}{n}} + C\frac{h\log(h^{-1})}{n}. \tag{E.31}$$

Next, we have the upper bound of the supreme of the expectation

$$\begin{aligned}
\sup_{z\in\mathcal{X}}\mathbb{E}[K_h(X_1 - z)\xi_i^2(z)] &= h^{-1}\int K\left(\frac{x-z}{h}\right)(f_1(x) - f_1(z))^2 p_{X_1}(x)dx \\
&= \int K(u)(f_1'(z)uh + o(uh))^2(p_{X_1}(z) + p_{X_1}'(z)uh + o(uh))du \\
&= [f'(z)]^2 p_{X_1}(z)\int u^2K(u)du\cdot h^2 + o(h^2) \leq Ch^2. \qquad (\text{E.32})
\end{aligned}$$

Combining (E.31) and (E.32) with Lemma H.4 with $t = \log n/\sqrt{n}$, with probability at least $1-1/n$,

$$\begin{aligned}
\sup_{z\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^n K_h(X_{i1} - z)\xi_i^2(z) &\leq V_n(z) + \sup_{z\in\mathcal{X}}\mathbb{E}[K_h(X_{11} - z)\xi_i^2(z)] \\
&\leq C\sqrt{\frac{h^3\log(nh^{-1})}{n}} + C\frac{h^{5/4}\log(nh^{-1})}{n^{3/4}} + C\frac{h\log^2 n}{n} + Ch^2 = O\left(h^2\right).
\end{aligned}$$

Similarly, we also have $\sup_{z\in\mathcal{X}} n^{-1}\sum_{i=1}^n K_h(X_{i1} - z)\zeta_i^2(z) = o_P(h^2)$.

E.4 Proof of Lemma A.4

For $j = 2, \dots, n$, we define the process

$$G_n(z,k,j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{h} K\left(\frac{X_{i1}-z}{h}\right) \psi_{jk}(X_{ji})\varepsilon_i.$$

Since ε_i are subgaussian random variables, we have $\mathbb{P}(\max_i |\varepsilon_i| > C\sqrt{\log n}) \le 1/n$. Conditioning on the event $\mathcal{A} = \{\max_i |\varepsilon_i| < C\sqrt{\log n}\}$, we can apply the Mc'Diarmid's inequality to obtain

$$\mathbb{P}\left(\max_{j,k}\sup_{z\in\mathcal{X}} G_n(z,k,j) - \mathbb{E}\left[\max_{j,k}\sup_{z\in\mathcal{X}} G_n(z,k,j) \,|\, \mathcal{A}\right] > t \,|\, \mathcal{A}\right) \le \exp\left(-C\frac{nh^2t^2}{\log^2 n}\right). \tag{E.33}$$

Next, we bound $\mathbb{E}\big[\max_{j,k}\sup_{z\in\mathcal{X}} G_n(z,k,j) \,|\, \mathcal{A}\big]$. Using Dudley's entropy integral (see Corollary 2.2.5 in van der Vaart and Wellner (1996)), conditioning on $\{X_{ij}\}_{i\in[n],j\in[d]}$, we have with probability $1-1/n$, there exists a constant C such that

$$\mathbb{E}\left[\max_{1\le j\le d}\max_{1\le k\le m}\sup_{z\in\mathcal{X}} G_n(z,k,j) \,|\, \mathcal{A}\right] \le \mathbb{E}\left[\int_0^{\sigma_n} \sqrt{\log N(\mathcal{G}'_h, L^2(\widehat{\mathbb{P}}_n), \epsilon)}d\epsilon \,|\, \mathcal{A}\right],$$

where $\widehat{\mathbb{P}}_n = n^{-1}\sum_{i=1}^n \delta_{X_{i1},\dots,X_{id}}$, $\sigma_n = \max_{1\le j\le d}\max_{1\le k\le m}\sup_{z\in\mathcal{X}} \widehat{\mathbb{P}}_n[K_h(\cdot - z)\psi_{jk}(\cdot)]^2$ and

$$\mathcal{G}'_h = \left\{g_z(x_1,x_2) = h^{-1}K(h^{-1}(x_1 - z))\psi_{jk}(x_2) \,|\, 1\le k\le m, z\in\mathcal{X}, x_1, x_2 \;\in\mathcal{X}\right\}.$$

From Lemma H.3 and similar to the previous computation on the covering number, for any measure Q, we have the uniform upper bound of covering number as

$$\sup_Q N\left(\mathcal{G}'_h, L^2(Q), \epsilon\right) \le dm\left(\frac{2||K||_{\mathrm{TV}}A}{h\epsilon}\right)^4.$$

Following a similar argument as in the proof of Lemma A.1, we bound the variance of process by Cauchy-Schwarz inequality as

$$\begin{aligned}
\sigma_P^2 &:= \mathbb{E}\left[\left(K_h\left(X_1 - z\right)\psi_{jk}(X_j)\Big|\mathcal{A}\right)^2\right] \\
&\le h^{-2}\mathbb{E}\left[K^2\left(h^{-1}(X_1 - z)\right)\mathbb{E}\left[\psi_{jk}^2(X_j)\,|\,X_1\right] \,|\, \mathcal{A}\right] \le b\|K\|_\infty^2(mh)^{-1},
\end{aligned}$$

and $g \le \|K\|_\infty h^{-1}$ for any $g \in \mathcal{G}'_h$. Since $m(nh)^{-1} = o(1)$, we have $n\sigma_P^2 \ge h^{-2}\log(dm(2\|K\|_{\mathrm{TV}}A/(h\sigma)))$. Therefore, by Lemma H.2, we derive that

$$\mathbb{E}\left[\int_0^{\sigma_n} \sqrt{\log N(\mathcal{G}'_h, L^2(\widehat{\mathbb{P}}_n), \epsilon)} d\epsilon \,\Big|\, \mathcal{A}\right] \le C\sigma_P\sqrt{\log(dm(h\sigma_P)^{-4})} \le C\sqrt{\frac{\log(dm^3h^{-2})}{mh}}$$
$$\text{and } \mathbb{E}\Big[\max_{j,k}\sup_{z\in\mathcal{X}} G_n(z,k,j)\,\Big|\,\mathcal{A}\Big] \le C\sqrt{\log(dm^3h^{-2})/(mh)}.$$

Choosing $t = C\log^2 n/(\sqrt{n}h)$ in (E.33), we have

$$\mathbb{P}\left(\max_{j,k}\sup_{z\in\mathcal{X}} G_n(z,k,j) > C\sqrt{\frac{\log(dm^3h^{-2})}{mh}} + \frac{\log^2 n}{\sqrt{n}h}\right)$$
$$\le \mathbb{P}\left(\max_{j,k}\sup_{z\in\mathcal{X}} G_n(z,k,j) > C\sqrt{\frac{\log(dm^3h^{-2})}{mh}} + \frac{\log^2 n}{\sqrt{n}h}\,\Big|\mathcal{A}\right) + \mathbb{P}(\mathcal{A}^c) \le 2/n.$$

Therefore, when $m(nh)^{-1} = o(1)$, with probability $1-2/n$,

$$\sup_{z\in\mathcal{X}}\max_{j\ge 2}\frac{1}{n}\|\mathbf{\Psi}_{\bullet j}^T\mathbf{W}_z\boldsymbol{\varepsilon}\|_2 \le \sqrt{\frac{m}{n}}\max_{1\le j\le d}\max_{1\le k\le m}\sup_{z\in\mathcal{X}} G_n(z,k,j) \le 2C\sqrt{\frac{\log(dmh^{-1})}{nh}}.$$

When $j=1$, recalling that $\mathbf{\Psi}_{\bullet 1} = (1,\ldots,1)^T \in \mathbb{R}^n$, we have

$$\sup_{z\in\mathcal{X}}\frac{1}{n}\|\mathbf{\Psi}_{\bullet 1}^T\mathbf{W}_z\boldsymbol{\varepsilon}\|_2 = \sup_{z\in\mathcal{X}}\left|\frac{1}{n}\sum_{i=1}^n K_h(z-X_{i1})\varepsilon_i\right|,$$

and, similar to the case when $j\ge 2$, we can show that $\sup_{z\in\mathcal{X}} n^{-1}\|\mathbf{\Psi}_{\bullet 1}^T\mathbf{W}_z\boldsymbol{\varepsilon}\|_2 \le C\sqrt{\log(h^{-1})/(nh)}$ with probability $1-2/n$. This completes the proof.

F Auxiliary Lemmas for Bootstrap Confidence Bands

In this section, we describe the proof of these technical lemmas used in Section C. Section F.2 to Section F.6 provide the proofs of lemmas in Section C.1 supporting the proof of Theorem 3.7.

F.1 Proof of Lemma C.2

Recall the rate r_n of the estimated function shown in Theorem 3.2 is

$$r_n := \sqrt{\frac{s^2 \log(dmh^{-1})}{nm^{-2}h}} + \sqrt{\frac{s^3}{m^3} + \frac{s\log(dh^{-1})}{nm^{-5/2}}} + s\sqrt{m}h^2.$$

We first establish a lemma on the estimation error of $\widehat{\varepsilon}_i$.

Lemma F.1. Let $\widehat{\varepsilon}_i = Y_i - \widehat{f}(X_{i1}, \ldots, X_{id})$ for $i = 1, \ldots, n$. Under Assumption (**A4**), we have

$$\mathbb{P}\left(\max_{i\in[n]} |\widehat{\varepsilon}_i - \varepsilon_i| < 2Cr_n\sqrt{m}\right) \geq 1 - \frac{1}{n}.$$

If $h \asymp n^{-\delta}$, $m \asymp n^{\delta}$ for $\delta > 1/5$, we have $r_n\sqrt{m} = o(n^{-1/5})$.

We defer the proof of the lemma to the end of this subsection. With the rate of $\max_{i\in[n]} |\widehat{\varepsilon}_i - \varepsilon_i|$, we can first bound the rate of $\widehat{\sigma}^2 - \sigma^2$. Using the triangle inequality, we have

$$|\widehat{\sigma}^2 - \sigma^2| \leq \underbrace{\frac{1}{n}\sum_{i=1}^{n}(\widehat{\varepsilon}_i - \varepsilon_i)^2}_{\mathrm{I}} + \underbrace{\frac{2}{n}\sum_{i=1}^{n}|(\widehat{\varepsilon}_i - \varepsilon_i)\varepsilon_i|}_{\mathrm{II}} + \underbrace{\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \sigma^2\right|}_{\mathrm{III}}. \tag{F.1}$$

From Lemma F.1, we have the convergence rate of the noise estimator

$$\mathbb{P}(\mathrm{I} > 4cr_n^2 m) \leq \mathbb{P}\left(\max_{i\in[n]} |\widehat{\varepsilon}_i - \varepsilon_i|^2 > 4cr_n^2 m\right) \leq 1/n. \tag{F.2}$$

Under Assumption (**A4**), ε_i are subgaussian random variables with variance-proxy σ_ε^2. Using Berstein's inequality, we have

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - \sigma^2\right| > C_1\sqrt{\frac{\sigma_\varepsilon^2 \log n}{n}}\right) \leq \frac{2}{n} \quad \text{and} \quad \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| - \mathbb{E}|\varepsilon|\right| > c\sqrt{\frac{\sigma_\varepsilon^2 \log n}{n}}\right) \leq \frac{2}{n}. \tag{F.3}$$

Suppose n is large enough, so that $\mathbb{E}|\epsilon| \leq \sqrt{\sigma_\varepsilon^2 \log n/n}$. We now can bound the second term by

$$\begin{aligned}&\mathbb{P}(\mathrm{II} > 2c(c+1)\mathbb{E}|\varepsilon|r_n\sqrt{m})\\ &\leq \mathbb{P}\left(\max_{i\in[n]} |\widehat{\varepsilon}_i - \varepsilon_i| > 2cr_n\sqrt{m}\right) + \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n}|\varepsilon_i| > \mathbb{E}|\varepsilon| + c\sqrt{\frac{\sigma_\varepsilon^2 \log n}{n}}\right) \leq \frac{3}{n}.\end{aligned}$$

Applying the fact that $\mathbb{E}|\epsilon_i| \leq \sigma^2$, we have the upper bound of the third term as

$$\mathbb{P}\left(\mathrm{III} > 2c(c+1)\sigma^2 r_n\sqrt{m}\right) \leq 3/n. \tag{F.4}$$

Combining (F.2), (F.3), (F.4) with (F.1), we have the estimation rate of the variance of noise as

$$\mathbb{P}\left(|\widehat{\sigma}^2 - \sigma^2| \geq C_1 r_n\sqrt{m}\right) \leq 6/n. \tag{F.5}$$

Now we come back to prove Lemma F.1.

Proof of Lemma F.1. Recall that the estimator of the true function is

$$\widehat{f}(x_1, \ldots, x_d) = \widehat{f}_1(x_1) + \sum_{j=2}^{d}\sum_{k=1}^{m} \widehat{\boldsymbol{\beta}}_{jk}\psi_{jk}(x_j).$$

Similar to Lemma A.2, let $\delta_i = \sum_{j=2}^{d} f_j(X_{ji}) - f_{mj}(X_{ji})$ and the B-spline theory (see Lemma 1, Huang et al. (2010)) that $\delta_i^2 \leq sm^{-2\gamma}$. Define the event

$$\mathcal{E} = \left\{\sup_{z\in\mathcal{X}}\left\{\sqrt{m}|\widehat{a}_z - f_1(z)| + \sum_{i=2}^{d}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_2\right\} \leq Cr_n\right\}.$$

From Theorem 3.2, we have $\mathbb{P}(\mathcal{E}) \geq 1 - 1/n$. Conditioning the event $\mathcal{E}$, we have

$$\begin{aligned}
\max_{i\in[n]}|\widehat{\varepsilon}_i - \varepsilon_i| &= \max_{i\in[n]}\left|f(X_{i1}, \ldots, X_{id}) - \widehat{f}(X_{i1}, \ldots, X_{id})\right| \\
&\leq \max_{i\in[n]}\left|\widehat{f}_1(X_{i1}) - f_1(X_{i1})\right| + \max_{i\in[n]}\left|\sum_{j=2}^{d}\sum_{k=1}^{m}(\widehat{\boldsymbol{\beta}}_{jk} - \boldsymbol{\beta}_{jk})\psi_{jk}(X_{ij})\right| + \max_{i\in[n]}|\delta_i| \\
&\leq \sup_{z\in\mathcal{X}}\left|\widehat{f}_1(z) - f_1(z)\right| + \sqrt{m}\sum_{i=2}^{d}\|\widehat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|_2 + \sqrt{s}m^{-\gamma} \leq 2C\sqrt{m}r_n,
\end{aligned}$$

where the second inequality is because of Hölder inequality as well as the fact that $\psi_{jk} \leq 1$ for all j, k and the last inequality is since we are conditioning on $\mathcal{E}$. $\square$

F.2 Proof of Lemma 3.4

The high level idea of proving Lemma 3.4 is similar to the proof of Lemma E.1. We aim to bound the rate of $\sup_z \|\widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta}_z - \boldsymbol{\Sigma}_z \boldsymbol{\theta}_z\|_\infty$. Therefore, we consider the random variable

$$\widetilde{Z}_{kj} = \sup_{z\in\mathcal{X}} \|\widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta}_z - \boldsymbol{\Sigma}_z \boldsymbol{\theta}_z\|_\infty = \sup_{z\in\mathcal{X}} (\mathbb{E}_n - \mathbb{E})\Big[K_h(X_{i1}-z)\psi_{jk}(X_{ij}) \sum_{j'=1}^{d}\sum_{k'=1}^{m} \psi_{j'k'}(X_{ij'})(\boldsymbol{\theta}_z)_{j'k'}\Big].$$

Recall that when j or k equals to 1, $\psi_{jk} \equiv 1$. Similar to the proof of Lemma E.1, we have three cases: (1) $j = k = 1$, (2) only one of j or k equals to 1 and (3) neither of j, k equals to 1. We only analyze the hardest case (3) in this proof and we can deal with the first two cases through a similar procedure. For the minor differences among the analysis of these three cases, we refer to the proof of Lemma E.1.

We first study the covering number of the space

$$\overline{\mathcal{G}}_h = \Big\{ h^{-1}K(h^{-1}(x_1 - z))\psi_{jk}(x_j) \sum_{j'=1}^{d}\sum_{k'=1}^{m} \psi_{j'k'}(x_{j'})(\boldsymbol{\theta}_z)_{j'k'} \,\big|\, z \in \mathcal{X}, j, k \in [d] \Big\}.$$

Since $\overline{\mathcal{G}}_h$ can be decomposed into a production of a few functions, we aim to apply Lemma H.1 to bound its covering number. Lemma H.3 gives us the covering number of $\{h^{-1}K(h^{-1}(\cdot - z)) \,|\, z \in \mathcal{X}\}$, it remains to bound the covering number of

$$\overline{\mathcal{G}}_h^{(1)} = \Big\{ \overline{g}_z(\boldsymbol{x}) := \sum_{j'=1}^{d}\sum_{k'=1}^{m} \psi_{j'k'}(x_{j'})(\boldsymbol{\theta}_z)_{j'k'} \,\big|\, z \in \mathcal{X} \Big\}.$$

Given any $z \in \mathcal{X}$, we can find a $\widetilde{z}$ such that $|z - \widetilde{z}| \le \epsilon$. We then have given any measure Q,

$$\begin{aligned}\|\overline{g}_z - \overline{g}_{\widetilde{z}}\|_{L^2(Q)}^2 &= \mathbb{E}_Q\Big[\sum_{j'=1}^{d}\sum_{k'=1}^{m} \psi_{j'k'}(X_{ij'})[(\boldsymbol{\theta}_z)_{j'k'} - (\boldsymbol{\theta}_{\widetilde{z}})_{j'k'}]\Big]^2 \\ &\le L^2\|\boldsymbol{\theta}_z - \boldsymbol{\theta}_{\widetilde{z}}\|_1^2 \le L^2 d\|\boldsymbol{\theta}_z - \boldsymbol{\theta}_{\widetilde{z}}\|_2^2 \le 2L^2 dm\rho_{\min}^{-2}(B/b)L\rho_{\max}\cdot\epsilon^2,\end{aligned}$$

where the last inequality is due to Lemma F.4. Therefore,

$$\sup_Q \Big(\overline{\mathcal{G}}_h^{(1)}, L^2(Q), \epsilon\Big) \le \sqrt{2L^2 dm\rho_{\min}^{-2}(B/b)L|\mathcal{X}|^2\rho_{\max}}/\epsilon. \tag{F.6}$$

According to Corollary 8 in Chapter XI of de Boor (2001), we have

$$\sup_{z,\boldsymbol{x}} \Big| \sum_{j'=1}^{d} \sum_{k'=1}^{m} \psi_{j'k'}(x_{j'})(\boldsymbol{\theta}_z)_{j'k'} \Big| \leq L \sum_{j=1}^{d} \sup_{z} \|(\boldsymbol{\theta}_z)_{j\bullet}\|_\infty = L\sqrt{d} \sup_{z} \|\boldsymbol{\theta}_z\|_2 \leq 4L\rho_{\min}^{-1} m\sqrt{d}, \quad \text{(F.7)}$$

where the last inequality is due to Lemma F.3. By Lemma H.1, combining (F.6), (F.7) and Lemma H.3, we have

$$\sup_{Q} N\left(\overline{\mathcal{G}}_h, L^2(Q), \epsilon\right) \leq d^2 \left(\frac{C|\mathcal{X}|md}{\rho_{\min}^2 bh\epsilon}\right)^5. \quad \text{(F.8)}$$

We then consider the envelop function of $\overline{\mathcal{G}}_h$ as

$$\overline{F}(\boldsymbol{x}) = 4h^{-1}\|K\|_\infty \sup_{z} \Big| \sum_{j'=1}^{d} \sum_{k'=1}^{m} \psi_{j'k'}(x_{j'})(\boldsymbol{\theta}_z)_{j'k'} \Big|.$$

In order to study $\overline{F}(\boldsymbol{x})$, we define $\overline{\boldsymbol{\Sigma}} = \mathbb{E}[\boldsymbol{\Psi}_{1\bullet}\boldsymbol{\Psi}_{1\bullet}^T]$ and $\overline{\boldsymbol{\theta}} = \overline{\boldsymbol{\Sigma}}^{-1}\mathbf{e}_1$. We decompose $\overline{F}(\boldsymbol{x})$ into

$$\overline{F}^{(1)}(\boldsymbol{x}) = 4h^{-1}\|K\|_\infty \Big| \sum_{j'=1}^{d} \sum_{k'=1}^{m} \psi_{j'k'}(x_{j'})(\overline{\boldsymbol{\theta}})_{j'k'} \Big| \text{ and}$$
$$\overline{F}^{(2)}(\boldsymbol{x}) = 4h^{-1}\|K\|_\infty \sup_{z} \Big| \sum_{j'=1}^{d} \sum_{k'=1}^{m} \psi_{j'k'}(x_{j'})(\boldsymbol{\theta}_z - \overline{\boldsymbol{\theta}})_{j'k'} \Big|.$$

According to Lemma F.2, we have

$$\|\overline{F}^{(1)}\|_{L^2(\mathbb{P})}^2 \leq 4mh^{-1}\|K\|_\infty \|\overline{\boldsymbol{\theta}}\|_2^2 \leq 4h^{-1}\|K\|_\infty \rho_{\min}.$$

Similarly, we also have

$$\|\overline{F}^{(2)}\|_{L^2(\mathbb{P})}^2 \leq 4mh^{-1}\|K\|_\infty \sup_{z} \|\boldsymbol{\theta}_z - \overline{\boldsymbol{\theta}}\|_2^2 \leq 8h^{-1}\|K\|_\infty \rho_{\min}.$$

Therefore, we have $\sigma_P^2 \leq \|F\|_{L^2(\mathbb{P})}^2 \leq 32h^{-1}\|K\|_\infty \rho_{\min}$. By Lemma H.2 we have

$$\mathbb{E}\Big[\max_{k,j} \widetilde{Z}_{kj}\Big] \leq C_1 \sqrt{\frac{(\log^2 d)\log(|\mathcal{X}|/\rho_{\min})}{nh}}. \quad \text{(F.9)}$$

We can also apply Lemma H.5 to obtain

$$\mathbb{P}\Big(\sqrt{n}\max_{k,j}\widetilde{Z}_{kj} \geq 2\sqrt{n}\mathbb{E}[\max_{k,j}\widetilde{Z}_{kj}] + Ch^{-1/2}\sqrt{t} + Ch^{-1/2}t\Big) \leq t^{-1}. \tag{F.10}$$

Combining (F.9) with (F.10), we have

$$\max_{k,j}\widetilde{Z}_{kj} = O_P\Big(\sqrt{(\log^2 d)\log(|\mathcal{X}|/\rho_{\min})/nh}\Big).$$

Finally, we finish the proof of the lemma by

$$\big\|\widehat{\mathbf{\Sigma}}_z\boldsymbol{\theta}_z - \mathbf{e}_1\big\|_{2,\infty} \leq \sqrt{m}\big\|\widehat{\mathbf{\Sigma}}_z\boldsymbol{\theta}_z - \mathbf{e}_1\big\|_{\infty} = O_P\Big(\sqrt{m(\log^2 d)\log(|\mathcal{X}|/\rho_{\min})/nh}\Big).$$

F.3 Auxiliary Lemmas for Constraint Rate

In this section, we prove some auxiliary lemmas needed in the proof of Lemma 3.4.

Lemma F.2. Under Assumptions (**A1**), (**A2**) and (**A6**), there exists a constant $\rho_{\max} < \infty$ such that for any $\boldsymbol{\beta}_+ \in \mathbb{R}^{1+(d-1)m}$,

$$\frac{\boldsymbol{\beta}_+^T\mathbb{E}[\mathbf{\Psi}_{1\bullet}\mathbf{\Psi}_{1\bullet}^T]\boldsymbol{\beta}_+}{\|\boldsymbol{\beta}_+\|_2^2} \leq \frac{3\rho_{\max}}{2m}. \tag{F.11}$$

Proof. We first derive some inequalities from (3.10). Given any $j \neq k \geq 2$, let $u_j(x_j) = \sum_{s=1}^m \boldsymbol{\beta}_{js}\psi_{js}(x_j)$, $\Delta_{jk}(x_j,x_k) := |p_{j,k}(x_j,x_k) - p_j(x_j)p_j(x_k)|$, and we have

$$\begin{aligned}
&\big|\boldsymbol{\beta}_+^T\mathbb{E}[\mathbf{\Psi}_{1j}\mathbf{\Psi}_{1k}^T]\boldsymbol{\beta}_+\big| = \Big|\iint u_j(x_j)u_k(x_k)p_{jk}(x_j,x_k)dx_jdx_k\Big| \\
&\leq \Big|\int u_j(x_j)p_j(x_j)dx_j\int u_k(x_k)p_k(x_k)dx_k\Big| + \iint \big|u_j(x_j)u_k(x_k)\big|\Delta_{jk}(x_j,x_k)dx_jdx_k \\
&\leq m^{-1}\|\boldsymbol{\beta}_j\|_2\|\boldsymbol{\beta}_k\|_2\cdot\|\Delta_{jk}\|_2,
\end{aligned} \tag{F.12}$$

where the last inequality is due to $\mathbb{E}[\psi_{jk}(X_j)] = 0$ and (A.10). Therefore, we have

$$\begin{aligned}
\boldsymbol{\beta}_+^T\mathbb{E}[\mathbf{\Psi}_{1\bullet}\mathbf{\Psi}_{1\bullet}^T]\boldsymbol{\beta}_+ &= \sum_{j=1}^d \boldsymbol{\beta}_j^T\mathbb{E}[\mathbf{\Psi}_{1j}\mathbf{\Psi}_{1j}^T]\boldsymbol{\beta}_j + \sum_{j\neq k}\boldsymbol{\beta}_j^T\mathbb{E}[\mathbf{\Psi}_{1j}\mathbf{\Psi}_{1k}^T]\boldsymbol{\beta}_k \\
&\leq \frac{\rho_{\max}}{m}\sum_{j=1}^d\|\boldsymbol{\beta}_j\|_2^2 + \sum_{j\neq k} m^{-1}\|\boldsymbol{\beta}_j\|_2\|\boldsymbol{\beta}_k\|_2\cdot\|\Delta_{jk}\|_2 \leq \frac{3\rho_{\max}}{2m}\|\boldsymbol{\beta}\|_2^2,
\end{aligned}$$

where the last inequality is due to Assumption (**A6**) and the Gershgorin circle theorem. $\square$

Lemma F.3. Under Assumptions (**A1**), (**A2**), (**A3**) and (**A6**), there exists a constant $\rho_{\max} < \infty$ such that for any $z \in \mathcal{X}$ and any $\boldsymbol{\beta}_+ \in \mathbb{R}^{1+(d-1)m}$,

$$\frac{\rho_{\min}}{2m} \le \frac{\boldsymbol{\beta}_+^T \boldsymbol{\Sigma}_z \boldsymbol{\beta}_+}{\|\boldsymbol{\beta}_+\|_2^2} \le \frac{3\rho_{\max}}{2m} \text{ and } \sup_z \|\boldsymbol{\theta}_z\|_2 \le \frac{2m}{\rho_{\min}}. \tag{F.13}$$

Proof. We first derive some inequalities from (3.10). Denote $\Delta_{1jk}(x_1, x_j, x_k) := |p_{1,j,k}(x_1, x_j, x_k) - p_1(x_1)p_j(x_j)p_k(x_k)|$. Given any $j \neq k \ge 2$, let $u_j(x_j) = \sum_{s=1}^m \boldsymbol{\beta}_{js}\psi_{js}(x_j)$, similar to (F.12), we have

$$\begin{aligned}\left|\boldsymbol{\beta}_j^T \mathbb{E}[K_h(X_1 - z)\boldsymbol{\Psi}_{1j}\boldsymbol{\Psi}_{1k}^T]\boldsymbol{\beta}_k\right| &= \left|\iiint K_h(x_1 - z)u_j(x_j)u_k(x_k)p_{1,j,k}(x_1, x_j, x_k)dx_1dx_jdx_k\right| \\ &\le m^{-1}\|\boldsymbol{\beta}_j\|_2\|\boldsymbol{\beta}_k\|_2\|\Delta_{1jk}\|_2. \end{aligned}\tag{F.14}$$

Therefore, we have

$$\begin{aligned}\boldsymbol{\beta}_+^T \boldsymbol{\Sigma}_z \boldsymbol{\beta}_+ &= \sum_{j=1}^d \boldsymbol{\beta}_j^T \mathbb{E}[K_h(X_1 - z)\boldsymbol{\Psi}_{1j}\boldsymbol{\Psi}_{1j}^T]\boldsymbol{\beta}_j + \sum_{j \neq k} \boldsymbol{\beta}_j^T \mathbb{E}[K_h(X_1 - z)\boldsymbol{\Psi}_{1j}\boldsymbol{\Psi}_{1k}^T]\boldsymbol{\beta}_k \\ &\le \frac{\rho_{\max}}{m}\sum_{j=1}^d \|\boldsymbol{\beta}_j\|_2^2 + \sum_{j \neq k} m^{-1}\|\boldsymbol{\beta}_j\|_2\|\boldsymbol{\beta}_k\|_2\|\Delta_{1jk}\|_2 \le \frac{3\rho_{\max}}{2m}\|\boldsymbol{\beta}\|_2^2,\end{aligned}$$

where the first inequality is due to (E.9) and the last inequality is due to the Gershgorin circle theorem. Similarly, we apply the Gershgorin circle theorem to obtain a lower bound as

$$\begin{aligned}\boldsymbol{\beta}_+^T \boldsymbol{\Sigma}_z \boldsymbol{\beta}_+ &= \sum_{j=1}^d \boldsymbol{\beta}_j^T \mathbb{E}[K_h(X_1 - z)\boldsymbol{\Psi}_{1j}\boldsymbol{\Psi}_{1j}^T]\boldsymbol{\beta}_j + \sum_{j \neq k} \boldsymbol{\beta}_j^T \mathbb{E}[K_h(X_1 - z)\boldsymbol{\Psi}_{1j}\boldsymbol{\Psi}_{1k}^T]\boldsymbol{\beta}_k \\ &\ge \frac{\rho_{\min}}{m}\sum_{j=1}^d \|\boldsymbol{\beta}_j\|_2^2 - \sum_{j \neq k} m^{-1}\|\boldsymbol{\beta}_j\|_2\|\boldsymbol{\beta}_k\|_2\|\Delta_{1jk}\|_2 \ge \frac{\rho_{\min}}{2m}\|\boldsymbol{\beta}\|_2^2,\end{aligned}$$

where the first inequality is due to Assumption (**A3**). $\square$

The following lemma shows the Lipschitz properties of $\boldsymbol{\theta}_z$.

Lemma F.4. Under Assumptions (**A1**) - (**A6**), we have

$$\|\boldsymbol{\theta}_z - \boldsymbol{\theta}_{z'}\|_2 \le 2m\rho_{\min}^{-2}(B/p_1(z))L\rho_{\max} \cdot |z - z'|,$$

where C_K is a constant only depending on the kernel K.

Proof. The idea of proving this lemma is similar to Lemma F.3. Given any $j \neq k \geq 2$, again let $u_j(x_j) = \sum_{s=1}^m \boldsymbol{\beta}_{js}\psi_{js}(x_j)$ and we have

$$\begin{aligned}
&\big|\boldsymbol{\beta}_j^T \mathbb{E}[(K_h(X_1 - z) - K_h(X_1 - z'))\boldsymbol{\Psi}_{1j}\boldsymbol{\Psi}_{1k}^T]\boldsymbol{\beta}_k\big| \\
&= \big| \int (K_h(x_1 - z) - K_h(x_1 - z'))u_j(x_j)u_k(x_k)p_{1,j,k}(x_1, x_j, x_k)dx_1dx_jdx_k\big| \\
&\leq \big| \int K(x_1) \sup_u \mathbb{E}[u_j(X_j)u_k(X_k)|X_1 = u](p_1(z + x_1h) - p_1(z' + x_1h))dx_1\big| \\
&\leq b^{-1}L|z - z'|m^{-1}\|\boldsymbol{\beta}_j\|_2\|\boldsymbol{\beta}_k\|_2\|\Delta_{1jk}\|_2.
\end{aligned}$$

Similarly, we also have

$$\begin{aligned}
&\big|\boldsymbol{\beta}_j^T \mathbb{E}[(K_h(X_1 - z) - K_h(X_1 - z'))\boldsymbol{\Psi}_{1j}\boldsymbol{\Psi}_{1j}^T]\boldsymbol{\beta}_j\big| \\
&\leq \big| \int K(x_1) \sup_u \mathbb{E}[u_j(X_j)^2|X_1 = u](p_1(z + x_1h) - p_1(z' + x_1h))dx_1\big| \\
&\leq (B/b)L|z - z'|m^{-1}\|\boldsymbol{\beta}_j\|_2^2.
\end{aligned}$$

Therefore, for any $\boldsymbol{\beta}_+ \in \mathbb{R}^{1+(d-1)m}$ and z, z',

$$\begin{aligned}
\boldsymbol{\beta}_+^T(\boldsymbol{\Sigma}_z - \boldsymbol{\Sigma}_{z'})\boldsymbol{\beta}_+ &\leq (B/b)L|z - z'| \cdot m^{-1}\sum_{j=1}^d \|\boldsymbol{\beta}_j\|_2^2 + b^{-1}L|z - z'| \cdot \sum_{j\neq k} m^{-1}\|\boldsymbol{\beta}_j\|_2\|\boldsymbol{\beta}_k\|_2\|\Delta_{1jk}\|_2 \\
&\leq \frac{2(B/b)L\rho_{\max}}{m}\|\boldsymbol{\beta}_+\|_2^2 \cdot |z - z'|.
\end{aligned}$$

Therefore, combining with Lemma F.3, we can apply the matrix inverse perturbation inequality (see e.g., Demmel (1992)) and have

$$\|\boldsymbol{\theta}_z - \boldsymbol{\theta}_{z'}\|_2 \leq \|\boldsymbol{\Sigma}_z^{-1}\|_2^2\|\boldsymbol{\Sigma}_z - \boldsymbol{\Sigma}_{z'}\|_2 \leq 2m\rho_{\min}^{-2}(B/b)L\rho_{\max} \cdot |z - z'|.$$

□

F.4 Proof of Lemma C.4

Applying the fact that $\widehat{\boldsymbol{\theta}}_z^T \widehat{\boldsymbol{\Sigma}}_z \widehat{\boldsymbol{\theta}}_z \leq \boldsymbol{\theta}_z^T \widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta}_z$, we have the following inequality

$$\begin{aligned}
\left(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z\right)^T \widehat{\boldsymbol{\Sigma}}_z \left(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z\right) &= \widehat{\boldsymbol{\theta}}_z^T \widehat{\boldsymbol{\Sigma}}_z \widehat{\boldsymbol{\theta}}_z - 2\widehat{\boldsymbol{\theta}}_z^T \widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta}_z + \boldsymbol{\theta}_z^T \widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta}_z \\
&\leq 2\boldsymbol{\theta}_z^T \widehat{\boldsymbol{\Sigma}}_z \boldsymbol{\theta}_z - 2\left(\widehat{\boldsymbol{\theta}}_z^T \widehat{\boldsymbol{\Sigma}}_z - e_i\right)\boldsymbol{\theta}_z - 2e_i^T \boldsymbol{\theta}_z \\
&= 2\boldsymbol{\theta}_z^T \left(\widehat{\boldsymbol{\Sigma}}_z - \boldsymbol{\Sigma}_z\right)\boldsymbol{\theta}_z - 2\left(\widehat{\boldsymbol{\theta}}_z^T \widehat{\boldsymbol{\Sigma}}_z - e_i\right)\boldsymbol{\theta}_z \\
&\leq 2\|\boldsymbol{\theta}_z\|_1^2 \|\widehat{\boldsymbol{\Sigma}}_z - \boldsymbol{\Sigma}_z\|_{\max} + 2\|\widehat{\boldsymbol{\Sigma}}_z \widehat{\boldsymbol{\theta}}_z - \mathbf{e}_1\|_{2,\infty} \|\boldsymbol{\theta}_z\|_1. \qquad \text{(F.15)}
\end{aligned}$$

We now study the rate of $\boldsymbol{\theta}_z$ in this subsection. We separate $\boldsymbol{\Sigma}_z$ into four blocks such that

$$\boldsymbol{\Sigma}_z = \begin{pmatrix} \boldsymbol{\Sigma}_z^{(1,1)} & \boldsymbol{\Sigma}_z^{(2,1)T} \\ \boldsymbol{\Sigma}_z^{(2,1)} & \boldsymbol{\Sigma}_z^{(2,2)} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_z^{(1,1)} \in \mathbb{R}$, $\boldsymbol{\Sigma}_z^{(2,1)} \in \mathbb{R}^{(d-1)m}$ and $\boldsymbol{\Sigma}_z^{(2,2)} \in \mathbb{R}^{(d-1)m \times (d-1)m}$. By Lemma F.3, both $[\boldsymbol{\Sigma}_z^{(1,1)}]^{-1}$ and $[\boldsymbol{\Sigma}_z^{(2,2)}]^{-1}$ exist for any $z \in \mathcal{X}$. By the inversion formula of a block matrix, we have

$$\boldsymbol{\Sigma}_z^{-1} = \begin{pmatrix} \boldsymbol{\Theta}_z^{(1,1)} & \boldsymbol{\Theta}_z^{(2,1)T} \\ \boldsymbol{\Theta}_z^{(2,1)} & \boldsymbol{\Theta}_z^{(2,2)} \end{pmatrix},$$

where the concrete formulations of these four submatrices are

$$\begin{aligned}
\boldsymbol{\Theta}_z^{(1,1)} &= \left(\boldsymbol{\Sigma}_z^{(1,1)} - [\boldsymbol{\Sigma}_z^{(2,1)}]^T [\boldsymbol{\Sigma}_z^{(2,2)}]^{-1} \boldsymbol{\Sigma}_z^{(2,1)}\right)^{-1}, \\
\boldsymbol{\Theta}_z^{(2,1)} &= -\boldsymbol{\Theta}_z^{(1,1)} [\boldsymbol{\Sigma}_z^{(2,2)}]^{-1} \boldsymbol{\Sigma}_z^{(2,1)}, \\
\boldsymbol{\Theta}_z^{(2,2)} &= [\boldsymbol{\Sigma}_z^{(2,2)}]^{-1} - \boldsymbol{\Theta}_z^{(2,1)} [\boldsymbol{\Sigma}_z^{(2,1)}]^T [\boldsymbol{\Sigma}_z^{(2,2)}]^{-1}.
\end{aligned}$$

In order to bound $\boldsymbol{\theta}_z = (\boldsymbol{\Theta}_z^{(1,1)}, \boldsymbol{\Theta}_z^{(2,1)T})^T$, we first bound the ℓ_1 norm of the second part

$$\|\boldsymbol{\Sigma}_z^{(2,1)}\|_1 = \sum_{j=2}^{d} \sum_{k=1}^{m} \left|\mathbb{E}[K_h(X_1 - z)\psi_{jk}(X_j)]\right|.$$

Following a similar analysis as (F.12), we can bound the norm by $\left\|\boldsymbol{\Sigma}_z^{(2,1)}\right\|_1 \leq \frac{\rho_{\max}}{2}$ and $\left\|[\boldsymbol{\Sigma}_z^{(2,2)}]^{-1} \boldsymbol{\Sigma}_z^{(2,1)}\right\|_1 \leq$

$(\rho_{\max}/\rho_{\min}) \cdot m/2$. In fact, by Lemma F.3, $\mathbf{\Theta}_z^{(1,1)} > 0$. Combining with $\mathbf{\Sigma}_z^{(1,1)} = \mathbb{E}[K_h(X_1 - z)] = p_1(z) + o(1)$, we can have $\mathbf{\Sigma}_z^{(2,1)T}[\mathbf{\Sigma}_z^{(2,2)}]^{-1}\mathbf{\Sigma}_z^{(2,1)} \le p_1(z) + o(1)$ for any $z \in \mathcal{X}$.

Summarizing the inequalities above, we have

$$\begin{aligned}
\sup_{z\in\mathcal{X}} \|\boldsymbol{\theta}_z\|_1 &\le \sup_{z\in\mathcal{X}} |\mathbf{\Theta}_z^{(1,1)}| + \sup_{z\in\mathcal{X}} \|\mathbf{\Theta}_z^{(2,1)}\|_1 \\
&= \sup_{z\in\mathcal{X}} \left|\mathbf{\Sigma}_z^{(1,1)} - \mathbf{\Sigma}_z^{(2,1)T}[\mathbf{\Sigma}_z^{(2,2)}]^{-1}\mathbf{\Sigma}_z^{(2,1)}\right|^{-1} + \sup_{z\in\mathcal{X}} \left\|\mathbf{\Theta}_z^{(1,1)}[\mathbf{\Sigma}_z^{(2,2)}]^{-1}\mathbf{\Sigma}_z^{(2,1)}\right\|_1 \\
&= \sup_{z\in\mathcal{X}} \left\{(p_1(z) + O(1))^{-1} + (p_1(z) + O(1))^{-1} \cdot O(\rho_{\min}^{-1} m^{3/2})\right\} \le C(b\rho_{\min})^{-1} m. \quad (\text{F.16})
\end{aligned}$$

Plugging (3.11), (F.16) and (E.8) into (F.15), we prove the lemma.

F.5 Proof of Lemma C.1

We can expand the difference between two processes as

$$\widetilde{\mathbb{H}}_n'(z) - \widetilde{Z}_n'(z) = \underbrace{\sqrt{n^{-1}h}\textstyle\sum_{i=1}^n K_h(X_{i1} - z)\eta_i' \mathbf{\Psi}_{i\bullet}^T \widehat{\boldsymbol{\theta}}_z}_{T_1(z)} + \underbrace{\sqrt{nh}(\mathbf{e}_1^T - \widehat{\boldsymbol{\theta}}_z^T \widehat{\mathbf{\Sigma}}_z)(\widehat{\boldsymbol{\beta}}_+ - \boldsymbol{\beta}_+)}_{T_2(z)},$$

where η_i' is defined as

$$\eta_i' = f(X_{1i}, \ldots, X_{di}) - \sum_{j=1}^d f_{mj}(X_{ji}) \qquad \text{for any } i \in [n]. \quad (\text{F.17})$$

Using Theorem 3.2 and Lemma 3.4, with probability $1 - c/n$, we have

$$\begin{aligned}
&\sup_{z\in\mathcal{X}} |T_2(z)| \le \sqrt{nh}\|\widehat{\mathbf{\Sigma}}_z\boldsymbol{\theta}_z - \mathbf{e}_1\|_{2,\infty}\|\widehat{\boldsymbol{\beta}}_+ - \boldsymbol{\beta}_+\|_{2,1} \\
&\le C\sqrt{nh}\left(\sqrt{\frac{m\log^2 d}{nh}}\right) \cdot sm\left(\sqrt{\frac{\log(dmh^{-1})}{nh}} + \frac{\sqrt{s}}{m^{5/2}} + \frac{m^{3/2}\log(dh^{-1})}{n} + \frac{h^2}{\sqrt{m}}\right). \quad (\text{F.18})
\end{aligned}$$

Since $mh = o(1)$ and $h \asymp n^{-\delta}$ for $\delta > 1/5$, we have $\sup_{z\in\mathcal{X}} |T_2(z)| = o_P(n^{-1/10})$.

To bound $T_1(z)$, we first apply the triangle inequality and Cauchy-Schwartz inequality to decompose $T_1(z)$ into three smaller fragments

$$T_1(z) = \sqrt{h/n}\sum_{i=1}^{n} K_h(X_{i1}-z)\eta_i'\boldsymbol{\Psi}_{i\bullet}^T(\widehat{\boldsymbol{\theta}}_z-\boldsymbol{\theta}_z) + \sqrt{h/n}\sum_{i=1}^{n} K_h(X_{i1}-z)\eta_i'\boldsymbol{\Psi}_{i\bullet}^T\boldsymbol{\theta}_z$$
$$\leq \sqrt{nh}\cdot T_{11}^{1/2}(z)\cdot T_{12}^{1/2}(z) + \sqrt{nh}\cdot T_{13}(z),$$

where the three processes T_{11}, T_{12} and T_{13} are defined as follows

$$T_{11}(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1}-z)(\boldsymbol{\Psi}_{i\bullet}^T(\widehat{\boldsymbol{\theta}}_z-\boldsymbol{\theta}_z))^2, \quad T_{12}(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1}-z)(\eta_i')^2$$
$$\text{and} \quad T_{13}(z) = \frac{1}{n}\sum_{i=1}^{n} K_h(X_{i1}-z)\eta_i'\boldsymbol{\Psi}_{i\bullet}^T\boldsymbol{\theta}_z.$$

From Lemma C.4, we can bound the supreme of $T_{11}(z)$ by

$$\sup_{z\in\mathcal{X}}|T_{11}(z)| = \sup_{z\in\mathcal{X}}\Big(\widehat{\boldsymbol{\theta}}_z-\boldsymbol{\theta}_z\Big)^T\widehat{\boldsymbol{\Sigma}}_z\Big(\widehat{\boldsymbol{\theta}}_z-\boldsymbol{\theta}_z\Big)$$
$$\leq C\inf_{z\in\mathcal{X}}\frac{\log(D_n/p_1(z))}{p_1^2(z)}\cdot m\Bigg(\sqrt{\frac{m\log(dm)}{nh}}+\frac{m}{nh}+\sqrt{\frac{\log(1/h)}{nh}}\Bigg). \tag{F.19}$$

Let $\boldsymbol{\delta}, \boldsymbol{\xi}_z$ and $\boldsymbol{\zeta}_z$ be as defined in Lemma A.2 and Lemma A.3. From those two lemmas, with probability $1-c/n$, we have

$$\sup_{z\in\mathcal{X}}|T_{12}(z)| \leq \sup_{z\in\mathcal{X}}\frac{2}{n}\|\mathbf{W}_z^{1/2}\boldsymbol{\delta}\|_2^2 + \sup_{z\in\mathcal{X}}\frac{2}{n}\|\mathbf{W}_z^{1/2}(\boldsymbol{\xi}_z+\boldsymbol{\zeta}_z)\|_2^2 \leq C(sm^{-4}+h^2). \tag{F.20}$$

Lemma A.2 and Lemma A.3 also give us

$$\begin{aligned}\sup_{z\in\mathcal{X}}|T_{13}(z)| &\leq \sup_{z\in\mathcal{X}}\frac{1}{n}\|\boldsymbol{\Psi}^T\mathbf{W}_z(\boldsymbol{\delta}+\boldsymbol{\xi}_z+\boldsymbol{\zeta}_z)\|_{2,\infty}\|\boldsymbol{\theta}_z\|_1 \\ &\leq Cm\Bigg(\sqrt{s}\cdot m^{-5/2}+\sqrt{\frac{h\log(dh^{-1})}{n}}+\frac{m^{3/2}\log(dh^{-1})}{n}+\frac{h^2}{\sqrt{m}}\Bigg).\end{aligned} \tag{F.21}$$

Combining (F.19), (F.20) with (F.21), if the scaling condition of Theorem 3.7 is satisfied, there exists a constant c such that $\mathbb{P}(\sup_{z\in\mathcal{X}}|T_1(z)| > Cn^{-c}) < C/n$. Combining this inequality with the rate of $\sup_{z\in\mathcal{X}}|T_1(z)|$ in (F.18), we have our lemma proved.

F.6 Proof of Lemma C.3

We first bound the difference between $\widehat{\boldsymbol{\theta}}_z^T \boldsymbol{\Sigma}_z' \widehat{\boldsymbol{\theta}}_z$ and $\boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z$ by applying triangle inequality and Cauchy-Schwartz inequality

$$\begin{aligned}
&\left|\widehat{\boldsymbol{\theta}}_z^T \boldsymbol{\Sigma}_z' \widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z\right| \\
&\leq (\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T \boldsymbol{\Sigma}_z' (\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z) + 2(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z \\
&\leq h^{-1}(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T \widehat{\boldsymbol{\Sigma}}_z (\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z) + 2h^{-1/2}\sqrt{(\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T \widehat{\boldsymbol{\Sigma}}_z (\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)}\sqrt{\boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z}.
\end{aligned} \tag{F.22}$$

From Lemma C.4, we have the desired upper bound in the lemma that

$$\sup_{z\in\mathcal{X}} (\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z)^T \widehat{\boldsymbol{\Sigma}}_z (\widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z) \leq m\left(\sqrt{\frac{m\log(dm)}{nh}} + \frac{m}{nh} + \sqrt{\frac{\log(1/h)}{nh}}\right). \tag{F.23}$$

The following lemma gives us a bound on the term $\boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z$.

Lemma F.5. Under Assumption (**A1**), for any $z \in \mathcal{X}$,

$$\mathbb{E}[\boldsymbol{\Sigma}_z'] = h^{-1}\lambda(K)\big[\boldsymbol{\Sigma}_z + o(h)\big],$$

where $\lambda(K) = \int K^2(u)du$. Furthermore, with probability at least $1 - c/n$,

$$\sup_{z\in\mathcal{X}} ||\boldsymbol{\Sigma}_z' - \mathbb{E}[\boldsymbol{\Sigma}_z']||_{\max} \leq C\left(\frac{1}{nh^2} + \frac{1}{\sqrt{nh^3}} + \sqrt{\frac{\log(dm)}{nmh^3}}\right).$$

We defer the proof of the lemma to the end of the section. Using Lemma F.5, we have

$$\begin{aligned}
\boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z &\geq \boldsymbol{\theta}_z^T \mathbb{E}[\boldsymbol{\Sigma}_z'] \boldsymbol{\theta}_z - \|\boldsymbol{\Sigma}_z' - \mathbb{E}[\boldsymbol{\Sigma}_z']\|_{\max} \|\boldsymbol{\theta}_z\|_1^2 \\
&\geq h^{-1}\lambda(K)\boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z \boldsymbol{\theta}_z - \|\boldsymbol{\Sigma}_z' - \mathbb{E}[\boldsymbol{\Sigma}_z']\|_{\max} \|\boldsymbol{\theta}_z\|_1^2 - o(1) \\
&\geq h^{-1}\lambda(K)\mathbf{e}_1^T \boldsymbol{\theta}_z - Cm^2\left(\frac{1}{nh^2} + \frac{1}{\sqrt{nh^3}} + \sqrt{\frac{\log(dm)}{nmh^3}}\right) - o(1).
\end{aligned} \tag{F.24}$$

We can also bound from the other direction as

$$\boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z \leq h^{-1}\lambda(K)\mathbf{e}_1^T \boldsymbol{\theta}_z + Cm^2\left(\frac{1}{nh^2} + \frac{1}{\sqrt{nh^3}} + \sqrt{\frac{\log(dm)}{nmh^3}}\right) + o(1). \tag{F.25}$$

Combining (F.22), (F.23), (F.24) and (F.25), if $mh = o(1), h = n^{-\delta}$ for $\delta > 1/5$, there exists a constant c such that for any $z \in \mathcal{X}$,

$$\begin{aligned}\widehat{\boldsymbol{\theta}}_z^T \boldsymbol{\Sigma}_z' \widehat{\boldsymbol{\theta}}_z &\geq \boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z - \left| \widehat{\boldsymbol{\theta}}_z^T \boldsymbol{\Sigma}_z' \widehat{\boldsymbol{\theta}}_z - \boldsymbol{\theta}_z^T \boldsymbol{\Sigma}_z' \boldsymbol{\theta}_z \right| \\ &= h^{-1} \left(\lambda(K) \mathbf{e}_1^T \boldsymbol{\theta}_z + o(1) \right) \geq c h^{-1} \mathbf{e}_1^T \boldsymbol{\theta}_z\end{aligned}$$

Similarly, we also have $\widehat{\boldsymbol{\theta}}_z^T \boldsymbol{\Sigma}_z' \widehat{\boldsymbol{\theta}}_z \leq C h^{-1} \mathbf{e}_1^T \boldsymbol{\theta}_z$. The proof will be done once we prove Lemma F.5.

Proof of Lemma F.5. For any $j, j' \in [d]$ and $k, k' \in [m]$, we have

$$\begin{aligned}\mathbb{E}[\boldsymbol{\Sigma}_z']_{jj'kk'} &= \int K_h^2(x - z)(\psi_{jk}(x_j)\psi_{j'k'}(x_{j'}))p_{1,j,j'}(x_1, x_j, x_{j'})dx_1 dx_j dx_{j'} \\ &= h^{-1} \int K^2(u)(\psi_{jk}(x_j)\psi_{j'k'}(x_{j'}))p_{1,j,j'}(z + uh, x_j, x_{j'})du dx_j dx_{j'} \\ &= h^{-1}\lambda(K) \int K(u)(\psi_{jk}(x_j)\psi_{j'k'}(x_{j'}))(p_{1,j,j'}(z, x_j, x_{j'}) + o(h))du dx_j dx_{j'} \\ &= h^{-1}\lambda(K) \left[\boldsymbol{\Sigma}_z + o(h) \right]_{jj'kk'}.\end{aligned}$$

The second part of the proof is similar to the proof of Lemma E.1. Consider the random variable

$$Z_{kk'jj'} = \sup_{z \in \mathcal{X}} (\mathbb{E}_n - \mathbb{E})[K_h^2(X_{i1} - z)\psi_k(X_{ij})\psi_{k'}(X_{ij'})].$$

Define the following two function classes

$$\begin{aligned}&\mathcal{G}_h = \left\{ g_z(x_1, x_2, x_3) = h^{-2}K^2(h^{-1}(x_1 - z))\psi_k(x_2)\psi_{k'}(x_3) \,\Big|\, z \in \mathcal{X}, x_1, x_2, x_3 \in \mathcal{X} \right\} \text{ and} \\ &\mathcal{F}_h^2 = \left\{ h^{-2}K^2(h^{-1}(\cdot - z)) \,\middle|\, z \in \mathcal{X} \right\}.\end{aligned}$$

Using Lemma H.3, we bound the covering number by

$$\sup_Q N\left(\mathcal{F}_h, L^2(Q), \epsilon\right) \leq \left(\frac{8||K||_{\mathrm{TV}}^2 A^2}{h^2 \epsilon} \right)^8,$$

where Q is any measure on $\mathbb{R}$. Therefore the covering number for $\mathcal{G}_h$ satisfies

$$N\left(\mathcal{G}_h, L^2(\mathbb{P}), \epsilon\right) \leq \left(\frac{8||K||_{\mathrm{TV}}^2 A^2}{h^2 \epsilon} \right)^8.$$

The envelope of $\mathcal{G}_h$ is $U = 4h^{-2}\|K\|_\infty$ and we bound the variance of the process by

$$\begin{aligned}\sigma_P^2 &:= \mathbb{E}\left[\left(K_h^2\left(X_1 - z\right)\left(\psi_k(X_j)\psi_{k'}(X_{j'})\right)^2\right]\\&= h^{-3}\mathbb{E}\left[K^2\left(h^{-1}(X_1 - z)\right)\mathbb{E}\left[(\psi_k^2(X_j)\psi_{k'}^2(X_{j'}) \mid X_1\right]\right] \le Cm^{-2}h^{-3}.\end{aligned}$$

Using Lemma H.2, we obtain the upper bound of the expectation

$$\mathbb{E}[Z_{kk'jj'}] \le C_1\sqrt{\frac{\log(C_2 m)}{nm^2h^3}}.$$

As $|Z_{kk'jj'}| \le 4h^{-2}$ and $\sigma^2 \le Cm^{-2}h^{-3}$, Lemma H.4 gives us

$$\mathbb{P}\left(Z_{kk'jj'} \ge \mathbb{E}[Z_{kk'jj'}] + t\sqrt{Cm^{-2}h^{-3} + 4h^{-2}\mathbb{E}[Z_{kk'jj'}]} + 4t^2h^{-2}/3\right) \le \exp(-nt^2).$$

By letting $t = 3\sqrt{\log(dm)/n}$, we obtain

$$\sup_{z\in\mathcal{X}}\max_{j,j'\ge 2}\left|\boldsymbol{\Sigma}_z'(j,j') - \mathbb{E}\boldsymbol{\Sigma}_z'(j,j')\right| = O_P\left(\frac{1}{nh^2} + \sqrt{\frac{\log(dm)}{nm^2h^3}}\right). \tag{F.26}$$

We also define the empirical process

$$\bar{Z}_{kj} = \sup_{z\in\mathcal{X}}\frac{1}{n}\sum_{i=1}^n K_h(X_{i1} - z)\psi_k(X_{ij}) - \mathbb{E}\left[K_h(X_1 - z)\psi_k(X_j)\right].$$

As above, we can show that the suprema of the empirical process has the convergence rate as

$$\sup_{z\in\mathcal{X}}\max_{j\ge 2}\left|\boldsymbol{\Sigma}_z'(j,1) - \mathbb{E}\boldsymbol{\Sigma}_z'(j,1)\right| = O_P\left(\frac{1}{nh^2} + \sqrt{\frac{\log(dm)}{nmh^3}}\right). \tag{F.27}$$

Finally, we have the following upper bound

$$\begin{aligned}\sup_{z\in\mathcal{X}}\left|\boldsymbol{\Sigma}_z'(1,1) - \mathbb{E}\boldsymbol{\Sigma}_z'(1,1)\right| &\le \sup_{z\in\mathcal{X}}\left|\frac{1}{n}\sum_{i=1}^n K_h^2(X_{i1} - z) - \mathbb{E}[K_h^2(X_1 - z)]\right|\\&\le C\left(\frac{1}{\sqrt{nh^3}} + \frac{1}{nh^2}\right).\end{aligned} \tag{F.28}$$

Combining (F.26), (F.27) and (F.28), with probability at least $1 - c/n$, we have

$$\sup_{z \in \mathcal{X}} \|\boldsymbol{\Sigma}'_z - \mathbb{E}[\boldsymbol{\Sigma}'_z]\|_{\max} \leq C \left(\frac{1}{nh^2} + \frac{1}{\sqrt{nh^3}} + \sqrt{\frac{\log(dm)}{nmh^3}} \right),$$

which completes the proof of the Lemma. □

G Proof of Proposition 3.3: Examples for Assumption (A6)

In this section, we give concrete examples under which Assumption (**A6**) is satisfied. Given some $\rho \in [0, 1/2]$, denote $\mathbf{M}(\rho, p)$ as a $p \times p$ 3-banded matrix with $[\mathbf{M}(\rho, p)]_{kk} = 1 + \rho$ for all $k \in [d]$, $[\mathbf{M}(\rho, p)]_{st} = \rho$ if $0 < |s - t| \leq 1$ and $[\mathbf{M}(\rho, p)]_{st} = 0$ if $|s - t| > 1$.

For any $j \in \{2, \ldots, d\}$, we consider a covariance matrix $\boldsymbol{\Sigma}^{(j)}$ such that $\boldsymbol{\Sigma}^{(2)} = \text{diag}(\mathbf{I}_2, \mathbf{M}(\rho, d - 2)) + (\rho/d)(\mathbf{e}_1\mathbf{e}_2^T + \mathbf{e}_2^T\mathbf{e}_1)$ and

$$\boldsymbol{\Sigma}^{(j)} = \text{diag}(\mathbf{I}_2, \mathbf{M}(\rho, j - 3), 1 + \rho, \mathbf{M}(\rho, d - j)) + (\rho/d)(\mathbf{e}_1\mathbf{e}_j^T + \mathbf{e}_j^T\mathbf{e}_1) \text{ for } j > 2,$$

where $\mathbf{e}_j$ is the j-th canonical basis in $\mathbb{R}^d$ and $\delta_j \in \{0, 1\}$. We assume the covariances are sparse in the sense that $J := \sum_{j=2}^d \delta_j = O(1)$ and $\delta_2 = 1$. Given some $\pi \in (0, 1/2)$, we consider the mixture distribution

$$p(\boldsymbol{x}) = \frac{1 - \pi}{\pi^2(1 - \pi^{d-1})} \sum_{j=2}^d \frac{\pi^j}{\sqrt{(2\pi)^d |\boldsymbol{\Sigma}^{(j)}|}} \exp\left(-\frac{1}{2} \boldsymbol{x}^T (\boldsymbol{\Sigma}^{(j)})^{-1} \boldsymbol{x} \right). \tag{G.1}$$

We can easily check that

$$\sum_{j=2}^d \text{Cov}(X_1, X_j) \geq \pi^2 \rho, \text{Cov}(X_j, X_k) \geq \pi^2 \rho \text{ for all } |j - k| \leq 1.$$

We first show a property for 3-tuple joint density p_{1jk}. For simplicity, we denote $\boldsymbol{w} = (x_1, x_j, x_k)^T$ and $p_{1,j,k} \propto \exp(-\frac{1}{2}\boldsymbol{w}^T \mathbf{S}^{-1} \boldsymbol{w})$. Denote $\mathbf{S}_D = \text{diag}(\mathbf{S})$. We have

$$\delta(\mathbf{S})^2 := \|p_{1jk} - p_1 p_j p_k\|_2^2 = \frac{1}{\sqrt{(2\pi)^3}} \left[\frac{1}{\sqrt{8|\mathbf{S}|}} - \frac{2}{\sqrt{|\mathbf{S}^{-1} - \mathbf{S}_D^{-1}||\mathbf{S}||\mathbf{S}_D|}} + \frac{1}{\sqrt{8|\mathbf{S}_D|}} \right], \tag{G.2}$$

where the right hand side of the equality is obtained through integrating normal density functions and we omit the details.

Given the joint density $p(\boldsymbol{x})$ in (G.1), we have

$$\|p_{1jk} - p_1 p_j p_k\|_2 = \frac{1-\pi}{\pi^2(1-\pi^{d-1})} \sum_{s=2}^{d} \pi^s \delta(\mathbf{S}_{(s,j,k)}), \text{ where } \mathbf{S}_{(s,j,k)} = \begin{pmatrix} \mathbf{S}_{11}^{(s)} & \mathbf{S}_{1j}^{(s)} & \mathbf{S}_{1k}^{(s)} \\ \mathbf{S}_{j1}^{(s)} & \mathbf{S}_{jj}^{(s)} & \mathbf{S}_{jk}^{(s)} \\ \mathbf{S}_{k1}^{(s)} & \mathbf{S}_{kj}^{(s)} & \mathbf{S}_{kk}^{(s)} \end{pmatrix}.$$

Suppose $j < k$, and have three cases:

$$\mathbf{S}_{(s,j,k)} = \begin{cases} (1+\rho)\mathbf{I}_3 + (\rho/d)\begin{pmatrix} 0&1&0\\1&0&0\\0&0&0 \end{pmatrix}, & \text{if } s = j; \\ (1+\rho)\mathbf{I}_3 + (\rho/d)\begin{pmatrix} 0&0&0\\0&0&1\\0&1&0 \end{pmatrix}, & \text{if } s = k; \\ (1+\rho)\mathbf{I}_3 + \rho\begin{pmatrix} 0&0&0\\0&0&1\\0&1&0 \end{pmatrix}, & \text{if } s \neq j,k, \text{ and } j > 2, k = j+1; \\ (1+\rho)\mathbf{I}_3, & \text{otherwise.} \end{cases} \tag{G.3}$$

For case (i) in (G.3), applying (G.2) and we have when $\rho \in [0, 1/2]$,

$$\|p_{1jk} - p_1 p_j p_k\|_2^2 = \frac{1}{\sqrt{8(2\pi)^2}} \left[\frac{1}{\sqrt{(1+\rho)^2 - (\rho/d)^2}} - \frac{4}{\sqrt{4(1+\rho)^2 - (\rho/d)^2}} + \frac{1}{1+\rho} \right] \leq (\rho/d)^2.$$

Case (ii) in (G.3) is the same. For case (iii) in (G.3), they are just case (i) with $d = 1$, i.e., $\|p_{1jk} - p_1 p_j p_k\|_2^2 \leq \rho^2$. For case (iv), $\|p_{1jk} - p_1 p_j p_k\|_2^2 = 0$.

Summarizing the results above, since $\rho, \pi \in (0, 1/2)$, we have for any $k > 0$,

$$\begin{aligned} \sum_{j \neq k} \|p_{1jk} - p_1 p_j p_k\|_2 &= \frac{1-\pi}{\pi^2(1-\pi^{d-1})} \sum_{j \neq k} \sum_{s=2}^{d} \pi^s \delta(\mathbf{S}_{(s,j,k)}) \\ &\leq \frac{1-\pi}{\pi^2(1-\pi^{d-1})} \left[\sum_{s=2}^{d} \pi^s \rho + 2\rho\pi^2 \right] \leq 5\rho \leq \rho_{\min}/(2B), \end{aligned}$$

if $\rho/\pi \leq \rho_{\min}/(6B)$.

For the bivariate density $\|p_{jk} - p_j p_k\|_2$, we can similarly get

$$\sum_{j \geq 2} \|p_{1j} - p_1 p_j\|_2 \leq \frac{1-\pi}{\pi^2(1-\pi^{d-1})} \rho\pi^2 \leq 2\rho \leq \rho_{\min}/(2B).$$

Therefore, if we choose $\pi = 1/3$, then if $\rho \le \rho_{\min}/(18B)$, our example above satisfies

$$\sum_{j=2}^{d} \mathrm{Cov}(X_1, X_j) \ge \rho/9 \text{ and } \mathrm{Cov}(X_j, X_k) \ge \rho/9 \text{ for all } |j-k| \le 1,$$

and Assumption (**A6**) holds.

H Results on Empirical Processes

Lemma H.1 (Lemma H.2, Lu et al. (2015))**.** Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two function classes satisfying

$$N(\mathcal{F}_1, \|\cdot\|_{L_2(Q)}, a_1\epsilon) \le C_1\epsilon^{-v_1} \qquad \text{and} \qquad N(\mathcal{F}_2, \|\cdot\|_{L_2(Q)}, a_2\epsilon) \le C_2\epsilon^{-v_2}$$

for some $C_1, C_2, a_1, a_2, v_1, v_2 > 0$ and any $0 < \epsilon < 1$. Define $\|\mathcal{F}_\ell\|_\infty = \sup\{\|f\|_\infty, f \in \mathcal{F}_\ell\}$ for $\ell = 1, 2$ and $U = \|\mathcal{F}_1\|_\infty \vee \|\mathcal{F}_2\|_\infty$. For the function classes $\mathcal{F}_\times = \{f_1 f_2 \,|\, f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$ and $\mathcal{F}_+ = \{f_1 + f_2 \,|\, f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$, we have for any $\epsilon \in (0,1)$,

$$N(\mathcal{F}_\times, \|\cdot\|_{L_2(Q)}, \epsilon) \le C_1C_2\left(\frac{2a_1U}{\epsilon}\right)^{v_1}\left(\frac{2a_2U}{\epsilon}\right)^{v_2};$$
$$N(\mathcal{F}_+, \|\cdot\|_{L_2(Q)}, \epsilon) \le C_1C_2\left(\frac{2a_1}{\epsilon}\right)^{v_1}\left(\frac{2a_2}{\epsilon}\right)^{v_2}.$$

Lemma H.2 (Corollary 5.1, Chernozhukov et al. (2014b))**.** Assume that the functions in $\mathcal{F}$ defined on $\mathcal{X}$ are uniformly bounded by an envelope function $F(\cdot)$ such that $|f(x)| \le F(x)$ for all $x \in \mathcal{X}$ and $f \in \mathcal{F}$. Define $\sigma_P^2 = \sup_{f\in\mathcal{F}} \mathbb{E}[f^2]$. Let Q be any measure over $\mathcal{X}$. If for some $A \ge e, V \ge 0$ and for all $\varepsilon > 0$, the covering entropy satisfies

$$\sup_Q N(\mathcal{F}, L^2(Q); \epsilon) \le \left(\frac{A\|F\|_{L^2(Q)}}{\varepsilon}\right)^V,$$

then for any i.i.d. subgaussian mean zero random variables $\varepsilon_1, \ldots, \varepsilon_n$ there exits a universal constant C such that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n (f(X_{i1}) - \mathbb{E}f(X))\right] \le C\left[\sqrt{\frac{V}{n}}\sigma_P\sqrt{\log\frac{A\|F\|_{L^2(\mathbb{P})}}{\sigma_P}} + \frac{V\|F\|_{L^2(\mathbb{P})}}{\sqrt{n}}\log\frac{A\|F\|_{L^2(\mathbb{P})}}{\sigma_P}\right].$$

Lemma H.3 (Lemma 3, Giné and Nickl (2009))**.** Let $K : \mathbb{R} \mapsto \mathbb{R}$ be a bounded variation function. Define the function class $\mathcal{F}_h = \{K((t - \cdot)/h) \,|\, t \in \mathbb{R}\}$. There exists $A < \infty$ such that for all probability measures Q on $\mathbb{R}$, we have

$$\sup_{Q} N(\mathcal{F}_h, L^2(Q), \epsilon) \leq \left(\frac{2\|K\|_{\mathrm{TV}} A}{\epsilon}\right)^4, \text{ for any } \epsilon \in (0, 1).$$

Lemma H.4 (Bousquet (2002))**.** Let $X_1, \ldots, X_n$ be independent random variables and $\mathcal{F}$ is a function class such that there exist η_n and τ_n^2 satisfying

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq \eta_n \quad \text{and} \quad \sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \mathrm{Var}(f(X_{i1})) \leq \tau_n^2.$$

Define the random variable Z being the suprema of an empirical process

$$Z = \sup_{f \in \mathcal{F}} \left| \frac{1}{n} \sum_{i=1}^{n} (f(X_{i1}) - \mathbb{E} f(X_{i1})) \right|. \tag{H.1}$$

Then for any $z > 0$, we have the following concentration inequality on the suprema

$$\mathbb{P}\left(Z \geq \mathbb{E}Z + z\sqrt{2(\tau_n^2 + 2\eta_n \mathbb{E}Z)} + 2z^2\eta_n/3\right) \leq \exp(-nz^2).$$

The following lemma gives the deviation inequality when $\mathcal{F}$ is not universally bounded.

Lemma H.5 (Theorem 5.1, Chernozhukov et al. (2014b))**.** Let $F(\cdot)$ be the envelope function of $\mathcal{F}$ such that $F \in L^2(\mathbb{P})$ and Z is defined in (H.1), For every $t \geq 1$, there exists a universal constant C such that

$$\mathbb{P}\left(Z \geq 2\mathbb{E}Z + C(\sigma_P + \|F\|_{L^2(\mathbb{P})})z + \|F\|_{L^2(\mathbb{P})} z^2\right) \leq 1/z^2.$$

References

Bousquet, O. (2002). A Bennett concentration inequality and its application to suprema of empirical processes. *Comptes Rendus Mathematique* **334** 495–500.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Ann. Stat.* **42** 1787–1818.

Chernozhukov, V., Chetverikov, D. and Kato, K. (2014b). Gaussian approximation of suprema of empirical processes. *Ann. Stat.* **42** 1564–1597.

de Boor, C. (2001). *A practical guide to splines*, vol. 27 of *Applied Mathematical Sciences*. Revised ed. Springer-Verlag, New York.

Demmel, J. (1992). The componentwise distance to the nearest singular matrix. *SIAM Journal on Matrix Analysis and Applications* **13** 10–19.

Farrell, M. H. (2013). Robust inference on average treatment effects with possibly more covariates than observations. *arXiv preprint arXiv:1309.4686* .

Friedman, J. H., Hastie, T. J. and Tibshirani, R. J. (2010). A note on the group lasso and a sparse group lasso. *ArXiv e-prints, arXiv:1001.0736* .

Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. In *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, vol. 38. Elsevier.

Giné, E. and Nickl, R. (2009). An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probab. Theory Related Fields* **143** 569–596.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13–30.

Huang, J., Horowitz, J. L. and Wei, F. (2010). Variable selection in nonparametric additive models. *Ann. Stat.* **38** 2282–2313.

Lu, J., Kolar, M. and Liu, H. (2015). Post-regularization inference for dynamic nonparanormal graphical models. *arXiv preprint arXiv:1512.08298* .

Qin, Z., Scheinberg, K. and Goldfarb, D. (2013). Efficient block-coordinate descent algorithms for the group lasso. *Mathematical Programming Computation* **5** 143–169.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Yuan, M. and Lin, Y. (2007). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. B* **68** 49–67.

Zhou, S., Shen, X. and Wolfe, D. A. (1998). Local asymptotics for regression splines and confidence regions. *Ann. Stat.* **26** 1760–1782.