Supplemental Appendix to "Strong Approximations for Empirical Processes Indexed by Lipschitz Functions"

Matias D. Cattaneo[1] Ruiqi (Rae) Yu[1*]

June 6, 2024

Abstract

This supplement appendix reports some additional results not discussed in the main paper to conserve space, and provides all the technical proofs.

[1]Department of Operations Research and Financial Engineering, Princeton University
[*]Corresponding author: rae.yu@princeton.edu

Contents

SA-I Additional Results

This section presents additional results not reported in the paper to conserve space and streamline the presentation.

SA-I.1 General Empirical Process

The following corollaries provide additional results for their counterparts in Section 3.1 of the paper. In particular, the results reported here allow for exponentially decaying tails, and for a more general expression under polynomial entropy condition.

Corollary SA.1 (VC-Type Bounded Functions)**.** *Suppose the conditions of Corollary 1 hold. Then,*

$$\mathsf{S}_n(t) = \mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\mathcal{H}}(t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n))\mathtt{TV}_{\mathcal{H}}} + \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n))$$

in Theorem 1.

Corollary SA.2 (VC-Type Lipschitz Functions)**.** *Suppose the conditions of Corollary 2 hold. Then,*

$$\begin{aligned}\mathsf{S}_n(t) = \min\left\{\mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\mathcal{H}}}, \mathsf{l}_{n,d}\sqrt{\mathtt{L}_{\mathcal{H}}}\right\}&\sqrt{(t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n))\mathtt{TV}_{\mathcal{H}}}\\ &+ \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n))\end{aligned}$$

in Theorem 1.

Corollary SA.3 (Polynomial-Entropy Functions)**.** *Suppose the conditions of Corollary 2 hold. Then,*

$$\mathsf{S}_n(t) = \mathtt{a}_{\mathcal{H}}(2 - \mathtt{b}_{\mathcal{H}})^{-2}\min\{\mathsf{S}_n^{bdd}(t), \mathsf{S}_n^{lip}(t), \mathsf{S}_n^{err}(t)\}$$

in Theorem 1, where

$$\begin{aligned}\mathsf{S}_n^{bdd}(t) &= \mathsf{m}_{n,d}\sqrt{d\mathtt{c}_1\mathtt{M}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}}}(\sqrt{t} + (\mathsf{m}_{n,d}^2\mathtt{M}_{\mathcal{H}}^{-1}\mathtt{TV}_{\mathcal{H}})^{-\frac{\mathtt{b}_{\mathcal{H}}}{4}})\\ &\quad+ \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{d^3\mathtt{c}_3\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + (\mathsf{m}_{n,d}^2\mathtt{M}_{\mathcal{H}}^{-1}\mathtt{TV}_{\mathcal{H}})^{-\frac{\mathtt{b}_{\mathcal{H}}}{2}}),\\ \mathsf{S}_n^{lip}(t) &= \mathsf{l}_{n,d}\sqrt{d\mathtt{c}_1\mathtt{c}_2\mathtt{L}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}}}(\sqrt{t} + (\mathsf{l}_{n,d}^2\mathtt{M}_{\mathcal{H}}^{-2}\mathtt{L}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}})^{-\frac{\mathtt{b}_{\mathcal{H}}}{4}})\\ &\quad+ \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{d^3\mathtt{c}_3\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + (\mathsf{l}_{n,d}^2\mathtt{M}_{\mathcal{H}}^{-2}\mathtt{L}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}})^{-\frac{\mathtt{b}_{\mathcal{H}}}{2}}),\\ \mathsf{S}_n^{err}(t) &= \min\{\mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\mathcal{H}}}, \mathsf{l}_{n,d}\sqrt{\mathtt{c}_2\mathtt{L}_{\mathcal{H}}}\}\sqrt{d\mathtt{c}_1\mathtt{TV}_{\mathcal{H}}}(\sqrt{t} + n^{\frac{\mathtt{b}_{\mathcal{H}}}{2(\mathtt{b}_{\mathcal{H}}+2)}})\\ &\quad+ \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{d^3\mathtt{c}_3\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + n^{\frac{\mathtt{b}_{\mathcal{H}}}{\mathtt{b}_{\mathcal{H}}+2}}) + n^{-\frac{1}{b+2}}\mathtt{M}_{\mathcal{H}}\sqrt{t}.\end{aligned}$$

SA-I.2 Multiplicative-Separable Empirical Process

This section considers uniform Gaussian strong approximation for the following multiplicative-separable empirical process:

$$M_n(g, r) := \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(g(\mathbf{x}_i)r(y_i) - \mathbb{E}[g(\mathbf{x}_i)r(y_i)]), \qquad (g, r) \in \mathcal{G} \times \mathcal{R}. \tag{SA-1}$$

For example, the local empirical process discussed in Section 4 can also be represented as $(M_n(g,r):(g,r)\in \mathcal{G}\times\mathcal{R})$ with $\mathcal{G}=\{b^{-d/2}K((\cdot-\mathbf{x})/b):\mathbf{x}\in\mathcal{X}\}$ and $\mathcal{R}=\{\mathrm{Id}\}$, but calculated based on a centered sample $((\mathbf{x}_i,y_i'):1\le i\le n)$, with $y_i'=y_i-\mathbb{E}[y_i|\mathbf{x}_i]$.

The results and proof techniques for the multiplicative-separable empirical process are similar to those for the residual-based empirical process studied in the paper, but we report them here for completeness.

Theorem SA.1. *Suppose Assumption B holds with $\mathcal{X}=[0,1]^d$, and the following two conditions hold.*

(i) *$\mathcal{G}$ is a real-valued pointwise measurable class of functions on $(\mathcal{X},\mathcal{B}(\mathcal{X}),\mathbb{P}_X)$ such that $J(\mathcal{G},\mathtt{M}_{\mathcal{G}},1)<\infty$.*

(ii) *$\mathcal{R}$ be a real-valued pointwise measurable class of functions on $(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathbb{P}_Y)$ such that $J(\mathcal{R},M_{\mathcal{R}},1)<\infty$. Furthermore, one of the following holds:*

   (a) *$\mathtt{M}_{\mathcal{R}}\lesssim 1$ and $\mathtt{pTV}_{\mathcal{R}}\lesssim 1$, and set $\alpha=0$, or*

   (b) *$M_{\mathcal{R}}(y)\lesssim 1+|y|^\alpha$ and $\mathtt{pTV}_{\mathcal{R},(-|y|,|y|)}\lesssim 1+|y|^\alpha$ for all $y\in\mathbb{R}$ and for some $\alpha>0$, and $\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}[\exp(y_i)|\mathbf{x}_i=\mathbf{x}]\le 2$.*

Then, on a possibly enlarged probability space, there exists a sequence of mean-zero Gaussian processes $(Z_n^M(g,r):(g,r)\in\mathcal{G}\times\mathcal{R}))$ with almost surely continuous trajectory such that:

- $\mathbb{E}[M_n(g_1,r_1)M_n(g_2,r_2)]=\mathbb{E}[Z_n^M(g_1,r_1)Z_n^M(g_2,r_2)]$ *for all* $(g_1,r_1),(g_2,r_2)\in\mathcal{G}\times\mathcal{R}$, *and*
- $\mathbb{P}[\|M_n-Z_n^M\|_{\mathcal{G}\times\mathcal{R}}>C_1C_\alpha\mathsf{T}_n^M(t)]\le C_2e^{-t}$ *for all* $t>0$,

where C_1 and C_2 are universal constants, $C_\alpha=\max\{1+(2\alpha)^{\frac{\alpha}{2}},1+(4\alpha)^\alpha\}$ and

$$\mathsf{T}_n^M(t)=\min_{\delta\in(0,1)}\{\mathsf{A}_n^M(t,\delta)+\mathsf{F}_n^M(t,\delta)\}$$

with

$$\begin{aligned}\mathsf{A}_n^M(t,\delta):=&\sqrt{d}\min\Big\{\Big(\frac{\mathtt{c}_1^d\mathtt{E}_{\mathcal{G}}\mathtt{TV}_{\mathcal{G}}^d\mathtt{M}_{\mathcal{G}}^{d+1}}{n}\Big)^{\frac{1}{2(d+1)}},\Big(\frac{\mathtt{c}_1^d\mathtt{c}_2^d\mathtt{E}_{\mathcal{G}}^2\mathtt{M}_{\mathcal{G}}^2\mathtt{TV}_{\mathcal{G}}^d\mathtt{L}_{\mathcal{G}}^d}{n^2}\Big)^{\frac{1}{2(d+2)}}\Big\}(t+\log(n\mathtt{N}_{\mathcal{G}}(\delta/2)\mathtt{N}_{\mathcal{R}}(\delta/2)N^*))^{\alpha+1}\\&+\sqrt{\frac{\min\{\mathtt{M}_{\mathcal{G}}^2(M^*+N^*),\mathtt{M}_{\mathcal{G}}(\mathtt{c}_3\mathtt{K}_{\mathcal{G}}\mathtt{M}_{\mathcal{V}_{\mathcal{R}}}+\mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{V}_{\mathcal{R}}}+\mathtt{M}_{\mathcal{G}})\}}{n}}(\log n)^\alpha(t+\log(n\mathtt{N}_{\mathcal{G}}(\delta/2)\mathtt{N}_{\mathcal{R}}(\delta/2)N^*))^{\alpha+1},\\\mathsf{F}_n^M(t,\delta):=&J(\delta)\mathtt{M}_{\mathcal{G}}+\frac{(\log n)^{\alpha/2}\mathtt{M}_{\mathcal{G}}J^2(\delta)}{\delta^2\sqrt{n}}+\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}\sqrt{t}+(\log n)^\alpha\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}t^\alpha,\end{aligned}$$

and

$$\begin{aligned}\mathcal{V}_{\mathcal{R}}&:=\{\theta(\cdot,r):\mathbf{x}\mapsto\mathbb{E}[r(y_i)|\mathbf{x}_i=\mathbf{x}],\mathbf{x}\in\mathcal{X},r\in\mathcal{R}\},\\J(\delta)&:=\sqrt{2}J(\mathcal{G},\mathtt{M}_{\mathcal{G}},\delta/\sqrt{2})+\sqrt{2}J(\mathcal{R},M_{\mathcal{R}},\delta/\sqrt{2}),\\M^*&:=\Big\lfloor\log_2\min\Big\{\Big(\frac{n\mathtt{TV}_{\mathcal{G}}}{\mathtt{E}_{\mathcal{G}}}\Big)^{\frac{d}{d+1}},\Big(\frac{n\mathtt{L}_{\mathcal{G}}\mathtt{TV}_{\mathcal{G}}}{\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}\Big)^{\frac{d}{d+2}}\Big\}\Big\rfloor,\\N^*&:=\Big\lceil\log_2\max\Big\{\Big(\frac{n\mathtt{M}_{\mathcal{G}}^{d+1}}{\mathtt{E}_{\mathcal{G}}\mathtt{TV}_{\mathcal{G}}^d}\Big)^{\frac{1}{d+1}},\Big(\frac{n^2\mathtt{M}_{\mathcal{G}}^{2d+2}}{\mathtt{TV}_{\mathcal{G}}^d\mathtt{L}_{\mathcal{G}}^d\mathtt{E}_{\mathcal{G}}^2}\Big)^{\frac{1}{d+2}}\Big\}\Big\rceil.\end{aligned}$$

Corollary SA.4 (VC-Type Lipschitz Functions)**.** *Suppose the conditions of Theorem SA.1 hold. In addition, assume that $\mathcal{G}$ is a VC-type class with respect to envelope function $\mathtt{M}_{\mathcal{G}}$ with constant $\mathtt{c}_{\mathcal{G}}\ge e$ and exponent*

$\mathtt{d}_{\mathcal{G}} \geq 1$, *and* $\mathcal{R}$ *is a VC-type class with respect to* $M_{\mathcal{R}}$ *with constant* $\mathtt{c}_{\mathcal{R}} \geq e$ *and exponent* $\mathtt{d}_{\mathcal{R}} \geq 1$. *Suppose there exists a constant* $\mathtt{c}_4$ *such that* $|\log_2 \mathtt{E}_{\mathcal{G}}| + |\log_2 \mathtt{TV}| + |\log_2 \mathtt{M}_{\mathcal{G}}| \leq \mathtt{c}_4 \log_2 n$, *where* $\mathtt{TV} = \max\{\mathtt{TV}_{\mathcal{G}}, \mathtt{TV}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}}\}$ *with* $\mathcal{V}_{\mathcal{R}} := \{\theta(\cdot, r) : \mathbf{x} \mapsto \mathbb{E}[r(y_i)|\mathbf{x}_i = \mathbf{x}], \mathbf{x} \in \mathcal{X}, r \in \mathcal{R}\}$. *Then,*

$$\begin{aligned}\mathsf{T}_n^M(t) = \sqrt{d}\min\Big\{\Big(\frac{\mathtt{c}_1^d \mathtt{E}_{\mathcal{G}} \mathtt{TV}_{\mathcal{G}}^d \mathtt{M}_{\mathcal{G}}^{d+1}}{n}\Big)^{\frac{1}{2(d+1)}}, \Big(\frac{\mathtt{c}_1^d \mathtt{c}_2^d \mathtt{E}_{\mathcal{G}}^2 \mathtt{M}_{\mathcal{G}}^2 \mathtt{TV}_{\mathcal{G}}^d \mathtt{L}_{\mathcal{G}}^d}{n^2}\Big)^{\frac{1}{2(d+2)}}\Big\}(t + \mathtt{c}_4\log_2(n) + \mathtt{d}\log(\mathtt{c}n))^{\alpha+1} \\ + \sqrt{\frac{\min\{\mathtt{c}_3\log_2(n)\mathtt{M}_{\mathcal{G}}^2, \mathtt{M}_{\mathcal{G}}(\mathtt{c}_3\mathtt{K}_{\mathcal{G}}\mathtt{M}_{\mathcal{V}_{\mathcal{R}}} + \mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{V}_{\mathcal{R}}} + \mathtt{M}_{\mathcal{G}})\}}{n}}(\log n)^{\alpha}(t + \mathtt{c}_4\log_2(n) + \mathtt{d}\log(\mathtt{c}n))^{\alpha+1}.\end{aligned}$$

in Theorem SA.1, where $\mathtt{c} = \mathtt{c}_{\mathcal{G}}\mathtt{c}_{\mathcal{R}}$, $\mathtt{d} = \mathtt{d}_{\mathcal{G}} + \mathtt{d}_{\mathcal{R}}$.

Theorem SA.2. *Suppose* $(\mathbf{z}_i = (\mathbf{x}_i, y_i), 1 \leq i \leq n)$ *are i.i.d. random variables taking values in* $(\mathcal{X} \times \mathbb{R}, \mathcal{B}(\mathcal{X} \times \mathbb{R}))$ *with* $\mathcal{X} \subseteq \mathbb{R}^d$, *and the following conditions hold.*

(i) $\mathcal{G}$ *is a class of functions on* $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$ *such that* $\mathtt{M}_{\mathcal{G}} < \infty$ *and* $\mathcal{G} \subseteq \operatorname{Span}\{\mathbb{1}_{\Delta_l} : 0 \leq l < L\}$, *where* $\{\Delta_l : 0 \leq l < L\}$ *forms a quasi-uniform partition of* $\mathcal{X}$ *in the sense that*

$$\mathcal{X} \subseteq \sqcup_{0 \leq l < L}\Delta_l \qquad \text{and} \qquad \frac{\max_{0\leq l<L}\mathbb{P}_X(\Delta_l)}{\min_{0\leq l<L}\mathbb{P}_X(\Delta_l)} \leq \rho < \infty.$$

In addition, $J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, 1) < \infty$.

(ii) $\mathcal{R}$ *is a real-valued pointwise measurable class of functions on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$, *such that* $J(\mathcal{R}, M_{\mathcal{R}}, 1) < \infty$. *Furthermore, one of the following holds:*

  (a) $\mathtt{M}_{\mathcal{R}} \lesssim 1$ *and* $\mathtt{pTV}_{\mathcal{R}} \lesssim 1$, *and set* $\alpha = 0$, *or*

  (b) $M_{\mathcal{R}}(y) \lesssim 1+|y|^{\alpha}$, $\mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \lesssim 1+|y|^{\alpha}$ *for all* $y \in \mathbb{R}$ *and for some* $\alpha > 0$, *and* $\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

(iii) *There exists a constant* $\mathtt{c}_5$ *such that* $|\log_2 \mathtt{E}_{\mathcal{G}}| + |\log_2 \mathtt{M}_{\mathcal{G}}| + |\log_2 L| \leq \mathtt{c}_5 \log_2 n$.

Then, on a possibly enlarged probability space, there exists mean-zero Gaussian processes $(Z_n^M(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ *with almost sure continuous trajectory such that:*

- $\mathbb{E}[M_n(g_1, r_1)M_n(g_2, r_2)] = \mathbb{E}[Z_n^M(g_1, r_1)Z_n^M(g_2, r_2)]$ *for all* $(g_1, r_1), (g_2, r_2) \in \mathcal{G} \times \mathcal{R}$, *and*
- $\mathbb{P}[\|M_n - Z_n^M\|_{\mathcal{G}\times\mathcal{R}} > C_1 C_\alpha C_\rho \min_{\delta\in(0,1)}(\mathsf{H}_n^M(t, \delta) + \mathsf{F}_n^M(t, \delta))] \leq C_2 e^{-t} + Le^{-C_\rho n/L}$ *for all* $t > 0$,

where C_1 *and* C_2 *are universal constants,* $C_\alpha = \max\{1 + (2\alpha)^{\frac{\alpha}{2}}, 1 + (4\alpha)^{\alpha}\}$, C_ρ *is a constant that only depends on* ρ,

$$\begin{aligned}\mathsf{H}_n^M(t, \delta) := \sqrt{\frac{L\mathtt{M}_{\mathcal{G}}\mathtt{E}_{\mathcal{G}}}{n}}\,(t + \log \mathtt{N}_{\mathcal{G}}(\delta/2) + \log \mathtt{N}_{\mathcal{R}}(\delta/2) + \log_2 N^*)^{\alpha+\frac{1}{2}} \\ + \sqrt{\frac{\min\{L + N^*, \mathtt{S}_{\mathcal{G}}^2\}}{n}}\mathtt{M}_{\mathcal{G}}(\log n)^{\alpha}\,(t + \log \mathtt{N}_{\mathcal{G}}(\delta/2) + \log \mathtt{N}_{\mathcal{R}}(\delta/2) + \log_2 N^*)^{\alpha+1},\end{aligned}$$

with $\mathtt{c} = \mathtt{c}_{\mathcal{G}}\mathtt{c}_{\mathcal{R}}$, $\mathtt{d} = \mathtt{d}_{\mathcal{G}} + \mathtt{d}_{\mathcal{R}}$, $N^* = \Big\lceil \log_2\Big(\frac{n\mathtt{M}_{\mathcal{G}}}{2^L\mathtt{E}_{\mathcal{G}}}\Big)\Big\rceil$, $\mathtt{S}_{\mathcal{G}} = \sup_{g\in\mathcal{G}}\sum_{l=1}^{L}\mathbb{1}(\operatorname{Supp}(g) \cap \Delta_l \neq \emptyset)$.

SA-I.3 Residual-Based Empirical Process

The following theorem presents a generalization of Theorem 3 in the paper.

Theorem SA.3. *Suppose Assumption B holds with $\mathcal{X} = [0,1]^d$, and the following two conditions hold.*

(i) $\mathcal{G}$ *is a real-valued pointwise measurable class of functions on $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$ such that $J(\mathcal{G}, \mathtt{M}_\mathcal{G}, 1) < \infty$.*

(ii) $\mathcal{R}$ *be a real-valued pointwise measurable class of functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$ such that $J(\mathcal{R}, M_\mathcal{R}, 1) < \infty$. Furthermore, one of the following holds:*

  (a) $\mathtt{M}_\mathcal{R} \lesssim 1$ *and* $\mathtt{pTV}_\mathcal{R} \lesssim 1$, *and set* $\alpha = 0$, *or*

  (b) $M_\mathcal{R}(y) \lesssim 1+|y|^\alpha$, $\mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \lesssim 1+|y|^\alpha$ *for all $y \in \mathbb{R}$ and for some $\alpha > 0$, and* $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

Then, on a possibly enlarged probability space, there exists a sequence of mean-zero Gaussian processes $(Z_n^R(g,r) : (g,r) \in \mathcal{G}\times\mathcal{R}))$ with almost surely continuous trajectory such that:

1. $\mathbb{E}[R_n(g_1,r_1)R_n(g_2,r_2)] = \mathbb{E}[Z_n^R(g_1,r_1)Z_n^R(g_2,r_2)]$ *for all* $(g_1,r_1),(g_2,r_2) \in \mathcal{G}\times\mathcal{R}$.

2. $\mathbb{P}[\|R_n - Z_n^R\|_{\mathcal{G}\times\mathcal{R}} > C_1 C_\alpha \mathsf{T}_n^R(t)] \leq C_2 e^{-t}$ *for all* $t > 0$,

where C_1 and C_2 are universal constants, $C_\alpha = \max\{1+(2\alpha)^{\frac{\alpha}{2}}, 1+(4\alpha)^\alpha\}$, and and

$$\mathsf{T}_n^R(t) = \min_{\delta\in(0,1)}\{\mathsf{A}_n^R(t,\delta) + \mathsf{F}_n^R(t,\delta)\}$$

with

$$\begin{aligned}\mathsf{A}_n^R(t,\delta) &:= \sqrt{d}\min\Big\{\Big(\frac{\mathtt{c}_1^d \mathtt{E}_\mathcal{G}\mathtt{TV}^d\mathtt{M}_\mathcal{G}^{d+1}}{n}\Big)^{\frac{1}{2(d+1)}}, \Big(\frac{\mathtt{c}_1^d\mathtt{c}_2^d\mathtt{E}_\mathcal{G}^2\mathtt{M}_\mathcal{G}^2\mathtt{TV}^d\mathtt{l}^d}{n^2}\Big)^{\frac{1}{2(d+2)}}\Big\}(t+\log(n\mathtt{N}_\mathcal{G}(\delta/2)\mathtt{N}_\mathcal{R}(\delta/2)N_*))^{\alpha+1}\\ &\quad + \frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}}(\log n)^\alpha(t+\log(n\mathtt{N}_\mathcal{G}(\delta/2)\mathtt{N}_\mathcal{R}(\delta/2)N_*))^{\alpha+1},\\ \mathsf{F}_n^R(t,\delta) &:= J(\delta)\mathtt{M}_\mathcal{G} + \frac{\log(n)\mathtt{M}_\mathcal{G}J^2(\delta)}{\delta^2\sqrt{n}} + \frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}}\sqrt{t} + (\log n)^\alpha\frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}}t^\alpha,\end{aligned}$$

and

$$\begin{aligned}\mathtt{TV} &:= \max\{\mathtt{TV}_\mathcal{G}, \mathtt{TV}_{\mathcal{G}\times\mathcal{V}_\mathcal{R}}\}, \qquad \mathtt{L} := \max\{\mathtt{L}_\mathcal{G}, \mathtt{L}_{\mathcal{G}\times\mathcal{V}_\mathcal{R}}\},\\ \mathcal{V}_\mathcal{R} &:= \{\theta(\cdot,r) : \mathbf{x}\mapsto\mathbb{E}[r(y_i)|\mathbf{x}_i=\mathbf{x}], \mathbf{x}\in\mathcal{X}, r\in\mathcal{R}\},\\ M_* &:= \Big\lfloor \log_2\min\Big\{\Big(\frac{n\mathtt{TV}}{\mathtt{E}_\mathcal{G}}\Big)^{\frac{d}{d+1}}, \Big(\frac{n\mathtt{L}\mathtt{TV}_\mathcal{G}}{\mathtt{E}_\mathcal{G}\mathtt{M}_\mathcal{G}}\Big)^{\frac{d}{d+2}}\Big\}\Big\rfloor,\\ N_* &:= \Big\lceil \log_2\max\Big\{\Big(\frac{n\mathtt{M}_\mathcal{G}^{d+1}}{\mathtt{E}_\mathcal{G}\mathtt{TV}^d}\Big)^{\frac{1}{d+1}}, \Big(\frac{n^2\mathtt{M}_\mathcal{G}^{2d+2}}{\mathtt{TV}^d\mathtt{L}^d\mathtt{E}_\mathcal{G}^2}\Big)^{\frac{1}{d+2}}\Big\}\Big\rceil.\end{aligned}$$

The following theorem presents a generalization of Theorem 4 in the paper.

Theorem SA.4. *Suppose $(\mathbf{z}_i = (\mathbf{x}_i, y_i), 1 \leq i \leq n)$ are i.i.d. random variables taking values in $(\mathcal{X}\times\mathbb{R}, \mathcal{B}(\mathcal{X}\times\mathbb{R}))$ with $\mathcal{X}\subseteq\mathbb{R}^d$, and the following conditions hold.*

(i) $\mathcal{G}$ *is a class of functions on* $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$ *such that* $\mathtt{M}_\mathcal{G} < \infty$ *and* $\mathcal{G} \subseteq \mathrm{Span}\{\mathbb{1}_{\Delta_l} : 0 \leq l < L\}$*, where* $\{\Delta_l : 0 \leq l < L\}$ *forms a quasi-uniform partition of* $\mathcal{X}$ *in the sense that*

$$\mathcal{X} \subseteq \sqcup_{0 \leq l < L} \Delta_l \qquad \text{and} \qquad \frac{\max_{0 \leq l < L} \mathbb{P}_X(\Delta_l)}{\min_{0 \leq l < L} \mathbb{P}_X(\Delta_l)} \leq \rho < \infty.$$

In addition, $J(\mathcal{G}, \mathtt{M}_\mathcal{G}, 1) < \infty$.

(ii) $\mathcal{R}$ *is a real-valued pointwise measurable class of functions on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$*, such that* $J(\mathcal{R}, M_\mathcal{R}, 1) < \infty$*. Furthermore, one of the following holds:*

(a) $\mathtt{M}_\mathcal{R} \lesssim 1$ *and* $\mathtt{pTV}_\mathcal{R} \lesssim 1$*, and set* $\alpha = 0$*, or*

(b) $M_\mathcal{R}(y) \lesssim 1 + |y|^\alpha$, $\mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \lesssim 1 + |y|^\alpha$ *for all* $y \in \mathbb{R}$ *and for some* $\alpha > 0$*, and* $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[\exp(y_i) | \mathbf{x}_i = \mathbf{x}] \leq 2$.

(iii) *There exists a constant* $\mathtt{c}_5$ *such that* $|\log_2 \mathtt{E}_\mathcal{G}| + |\log_2 \mathtt{M}_\mathcal{G}| + |\log_2 L| \leq \mathtt{c}_5 \log_2 n$.

Then, on a possibly enlarged probability space, there exists mean-zero Gaussian processes $(Z_n^R(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ *with almost sure continuous trajectory such that:*

- $\mathbb{E}[R_n(g_1, r_1) R_n(g_2, r_2)] = \mathbb{E}[Z_n^R(g_1, r_1) Z_n^R(g_2, r_2)]$ *for all* $(g_1, r_1), (g_2, r_2) \in \mathcal{G} \times \mathcal{R}$*, and*
- $\mathbb{P}[\|R_n - Z_n^R\|_{\mathcal{G} \times \mathcal{R}} > C_1 C_\alpha (C_\rho \min_{\delta \in (0,1)} (\mathsf{H}_n^R(t, \delta) + \mathsf{F}_n^R(t, \delta)) + \mathsf{W}_n(t))] \leq C_2 e^{-t} + L e^{-C_\rho n / L}$ *for all* $t > 0$,

where C_1 *and* C_2 *are universal constants,* $C_\alpha = \max\{1 + (2\alpha)^{\frac{\alpha}{2}}, 1 + (4\alpha)^\alpha\}$*,* C_ρ *is a constant that only depends on* ρ*,*

$$\begin{aligned}
\mathsf{H}_n^R(t, \delta) &:= \sqrt{\frac{L \mathtt{M}_\mathcal{G} \mathtt{E}_\mathcal{G}}{n}} \left(t + \log \mathtt{N}_\mathcal{G}(\delta/2) + \log \mathtt{N}_\mathcal{R}(\delta/2) + \log_2 N^*\right)^{\alpha + \frac{1}{2}} \\
&\quad + \frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}} (\log n)^\alpha \left(t + \log \mathtt{N}_\mathcal{G}(\delta/2) + \log \mathtt{N}_\mathcal{R}(\delta/2) + \log_2 N^*\right)^{\alpha + 1}, \\
\mathsf{W}_n(t) &:= \mathbb{1}(\mathrm{card}(\mathcal{R}) > 1) \sqrt{\mathtt{M}_\mathcal{G} \mathtt{E}_\mathcal{G}} \Big(\max_{0 \leq l < L} \|\Delta_l\|_\infty\Big) \mathtt{L}_{\mathcal{V}_\mathcal{R}} \sqrt{t + \log \mathtt{N}_\mathcal{G}(\delta/2) + \log \mathtt{N}_\mathcal{R}(\delta/2) + \log_2 N^*}.
\end{aligned}$$

with $\mathcal{V}_\mathcal{R} := \{\theta(\cdot, r) : \mathbf{x} \mapsto \mathbb{E}[r(y_i) | \mathbf{x}_i = \mathbf{x}], \mathbf{x} \in \mathcal{X}, r \in \mathcal{R}\}$.

SA-I.4 Local Polynomial Estimators

The following lemma provides the sufficient conditions for the results discussed in Section 4.1 in the paper.

Lemma SA.1. *Consider the setup of Section 4.1, and assume the following regularity conditions hold:*

(a) *Assumption B holds.*

(b) $\mathbf{x} \mapsto \theta(\mathbf{x}; r)$ *is* $(\mathfrak{p} + 1)$*-times continuously differentiable with bounded* $(\mathfrak{p} + 1)$*th partial derivatives uniformly over* $\mathbf{x} \in \mathcal{X}$ *and* $r \in \mathcal{R}_l$, $l = 1, 2$*, for some* $\mathfrak{p} \geq 0$.

(c) $K : \mathbb{R}^d \to \mathbb{R}$ *is non-negative, Lipschitz, and compact supported.*

If $(nb^d)^{-1}\log n \to 0$, *then*

$$\sup_{\mathbf{x}\in\mathcal{X},r\in\mathcal{R}_2} \left|\mathbf{e}_1^\top(\widehat{\mathbf{H}}_{\mathbf{x}}^{-1} - \mathbf{H}_{\mathbf{x}}^{-1})\mathbf{S}_{\mathbf{x},r}\right| = O((nb^d)^{-1}\log n) \qquad a.s., \quad and$$

$$\sup_{\mathbf{x}\in\mathcal{X},r\in\mathcal{R}_l} \left|\mathbb{E}[\widehat{\theta}(\mathbf{x},r)|\mathbf{x}_1,\cdots,\mathbf{x}_n] - \theta(\mathbf{x},r)\right| = O(b^{1+\mathfrak{p}}) \qquad a.s., \quad l=1,2.$$

If, in addition, $\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}[\exp(y_i)|\mathbf{x}_i=\mathbf{x}]\leq 2$, *then*

$$\sup_{\mathbf{x}\in\mathcal{X},r\in\mathcal{R}_1} \left|\mathbf{e}_1^\top(\widehat{\mathbf{H}}_{\mathbf{x}}^{-1} - \mathbf{H}_{\mathbf{x}}^{-1})\mathbf{S}_{\mathbf{x},r}\right| = O((nb^d)^{-1}\log n + (nb^d)^{-3/2}(\log n)^{5/2}) \qquad a.s.$$

Notice that aside for the condition $\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}[\exp(y_i)|\mathbf{x}_i=\mathbf{x}]\leq 2$ for $\mathcal{R}_1$, the other assumptions in Theorem 3 are satisfied in this example.

The following two examples provides the omitted details concerning uniform Gaussian strong approximation rates obtained via other methods, which are discussed in Section 4.1 of the paper.

Example SA.1 (Strong Approximation via Rio (1994))**.** *Consider the setup of Section 4.1, and assume the following regularity conditions hold:*

(a) $(\mathbf{x}_i,y_i)=(\mathbf{x}_i,\varphi(\mathbf{x}_i,u_i))$, *where* $\mathbf{z}_i=(\mathbf{x}_i,u_i)$ *satisfies Assumption A and* $\mathtt{M}_{\{\varphi\}}<\infty$, $\sup_{g\in\mathcal{G}}\mathtt{TV}_{\{\varphi\},\mathrm{supp}(g)\times[0,1]}\lesssim \sup_{g\in\mathcal{G}}\mathfrak{m}(\mathrm{Supp}(g)\times[0,1])<\infty$ *and* $\mathtt{K}_{\{\varphi\}}<\infty$.

(b) $\sup_{g\in\mathcal{G}}\mathtt{TV}_{\mathcal{V}_{\mathcal{R}_l},\mathrm{supp}(g)}\lesssim\sup_{g\in\mathcal{G}}\mathfrak{m}(\mathrm{Supp}(g))<\infty$ *and* $\mathtt{K}_{\{\theta(\cdot,r):r\in\mathcal{R}_l\}}<\infty$, *for* $l=1,2$.

(c) $K:\mathbb{R}^d\to\mathbb{R}$ *is non-negative, Lipschitz, and compactly supported.*

For $\mathcal{R}_1$, *take* $\mathcal{H}_1=\{h\circ\phi_Z^{-1}:h\in\widetilde{\mathcal{H}}_1\}$, *where* $\widetilde{\mathcal{H}}_1:=\{(\mathbf{x},u)\in\mathcal{X}\times[0,1]\mapsto g(\mathbf{x})\varphi(\mathbf{x},u)-g(\mathbf{x})\theta(\mathbf{x},\mathrm{Id}):g\in\mathcal{G}\}$, ϕ_Z *is the Rosenblatt transformation (Lemma SA.12) based on the Lebesgue density of* $\mathbf{z}_i=(\mathbf{x}_i,u_i)$, *and* $\mathcal{G}=\{b^{-d/2}\mathcal{K}_{\mathbf{x}}(\frac{\cdot-\mathbf{x}}{b}):\mathbf{x}\in\mathcal{X}\}$ *with* $\mathcal{K}_{\mathbf{x}}(\mathbf{u})=\mathbf{e}_1^\top\mathbf{H}_{\mathbf{x}}^{-1}\mathbf{p}(\mathbf{u})K(\mathbf{u})$. *Then, using the notation introduced in the paper,*

$$\begin{aligned}
\mathtt{M}_{\mathcal{H}_1} &= \mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}\lesssim b^{-d/2},\\
\mathtt{TV}_{\mathcal{H}_1} &= \frac{\overline{f}_Z^2}{\underline{f}_Z}(\mathtt{TV}_{\mathcal{G}}+\mathtt{M}_{\mathcal{G}}\sup_{g\in\mathcal{G}}\mathfrak{m}(\mathrm{Supp}(g)))\lesssim\frac{\overline{f}_Z^2}{\underline{f}_Z}b^{d/2-1},\\
\mathtt{K}_{\mathcal{H}_1} &\leq (2\sqrt{d})^{d-1}\frac{\overline{f}_Z^{d+1}}{\underline{f}_Z^d}(\mathtt{K}_{\mathcal{G}}+\mathtt{M}_{\mathcal{G}}\mathtt{K}_{\{\varphi\}}+\mathtt{M}_{\mathcal{G}}\mathtt{K}_{\mathcal{V}_1})\lesssim(2\sqrt{d})^{d-1}\frac{\overline{f}_Z^{d+1}}{\underline{f}_Z^d}b^{-d/2},\\
\mathtt{N}_{\mathcal{H}_1}(\varepsilon) &\lesssim \varepsilon^{-d-1}.
\end{aligned} \tag{SA-2}$$

Rio (1994) implies that $(X_n(h):h\in\mathcal{H}_1)=(\sqrt{nb^d}\mathbf{e}_1^\top\mathbf{H}_{\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x},r}:\mathbf{x}\in[0,1]^d,r\in\mathcal{R}_1)$ *admits a uniform Gaussian strong approximation with rate*

$$\mathsf{S}_n(t)=C_{d,\varphi,1}\sqrt{\frac{d\overline{f}_Z^2}{\underline{f}_Z}}(nb^{d+1})^{-1/(2d+2)}\sqrt{t+(d+1)\log n}+C_{d,\varphi,1}\frac{(2\sqrt{d})^{d-1}\overline{f}_Z^{d+1}}{\underline{f}_Z^d}(nb^d)^{-1/2}(t+(d+1)\log n),$$

where $C_{d,\varphi,1}$ *is a quantity that only depends on* d *and* φ.

For $\mathcal{R}_2$, *take* $\mathcal{H}_2=\{h\circ\phi_Z^{-1}:h\in\widetilde{\mathcal{H}}_2\}$, *where* $\widetilde{\mathcal{H}}_2:=\{(\mathbf{x},u)\in\mathcal{X}\times[0,1]\mapsto g(\mathbf{x})r\circ\varphi(\mathbf{x},u)-g(\mathbf{x})\theta(\mathbf{x},r):g\in\mathcal{G},r\in\mathcal{R}_2\}$. *Suppose* φ *is continuously differentiable with* $\min_{(\mathbf{x},u)\in[0,1]^{d+1}}|\partial_u\varphi(\mathbf{x},u)|>0$. *Then, using* *the notation introduced in the paper,*

$$\begin{aligned}
\mathtt{M}_{\mathcal{H}_2} &= \mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}} \lesssim b^{-d/2},\\
\mathtt{TV}_{\mathcal{H}_2} &\leq \frac{\overline{f}_Z^2}{\underline{f}_Z}(\mathtt{TV}_{\mathcal{G},[0,1]^d} + \mathtt{E}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}} \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{supp}(g)))\frac{\max_{(\mathbf{x},u)\in[0,1]^{d+1}}|\partial_u\varphi(\mathbf{x},u)|}{\min_{(\mathbf{x},u)\in[0,1]^{d+1}}|\partial_u\varphi(\mathbf{x},u)|} \lesssim \frac{\overline{f}_Z^2}{\underline{f}_Z} b^{d/2-1},\\
\mathtt{N}_{\mathcal{H}_2}(\varepsilon) &\lesssim \varepsilon^{-d-1}.
\end{aligned}$$

Rio (1994) implies that $(X_n(h) : h \in \mathcal{H}_2) = (\sqrt{nb^d}\mathbf{e}_1^\top \mathbf{H}_{\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x},r} : \mathbf{x} \in [0,1]^d, r \in \mathcal{R}_2)$ *admits a Gaussian strong approximation with rate function*

$$\mathsf{S}_n(t) = C_{d,\varphi,2}\sqrt{\frac{d\overline{f}_Z^2}{\underline{f}_Z}}(nb^{d+1})^{-1/(2d+2)}\sqrt{t + (d+1)\log n} + C_{d,\varphi,2}\sqrt{\frac{\log n}{nb^d}}(t + (d+1)\log n),$$

where $C_{d,\varphi,2}$ *is a quantity that only depends on* d *and* φ*.*

The strong approximation rates stated in Section 4.1 now follow directly from the strong approximation results above. ▲

Example SA.2 (Strong Approximation via Theorem 1)**.** *Consider the setup of Section 4.1, and assume the following regularity conditions hold:*

(a) $(\mathbf{x}_i, y_i) = (\mathbf{x}_i, \varphi(\mathbf{x}_i, u_i))$*, where* $\mathbf{z}_i = (\mathbf{x}_i, u_i)$ *satisfies Assumption A and* $\mathtt{M}_{\{\varphi\}} < \infty$, $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\varphi\},\mathrm{supp}(g)} \lesssim \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)) < \infty$, $\mathtt{K}_{\{\varphi\}} < \infty$*, and* $\mathtt{L}_{\{\varphi\}} < \infty$*.*

(b) $\sup_{r\in\mathcal{R}_\ell} \sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}} |\theta(\mathbf{x}, r) - \theta(\mathbf{y}, r)|/\|\mathbf{x} - \mathbf{y}\|_\infty < \infty$ *for* $\ell = 1, 2$*.*

(c) $K : \mathbb{R}^d \to \mathbb{R}$ *is non-negative, Lipschitz, and compactly supported.*

Then, Equations (SA-2) hold, and

$$\mathtt{L}_{\mathcal{H}_1} \lesssim \mathtt{L}_{\widetilde{\mathcal{H}}_1}\frac{\overline{f}_Z}{\underline{f}_Z} \lesssim (\mathtt{L}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}} + \mathtt{M}_{\mathcal{G}}\mathtt{L}_{\{\varphi\}} + \mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{V}_1})\frac{\overline{f}_Z}{\underline{f}_Z} \lesssim b^{-d/2-1}\frac{\overline{f}_Z}{\underline{f}_Z}.$$

Theorem 1 implies $(X_n(h) : h \in \mathcal{H}_1) = (\sqrt{nb^d}\mathbf{e}_1^\top \mathbf{H}_{\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x},r} : \mathbf{x} \in [0,1]^d, r \in \mathcal{R}_1)$ *admits a uniform Gaussian strong approximation with rate*

$$\mathsf{S}_n(t) = C_{d,\varphi,3}\sqrt{\frac{d\overline{f}_Z^3}{\underline{f}_Z^2}}(nb^{d+1})^{-1/(d+1)}\sqrt{t + (d+1)\log n} + C_{d,\varphi,3}\frac{(2\sqrt{d})^{d-1}\overline{f}_Z^{d+1}}{\underline{f}_Z^d}(nb^d)^{-1/2}(t + (d+1)\log n).$$

where $C_{d,\varphi,3}$ *is a quantity that only depends on* d *and* φ*.*

The strong approximation rate stated in Section 4.1 in the paper now follow directly from the strong approximation result above. ▲

Example SA.3 (Strong Approximation via Theorem 3)**.** *Consider the setup of Section 4.1 and assume the following regularity conditions hold:*

(a) *Assumption B holds.*

(b) $\sup_{r \in \mathcal{R}_\ell} \sup_{\mathbf{x},\mathbf{y} \in \mathcal{X}} |\theta(\mathbf{x}, r) - \theta(\mathbf{y}, r)|/\|\mathbf{x} - \mathbf{y}\|_\infty < \infty$ *for* $\ell = 1, 2$.

(c) $K : \mathbb{R}^d \to \mathbb{R}$ *is non-negative, Lipschitz, and compact supported.*

Recall that $\mathcal{G} = \{b^{-d/2}\mathcal{K}_{\mathbf{x}}(\frac{\cdot - \mathbf{x}}{b}) : \mathbf{x} \in \mathcal{X}\}$. *Then, using the notation introduced in the paper,*

$$\mathtt{M}_{\mathcal{G}} \lesssim b^{-d/2}, \quad \mathtt{E}_{\mathcal{G}} \lesssim b^{d/2}, \quad \mathtt{TV}_{\mathcal{G}} \lesssim b^{d/2-1}, \quad \mathtt{L}_{\mathcal{G}} \lesssim b^{-d/2-1}, \quad \mathtt{N}_{\mathcal{G}}(\varepsilon) \lesssim \varepsilon^{-d-1}.$$

Theorem 3 implies that $(R_n(g, r) : g \in \mathcal{G}, r \in \mathcal{R}_1) = (\sqrt{nb^d}\mathbf{e}_1^\top \mathbf{H}_{\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x},r} : \mathbf{x} \in [0, 1]^d, r \in \mathcal{R}_1)$ *admits a uniform Gaussian strong approximation with rate function*

$$\mathsf{S}_n(t) = \left(\frac{\overline{f}_X^3}{\underline{f}_X^2}\right)^{\frac{d}{2(d+2)}} \sqrt{d}(nb^d)^{-1/(d+2)}(t + (d + 1)\log n)^{3/2} + (nb^d)^{-1/2}(t + (d + 1)\log n).$$

If, in addition, $\sup_{\mathbf{x} \in [0,1]^d} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$, *then Theorem 3 implies* $(R_n(g, r) : g \in \mathcal{G}, r \in \mathcal{R}_1) = (\sqrt{nb^d}\mathbf{e}_1^\top \mathbf{H}_{\mathbf{x}}^{-1}\mathbf{S}_{\mathbf{x},r} : \mathbf{x} \in [0, 1]^d, r \in \mathcal{R}_1)$ *admits a uniform Gaussian strong approximation with rate function*

$$\mathsf{S}_n(t) = \left(\frac{\overline{f}_X^3}{\underline{f}_X^2}\right)^{\frac{d}{2(d+2)}} \sqrt{d}(nb^d)^{-1/(d+2)}(t + (d + 1)\log n)^{5/2} + (nb^d)^{-1/2}(t + (d + 1)\log n).$$

The strong approximation rate stated in Section 4.1 in the paper now follow directly from the strong approximation result above. ▲

SA-I.5 Haar Basis Regression Estimators

The following lemma gives precise regularity conditions for Example 3 in the paper.

Lemma SA.2 (Haar Basis Regression Estimators)**.** *Consider the setup in Example 3, and assume the following regularity conditions hold:*

(a) *Assumption B holds with* $[0, 1]^d$.

(b) $\sup_{r \in \mathcal{R}_\ell} \sup_{\mathbf{x},\mathbf{y} \in \mathcal{X}} |\theta(\mathbf{x}, r) - \theta(\mathbf{y}, r)|/\|\mathbf{x} - \mathbf{y}\|_\infty < \infty$ *for* $\ell = 1, 2$.

(c) $K : \mathbb{R}^d \to \mathbb{R}$ *is non-negative, Lipschitz, and compact supported.*

If $\log(nL)L/n \to 0$, *then*

$$\sup_{r \in \mathcal{R}_2} \sup_{\mathbf{x} \in \mathcal{X}} \left|\mathbf{p}(\mathbf{x})^\top(\widehat{\mathbf{Q}}^{-1} - \mathbf{Q}^{-1})\mathbf{T}_r\right| = O(\log(nL)L/n) \qquad a.s., \quad and$$

$$\sup_{r \in \mathcal{R}_\ell} \sup_{\mathbf{x} \in \mathcal{X}} \left|\mathbb{E}[\check{\theta}(\mathbf{x}, r)|\mathbf{x}_1, \cdots, \mathbf{x}_n] - \theta(\mathbf{x}, r)\right| = O\big(\max_{0 \leq l < L} \|\Delta_l\|_\infty\big) \qquad a.s., \quad l = 1, 2.$$

If, in addition, $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$, *then*

$$\sup_{r \in \mathcal{R}_2} \sup_{\mathbf{x} \in \mathcal{X}} \left|\mathbf{p}(\mathbf{x})^\top(\widehat{\mathbf{Q}}^{-1} - \mathbf{Q}^{-1})\mathbf{T}_r\right| = O(\log(nL)L/n + (\log n)(\log(nL)L/n)^{3/2}) \qquad a.s.$$

SA-II General Empirical Process: Proofs

We first introduce quasi-dyadic expansions of $\mathbb{R}^d$, and the associated $L_2(\mathbb{R}^d)$ projection of functions onto the class of piecewise constant functions on those cells. This enables us to couple a general empirical process indexed by piecewise constant functions with a Gaussian process. We then present a sequence of technical lemmas that bound the different approximation error terms discussed in Section 3 with different levels of generality. The proofs of these preliminary lemmas can be found in the supplemental appendix.

SA-II.1 Cell Expansions

Definition SA.1 (Quasi-Dyadic Expansion of $\mathbb{R}^d$). *A collection of Borel measurable sets in* $\mathbb{R}^d$, $\mathscr{C}_K(\mathbb{P},\rho) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{K-j}, 0 \leq j \leq K\}$, *is called a* quasi-dyadic expansion of $\mathbb{R}^d$ of depth K with respect to probability measure $\mathbb{P}$ *if the following three conditions hold:*

1. $\mathcal{C}_{j,k} = \mathcal{C}_{j-1,2k} \sqcup \mathcal{C}_{j-1,2k+1}$, *for all* $0 \leq k < 2^{K-j}, 1 \leq j \leq K$,
2. $\mathbb{P}(\mathcal{C}_{K,0}) = 1$, *and*
3. $\max_{0\leq k<2^K} \mathbb{P}(\mathcal{C}_{0,k})/\min_{0\leq k<2^K} \mathbb{P}(\mathcal{C}_{0,k}) \leq \rho$.

When $\rho = 1$, $\mathscr{C}_K(\mathbb{P},1)$ *is called a* dyadic expansion of $\mathbb{R}^d$ of depth K with respect to $\mathbb{P}$.

This definition implies $\frac{1}{2}\frac{2}{1+\rho} \leq \mathbb{P}(\mathcal{C}_{j-1,2k})/\mathbb{P}(\mathcal{C}_{j,k}) \leq \frac{1}{2}\frac{2\rho}{1+\rho}$ for all $0 \leq k < 2^{K-j}, 1 \leq j \leq K$, since each $\mathcal{C}_{j-1,l}$ is a disjoint union of 2^{j-1} cells of the form $\mathcal{C}_{0,k}$, which implies the third condition in Definition SA.1. Furthermore, in the special case that $\rho = 1$, $\mathbb{P}(\mathcal{C}_{j-1,2k}) = \mathbb{P}(\mathcal{C}_{j-1,2k+1}) = \frac{1}{2}\mathbb{P}(\mathcal{C}_{j,k})$, that is, the child level cells are obtained by splitting the parent level cells *dyadically in probability*.

The next definition specializes the dyadic expansion scheme to axis-aligned splits.

Definition SA.2 (Axis-Aligned Quasi-Dyadic Expansion of $\mathbb{R}^d$). *A collection of Borel measurable sets in* $\mathbb{R}^d$, $\mathcal{A}_K(\mathbb{P},\rho) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{K-j}, 0 \leq j \leq K\}$, *is an* axis-aligned quasi-dyadic expansion of $\mathbb{R}^d$ of depth K with respect to probability measure $\mathbb{P}$ *if it can be constructed via the following procedure:*

1. Initialization ($q = 0$): *Take* $\mathcal{C}_{K-q,0} = \mathcal{X}$ *where* $\mathcal{X} \subseteq \mathbb{R}^d$ *is the support of* $\mathbb{P}$.
2. Iteration ($q = 1, \ldots, K$): *Given* $\mathcal{C}_{K-l,k}$ *for* $0 \leq l \leq q-1, 0 \leq k < 2^l$, *take* $s = (q \mod d) + 1$, *and construct* $\mathcal{C}_{K-q,2k} = \mathcal{C}_{K-q+1,k} \cap \{\mathbf{x} \in \mathbb{R} : \mathbf{e}_s^\top \mathbf{x} \leq c_{K-q+1,k}\}$ *and* $\mathcal{C}_{K-q,2k+1} = \mathcal{C}_{K-q+1,k} \cap \{\mathbf{x} \in \mathbb{R} : \mathbf{e}_s^\top \mathbf{x} > c_{K-j+1,k}\}$ *such that* $\mathbb{P}(\mathcal{C}_{K-q,2k})/\mathbb{P}(\mathcal{C}_{K-q+1,k}) \in [\frac{1}{1+\rho}, \frac{\rho}{1+\rho}]$ *for all* $0 \leq k < 2^{q-1}$. *Continue until* $(\mathcal{C}_{0,k} : 0 \leq k < 2^K)$ *has been constructed.*

When $\rho = 1$ *and* $\mathbb{P}$ *is continuous,* $\mathcal{A}_K(\mathbb{P},\rho)$ *is unique.*

SA-II.2 Projection onto Piecewise Constant Functions

For a quasi-dyadic expansion $\mathscr{C}_K(\mathbb{P},\rho)$, the mean square projection from $L_2(\mathbb{R}^d)$ to the associated span of the terminal cells $\mathscr{E}_K := \operatorname{Span}\{\mathbb{1}_{\mathcal{C}_{0,k}} : 0 \leq k < 2^K\}$ is

$$\Pi_0(\mathscr{C}_K(\mathbb{P},\rho))[h] := \sum_{0\leq k<2^K} \frac{\mathbb{1}_{\mathcal{C}_{0,k}}}{\mathbb{P}(\mathcal{C}_{0,k})} \int_{\mathcal{C}_{0,k}} h(\mathbf{u}) d\mathbb{P}(\mathbf{u}), \qquad h \in L_2(\mathbb{R}^d). \tag{SA-3}$$

$\Pi_0(\mathscr{C}_K(\mathbb{P},\rho))[h]$ is a linear combination of a Haar-type basis, which gives the following orthogonal decomposition.

Lemma SA.3. *For any* $h \in L_2(\mathbb{R}^d)$,

$$\Pi_0(\mathcal{C}_K(\mathbb{P},\rho))[h] = \beta_{K,0}(h)e_{K,0} + \sum_{1 \le j \le K} \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(h)\widetilde{e}_{j,k},$$

where

$$\beta_{j,k}(h) := \frac{1}{\mathbb{P}(\mathcal{C}_{j,k})} \int_{\mathcal{C}_{j,k}} h(\mathbf{u}) d\mathbb{P}(\mathbf{u}), \qquad \widetilde{\beta}_{j,k}(h) := \beta_{j-1,2k}(h) - \beta_{j-1,2k+1}(h),$$

$$e_{j,k} := \mathbb{1}_{\mathcal{C}_{j,k}}, \qquad \widetilde{e}_{j,k} := \frac{\mathbb{P}(C_{j-1,2k+1})}{\mathbb{P}(C_{j,k})} e_{j-1,2k} - \frac{\mathbb{P}(\mathcal{C}_{j-1,2k})}{\mathbb{P}(\mathcal{C}_{j,k})} e_{j-1,2k+1},$$

for $0 \le k < 2^{K-j}, 1 \le j \le K$.

To save notation, we will use Π_0 as a short hand for $\Pi_0(\mathcal{C}_K(\mathbb{P},\rho))$ in what follows. In the special case of axis aligned quasi-dyadic expansion, we use $\Pi_{\mathcal{A}_{Kd}}$ as a short hand for $\Pi_0(\mathcal{A}_K(\mathbb{P},\rho))$.

SA-II.3 Strong Approximation Constructions

Suppose $(\widetilde{\xi}_{j,k} : 0 \le k < 2^{K-j}, 1 \le j \le K)$ are i.i.d. standard Gaussian random variables. Take $F_{(j,k),m}$ to be the cumulative distribution function of $(S_{j,k} - mp_{j,k})/\sqrt{mp_{j,k}(1-p_{j,k})}$, where $p_{j,k} = \mathbb{P}(\mathcal{C}_{j-1,2k})/\mathbb{P}(\mathcal{C}_{j,k})$ and $S_{j,k}$ is a $\mathrm{Bin}(m, p_{j,k})$ random variable, and $G_{(j,k),m}(t) = \sup\{x : F_{(j,k),m}(x) \le t\}$. We define $U_{j,k}, \widetilde{U}_{j,k}$'s via the following iterative scheme:

1. *Initialization:* Take $U_{K,0} = n$.
2. *Iteration:* Suppose we have define $U_{l,k}$ for $j < l \le K, 0 \le k < 2^{K-l}$, then solve for $U_{j,k}$'s such that

$$\begin{aligned} \widetilde{U}_{j,k} &= \sqrt{U_{j,k}p_{j,k}(1-p_{j,k})}G_{(j,k),U_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k}), \\ \widetilde{U}_{j,k} &= (1-p_{j,k})U_{j-1,2k} - p_{j,k}U_{j-1,2k+1} = U_{j-1,2k} - p_{j,k}U_{j,k}, \\ U_{j-1,2k} &+ U_{j-1,2k+1} = U_{j,k}, \quad 0 \le k < 2^{K-j}. \end{aligned}$$

   Continue till we have defined $U_{0,k}$ for $0 \le k < 2^K$.

Then $\{U_{j,k} : 0 \le j \le K, 0 \le k < 2^{K-j}\}$ have the same joint distribution as $\{\sum_{i=1}^n e_{j,k}(\mathbf{x}_i) : 0 \le j \le K, 0 \le k < 2^{K-j}\}$. By Vorob'ev–Berkes–Philipp theorem (Dudley, 2014, Theorem 1.31), $\{\widetilde{\xi}_{j,k} : 0 \le k < 2^{K-j}, 1 \le j \le K\}$ can be constructed on a possibly enlarged probability space such that the previously constructed $U_{j,k}$ satisfies $U_{j,k} = \sum_{i=1}^n e_{j,k}(\mathbf{x}_i)$ almost surely for all $0 \le j \le K, 0 \le k < 2^{K-j}$. We will show $\widetilde{\xi}_{j,k}$'s can be given as a Brownian bridge indexed by $\widetilde{e}_{j,k}$'s.

Lemma SA.4. *Suppose* $\mathcal{H}$ *is a class of real-valued pointwise measurable functions on* $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$ *such that* $\mathtt{M}_{\mathcal{H}} < \infty$ *and* $J(1, \mathcal{H}, \mathtt{M}_{\mathcal{H}}) < \infty$*, and* $\mathcal{C}_K$ *is a quasi-dyadic expansion of* $\mathbb{R}^d$ *of depth* K *with respect to* $\mathbb{P}_X$*. Then,* $\mathcal{H} \cup \Pi_0\mathcal{H} \cup \mathcal{E}_K$ *is* $\mathbb{P}_X$*-pregaussian.*

Then by Skorohod Embedding lemma (Dudley, 2014, Lemma 3.35), on a possibly enlarged probability space, we can construct a Brownian bridge $(Z_n^X(h) : h \in \mathcal{H})$ that satisfies

$$\widetilde{\xi}_{j,k} = \frac{\mathbb{P}(\mathcal{C}_{j,k})}{\sqrt{\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})}} Z_n^X(\widetilde{e}_{j,k}),$$

for $0 \leq k < 2^{K-j}, 1 \leq j \leq K$. Moreover, call

$$V_{j,k} := \sqrt{n} Z_n^X(e_{j,k}), \qquad \widetilde{V}_{j,k} := \sqrt{n} Z_n^X(\widetilde{e}_{j,k}), \qquad \widetilde{\xi}_{j,k} := \frac{\mathbb{P}(\mathcal{C}_{j,k})}{\sqrt{n\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})}} \widetilde{V}_{j,k}.$$

for $0 \leq k < 2^{K-j}, 1 \leq j \leq K$. Notice that for all $h \in \mathcal{E}_K$, we have

$$\sqrt{n} X_n(h) = \sum_{j=1}^{K} \sum_{0 \leq k < 2^{K-j}} \widetilde{\beta}_{j,k}(h) \widetilde{U}_{j,k}, \quad \sqrt{n} Z_n^X(h) = \sum_{j=1}^{K} \sum_{0 \leq k < 2^{K-j}} \widetilde{\beta}_{j,k}(h) \widetilde{V}_{j,k}.$$

The difference between $X_n(h)$ and $Z_n^X(h)$ will rely on the coefficient $\widetilde{\beta}_{j,k}(h)$ and the coupling between $\widetilde{U}_{j,k}$ and $\widetilde{V}_{j,k}$, which is the essence of Theorem 2.1 in Rio (1994). Although Theorem 2.1 in Rio (1994) is stated for i.i.d uniformly distributed on $[0, 1]$ random variables, the underlying process only depends through the counts of the random variables taking values in each interval of the form $[k2^{-j}, (k+1)2^{-j})$, which have the same distribution as the counts of $\mathbf{x}_i$'s in $\mathcal{C}_{j,k}$'s. Hence, we have a direct corollary of Theorem 2.1 in Rio (1994) as follows:

Lemma SA.5. *Given a dyadic expansion $\mathcal{C}_K(\mathbb{P}_X, 1)$, for any $g \in \mathcal{E}_K$ and any $t > 0$,*

$$\mathbb{P}\left(\sqrt{n}\left|X_n(g) - Z_n^X(g)\right| \geq 24\sqrt{\|g\|_{\mathcal{E}_K}^2 x} + 4\sqrt{\mathtt{C}_{\{g\}} x}\right) \leq 2\exp(-x),$$

where $\|g\|_{\mathcal{E}_K}^2 = \sum_{j=1}^{K} \sum_{0 \leq k < 2^{K-j}} \widetilde{\beta}_{j,k}^2(g)$, and

$$\mathtt{C}_{\mathcal{F}} = \sup_{f \in \mathcal{F}} \min \left\{ \sup_{(j,k)} \left[\sum_{l<j} (j-l)(j-l+1) 2^{l-j} \sum_{m:\mathcal{C}_{l,m} \subseteq \mathcal{C}_{j,k}} \widetilde{\beta}_{l,m}^2(f) \right], \|f\|_\infty^2 K \right\}.$$

The above lemma relies on coupling of $\mathrm{Bin}(m, 1/2)$ random variables with Gaussian random variables. The coupling also holds for $\mathrm{Bin}(m, p)$ with the error term only depending on how far away p is bounded away from 0 and 1:

Lemma SA.6. *Suppose $X \sim \mathrm{Bin}(n, p)$ where $0 < \underline{p} < p < \overline{p} < 1$. Then there exists a standard Gaussian random variable $Z \sim N(0, 1)$ and constants $c_0, c_1, c_2, c_3 > 0$ only depending on $\underline{p}$ and $\overline{p}$ such that whenever the event $A = \{|X - np| \leq c_1 n\}$ occurs and $c_0\sqrt{n} \geq 1$, we have*

$$\left|X - np - \sqrt{np(1-p)}Z\right| \leq c_2 Z^2 + c_3,$$
$$|X - np| \leq \frac{1}{c_0} + 2\sqrt{np(1-p)}|Z|.$$

In particular, we can take $c_0 > 0$ to be the solution of

$$60 c_0 \overline{p} \left(\sqrt{\frac{1-\underline{p}}{\underline{p}}}\right)^3 \exp\left(2\sqrt{\frac{1-\underline{p}}{\underline{p}}} c_0\right) + 60 c_0 (1-\underline{p}) \left(\sqrt{\frac{\overline{p}}{1-\overline{p}}}\right)^3 = 1,$$

and take $c_1 = 15c_0\sqrt{\underline{p}(1-\overline{p})}$, $c_2 = 1/(15c_0), c_3 = 1/c_0$ *and* Z *can be taken via quantile transformation, that is, define* $F(x) = \mathbb{P}(X - np < \sqrt{np(1-p)}x)$ *and let* Φ *be the cumulative distribution function of a* $N(0,1)$ *random variable, then* Z *can be defined via* $Z := \Phi^{-1} \circ F\left((X-np)/\sqrt{np(1-p)}\right)$.

This enables the following strong approximation for the quasi-dyadic case:

Lemma SA.7. *Given a quasi-dyadic expansion* $\mathcal{C}_K(\mathbb{P}_X, \rho)$, $\rho > 1$, *for any* $g \in \mathcal{E}_K$ *and any* $t > 0$,

$$\begin{aligned}&\mathbb{P}\left(\sqrt{n}\left|X_n(g) - Z_n^X(g)\right| \geq c_\rho\sqrt{\|g\|_{\mathcal{E}_K}^2 x} + c_\rho\sqrt{\mathtt{C}_{\{g\}}x}\right)\\ &\leq 2\exp(-x) + 2^{K+2}\exp\left(-c_\rho n 2^{-K}\right),\end{aligned}$$

where $\|g\|_{\mathcal{E}_K}^2 = \sum_{j=1}^K \sum_{0 \leq k < 2^{K-j}} \widetilde{\beta}_{j,k}^2(g)$, c_ρ *is a constant that only depends on* ρ *and* $\mathtt{C}_{\{g\}}$ *is defined in Lemma SA.5.*

SA-II.4 Meshing Error

For $0 < \delta \leq 1$, consider the $(\delta\mathtt{M}_{\mathcal{H}})$-net of $(\mathcal{H}, e_{\mathbb{P}})$, with cardinality no larger than $\mathtt{N}_{\mathcal{H}}(\delta)$: define $\pi_{\mathcal{H}_\delta} : \mathcal{H} \mapsto \mathcal{H}$ such that $\|\pi_{\mathcal{H}_\delta}(h) - h\|_{\mathbb{P}_X,2} \leq \delta\mathtt{M}_{\mathcal{H}}$ for all $h \in \mathcal{H}$.

Lemma SA.8. *For all* $t > 0$ *and* $0 < \delta < 1$,

$$\begin{aligned}\mathbb{P}\left[\|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}} > C\mathsf{F}_n(t,\delta)\right] &\leq \exp(-t),\\ \mathbb{P}\left[\|Z_n^X \circ \pi_{\mathcal{H}_\delta} - Z_n^X\|_{\mathcal{H}} > C(\mathtt{M}_{\mathcal{H}}J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}}) + \delta\mathtt{M}_{\mathcal{H}}\sqrt{t})\right] &\leq \exp(-t),\end{aligned}$$

where C *is a universal constant.*

SA-II.5 L2 Projection Error

For X_n, Z_n^X, and Π_0 as defined above, and for $\mathcal{H}_\delta$ a δ-net of $(\mathcal{H}, e_{\mathbb{P}_X})$ with cardinality no greater than $\mathtt{N}_{\mathcal{H}}(\delta)$, the following lemma controls the mean square projection onto piecewise constant functions.

Lemma SA.9. *Let* $\mathcal{C}_K(\mathbb{P}_X, \rho) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{K-j}, 0 \leq j \leq K\}, \rho \geq 1$ *be a quasi-dyadic expansion of* $\mathbb{R}^d$ *of depth* K. *Define*

$$\mathcal{V} = \cup_{0 \leq k < 2^K}\left(\mathcal{C}_{0,k} - \mathcal{C}_{0,k}\right).$$

Then for all $t > 0$,

$$\begin{aligned}\mathbb{P}\left[\|X_n - X_n \circ \Pi_0\|_{\mathcal{H}_\delta} > \sqrt{4\mathtt{V}_{\mathcal{H}_\delta}t} + \frac{4\mathtt{B}_{\mathcal{H}_\delta}}{3\sqrt{n}}t\right] &\leq 2\mathtt{N}_{\mathcal{H}}(\delta)e^{-t},\\ \mathbb{P}\left[\|Z_n^X - Z_n^X \circ \Pi_0\|_{\mathcal{H}_\delta} > \sqrt{4\mathtt{V}_{\mathcal{H}_\delta}t}\right] &\leq 2\mathtt{N}_{\mathcal{H}}(\delta)e^{-t},\end{aligned}$$

where

$$\mathtt{V}_{\mathcal{H}_\delta} =: \min\{2\mathtt{M}_{\mathcal{H}}, \mathtt{L}_{\mathcal{H}_\delta}\|\mathcal{V}\|_\infty\}\left(\sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\right)^2 2^K\mathfrak{m}(\mathcal{V})\|\mathcal{V}\|_\infty\mathtt{TV}_{\mathcal{H}_\delta}, \quad \mathtt{B}_{\mathcal{H}_\delta} =: \min\{2\mathtt{M}_{\mathcal{H}}, \mathtt{L}_{\mathcal{H}_\delta}\|\mathcal{V}\|_\infty\}.$$

In particular, if $\mathbf{x}_i \overset{i.i.d}{\sim} \text{Unif}([0,1]^d)$ and the cells $\mathcal{A}_K(\mathbb{P}_X, 1)$ are axis-aligned dyadic expansion of depth K, then

$$\mathbb{P}\Big[\|X_n - X_n \circ \Pi_{\mathcal{A}_{Kd}}\|_{\mathcal{H}_\delta} > \sqrt{4d \min\{2\mathtt{M}_{\mathcal{H}_\delta}, \mathtt{L}_{\mathcal{H}_\delta} 2^{-K}\} 2^{-K} \mathtt{TV}_{\mathcal{H}_\delta} t} + \frac{4\min\{2\mathtt{M}_{\mathcal{H}_\delta}, \mathtt{L}_{\mathcal{H}_\delta} 2^{-K}\}}{3\sqrt{n}} t\Big] \leq 2\mathtt{N}_{\mathcal{H}}(\delta) e^{-t},$$

$$\mathbb{P}\Big[\|Z_n^X - Z_n^X \circ \Pi_{\mathcal{A}_{Kd}}\|_{\mathcal{H}_\delta} > \sqrt{4d \min\{2\mathtt{M}_{\mathcal{H}_\delta}, \mathtt{L}_{\mathcal{H}_\delta} 2^{-K}\} 2^{-K} \mathtt{TV}_{\mathcal{H}_\delta} t}\Big] \leq 2\mathtt{N}_{\mathcal{H}}(\delta) e^{-t},$$

for all $t > 0$.

SA-II.6 Strong Approximation Errors

The next lemma controls the strong approximation error for projected processes.

Lemma SA.10. *Let $\mathcal{C}_K(\mathbb{P}_X, 1) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{K-j}, 0 \leq j \leq K\}$ be a dyadic expansion of $\mathbb{R}^d$ of depth K as in Definition SA.1. For each $1 \leq j \leq K$, define*

$$\mathcal{U}_j := \cup_{0 \leq k < 2^{K-j}} (\mathcal{C}_{j-1,2k+1} - \mathcal{C}_{j-1,2k}).$$

Suppose X_n, Z_n^X and Π_0 are as defined above and $\mathcal{H}_\delta$ is a δ-net of $(\mathcal{H}, e_{\mathbb{P}_X})$ with cardinality no greater than $\mathtt{N}_{\mathcal{H}}(\delta)$. Then for all $t > 0$,

$$\mathbb{P}\Big[\|X_n \circ \Pi_0 - Z_n^X \circ \Pi_0\|_{\mathcal{H}_\delta} > 48\sqrt{\frac{\mathcal{R}_K(\mathcal{H}_\delta)}{n}t} + 4\sqrt{\frac{\mathtt{C}_{\mathcal{H}_\delta}}{n}t}\Big] \leq 2\mathtt{N}_{\mathcal{H}}(\delta) e^{-t},$$

where $\mathcal{R}_K(\mathcal{H}_\delta)$ is defined to be

$$\sum_{j=1}^{K} \min\{\mathtt{M}_{\mathcal{H}_\delta}, \|U_j\|_\infty \mathtt{L}_{\mathcal{H}_\delta}\} 2^{K-j} \min\left\{\left(\sup_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})\right)^2 2^{2(K-j)} \|\mathcal{U}_j\|_\infty \mathfrak{m}(\mathcal{U}_j) \mathtt{TV}_{\mathcal{H}_\delta}, \|U_j\|_\infty \mathtt{L}_{\mathcal{H}_\delta}, \mathtt{E}_{\mathcal{H}_\delta}\right\}.$$

Lemma SA.11. *Let $\mathcal{C}_K(\mathbb{P}_X, \rho) = \{\mathcal{C}_{j,k} : 0 \leq j \leq K, 0 \leq k < 2^{K-j}\}$, $\rho > 1$ be an* approximate dyadic expansion of $\mathbb{R}^d$ of depth K *as in Definition SA.1. For each $1 \leq j \leq K$, define*

$$\mathcal{U}_j := \cup_{0 \leq k < 2^{K-j}} (\mathcal{C}_{j-1,2k+1} - \mathcal{C}_{j-1,2k}).$$

Suppose $\mathcal{H}$ is a class of real-valued pointwise measurable functions in $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$ such that $\mathtt{M}_{\mathcal{H}} < \infty$ and $J(1, \mathcal{H}, \mathtt{M}_{\mathcal{H}}) < \infty$. Suppose X_n, Z_n^X, Π_0, $\mathcal{H}_\delta$ and $\mathcal{R}_K$ are defined as in Lemma SA.10. Then for all $t > 0$,

$$\mathbb{P}\Big[\|X_n \circ \Pi_0 - Z_n^X \circ \Pi_0\|_{\mathcal{H}_\delta} > C_\rho\sqrt{\frac{\mathcal{R}_K(\mathcal{H}_\delta)}{n}t} + C_\rho\sqrt{\frac{\mathtt{C}_{\mathcal{H}_\delta}}{n}t}\Big] \leq 2\mathtt{N}_{\mathcal{H}}(\delta) e^{-t} + 2^K \exp\left(-C_\rho n 2^{-K}\right),$$

where C_ρ is a constant only depending on ρ.

SA-II.7 Rosenblatt Reduction

Lemma SA.12. *Suppose $X = (X_1, \dots, X_d)$ is a random variable taking values in $\mathbb{R}^d$ with Lebesgue density f_X supported on $[0,1]^d$. Define the Rosenblatt transformation ϕ_X based on density of $\mathbf{x}_i$ by*

$$\phi_X(x_1,\dots,x_d) = \begin{bmatrix} \mathbb{P}\left(X_1 \le x_1\right) \\ \mathbb{P}\left(X_2 \le x_2 | X_1 = x_1\right) \\ \vdots \\ \mathbb{P}\left(X_d \le x_d | X_1 = x_1, \dots, X_{d-1} = x_{d-1}\right) \end{bmatrix}, \qquad (x_1,\dots,x_d) \in [0,1]^d.$$

Define $\widetilde{\mathcal{H}} := \{h \circ \phi_X^{-1}\}$. Suppose $\mathbf{u}_i \overset{\text{i.i.d}}{\sim} \mathrm{Unif}([0,1]^d)$, $1 \le i \le n$. Then

$$\begin{aligned} &\mathtt{M}_{\widetilde{\mathcal{H}}} = \mathtt{M}_{\mathcal{H}}, \quad \mathtt{L}_{\widetilde{\mathcal{H}}} \le \mathtt{L}_{\mathcal{H}} \frac{\overline{f}_X}{\underline{f}_X}, \quad \mathtt{TV}_{\widetilde{\mathcal{H}}} \le \mathtt{TV}_{\mathcal{H}} \frac{\overline{f}_X^2}{\underline{f}_X}, \quad \mathtt{K}_{\widetilde{\mathcal{H}}} \le \mathtt{K}_{\mathcal{H}} (2\sqrt{d})^{d-1} \frac{\overline{f}_X^{d+1}}{\underline{f}_X^d} \\ &\mathtt{E}_{\widetilde{\mathcal{H}}} = \mathtt{E}_{\mathcal{H}}, \quad \mathtt{N}_{\widetilde{\mathcal{H}}}(\varepsilon) = \mathtt{N}_{\mathcal{H}}(\varepsilon), \forall 0 < \varepsilon < 1. \end{aligned}$$

SA-II.8 Proof of Lemma SA.3

First, we show that $\{e_{Kd,0}\} \cup \{\widetilde{e}_{j,k} : 1 \le j \le Kd, 0 \le k < 2^{Kd-j}\}$ is an orthogonal basis. For notational simplicity, denote $\mathcal{I} = \{(j,k) : 1 \le j \le Kd, 0 \le k < 2^{Kd-j}\}$. Let $(j,k) \in \mathcal{I}$. Then

$$\begin{aligned} \langle e_{Kd,0}, \widetilde{e}_{j,k} \rangle &= \int_{\mathbb{R}^d} \frac{\mathbb{P}(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}(\mathcal{C}_{j,k})} e_{j-1,2k}(\mathbf{u}) d\mathbf{u} - \int_{\mathbb{R}^d} \frac{\mathbb{P}(\mathcal{C}_{j-1,2k})}{\mathbb{P}(\mathcal{C}_{j,k})} e_{j-1,2k+1}(\mathbf{u}) d\mathbf{u} \\ &= \frac{\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j-1,2k})}{\mathbb{P}(\mathcal{C}_{j,k})} - \frac{\mathbb{P}(C_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}(\mathcal{C}_{j,k})} = 0. \end{aligned}$$

Now let $(j_1,k_1), (j_2,k_2) \in \mathcal{I}$.

Case 1: $j_1 = j_2$ and $k_1 \ne k_2$, then $\widetilde{e}_{j_1,k_1}$ and $\widetilde{e}_{j_2,k_2}$ have different support, hence $\langle \widetilde{e}_{j_1,k_1}, \widetilde{e}_{j_2,k_2} \rangle = 0$.

Case 2: $j_1 \ne j_2$ and w.l.o.g. we will assume $j_1 < j_2$. By (1) in Definition SA.1, either $\mathcal{C}_{j_1,k_1} \cap \mathcal{C}_{j_2,k_2} = \emptyset$ or $\mathcal{C}_{j_1,k_1} \subset \mathcal{C}_{j_2,k_2}$. In the first case, we also have $\langle \widetilde{e}_{j_1,k_1}, \widetilde{e}_{j_2,k_2} \rangle = 0$. In the second case, using (1) in Definition SA.1 again, either $\mathcal{C}_{j_1,k_1} \subseteq \mathcal{C}_{j_2-1,2k_2}$ or $\mathcal{C}_{j_1,k_1} \subseteq \mathcal{C}_{j_2-1,2k_2+1}$. W.l.o.g we assume $\mathcal{C}_{j_1,k_1} \subseteq \mathcal{C}_{j_2-1,2k_2}$. Then

$$\begin{aligned} \langle \widetilde{e}_{j_1,k_1}, \widetilde{e}_{j_2,k_2} \rangle &= \langle \widetilde{e}_{j_1,k_1}, \frac{\mathbb{P}(\mathcal{C}_{j_2-1,2k_2})}{\mathbb{P}(\mathcal{C}_{j_2,k_2})} e_{j_2-1,2k_2} \rangle \\ &= \frac{\mathbb{P}(\mathcal{C}_{j_2-1,2k_2})}{\mathbb{P}(\mathcal{C}_{j_2,k_2})} \int_{\mathbb{R}^d} \frac{\mathbb{P}(\mathcal{C}_{j_1-1,2k_1+1})}{\mathbb{P}(\mathcal{C}_{j_1,k_1})} e_{j_1-1,2k_1}(\mathbf{u}) d\mathbf{u} - \int_{\mathbb{R}^d} \frac{\mathbb{P}(\mathcal{C}_{j_1-1,2k_1})}{\mathbb{P}(\mathcal{C}_{j_1,k_1})} e_{j_1-1,2k_1+1}(\mathbf{u}) d\mathbf{u} \\ &= 0. \end{aligned}$$

This shows that $\{e_{Kd,0}\} \cup \{\widetilde{e}_{j,k} : 1 \le j \le Kd, 0 \le k < 2^{Kd-j}\}$ is an orthogonal basis for $\mathcal{E}_{Kd} \subseteq L_2(\mathbb{R}^d)$ and hence

$$\Pi_0 h = \frac{\langle h, e_{Kd,0} \rangle}{\langle e_{Kd,0}, e_{Kd,0} \rangle} e_{Kd,0} + \sum_{1 \le j \le Kd} \sum_{0 \le k < 2^{Kd-j}} \frac{\langle h, \widetilde{e}_{j,k} \rangle}{\langle \widetilde{e}_{j,k}, \widetilde{e}_{j,k} \rangle} \widetilde{e}_{j,k}, \quad \forall h \in L_2(\mathbb{R}^d).$$

The coefficients are given by

$$
\begin{aligned}
&\frac{\langle h, \widetilde{e}_{j,k}\rangle}{\langle \widetilde{e}_{j,k}, \widetilde{e}_{j,k}\rangle} := \frac{\int_{\mathbb{R}^d} h(\mathbf{u})\widetilde{e}_{j,k}(\mathbf{u})d\mathbf{u}}{\int_{\mathbb{R}^d} \widetilde{e}_{j,k}(\mathbf{u})\widetilde{e}_{j,k}(\mathbf{u})d\mathbf{u}}\\
=&\frac{\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j,k})^{-1}\beta_{j-1,2k}(h) - \mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j,k})^{-1}\beta_{j-1,2k+1}(h)}{\mathbb{P}(\mathcal{C}_{j-1,2k+1})^2\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j,k})^{-2} + \mathbb{P}(\mathcal{C}_{j-1,2k})^2\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j,k})^{-2}}\\
=&\frac{\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j,k})^{-1}\beta_{j-1,2k}(h) - \mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j,k})^{-1}\beta_{j-1,2k+1}(h)}{\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j,k})^{-1} + \mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})\mathbb{P}(\mathcal{C}_{j,k})^{-1}}\\
=&\beta_{j-1,2k}(h) - \beta_{j-1,2k+1}(h) = \widetilde{\beta}_{j,k}(h), \quad \forall 1 \le j \le Kd, 0 \le k < 2^{Kd}.
\end{aligned}
$$

Moreover,

$$\frac{\langle h, e_{Kd,0}\rangle}{\langle e_{Kd,0}, e_{Kd,0}\rangle} = \mathbb{P}(\mathcal{C}_{Kd,0})^{-1}\int_{\mathcal{C}_{Kd,0}} h(\mathbf{u})d\mathbb{P}(\mathbf{u}) = \beta_{Kd,0}(h).$$

The proves the claim. □

SA-II.9 Proof of Lemma SA.4

First, we will show that $\Pi_0\mathcal{H} \cup \mathcal{E}_{Kd}$ is a VC-type of class. Notice that all $h \in \Pi_0\mathcal{H} \cap \mathcal{E}_{Kd}$ can be written in the form $\sum_{0\le k<2^{Kd}} c_k e_{0,k}$ with $c_k \in [-\mathtt{M}_{\mathcal{H}}, \mathtt{M}_{\mathcal{H}}]$. Denote $D = 2^{Kd}$. For any $\varepsilon > 0$, $\|\sum_{0\le k<2^{Kd}} c_k e_{0,k} - \sum_{0\le k<2^{Kd}} d_k e_{0,k}\|_\infty \le \varepsilon\mathtt{M}_{\mathcal{H}}$ if $|c_k - d_k| \le \varepsilon\mathtt{M}_{\mathcal{H}}/D$ for all $0 \le k < D$. Hence

$$\sup_Q N(\Pi_0\mathcal{H} \cup \mathcal{E}_{Kd}, e_Q, \varepsilon\mathtt{M}_{\mathcal{H}}) \le \left(\frac{D}{\varepsilon}\right)^D, \qquad \forall 0 < \varepsilon \le 1,$$

where sup is taken over all discrete measures on $\mathcal{X}$. Moreover, we have assumed $J(1, \mathcal{H}, M_{\mathcal{H}}) < \infty$. By Kolmogorov's extension theorem, there exists a mean-zero Gaussian Z_n^X indexed by $\mathcal{H} \cup \Pi_0\mathcal{H} \cup \mathcal{E}_{Kd}$ with the same covariance structure as X_n. Since $\mathtt{M}_{\mathcal{H}} < \infty$, $\mathcal{H} \cup \Pi_0\mathcal{H} \cup \mathcal{E}_{Kd}$ is totally bounded for $e_{\mathbb{P}_X}$. By separability of $\mathcal{H}$ and Corollary 2.2.9 in van der Vaart and Wellner (2013), there exists a version of Z_n^X with uniformly $e_{\mathbb{P}_X}$-continuous sample path. Hence $\mathcal{H} \cup \Pi_0\mathcal{H} \cup \mathcal{E}_{Kd}$ is pre-Gaussian. □

SA-II.10 Proof of Lemma SA.5

Take $w_i \overset{i.i.d}{\sim} N(0,1)$, $1 \le i \le n$ and $I_{j,k} := [k2^{-j}, (k+1)2^{-j})$, $0 \le k < 2^{Kd-j}$, $0 \le j \le Kd$. Take B to be a Brownian bridge on $[0,1]$, that is, there exists a standard Wiener process W such that $B(t) = W(t) - tW(1)$ for all $t \in [0,1]$. Define

$$v_{j,k} := \sqrt{n}\int_0^1 \mathbb{1}(t \in I_{j,k})dB(t), \quad \widetilde{v}_{j,k} := v_{j-1,2k} - v_{j-1,2k+1}.$$

Take F_m to be the cumulative distribution function of $(S_m - \frac{1}{2}m)/\sqrt{m/4}$, where S_m is a $\mathrm{Bin}(m, 1/2)$ random variable, and $G_m(t) = \sup\{x : F_m(x) \le t\}$. Define $u_{j,k}$'s and $\widetilde{u}_{j,k}$'s, again via the iterative quantile transformation technique by:

1. *Initialization:* Take $u_{Kd,0} = n$.

2. *Iteration:* Suppose we have define $u_{l,k}$ for $0 \le k < 2^{Kd-l}, j < l \le Kd$, then solve for $u_{j,k}$'s such that

$$\widetilde{u}_{j,k} = \frac{1}{2}\sqrt{U_{j,k}} G_{U_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k}),$$
$$\widetilde{u}_{j,k} = \frac{1}{2}u_{j-1,2k} - \frac{1}{2}u_{j-1,2k+1} = u_{j-1,2k} - \frac{1}{2}u_{j,k},$$
$$u_{j-1,2k} + u_{j-1,2k+1} = u_{j,k}, \quad 0 \le k < 2^{Kd-j}.$$

   Continue till we have defined $u_{0,k}$ for $0 \le k < 2^{Kd}$.

Then $u_{j,k}$'s have the same joint distribution as $\sum_{i=1}^{n} \mathbb{1}(w_i \in I_{j,k})$'s. Hence by Skorohod Embedding lemma (Dudley, 2014, Lemma 3.35), on a rich enough probability space, we can take $(B(t) : 0 \le t \le 1)$ such that $u_{j,k} = \sum_{i=1}^{n} \mathbb{1}(w_i \in I_{j,k})$ almost surely, for all $0 \le k < 2^{Kd-j}$, $0 \le j \le Kd$.

Moreover, distribution of the process $\{(X_n(h), Z_n^X(h)) : h \in \mathcal{E}_{Kd}\}$ is the same as distribution of the process

$$\left(\frac{1}{\sqrt{n}} \sum_{j=1}^{Kd} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(h)\widetilde{u}_{j,k}, \frac{1}{\sqrt{n}} \sum_{j=1}^{Kd} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(h)\widetilde{v}_{j,k} \right), \quad h \in \mathcal{E}_{Kd},$$

since $\{(\widetilde{u}_{j,k}, \widetilde{v}_{j,k}) : 0 \le k < 2^{Kd-j}, 1 \le j \le Kd\}$ and $\{(\widetilde{U}_{j,k}, \widetilde{V}_{j,k}) : 0 \le k < 2^{Kd-j}, 1 \le j \le Kd\}$ have the same joint distribution and

$$(X_n(h), Z_n^X(h)) = \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{Kd} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(h)\widetilde{U}_{j,k}, \frac{1}{\sqrt{n}} \sum_{j=1}^{Kd} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(h)\widetilde{V}_{j,k} \right), \quad \forall h \in \mathcal{E}_{Kd}.$$

Following Section 3 in Rio (1994), we choose either $p_i = \frac{1}{2}\left(\frac{1}{Kd} + \frac{1}{i(i+1)}\right)$ or $p_i = \frac{1}{i(i+1)}$ and Theorem 2.1 in Rio (1994), we have for any $h \in \mathcal{E}_{Kd}$, for any $t > 0$, with probability at least $1 - 2\exp(-t)$,

$$\left| \sum_{j=1}^{Kd} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(h)\widetilde{u}_{j,k} - \sum_{j=1}^{Kd} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(h)\widetilde{v}_{j,k} \right| \le 24\sqrt{\sum_{j=1}^{n} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}^2(h)t} + \sqrt{\mathtt{C}_{\{h\}}}t.$$

Hence for any $h \in \mathcal{E}_{Kd}$, for any $t > 0$,

$$\mathbb{P}\left(\sqrt{n}\left|X_n(h) - Z_n^X(h)\right| \ge 24\sqrt{\sum_{j=1}^{n} \sum_{0 \le k < 2^{Kd-j}} \widetilde{\beta}_{j,k}^2(h)t} + \sqrt{\mathtt{C}_{\{h\}}}t \right) \le 2\exp(-t).$$

□

SA-II.11 Proof of Lemma SA.6

Take $X_j \overset{i.i.d}{\sim} \text{Bern}(p), 1 \leq j \leq n$ where $0 < \underline{p} < p < \overline{p} < 1$. Take $\xi_j = (X_j - p)/\sqrt{np(1-p)}$ and $S_n = \sum_{j=1}^n \xi_j$. Then for any $h \in \mathbb{R}$,

$$\begin{aligned} L(h) &:= \sum_{j=1}^n \mathbb{E}\left[|\xi_j|^3 \exp(|h\xi_j|)\right] \\ &= \sum_{j=1}^n \mathbb{E}\left[\left(\frac{X_j - p}{\sqrt{np(1-p)}}\right)^3 \exp\left(h\frac{X_j - p}{\sqrt{np(1-p)}}\right)\right] \\ &= np\left(\frac{1-p}{\sqrt{np(1-p)}}\right)^3 \exp\left(h\frac{1-p}{\sqrt{np(1-p)}}\right) - n(1-p)\left(\frac{p}{\sqrt{np(1-p)}}\right)^3 \exp\left(-h\frac{p}{\sqrt{np(1-p)}}\right). \end{aligned}$$

Take $c_0 > 0$ such that

$$60c_0\overline{p}\left(\sqrt{\frac{1-\underline{p}}{\underline{p}}}\right)^3 \exp\left(2\sqrt{\frac{1-\underline{p}}{\underline{p}}}c_0\right) + 60c_0(1-\underline{p})\left(\sqrt{\frac{\overline{p}}{1-\overline{p}}}\right)^3 = 1.$$

Then for any $n \in \mathbb{N}$ and $\lambda = c_0\sqrt{n}$,

$$60\lambda L(2\lambda) \leq 1.$$

Then by Lemma 2 in Sakhanenko (1996), whenever $c_0\sqrt{n} \geq 1$ and the event $A = \{|S_n| < c_0\sqrt{n}\}$ occurs,

$$|S_n - Z| \leq \frac{1}{c_0\sqrt{n}} + \frac{S_n^2}{60c_0\sqrt{n}}.$$

Moreover, by its proof, Z can be taken such that $Z = \Phi^{-1} \circ F(S_n)$. We then proceed as in the proof for Lemma 2 in Brown *et al.* (2010), where they show for each the coupling exits with c_0 to c_3 not depending on n. They did not give explicit dependency of c_0 to c_3, however. Take c_1 such that $c_1/(60c_0) < 1/2$. In particular, we can take $c_1 = 15c_0$. Then on the event $B = \{|S_n| < c_1\sqrt{n}\}$,

$$|S_n - Z| \leq \frac{1}{c_0\sqrt{n}} + |S_n|\frac{c_1\sqrt{n}}{60c_0\sqrt{n}} \leq \frac{1}{c_0\sqrt{n}} + \frac{1}{2}|S_n|.$$

Hence by triangle inequality, $|S_n| \leq \frac{2}{c_0\sqrt{n}} + 2|Z|$, and

$$|S_n - Z| \leq \frac{1}{c_0\sqrt{n}} + \frac{1}{60c_0\sqrt{n}}\left(\frac{2}{c_0\sqrt{n}} + 2|Z|\right)^2 \leq \frac{2}{c_0\sqrt{n}} + \frac{2}{15c_0\sqrt{n}}|Z|^2.$$

Recall $X = \sum_{i=1}^n X_i \sim \text{Bin}(n,p)$, whenever the event $C = \{|X - np| < c_1 n\sqrt{\underline{p}(1-\overline{p})}\}$ occurs and $c_0\sqrt{n} \geq 1$,

$$\left|X - np - \sqrt{np(1-p)}Z\right| \leq \frac{2}{c_0}\sqrt{p(1-p)} + \frac{2}{15c_0}\sqrt{p(1-p)}|Z|^2 \leq \frac{1}{c_0} + \frac{Z^2}{15c_0}.$$

Moreover, $|S_n| \leq \frac{2}{c_0\sqrt{n}} + 2|Z|$ implies

$$|X - np| \leq \frac{1}{c_0} + 2\sqrt{np(1-p)}|Z|.$$

$\square$

SA-II.12 Proof of Lemma SA.7

For notational simplicity, denote $\mathcal{I} = \{(j,k) \in \mathbb{Z} \times \mathbb{Z} : 1 \leq j \leq Kd, 0 \leq k < 2^{Kd-j}\}$ and $\mathcal{J} = \mathcal{I} \cup \{(0,k) : 0 \leq k < 2^{Kd}\}$.

Part 1: Construction of Strong Approximation

The construction will be essentially the same as in Section SA-II.3. By Lemma SA.4, there exists a mean-zero Gaussian process $(Z_n^X(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H} \cup \mathcal{E}_{Kd})$ with almost sure continuous path and the same covariance structure as $(X_n(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H} \cup \mathcal{E}_{Kd})$. For each $(j,k) \in \mathcal{J}$, we will take $V_{j,k} = \sqrt{n} Z_n^X(e_{j,k})$ and $\widetilde{V}_{j,k} = \sqrt{n} Z_n^X(\widetilde{e}_{j,k})$. By checking the covariance structures, we can show that if we define $\widetilde{\xi}_{j,k}$ such that $\widetilde{V}_{j,k} = \sqrt{n \frac{\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}(\mathcal{C}_{j,k})^2}} \widetilde{\xi}_{j,k}$, then $\widetilde{\xi}_{j,k} \overset{i.i.d}{\sim} N(0,1), (j,k) \in \mathcal{J}$. Take $F_{(j,k),m}$ to be the cumulative distribution function of $(S_{j,k} - mp_{j,k})/\sqrt{mp_{j,k}(1-p_{j,k})}$, where $p_{j,k} = \mathbb{P}(\mathcal{C}_{j-1,2k})/\mathbb{P}(\mathcal{C}_{j,k})$ and $S_{j,k}$ is a $\mathrm{Bin}(m, p_{j,k})$ random variable. Define $G_{(j,k),m}(t) = \sup\{x : F_{(j,k),m}(x) \leq t\}$.

We define $U_{j,k}, (j,k) \in \mathcal{I}$ via the following iterative scheme:

1. *Initialization:* Take $U_{Kd,0} = n$.

2. *Iteration:* Suppose we have define $U_{l,k}$ for $j < l \leq Kd, 0 \leq k < 2^{Kd-l}$, then solve for $U_{j,k}$'s such that

$$\widetilde{U}_{j,k} = G_{(j,k),U_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k}),$$
$$\widetilde{U}_{j,k} = \mathbb{P}(\mathcal{C}_{j-1,2k+1})/\mathbb{P}(\mathcal{C}_{j,k}) U_{j-1,2k} - \mathbb{P}(\mathcal{C}_{j-1,2k})/\mathbb{P}(\mathcal{C}_{j,k}) U_{j-1,2k+1} = U_{j-1,2k} - \mathbb{P}(\mathcal{C}_{j-1,2k+1})/\mathbb{P}(\mathcal{C}_{j,k}) U_{j,k},$$
$$U_{j-1,2k} + U_{j-1,2k+1} = U_{j,k}, \quad 0 \leq k < 2^{Kd-j}.$$

   Continue till we have defined $U_{0,k}$ for $0 \leq k < 2^{Kd}$.

$\{\widetilde{U}_{j,k} : (j,k) \in \mathcal{I}\}$ have the same joint distribution as $\{\sum_{i=1}^n e_{j,k}(\mathbf{x}_i) : (j,k) \in \mathcal{I}\}$. By Skorohod Embedding lemma (Dudley, 2014, Lemma 3.35), Z_n^X can be constructed on a possibly enlarged probability space such that the previously constructed $U_{j,k}$ satisfies $U_{j,k} = \sum_{i=1}^n e_{j,k}(\mathbf{x}_i)$ for all $(j,k) \in \mathcal{I}$. Take $\overline{p} = \rho$ and $\underline{p} = \rho^{-1}$. Take c_0 to be the positive solution of

$$60 c_0 \overline{p} \left(\sqrt{\frac{1-\underline{p}}{\underline{p}}} \right)^3 \exp\left(2 \sqrt{\frac{1-\underline{p}}{\underline{p}}} c_0 \right) + 60 c_0 (1-\underline{p}) \left(\sqrt{\frac{\overline{p}}{1-\overline{p}}} \right)^3 = 1,$$

and take $c_1 = 15 c_0 \sqrt{\underline{p}(1-\overline{p})}$, $c_2 = 1/(15c_0)$ and $c_3 = 1/c_0$. Define $A = \{|\widetilde{U}_{j,k}| \leq c_1 U_{j,k} \text{ for all } (j,k) \in \mathcal{I}\}$. Notice that we can always take $c_1 \leq 1$, since $|\widetilde{U}_{j,k}| \leq U_{j,k}$ a.s.. Using Lemma SA.6, whenever A occurs,

$$\begin{aligned} \left| \widetilde{U}_{j,k} - \sqrt{U_{j,k} \frac{\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}(\mathcal{C}_{j,k})^2}} \widetilde{\xi}_{j,k} \right| &< c_2 \widetilde{\xi}_{j,k}^2 + c_3, \\ \left| \widetilde{U}_{j,k} \right| &\leq 1/c_0 + 2 \sqrt{\frac{\mathbb{P}(\mathcal{C}_{j-1,2k})\mathbb{P}(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}(\mathcal{C}_{j,k})^2} U_{j,k} |\widetilde{\xi}_{j,k}|}, \quad \forall (j,k) \in \mathcal{I}. \end{aligned} \tag{SA-4}$$

Now, we bound $\mathbb{P}(A^c)$. By Chernoff's inequality for Binomial distribution, for all $(j,k) \in \mathcal{I}$,

$$\mathbb{P}\left(U_{j,k} \leq \frac{1}{2}\mathbb{E}[U_{j,k}]\right) \leq \exp\left(-\frac{\mathbb{E}[U_{j,k}]}{8}\right).$$

Moreover, $\rho^{-1} n 2^{j-Kd} \leq \mathbb{E}[U_{j,k}] \leq \rho n 2^{j-Kd}$. Hence

$$\mathbb{P}\left(U_{j,k} \leq \rho^{-1} n 2^{j-Kd}\right) \leq \exp\left(-\rho^{-1} n 2^{j-Kd}\right), \qquad \forall (j,k) \in \mathcal{I}.$$

Using Hoeffding's inequality and the fact that $\widetilde{U}_{j,k} = U_{j-1,2k} - \frac{\mathbb{P}(\mathcal{C}_{j-1,2k})}{\mathbb{P}(\mathcal{C}_{j-1,2k+1})} U_{j,k} = U_{j-1,2k} - \mathbb{E}[U_{j-1,2k} | U_{j,k}]$,

$$\mathbb{P}\left(\left|\widetilde{U}_{j,k}\right| \geq c_1 U_{j,k} \middle| U_{j,k} \geq \frac{1}{2}\rho^{-1} n 2^{-Kd+j}\right) \leq 2\exp\left(-\frac{c_1^2 n 2^{-Kd+j}}{3\rho}\right).$$

Putting together and using union bound,

$$\begin{aligned}
\mathbb{P}\left(A^c\right) &= \sum_{(j,k)\in\mathcal{I}} \mathbb{P}\left(\left|\widetilde{U}_{j,k}\right| > c_1 U_{j,k}\right) \\
&\leq \sum_{(j,k)\in\mathcal{I}} \mathbb{P}\left(U_{j,k} \leq \frac{1}{2}\rho^{-1} n 2^{-Kd+j}\right) + \mathbb{P}\left(\left|\widetilde{U}_{j,k}\right| \geq c_1 U_{j,k} \middle| U_{j,k} \geq \frac{1}{2}\rho^{-1} n 2^{-Kd+j}\right) \\
&\leq \sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} \exp(-\rho^{-1} n 2^{j-Kd}) + 2\exp\left(-\frac{c_1^2 n 2^{-Kd+j}}{3\rho}\right) \\
&\leq 4 \cdot 2^{Kd} \exp\left(-\min\left\{\frac{c_1^2}{3} \wedge 1\right\} \rho^{-1} n 2^{-Kd}\right).
\end{aligned}$$

Part 2: Bounding Strong Approximation Error

Next we will show that the proof of Theorem 2.1 in Rio (1994) still goes through for *approximate dyadic scheme*. In other words, we will show that the *approximate dyadic scheme* gives essentially the same Gaussian coupling rates as the *exact dyadic scheme*. Using the same notation as in Rio (1994) and define $\widetilde{p}_{j,k} = \mathbb{P}(\mathcal{C}_{j-1,2k})/\mathbb{P}(\mathcal{C}_{j,k})$ for notational simplicity, for $g \in L_2(\mathcal{X} \times \mathbb{R})$, define

$$\begin{aligned}
X(g) &= \sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(g) \widetilde{U}_{j,k}, \\
Y(g) &= \sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(g) \sqrt{U_{j,k} \widetilde{p}_{j,k}(1-\widetilde{p}_{j,k})} \widetilde{\xi}_{j,k}, \\
Z(g) &= \sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(g) \widetilde{V}_{j,k}, \\
\Delta(f) &= X(g) - Z(g), \Delta_1(g) = (X-Y)(g), \Delta_2(g) = (Y-Z)(g).
\end{aligned}$$

Claim 1: $\mathbb{E}[\exp(t\Delta_1(h))\mathbb{1}(A)] \le \prod_{j=1}^{Kd}\prod_{0\le k<2^{Kd-j}} \mathbb{E}[\cosh(t\widetilde{\beta}_{j,k}(h)(2+\widetilde{\xi}_{j,k}^2/4))]$. It then follows from the proof of Lemma 2.2 in Rio (1994) that for all $|t|<1$,

$$\log \mathbb{E}[\exp(4t\Delta_1(h))\mathbb{1}(A)] \le -\frac{83}{3}c_\rho^2\left(\sum_{j=1}^{Kd}\sum_{0\le k<2^{Kd-j}}\widetilde{\beta}_{j,k}^2(h)\right)\log(1-t^2).$$

Proof of Claim 1: Denote $\mathcal{F}_j = \sigma\left(\left\{\widetilde{\xi}_{l,k} : j < l \le Kd, 0 \le k < 2^{Kd-l}\right\}\right)$, for all $1 \le j < Kd$. In particular, $\sigma\left(\left\{U_{l,k} : j \le l \le Kd, 0 \le k < 2^{Kd-l}\right\}\right) \subseteq \mathcal{F}_j$. Then by Equation SA-4, for all $t \in \mathbb{R}$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(t\sum_{0\le k<2^{Kd-j}}\widetilde{\beta}_{j,k}(g)\left(\widetilde{U}_{j,k}-\sqrt{U_{j,k}\widetilde{p}_{j,k}(1-\widetilde{p}_{j,k})}\widetilde{\xi}_{j,k}\right)\right)\mathbb{1}(A)\middle|\mathcal{F}_j\right]\\
\le&\mathbb{E}\left[\prod_{0\le k<2^{Kd-j}}\cosh\left(t\widetilde{\beta}_{j,k}(g)(c_2\widetilde{\xi}_{j,k}^2+c_3)\right)\mathbb{1}(A)\middle|\mathcal{F}_j\right].
\end{aligned}$$

Then we will use the same induction argument in the proof of Lemma 2.2 in Rio (1994): Call

$$\begin{aligned}
S_j(t) &:= \exp\left(t\sum_{0\le k<2^{Kd-j}}\widetilde{\beta}_{j,k}(g)\left(\widetilde{U}_{j,k}-\sqrt{U_{j,k}\widetilde{p}_{j,k}(1-\widetilde{p}_{j,k})}\widetilde{\xi}_{j,k}\right)\right),\\
T_j(t) &:= \prod_{0\le k<2^{Kd-j}}\cosh\left(t\widetilde{\beta}_{j,k}(g)(c_2\widetilde{\xi}_{j,k}^2+c_3)\right).
\end{aligned}$$

So $\mathbb{E}[\exp(t\Delta_1)\mathbb{1}(A)] = \mathbb{E}\left[\prod_{j=1}^{Kd}S_j(t)\mathbb{1}(A)\right]$, $\prod_{j=1}^{Kd}\prod_{0\le k<2^{Kd-j}}\mathbb{E}[\cosh(t\widetilde{\beta}_{j,k}(2+\widetilde{\xi}_{j,k}^2/4))] = \mathbb{E}\left[\prod_{j=1}^{Kd}T_j(t)\right]$. By Equation SA-4, for all $1 \le j \le Kd$,

$$\mathbb{E}\left[S_j(t)\prod_{l=1}^{j-1}T_l(t)\mathbb{1}(A)\middle|\mathcal{F}_j\right] \le \mathbb{E}\left[\prod_{l=1}^{j}T_l(t)\mathbb{1}(A)\middle|\mathcal{F}_j\right].$$

It follows that

$$\begin{aligned}
\mathbb{E}[\exp(t\Delta_1)\mathbb{1}(A)] &= \mathbb{E}\left[\prod_{j=1}^{Kd}S_j(t)\mathbb{1}(A)\right] = \mathbb{E}\left[\mathbb{E}\left[S_1(t)\mathbb{1}(A)|\mathcal{F}_1\right]\prod_{j=2}^{Kd}S_j(t)\right] \le \mathbb{E}\left[\mathbb{E}\left[T_1(t)\mathbb{1}(A)|\mathcal{F}_1\right]\prod_{j=2}^{Kd}S_j(t)\right]\\
&= \mathbb{E}\left[\mathbb{E}\left[T_1(t)S_2(t)\mathbb{1}(A)|\mathcal{F}_2\right]\prod_{j=3}^{Kd}S_j(t)\right] \le \mathbb{E}\left[\mathbb{E}\left[T_1(t)T_2(t)\mathbb{1}(A)|\mathcal{F}_2\right]\prod_{j=3}^{Kd}S_j(t)\right] \le \cdots\\
&\le \mathbb{E}\left[\prod_{j=1}^{Kd}T_j(t)\mathbb{1}(A)\right] \le \mathbb{E}\left[\prod_{j=1}^{Kd}T_j(t)\right] = \prod_{j=1}^{Kd}\prod_{0\le k<2^{Kd-j}}\mathbb{E}[\cosh(t\widetilde{\beta}_{j,k}(h)(c_2\widetilde{\xi}_{j,k}^2+c_3))]\\
&\le \prod_{j=1}^{Kd}\prod_{0\le k<2^{Kd-j}}\mathbb{E}[\cosh(tc_\rho\widetilde{\beta}_{j,k}(h)(\widetilde{\xi}_{j,k}^2/4+2))]
\end{aligned}$$

where in the last line, we have used independence of $\widetilde{\xi}_{j,k} : 1 \le j \le Kd, 0 \le k < 2^{Kd-j}$. W.l.o.g, we will assume that $c_\rho\|g\|_\infty \le 1$. Since we know $\widetilde{\xi}_{j,k}, 1 \le j \le Kd, 0 \le k < 2^{Kd-j}$ are i.i.d standard Gaussian, the same upper bound worked out in Rio (1994) for the right hand side of the inequality also holds here, namely, for all $t < 1$,

$$\log \mathbb{E}[\exp(4t\Delta_1)\mathbb{1}(A)] \leq -\frac{83}{3} c_{\rho^2} \left(\sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} \widetilde{\beta}_{j,k}^2(h) \right) \log(1-t^2) =: h_{\Delta_1}(t) \tag{SA-5}$$

Claim 2: $\mathbb{E}[\exp(t\Delta_2)\mathbb{1}(A)] \leq \mathbb{E}[\exp(tc_\rho \Delta_3)]$ for all $t > 0$, where

$$\Delta_3(h) = \sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(h) \widetilde{\xi}_{j,k} \left(1 + \sum_{l=j}^{Kd} \sum_{0 \leq q < 2^{Kd-l}} 2^{-|j-l|/2} \left| \widetilde{\xi}_{l,q} \right| \mathbb{1}(\mathcal{C}_{l,q} \supseteq \mathcal{C}_{j,k}) \right), h \in L_2(\mathcal{X} \times \mathbb{R}),$$

and c_ρ is a constant that only depends on ρ.

Proof of Claim 2: Denote $p_{j,k} = \mathbb{P}(\mathcal{C}_{j,k})$. Then for any $g \in L_2(\mathbb{R}^d)$, we have

$$\Delta_2(g) = \sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} \widetilde{\beta}_{j,k}(g) \left(\sqrt{U_{j,k}} - \sqrt{\mathbb{E}[U_{j,k}]} \right) \sqrt{\frac{p_{j-1,2k} p_{j-1,2k+1}}{p_{j,k}^2}} \widetilde{\xi}_{j,k}.$$

We will use the same strategy as in Rio (1994) adapted to the quasi-dyadic case: Fix $0 \leq j \leq Kd, 0 \leq l < 2^{Kd-j}$, we will denote by l_k the unique integer in $[0, 2^{Kd-l})$ such that $\mathcal{C}_{l,k_l} \supseteq \mathcal{C}_{j,k}$. Then

$$\begin{aligned}
\sqrt{U}_{j,k} - \sqrt{\mathbb{E}[U_{j,k}]} &= \sum_{l=j}^{Kd-1} \sqrt{U_{l,k_l} \frac{p_{j,k}}{p_{l,k_l}}} - \sqrt{U_{l+1,k_{l+1}} \frac{p_{j,k}}{p_{l+1,k_{l+1}}}} \\
&= \sum_{l=j}^{Kd-1} \sqrt{\frac{p_{j,k}}{p_{l+1,k_{l+1}}}} \left(\sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}} U_{l,k_l}} - \sqrt{U_{l+1,k_{l+1}}} \right).
\end{aligned}$$

By Equation SA-4, when the event A holds,

$$\begin{aligned}
\left| \sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}} U_{l,k_l}} - \sqrt{U_{l+1,k_{l+1}}} \right| &\leq \frac{\left| \widetilde{U}_{l,k_l} \right|}{\sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}} U_{l,k_l}} + \sqrt{U_{l+1,k_{l+1}}}} \\
&\leq \frac{2\sqrt{\frac{p_{l+1,2k_l}}{p_{l,k_l}} \frac{p_{l+1,2k_l+1}}{p_{l,k_l}} U_{l,k_l}} \left| \widetilde{\xi}_{l,k_l} \right| + \min\left\{ c_0^{-1}, \widetilde{U}_{l,k_l} \right\}}{\sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}} U_{l,k_l}} + \sqrt{U_{l+1,k_{l+1}}}} \\
&\leq 2\sqrt{\frac{p_{l+1,2k_l+1}}{p_{l,k_l}}} \left| \widetilde{\xi}_{l,k_l} \right| + \frac{\min\left\{ c_0^{-1}, |\widetilde{U}_{l,k_l}| \right\}}{\sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}} U_{l,k_l}} + \sqrt{U_{l+1,k_{l+1}}}}.
\end{aligned}$$

For the first summand,

$$\sum_{l=j}^{Kd-1} \sqrt{\frac{p_{j,k}}{p_{l+1,k_{l+1}}}} 2\sqrt{\frac{p_{l+1,2k_l+1}}{p_{l,k_l}}} \left| \widetilde{\xi}_{l,k_l} \right| \lesssim c_\rho \sum_{l=j}^{Kd-1} 2^{-(l-j)/2} \left| \widetilde{\xi}_{l,k_l} \right|.$$

For the second summand, we separate it into two terms as in Rio (1994),

$$
\begin{aligned}
&\sum_{l=j}^{Kd-1}\sqrt{\frac{p_{j,k}}{p_{l+1,k_{l+1}}}}\frac{\min\left\{c_0^{-1},-\widetilde{U}_{l,k_l}\right\}}{\sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}}U_{l,k_l}}+\sqrt{U_{l+1,k_{l+1}}}}\mathbb{1}(\widetilde{U}_{l,k_l}\leq 0)\\
=&\sum_{l=j}^{Kd-1}\sqrt{\frac{p_{j,k}}{p_{l+1,k_{l+1}}}}\frac{\min\left\{c_0^{-1},-\widetilde{U}_{l,k_l}\right\}}{\sqrt{U_{l+1,k_{l+1}}-\widetilde{U}_{l,k_l}}+\sqrt{U_{l+1,k_{l+1}}}}\mathbb{1}(\widetilde{U}_{l,k_l}\leq 0)\lesssim c_\rho,
\end{aligned}
$$

since $\sup_{0\leq x\leq u}\min\{c_0^{-1},x\}/(\sqrt{u}+\sqrt{u+x})\lesssim 1$.

$$
\begin{aligned}
&\sum_{l=j}^{Kd-1}\sqrt{\frac{p_{j,k}}{p_{l+1,k_{l+1}}}}\frac{\min\left\{c_0^{-1},\widetilde{U}_{l,k_l}\right\}}{\sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}}U_{l,k_l}}+\sqrt{U_{l+1,k_{l+1}}}}\mathbb{1}(\widetilde{U}_{l,k_l}>0)\\
\leq&\sum_{l=j}^{Kd-1}\sqrt{\frac{p_{j,k}}{p_{l+1,k_{l+1}}}}\left(\sqrt{U_{l+1,k_{l+1}}}-\sqrt{\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}}U_{l,k_l}}\right)\mathbb{1}\Big(\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}}U_{l,k_l}\leq U_{l+1,k_{l+1}}\leq\frac{p_{l+1,k_{l+1}}}{p_{l,k_l}}U_{l,k_l}+c_0^{-1}\Big)\\
\leq&\sum_{l=j}^{Kd-1}\sqrt{\frac{p_{j,k}}{p_{l+1,k_{l+1}}}}\sqrt{c_0^{-1}}\lesssim 1.
\end{aligned}
$$

It follows that when the event A holds,

$$
\left|\sqrt{U_{j,k}}-\sqrt{\mathbb{E}[U_{j,k}]}\right|\leq c_\rho\left(1+\sum_{l=j}^{Kd-1}2^{-(l-j)/2}\sum_{0\leq q<2^{Kd-l}}\left|\widetilde{\xi}_{l,q}\right|\mathbb{1}(\mathcal{C}_{l,q})\supseteq\mathcal{C}_{j,k}\right).
$$

It then follows from induction argument similar to Claim 1 that for all $t>0$,,

$$
\mathbb{E}\left[\exp(t\Delta_2)\mathbb{1}(A)\right]\leq\mathbb{E}\left[\exp(tc_\rho\Delta_3)\right]. \tag{SA-6}
$$

Take $h_{\Delta_3}(t)=\log\left(\mathbb{E}\left[\exp(tc_\rho\Delta_3)\right]\right),t>0$. Combining Equation SA-5 and SA-6, for any $t>0$,

$$
\begin{aligned}
\mathbb{P}(\Delta_1\geq t,A)&\leq\inf_{u>0}\mathbb{P}(\exp(\Delta_1u)\geq\exp(tu),A)\leq\inf_{u>0}\exp(-tu)\mathbb{E}\left[\exp(\Delta_1u)\mathbb{1}(A)\right]\\
&\leq\exp\left(-\sup_{u>0}(tu-h_{\Delta_1}(u/4))\right)=\exp\left(-\sup_{u>0}\left(tu+\frac{83}{3}c_\rho^2\|h\|_{\mathcal{B}}^2\log\left(1-u^2/16\right)\right)\right),\\
\mathbb{P}\left(\Delta_2\geq t,A\right)&\leq\inf_{u>0}\exp(-tu)\mathbb{E}\left[\exp(\Delta_2u)\mathbb{1}(A)\right]\leq\exp\left(-\sup_{u>0}(tu-h_{\Delta_3}(u))\right).
\end{aligned}
$$

Since Δ_3 only depends on $\widetilde{\xi}_{j,k},1\leq j\leq Kd,0\leq k<2^{Kd-j}$, it follows from Lemma 2.3 and Lemma 2.4 in

Rio (1994) that for any $h \in \mathcal{H}$, for any $t > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sqrt{n}|X_n(h) - Z_n^X(h)| \geq c_\rho\sqrt{\|h\|_{\mathcal{B}}^2 t} + c_\rho(1+\sqrt{8Kd})\|h\|_\infty t\right) \\
&\leq \mathbb{P}\left(\sqrt{n}|X_n(h) - Z_n^X(h)| \geq c_\rho\sqrt{\|h\|_{\mathcal{B}}^2 t} + c_\rho(1+\sqrt{8Kd})\|h\|_\infty t, A\right) + \mathbb{P}(A^c) \\
&\leq \mathbb{P}\left(|\Delta_1(h) + \Delta_2(h)| \geq c_\rho\sqrt{\|h\|_{\mathcal{B}}^2 t} + c_\rho(1+\sqrt{8Kd})\|h\|_\infty t, A\right) + \mathbb{P}(A^c) \\
&\leq 2\exp(-t) + \mathbb{P}(A^c) \leq 2\exp(-t) + 4\cdot 2^{Kd}\exp\left(-\min\left\{\frac{c_1^2}{3}\wedge 1\right\}\rho^{-1}n2^{-Kd}\right),
\end{aligned}$$

where $\|h\|_{\mathcal{B}}^2 = \sum_{j=1}^{Kd}\sum_{0\leq k<2^{Kd-j}}|\widetilde{\beta}_{j,k}(h)|^2$. □

SA-II.13 Proof of Lemma SA.8

Take $\mathcal{L} := \{h - \pi_{\mathcal{H}_\delta}(h) : h \in \mathcal{H}\}$. Then $\sigma := \sup_{l\in\mathcal{L}}\|l\|_{\mathbb{P}^X,2} \leq \delta \mathtt{M}_{\mathcal{H}}$. Moreover, for all $0 < \varepsilon < \delta$,

$$\sup_Q N(\mathcal{L}, e_Q, \varepsilon\mathtt{M}_{\mathcal{H}}) \leq N(\varepsilon)N(\delta) \leq N(\varepsilon)^2,$$

where the supremum is taken over all finite discrete measures. Hence $\int_0^u \sqrt{1 + \sup_Q \log N(\mathcal{L}, \|\cdot\|_{Q,2}, \varepsilon\mathtt{M}_{\mathcal{H}})}d\varepsilon \leq 2J(u, \mathcal{H}, \mathtt{M}_{\mathcal{H}})$ for all $0 < u < \delta$. By Theorem 5.2 in Chernozhukov *et al.* (2014), we have

$$\mathbb{E}\left[\|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}}\right] \lesssim J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}})\mathtt{M}_{\mathcal{H}} + \frac{\mathtt{M}_{\mathcal{H}}J^2(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}})}{\delta^2\sqrt{n}}.$$

By Talagrand's inequality (Giné and Nickl, 2016, Theorem 3.3.9), for all $t > 0$,

$$\mathbb{P}\left(\|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}} \geq C\left\{J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}})\mathtt{M}_{\mathcal{H}} + \frac{\mathtt{M}_{\mathcal{H}}J^2(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}})}{\delta^2\sqrt{n}} + \delta\mathtt{M}_{\mathcal{H}}\sqrt{t} + \frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}t\right\}\right) \leq \exp(-t),$$

where C is an absolute constant. By Corollary 2.2.9 in van der Vaart and Wellner (2013),

$$\mathbb{E}\left[\|Z_n - Z_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}}\right] \lesssim J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}})\mathtt{M}_{\mathcal{H}_\delta}.$$

By pointwise separability and a concentration inequality for Gaussian suprema, for all $t > 0$,

$$\mathbb{P}\left(\|Z_n - Z_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}} \geq C'\left\{J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}})\mathtt{M}_{\mathcal{H}} + \delta\mathtt{M}_{\mathcal{H}}\sqrt{t}\right\}\right) \leq \exp(-t),$$

where C' is another absolute constant. □

SA-II.14 Proof of Lemma SA.9

Let $h \in \mathcal{H}$. Then almost surely, $|h(\mathbf{x}_i) - \Pi_0 h(\mathbf{x}_i)| \leq \min\{2\mathtt{M}_{\mathcal{H}}, \mathtt{L}_{\mathcal{H}_\delta}\|\mathcal{V}\|_\infty\} =: \mathtt{B}_{\mathcal{H}_\delta}$. Then

$$\begin{aligned}
\mathbb{E}\left[|h(\mathbf{x}_i) - \Pi_0 h(\mathbf{x}_i)|\right] &= \sum_{0\leq k<2^{Kd}}\int_{\mathcal{C}_{0,k}}|h(\mathbf{x}) - 2^{Kd}\int_{\mathcal{C}_{0,k}}h(\mathbf{y})f_X(\mathbf{y})d\mathbf{y}|f_X(\mathbf{x})d\mathbf{x} \\
&\leq \sum_{0\leq k<2^{Kd}} 2^{Kd}\int_{\mathcal{C}_{0,k}}\int_{\mathcal{C}_{0,k}}|h(\mathbf{x}) - h(\mathbf{y})|f_X(\mathbf{y})f_X(\mathbf{x})d\mathbf{y}d\mathbf{x}.
\end{aligned}$$

Using a change of variable $\mathbf{s} = \mathbf{y} - \mathbf{x}$ and the fact that f_X is bounded above, we have

$$\begin{aligned}
&\mathbb{E}\left[|h(\mathbf{x}_i) - \Pi_0 h(\mathbf{x}_i)|\right] \\
&\leq \sum_{0\leq k<2^{Kd}} 2^{Kd} \int_{\mathcal{C}_{0,k}-\mathcal{C}_{0,k}} \int_{\mathcal{C}_{0,k}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| f_X(\mathbf{x}+\mathbf{s}) f_X(\mathbf{x}) \mathbb{1}_{\mathcal{C}_{0,k}}(\mathbf{x}+\mathbf{s}) d\mathbf{x} d\mathbf{s} \\
&\leq \left(\sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\right)^2 2^{Kd} \int_{\mathcal{V}} \int_{\mathcal{X}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| \, d\mathbf{x} d\mathbf{s}.
\end{aligned}$$

Let ϕ be a real-valued non-negative Lebesgue measurable function on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} \phi(\mathbf{u}) d\mathbf{u} = 1$. Define $\phi_\varepsilon = \varepsilon^{-d}\phi(\cdot/\varepsilon)$ and $h_\varepsilon = h * \phi_\varepsilon$. Then

$$\begin{aligned}
\int_{\mathcal{X}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| \, d\mathbf{x} &= \lim_{\varepsilon\downarrow 0} \int_{\mathcal{X}} |h_\varepsilon(\mathbf{x}) - h_\varepsilon(\mathbf{x}+\mathbf{s})| d\mathbf{x} \leq \lim_{\varepsilon\downarrow 0} \int_{\mathcal{X}} \int_0^{\|\mathbf{s}\|} \|\nabla h_\varepsilon(\mathbf{x} + t\mathbf{s}/\|\mathbf{s}\|)\| dt d\mathbf{x} \\
&\leq \int_0^{\|\mathbf{s}\|} \lim_{\varepsilon\downarrow 0} \int_{\mathcal{X}} \|\nabla h_\varepsilon(\mathbf{x} + t\mathbf{s}/\|\mathbf{s}\|)\| d\mathbf{x} dt \leq \|\mathbf{s}\| \mathtt{TV}_{\{h\}}.
\end{aligned}$$

It follows that

$$\mathbb{E}[|h(\mathbf{x}_i) - \Pi_0 h(\mathbf{x}_i)|] \leq \left(\sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\right)^2 2^{Kd} \mathfrak{m}(\mathcal{V}) \|\mathcal{V}\|_\infty \mathtt{TV}_{\{h\}},$$

where in (1) we used Dominated Convergence Theorem, in (2) we used Lemma 1 in De Giorgi (1955) and the fact that each $\mathcal{C}_{0,k}$ is a d-dimensional cube with side-length at most Δ_{Kd}. It follows that

$$\mathbb{V}[h(\mathbf{x}_i) - \Pi_0 h(\mathbf{x}_i)] \leq \min\{2\mathtt{M}_{\mathcal{H}}, \mathtt{L}_{\mathcal{H}_\delta} \|\mathcal{V}\|_\infty\} \left(\sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\right)^2 2^{Kd} \mathfrak{m}(\mathcal{V}) \|\mathcal{V}\|_\infty \mathtt{TV}_{\mathcal{H}_\delta} =: \mathtt{V}_{\mathcal{H}_\delta}, \forall h \in \mathcal{H}_\delta.$$

Then by Bernstein inequality, for any $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(|X_n(h) - X_n(\Pi_0 h)| \geq t\right) &\leq 2\exp\left(-\frac{\frac{1}{2}t^2 n}{n\mathtt{V}_{\mathcal{H}_\delta} + \frac{1}{3}\mathtt{B}_{\mathcal{H}_\delta} t\sqrt{n}}\right) \\
&\leq 2\exp\left(-\frac{1}{2}\min\left\{\frac{\frac{1}{2}t^2 n}{n\mathtt{V}_{\mathcal{H}_\delta}}, \frac{\frac{1}{2}t^2 n}{\frac{1}{3}\mathtt{B}_{\mathcal{H}_\delta} t\sqrt{n}}\right\}\right).
\end{aligned}$$

Set $u = \frac{1}{2}\min\left\{\frac{\frac{1}{2}t^2 n}{n\mathtt{V}_{\mathcal{H}_\delta}}, \frac{\frac{1}{2}t^2 n}{\frac{1}{3}\mathtt{B}_{\mathcal{H}_\delta} t\sqrt{n}}\right\} > 0$, then either $t = 2\sqrt{\mathtt{V}_{\mathcal{H}_\delta}}\sqrt{u}$ or $t = \frac{4}{3}\frac{\mathtt{B}}{\sqrt{n}}u$. Hence $t \leq 2\sqrt{\mathtt{V}_{\mathcal{H}_\delta}}\sqrt{u} + \frac{4}{3}\frac{\mathtt{B}_{\mathcal{H}_\delta}}{\sqrt{n}}u$. It follows that for any $u > 0$, $\mathbb{P}(|X_n(h) - X_n(\Pi_0 h)| \geq 2\sqrt{\mathtt{V}_{\mathcal{H}_\delta}}\sqrt{u} + \frac{4}{3}\frac{\mathtt{B}_{\mathcal{H}_\delta}}{\sqrt{n}}u) \leq 2\exp(-u)$. The result for $\|X_n - X_n \circ \Pi_0\|_{\mathcal{H}_\delta}$ then follows from a union bound. The result for $\|Z_n - Z_n \circ \Pi_0\|_{\mathcal{H}_\delta}$ follows from the fact that $Z_n(h) - Z_n(\Pi_0 h)$ is a mean-zero Gaussian with variance $\mathbb{V}[X_n(h) - X_n(\Pi_0)]$ and a union bound argument. □

SA-II.15 Proof of Lemma SA.10

We employ the same strategy as in the proof of Theorem 1.1 from Rio (1994), except noting that incorporating Lipschitz condition can lead to tighter bound for strong approximation error.

For each $1 \leq j \leq Kd$, there exists unique integers $j_1, \ldots, j_d$ such that $0 \leq j_1 \leq \ldots \leq j_d \leq j_1 + 1$ and $\sum_{i=1}^d j_i = j$. In particular, there exists a unique $l := l(j) \in [d]$ such that either $l \leq d-1$ and $j_l < j_{l+1}$ or

$l = d$ and $j_d < j_1 + 1$. Recall $\widetilde{\beta}_{j,k}(h) = \mathbb{E}[h(\mathbf{x}_i)|\mathbf{x}_i \in \mathcal{C}_{j-1,2k}] - \mathbb{E}[h(\mathbf{x}_i)|\mathbf{x}_i \in \mathcal{C}_{j-1,2k+1}]$.

$$\begin{aligned}
\widetilde{\beta}_{j,k}(h) &= 2^{Kd-j} \int_{\mathcal{C}_{j-1,2k}} h(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} - 2^{Kd-j} \int_{\mathcal{C}_{j-1,2k+1}} h(\mathbf{y}) f_X(\mathbf{y}) d\mathbf{y} \\
&= 2^{Kd-j} \int_{\mathcal{C}_{j-1,2k}} \left(h(\mathbf{x}) - \left(2^{Kd-j} \int_{\mathcal{C}_{j-1,2k+1}} h(\mathbf{y}) f_X(\mathbf{y}) d\mathbf{y} \right) \right) f_X(\mathbf{x}) d\mathbf{x} \\
&= 2^{2(Kd-j)} \int_{\mathcal{C}_{j-1,2k}} \int_{\mathcal{C}_{j-1,2k+1}} (h(\mathbf{x}) - h(\mathbf{y})) f_X(\mathbf{x}) f_X(\mathbf{y}) d\mathbf{y} d\mathbf{x} \\
&= 2^{2(Kd-j)} \int_{\mathcal{C}_{j-1,2k}} \int_{\mathcal{C}_{j-1,2k+1} - \mathcal{C}_{j-1,2k}} (h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})) f_X(\mathbf{x}) f_X(\mathbf{x}+\mathbf{s}) \mathbb{1}_{\mathcal{C}_{j-1,2k+1}}(\mathbf{x}+\mathbf{s}) d\mathbf{s} d\mathbf{x}.
\end{aligned}$$

Since we have assumed f is bounded from above on $\mathcal{X}$,

$$\left| \widetilde{\beta}_{j,k}(h) \right| \leq 2^{2(Kd-j)} \left(\sup_{\mathbf{x} \in \mathcal{X}} f_X(\mathbf{x}) \right)^2 \int_{\mathcal{C}_{j-1,2k+1} - \mathcal{C}_{j-1,2k}} \int_{\mathcal{C}_{j-1,2k}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| d\mathbf{x} d\mathbf{s}.$$

Recall we define $\mathcal{U}_j = \cup_{0 \leq k < 2^{Kd-j}} (\mathcal{C}_{j-1,2k+1} - \mathcal{C}_{j-1,2k})$. Then

$$\sum_{0 \leq k < 2^{Kd-j}} \left| \widetilde{\beta}_{j,k}(h) \right| \leq \left(\sup_{\mathbf{x} \in \mathcal{X}} f_X(\mathbf{x}) \right)^2 2^{2(Kd-j)} \int_{\mathcal{U}_j} \int_{\sqcup_{0 \leq k < 2^{Kd-j}} \mathcal{C}_{j-1,2k}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| d\mathbf{x} d\mathbf{s}.$$

Then by similar smoothing argument as in the proof of Lemma SA.9,

$$\int_{\sqcup_{0 \leq k < 2^{Kd-j}} \mathcal{C}_{j-1,2k}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| d\mathbf{x} \leq \|s\| \mathtt{TV}_{\{h\}}.$$

It follows that

$$\sum_{0 \leq k < 2^{Kd-j}} \left| \widetilde{\beta}_{j,k}(h) \right| \leq \left(\sup_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \right)^2 2^{2(Kd-j)} \|\mathcal{U}_j\|_\infty \mathfrak{m}(\mathcal{U}_j) \mathtt{TV}_{\{h\}}.$$

Alternatively, it also holds that

$$\sum_{0 \leq k < 2^{Kd-j}} \left| \widetilde{\beta}_{j,k}(h) \right| \leq 2^{Kd-j} \int_{\sqcup_{0 \leq k < 2^{Kd-(j-1)}} \mathcal{C}_{j-1,k}} |h(\mathbf{x})| \, f_X(\mathbf{x}) d\mathbf{x} \leq 2^{Kd-j} \mathtt{E}_{\{h\}}.$$

Moreover, $|\widetilde{\beta}_{j,k}(h)| \leq \min\{\mathtt{M}_{\{h\}}, \|\mathcal{U}_j\|_\infty \mathtt{L}_{\{h\}}\}$, hence

$$\sum_{j=1}^{Kd} \sum_{0 \leq k < 2^{Kd-j}} |\widetilde{\beta}_{j,k}(h)|^2 \leq \sum_{j=1}^{Kd} \min\{\mathtt{M}_{\mathcal{H}_\delta}, \|U_j\|_\infty \mathtt{L}_{\mathcal{H}_\delta}\} \sum_{0 \leq k < 2^{Kd-j}} |\widetilde{\beta}_{j,k}(h)| \leq \mathcal{R}_{Kd}\left(\mathcal{H}_\delta\right),$$

where $\mathcal{R}_{Kd}\left(\mathcal{H}_\delta\right)$ is defined to be

$$\sum_{j=1}^{Kd} \min\{\mathtt{M}_{\mathcal{H}_\delta}, \|U_j\|_\infty \mathtt{L}_{\mathcal{H}_\delta}\} 2^{Kd-j} \min\left\{ \left(\sup_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) \right)^2 2^{2(Kd-j)} \|\mathcal{U}_j\|_\infty \mathfrak{m}(\mathcal{U}_j) \mathtt{TV}_{\mathcal{H}_\delta}, \|U_j\|_\infty \mathtt{L}_{\mathcal{H}_\delta}, \mathtt{E}_{\mathcal{H}_\delta} \right\}.$$

Applying Lemma SA.5, for any $h \in \mathcal{H}_\delta$,for any $t > 0$, with probability at least $1 - 2\exp(-t)$,

$$\left|X_n \circ \Pi_0(h) - Z_n^X \circ \Pi_0(h)\right| \leq 48\sqrt{\frac{\mathcal{R}_{Kd}\left(\mathcal{H}_\delta\right)}{n}t} + \sqrt{\frac{\mathtt{C}_{\mathcal{H}_\delta}}{n}t}.$$

The result then follows from the fact that $\mathrm{Card}(\mathcal{H}_\delta) \leq \mathtt{N}(\delta)$ and a union bound argument. □

SA-II.16 Proof of Lemma SA.11

This follows from Lemma SA.7 and the same bound for $\|g\|_{\mathcal{B}}$ as in Lemma SA.10. □

SA-II.17 Proof of Lemma SA.12

The first three equalities are self-evident. In what follows, we will use $f_{\mathcal{I}|\mathcal{J}}(\cdot|\cdot)$ as a shorthand for the conditional density $f_{\mathbf{X}_\mathcal{I}|\mathbf{X}_\mathcal{J}}(\cdot|\cdot)$ and use the notations $\overline{f}_X = \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})$, $\underline{f}_X = \inf_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})$. Then ϕ_X^{-1} is given by $\phi_X^{-1} : (u_1, \cdots, u_d) \mapsto (g^{-1}(u_1), g_{u_1}^{-1}(u_2), \cdots, g_{u_1,\cdots,u_{d-1}}^{-1}(u_d))$, where $g(x_1) = F_1(x_1)$ and $g_{u_1,\cdots,u_{i-1}}(x_i) = F_{X_i|X_1,\cdots,X_{i-1}}(x_i|g^{-1}(u_1), \cdots, g_{u_1,\cdots,u_{i-2}}^{-1}(u_{i-1}))$. Hence

$$\nabla\phi_X^{-1}(\mathbf{u}) = \begin{bmatrix} 1/f_1(x_1) & 0 & 0 & \cdots & 0 \\ * & 1/f_{2|1}(x_2|x_1) & 0 & \cdots & 0 \\ \cdots & & & & \\ * & * & * & \cdots & 1/f_{d|1,\ldots,d-1}(x_d|x_1,\cdots,x_{d-1}) \end{bmatrix}, \mathbf{u} \in [0,1]^d,$$

where $(x_1, \cdots, x_d) = \phi_X^{-1}(u_1, \cdots, u_d)$. But for any $m \in [n]$ and $(x_1, \ldots, x_m) \in I^m$, we have the relation $f_{[m]}(x_1, \ldots, x_m) = \int_{I^{d-m}} f_X(x_1, \ldots, x_m, \mathbf{u})d\mathbf{u} \in [\underline{f}_X, \overline{f}_X]$. Hence $\|\|\|\nabla\phi_X^{-1}\|_{\mathrm{op}}\|_\infty \leq \overline{f}_X\underline{f}_X^{-1}$.

The second to last inequality follows from the fact that for any $h \in \mathcal{H}$, $\|h\circ\phi_X^{-1}\|_{\mathrm{Lip}} \leq \|h\|_{\mathrm{Lip}}\|\|\|\nabla\phi_X^{-1}\|_{\mathrm{op}}\|_\infty$. To show the third inequality, take $l : \mathbb{R}^d \to \mathbb{R}$ to be a non-negative function such that $\int_{\mathbb{R}^d} l(\mathbf{x})d\mathbf{x} = 1$. For any $\varepsilon > 0$, define $l_\varepsilon(\cdot) = l(\cdot/\varepsilon)/\varepsilon^d$. Define $h_\varepsilon := h * l_\varepsilon$. Then for any $h \in \mathcal{H}$,

$$\begin{aligned}
\mathtt{TV}_{\{h\circ\phi_X^{-1}\}} &= \lim_{\varepsilon\downarrow 0}\int_{\mathbf{u}\in I^d}\|\nabla(h_\varepsilon\circ\phi_X^{-1})(\mathbf{u})\|d\mathbf{u} = \lim_{\varepsilon\downarrow 0}\int_{\mathbf{u}\in[0,1]^d}\|\left(\nabla\phi_X^{-1}(\mathbf{u})\right)^\top\nabla h_\varepsilon(\phi_X^{-1}(\mathbf{u}))\|d\mathbf{u} \\
&= \lim_{\varepsilon\downarrow 0}\int_{\mathbf{x}\in\mathcal{X}}\|\left(\nabla\phi_X^{-1}(\phi_X(\mathbf{x}))\right)^\top\nabla h_\varepsilon(\mathbf{x})\|\det\left(\nabla\phi_X(\mathbf{x})\right)d\mathbf{x} \\
&\leq \lim_{\varepsilon\downarrow 0}\int_{\mathbf{x}\in\mathcal{X}}\|\nabla h_\varepsilon(\mathbf{x})\|d\mathbf{x}\cdot\|\det(\nabla\phi_X)\|_\infty\cdot\|\|\|\nabla\phi_X^{-1}\|_{\mathrm{op}}\|_\infty \leq \mathtt{TV}_{\{h\}}\overline{f}_X\frac{\overline{f}_X}{\underline{f}_X}.
\end{aligned}$$

Moreover, let $\mathcal{C} \subseteq \mathbb{R}^d$ be a cube with edges of length a parallel to the coordinate axises. Then $\phi_X^{-1}(\mathcal{C})$ is contained in another cube $\mathcal{C}'$ with edges of length at most $2\sqrt{d}\|\|\|\nabla\phi_X^{-1}\|_{\mathrm{op}}\|_\infty a$. Hence for any $h \in \mathcal{H}$,

$$\begin{aligned}
\sup_{\varphi\in\mathcal{D}_d(\mathcal{C})}\int h(\mathbf{x})\,\mathrm{div}(\varphi)(\mathbf{x})d\mathbf{x}/\|\|\varphi\|_2\|_\infty =&\lim_{\varepsilon\downarrow 0}\int_{\mathcal{C}}\|\nabla(h_\varepsilon\circ\phi_X^{-1})(\mathbf{u})\|d\mathbf{u} \\
\leq&\lim_{\varepsilon\downarrow 0}\int_{\mathcal{C}'}\|\nabla(h_\varepsilon\circ\phi_X^{-1})(\phi_X(\mathbf{x}))\|\det(\nabla\phi_X(\mathbf{x}))d\mathbf{x} \\
\leq&\lim_{\varepsilon\downarrow 0}\int_{\mathcal{C}'}\|\nabla h_\varepsilon(\mathbf{x})\|d\mathbf{x}\cdot\|\det(\nabla\phi_X)\|_\infty\cdot\|\|\|\nabla\phi_X^{-1}\|_{\mathrm{op}}\|_\infty \\
\leq&(2\sqrt{d})^{d-1}\|\det(\nabla\phi_X)\|_\infty\cdot\|\|\|\nabla\phi_X^{-1}\|_{\mathrm{op}}\|_\infty^d a^{d-1}\mathtt{K}_{\{h\}},
\end{aligned}$$

where we have used the definition of $\mathtt{K}_{\{h\}}$ in the last line. Hence

$$\mathtt{K}_{\widetilde{\mathcal{H}}} \leq (2\sqrt{d})^{d-1}\overline{f}_X \left(\frac{\overline{f}_X}{\underline{f}_X}\right)^d \mathtt{K}_{\mathcal{H}}.$$

□

SA-II.18 Proof of Theorem 1

The proof proceeds by bounding each of the terms discussed in

$$\|X_n - Z_n^X\|_{\mathcal{H}} \leq \|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}} + \|X_n - Z_n^X\|_{\mathcal{H}_\delta} + \|Z_n^X \circ \pi_{\mathcal{H}_\delta} - Z_n^X\|_{\mathcal{H}}$$

and

$$\|X_n - Z_n^X\|_{\mathcal{H}_\delta} \leq \|X_n - \Pi_0 X_n\|_{\mathcal{H}_\delta} + \|\Pi_0 X_n - \Pi_0 Z_n^X\|_{\mathcal{H}_\delta} + \|\Pi_0 Z_n^X - Z_n^X\|_{\mathcal{H}_\delta},$$

and then balancing their contributions.

We first make a reduction via Rosenblatt transformation. Take $\mathbf{u}_i = \phi_X(\mathbf{x}_i)$ where ϕ_X is defined as in Lemma SA.12. And define $\widetilde{h} = h \circ \phi_X^{-1}$ for each $h \in \mathcal{H}$ and consider $\widetilde{\mathcal{H}} = \{\widetilde{h} : h \in \mathcal{H}\}$. Then

$$X_n(h) = \frac{1}{\sqrt{n}}\sum_{i=1}^n h(\mathbf{x}_i) - \mathbb{E}[h(\mathbf{x}_i)] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{h}(\mathbf{u}_i) - \mathbb{E}[\widetilde{h}(\mathbf{u}_i)] =: \widetilde{X}_n(\widetilde{h}), \quad \forall h \in \mathcal{H}.$$

Consider $\mathcal{E}_K$ that is an axis-aligned iterative splitting of depth K based on the law of $\mathbf{u}_i$ as given in Definition SA.2. By Lemma SA.4 and Lemma SA.12, $\widetilde{\mathcal{H}} \cup \Pi_0\widetilde{\mathcal{H}} \cup \mathcal{E}_K$ is pre-Gaussian, hence by the argument in Section SA-II.3, on a possibly enlarged probability space there exists a mean-zero Gaussian process Z_n^X indexed by $\widetilde{\mathcal{H}} \cup \Pi_0\widetilde{\mathcal{H}} \cup \mathcal{E}_K$ such that with almost sure continuous sample path such that

$$\mathbb{E}\left[Z_n^X(g), Z_n^X(f)\right] = \mathbb{E}\left[\widetilde{X}_n(g), \widetilde{X}_n(f)\right], \quad \forall g, f \in \widetilde{\mathcal{H}} \cup \Pi_0\widetilde{\mathcal{H}} \cup \mathcal{E}_K,$$

and $U_{j,k} = \sum_{i=1}^n e_{j,k}(\mathbf{x}_i)$ for all (j,k)'s. Let $\mathcal{H}_\delta$ be a $\delta\mathtt{M}_{\widetilde{\mathcal{H}}} = \delta\mathtt{M}_{\mathcal{H}}$-net of $\widetilde{\mathcal{H}}$ with cardinality no greater than $\mathtt{N}_{\widetilde{\mathcal{H}}}(\delta)$.

Since $\mathbf{u}_i \overset{i.i.d}{\sim} \mathrm{Unif}([0,1]^d)$ and the cells $\mathcal{A}_K(\mathbb{P}_U, 1)$ are obtained via *axis aligned dyadic expansion of depth K w.r.p. to $\mathbb{P}_U$ which is the law of* $\mathbf{u}_i$, we have $\mathcal{U}_j \subseteq [-2^{-\frac{K-j}{d}+1}, 2^{-\frac{K-j}{d}+1}]^d$. Then by Lemma SA.10, for all $t > 0$,

$$\mathbb{P}\left[\|\widetilde{X}_n \circ \Pi_0 - Z_n^X \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta} > 48\sqrt{\frac{\widetilde{\mathcal{R}}_K(\mathcal{H}_\delta)}{n}t} + \sqrt{\frac{\mathtt{C}_{\widetilde{\mathcal{H}}_\delta}}{n}t}\right] \leq 2\widetilde{N}(\delta)e^{-t},$$

where

$$\widetilde{\mathcal{R}}_K(\mathcal{H}_\delta) \leq \begin{cases} \min\{\mathtt{TV}_{\widetilde{\mathcal{H}}_\delta}\mathtt{M}_{\widetilde{\mathcal{H}}_\delta}, \mathtt{TV}_{\widetilde{\mathcal{H}}_\delta}\mathtt{L}_{\widetilde{\mathcal{H}}_\delta}\}, & \text{if } d = 1, \\ \min\{2^K\mathtt{TV}_{\widetilde{\mathcal{H}}_\delta}\mathtt{M}_{\widetilde{\mathcal{H}}_\delta}, K\mathtt{TV}_{\widetilde{\mathcal{H}}_\delta}\mathtt{L}_{\widetilde{\mathcal{H}}_\delta}\}, & \text{if } d = 2, \\ \min\{2^{K(d-1)}\mathtt{TV}_{\widetilde{\mathcal{H}}_\delta}\mathtt{M}_{\widetilde{\mathcal{H}}_\delta}, 2^{K(d-2)}\mathtt{TV}_{\widetilde{\mathcal{H}}_\delta}\mathtt{L}_{\widetilde{\mathcal{H}}_\delta}\} & \text{if } d \geq 3. \end{cases}$$

Moreover, from the bound on $\widetilde{\beta}_{j,k}$ from Lemma SA.10, we know for each (j,k),

$$\sum_{m:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}} \left|\widetilde{\beta}_{j,k}(\widetilde{h})\right| \leq 2^{2(K-l)} \int_{\mathcal{U}_l}\int_{\mathcal{C}_{j,k}} |h(\mathbf{x})-h(\mathbf{x}+\mathbf{s})| d\mathbf{x} d\mathbf{s}$$
$$\leq 2^{2(K-l)} \int_{\mathcal{U}_l} \|\mathbf{s}\| \mathtt{K}_{\widetilde{\mathcal{H}}} \|C_{j,k}\|_\infty^{d-1} d\mathbf{s} \leq 2^{2(K-l)} \operatorname{Vol}(\mathcal{U}_l)\|\mathcal{U}_l\|_\infty \|C_{j,k}\|_\infty^{d-1} \mathtt{K}_{\widetilde{\mathcal{H}}} \leq 2^{\frac{d-1}{d}(j-l)} \mathtt{K}_{\widetilde{\mathcal{H}}}.$$

It follows from the definition of $\mathtt{C}_{\widetilde{\mathcal{H}}}$ that $\mathtt{C}_{\widetilde{\mathcal{H}}} \leq \min\{\sqrt{K\mathtt{M}_{\widetilde{\mathcal{H}}}^2}, \sqrt{d^3\mathtt{M}_{\widetilde{\mathcal{H}}}\mathtt{K}_{\widetilde{\mathcal{H}}}}\}$. For projection error, by Lemma SA.9, for all $t>0$, with probability at least $1-2\widetilde{N}(\delta)e^{-t}$,

$$\|\widetilde{X}_n - \widetilde{X}_n \circ \Pi_0\|_{\mathcal{H}_\delta} \leq \sqrt{4d\min\{2\mathtt{M}_{\widetilde{\mathcal{H}}_\delta}, \mathtt{L}_{\widetilde{\mathcal{H}}_\delta} 2^{-K}\} 2^{-K} \mathtt{TV}_{\widetilde{\mathcal{H}}_\delta} t} + \frac{4\min\{2\mathtt{M}_{\widetilde{\mathcal{H}}_\delta}, \mathtt{L}_{\widetilde{\mathcal{H}}_\delta} 2^{-K}\}}{3\sqrt{n}} t,$$
$$\|Z_n^X - Z_n^X \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta} \leq \sqrt{4d\min\{2\mathtt{M}_{\widetilde{\mathcal{H}}_\delta}, \mathtt{L}_{\widetilde{\mathcal{H}}_\delta} 2^{-K}\} 2^{-K} \mathtt{TV}_{\widetilde{\mathcal{H}}_\delta} t}.$$

We balance the previous two errors by choosing $K = \lfloor d^{-1}\log_2 n \rfloor$ and get for all $t>0$, with probability at least $1-2\exp(-t)$,

$$\|\widetilde{X}_n - Z_n^X\|_{\mathcal{H}_\delta} \leq \min\left\{\mathtt{M}_{n,d}\sqrt{\mathtt{M}_{\widetilde{\mathcal{H}}}}, \mathtt{L}_{n,d}\sqrt{\mathtt{L}_{\widetilde{\mathcal{H}}}}\right\} \sqrt{(t+\log \widetilde{\mathtt{N}}_{\widetilde{\mathcal{H}}}(\delta)) d\mathtt{TV}_{\widetilde{\mathcal{H}}}} + \sqrt{\frac{\min\{K, d^3\frac{\mathtt{K}_{\widetilde{\mathcal{H}}}}{\mathtt{M}_{\widetilde{\mathcal{H}}}}\}}{n}} (t+\log\widetilde{\mathtt{N}}_{\widetilde{\mathcal{H}}}(\delta))\mathtt{M}_{\widetilde{\mathcal{H}}}.$$

Moreover by Lemma SA.8 we bound fluctuation off-the-net by, for all $t>0$,

$$\mathbb{P}\left[\|\widetilde{X}_n - \widetilde{X}_n \circ \pi_{\widetilde{\mathcal{H}}_\delta}\|_{\widetilde{\mathcal{H}}} > C\widetilde{\mathtt{F}}_n(t,\delta)\right] \leq \exp(-t),$$
$$\mathbb{P}\left[\|Z_n^X \circ \pi_{\widetilde{\mathcal{H}}_\delta} - Z_n^X\|_{\widetilde{\mathcal{H}}} > C(\mathtt{M}_{\widetilde{\mathcal{H}}} J(\delta, \widetilde{\mathcal{H}}, \mathtt{M}_{\widetilde{\mathcal{H}}}) + \delta \mathtt{M}_{\widetilde{\mathcal{H}}}\sqrt{t})\right] \leq \exp(-t),$$

where

$$\widetilde{\mathtt{F}}_n(t,\delta) := J(\delta, \widetilde{\mathcal{H}}, \mathtt{M}_{\widetilde{\mathcal{H}}})\mathtt{M}_{\widetilde{\mathcal{H}}} + \frac{\log(n)\mathtt{M}_{\widetilde{\mathcal{H}}} J^2(\delta, \widetilde{\mathcal{H}}, \mathtt{M}_{\widetilde{\mathcal{H}}})}{\delta^2\sqrt{n}} + \delta\mathtt{M}_{\widetilde{\mathcal{H}}}\sqrt{t} + \frac{\mathtt{M}_{\widetilde{\mathcal{H}}}}{\sqrt{n}} t.$$

The result then follows from the relation between $\mathcal{H}$ quantities and $\widetilde{\mathcal{H}}$ quantities in Lemma SA.12 and the decomposition that

$$\|X_n - Z_n^X\|_{\mathcal{H}} = \|\widetilde{X}_n - Z_n^X\|_{\widetilde{\mathcal{H}}} \leq \|\widetilde{X}_n - \widetilde{X}_n \circ \pi_{\widetilde{\mathcal{H}}_\delta}\|_{\widetilde{\mathcal{H}}} + \|Z_n^X - Z_n^X \circ \pi_{\widetilde{\mathcal{H}}_\delta}\|_{\widetilde{\mathcal{H}}}$$
$$+ \|\widetilde{X}_n - \widetilde{X}_n \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta} + \|Z_n^X - Z_n^X \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta}$$
$$+ \|\widetilde{X}_n \circ \Pi_0 - Z_n^X \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta},$$

where we have abused the notation to mean the same thing by $Z_n^X(h)$ and $Z_n^X(\widetilde{h})$. □

SA-II.19 Proof of Theorem 2

Suppose $2^K \leq L < 2^{K+1}$. For each $l \in [d]$, we can divide at most 2^K cells into two intervals of equal measure under $\mathbb{P}_X$ such that we get a new partition of $\mathcal{X} = \sqcup_{0 \leq j < 2^{K+1}} \Delta_l'$ and satisfies

$$\frac{\max_{0\leq l<2^{K+1}} \mathbb{P}_X(\Delta_l')}{\min_{0\leq l<2^{K+1}} \mathbb{P}_X(\Delta_l')} \leq 2\rho.$$

By construction, there exists an axis-aligned quasi-dyadic expansion $\mathcal{A}_{K+1}(\mathbb{P}_X, 2\rho) = \{\mathcal{C}_{j,k} : 0 \leq j \leq K+1, 0 \leq k < 2^{K+1-j}\}$ such that

$$\left\{\mathcal{C}_{0,k} : 0 \leq k < 2^{K+1}\right\} = \left\{\Delta'_l : 0 \leq l < 2^{K+1}\right\},$$

and $\mathcal{H} \subseteq \mathrm{Span}\{\mathbb{1}_{\Delta_j} : 0 \leq j < L\} \subseteq \mathrm{Span}\{\mathcal{C}_{0,k} : 0 \leq k < 2^{K+1}\}$. Now we consider the term $\mathtt{C}_{\mathcal{H}}$ from Lemma SA.7. Let $h \in \mathcal{H}$. By definition of S and the step of splitting each cell into at most two, there exists $l_1, \cdots, l_{2S} \in \{0, \cdots, 2^{K+1}-1\}$ such that $h = \sum_{q=1}^{2S} c_q \mathbb{1}(\Delta'_{l_q})$ where $|c_q| \leq \mathtt{M}_{\{h\}}$. Fix (j,k). Let (l,m) be an index such that $\mathcal{C}_{l,m} \subseteq \mathcal{C}_{j,k}$. Since each Δ'_{l_q} belongs to at most one $\mathcal{C}_{l-1,k}$, $\widetilde{\beta}_{l,m}(\mathbb{1}(\Delta'_{l_q})) = 0$ if Δ'_{l_q} is not contained in $\mathcal{C}_{l,m}$ and $\widetilde{\beta}_{l,m}(\mathbb{1}(\Delta'_{l_q})) = 2^{-l+1}$ if $\Delta'_{l_q} \subseteq \mathcal{C}_{l,m}$. Hence

$$\sum_{m:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}} \left|\widetilde{\beta}_{l,m}(h)\right|^2 \leq 2S\sum_{q=1}^{2S}\sum_{m:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}} \left(c_q\widetilde{\beta}_{l,k}(\mathbb{1}(\Delta_{l_q}))\right)^2 \leq 2S\sum_{q=1}^{2S} c_q^2 2^{-2l} \leq 4S^2\mathtt{M}_{\mathcal{H}}^2 2^{-2l}.$$

It follows that

$$\mathtt{C}_{\mathcal{H}} = \sup_{h\in\mathcal{H}}\min\left\{\sup_{(j,k)}\left[\sum_{l<j}(j-l)(j-l+1)2^{l-j}\sum_{m:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}}\widetilde{\beta}_{l,m}^2(h)\right], \mathtt{M}_{\mathcal{H}}^2(K+1)\right\} \lesssim \mathtt{M}_{\mathcal{H}}^2\min\{K, S^2\}.$$

Then apply Lemma SA.7, we get there exists a mean-zero Gaussian process Z_n^X with the same covariance structure as X_n such that with probability at least $1 - 2\exp(-t) - 2^{K+1}\exp(-C_\rho n 2^{-K-1})$,

$$\begin{aligned}\|X_n - Z_n^X\|_{\mathcal{H}} \leq \min_{\delta\in(0,1)}\Bigg\{&C_\rho\sqrt{\frac{2^{K+2}\mathtt{M}_{\mathcal{H}}\mathtt{E}_{\mathcal{H}}}{n}(t+\log N_{\mathcal{H}}(\delta))} + C_\rho\sqrt{\frac{\min\{K,S^2\}}{n}}\mathtt{M}_{\mathcal{H}}(t+\log N_{\mathcal{H}}(\delta))\\ &+ F_n(t,\delta)\Bigg\},\end{aligned}$$

where $K \leq \log_2(L)$. □

SA-II.20 Proof of Corollary SA.1

Take $\delta = n^{-1/2}$. Under the VC-type class condition, $\log \mathtt{N}_{\mathcal{H}}(n^{-1}) \leq \log(\mathtt{c}_{\mathcal{H}}) + \mathtt{d}_{\mathcal{H}}\log(n) \leq \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}} n)$, where the last inequality holds since $\mathtt{c}_{\mathcal{H}} \geq e$ and $\mathtt{d}_{\mathcal{H}} > 0$. This gives

$$\mathsf{A}_n(t, n^{-1/2}) \leq \mathtt{m}_{n,d}\sqrt{d\mathtt{c}_1(t+\mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n))\mathtt{M}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}}} + \min\left\{\sqrt{\log(n)\mathtt{M}_{\mathcal{H}}}, \sqrt{d^3\mathtt{c}_3\mathtt{K}_{\mathcal{H}}}\right\}\sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}(t+\mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n)).$$

Moreover, $J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}}) \leq \int_0^\delta \sqrt{1+\mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}\varepsilon^{-1})}d\varepsilon \leq 3\delta\sqrt{\mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}/\delta)}$. It follows that

$$\mathsf{F}_n(t, n^{-1/2}) \leq \frac{3\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}\mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n) + \frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}(\sqrt{t}+t).$$

The result then follows from Theorem 1. □

SA-II.21 Proof of Corollary SA.2

The result follows by taking $\delta = n^{-1/2}$ and apply Theorem 1, with calculations similar to Corollary SA.1. □

SA-II.22 Proof of Corollary SA.3

Under the polynomial entropy condition, $\log \mathtt{N}_{\mathcal{H}}(\delta) \leq \mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}}$, $J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}}) \leq \sqrt{\mathtt{a}_{\mathcal{H}}}(2-\mathtt{b}_{\mathcal{H}})^{-1}\delta^{-\mathtt{b}_{\mathcal{H}}/2+1}$,

$$\mathsf{A}_n(t,\delta) \leq \min\{\mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\mathcal{H}}}, \mathsf{l}_{n,d}\sqrt{\mathtt{c}_2\mathtt{L}_{\mathcal{H}}}\}\sqrt{\mathtt{TV}_{\mathcal{H}}(t+\mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}})} + \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{d^3\mathtt{c}_3\mathtt{K}_{\mathcal{H}}+\mathtt{M}_{\mathcal{H}}}\}(t+\mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}}),$$

$$\mathsf{F}_n(t,\delta) \leq \mathtt{a}_{\mathcal{H}}(2-\mathtt{b}_{\mathcal{H}})^{-2}\left(\mathtt{M}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}/2+1} + \frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}\delta^{-\mathtt{b}_{\mathcal{H}}} + \delta\mathtt{M}_{\mathcal{H}}\sqrt{t} + \frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}t\right).$$

Notice that the two terms $\frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}\delta^{-\mathtt{b}_{\mathcal{H}}}$ and $\frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}t$ in $\mathsf{F}_n(t,\delta)$ are dominated by terms in $\mathsf{A}_n(t,\delta)$. And when $\delta \leq n^{-1/2}$, the third term $\delta\mathtt{M}_{\mathcal{H}}\sqrt{t}$ is also dominated by terms in $\mathsf{A}_n(t,\delta)$. To choose δ that balance A_n and F_n, we consider the following three cases:

Case 1: Choose δ such that $\mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}}} \asymp \mathtt{M}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}/2+1}$. Notice that this choice also makes $\delta\mathtt{M}_{\mathcal{H}}\sqrt{t} \leq \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{d^3\mathtt{c}_3\mathtt{K}_{\mathcal{H}}+\mathtt{M}_{\mathcal{H}}}\}(t+\mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}})$. Plug in $\delta_* = \mathsf{m}_{n,d}\sqrt{\mathtt{TV}_{\mathcal{H}}/\mathtt{M}_{\mathcal{H}}}$ into A_n, we get $\mathsf{A}_n(t,\delta_*) + \mathsf{F}_n(t,\delta_*) \leq \mathsf{S}_n^{bdd}(t)$.

Case 2: Choose δ such that $\mathsf{l}_{n,d}\sqrt{\mathtt{L}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}}} \asymp \mathtt{M}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}/2+1}$. Again, this choice of δ makes $\delta\mathtt{M}_{\mathcal{H}}\sqrt{t} \leq \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{d^3\mathtt{c}_3\mathtt{K}_{\mathcal{H}}+\mathtt{M}_{\mathcal{H}}}\}(t+\mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}})$. Plug in $\delta_* = \mathsf{l}_{n,d}\sqrt{\mathtt{L}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}}/\mathtt{M}_{\mathcal{H}}^2}$ into A_n, we get $\mathsf{A}_n(t,\delta_*) + \mathsf{F}_n(t,\delta_*) \leq \mathsf{S}_n^{lip}(t)$.

Case 3: Choose δ such that $\mathtt{M}_{\mathcal{H}}n^{-1/2}\delta^{-\mathtt{b}_{\mathcal{H}}} \asymp \mathtt{M}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}/2+1}$. Plug in $\delta_* = n^{-1/(\mathtt{b}_{\mathcal{H}}+2)}$, we get $\mathsf{A}_n(t,\delta_*) + \mathsf{F}_n(t,\delta_*) \leq \mathsf{S}_n^{err}(t)$.

□

SA-II.23 Proof of Corollary 1

The result follows from Corollary SA.1, taking $t = \log n$. □

SA-II.24 Proof of Corollary 2

The result follows from Corollary SA.2, taking $t = \log n$. □

SA-II.25 Proof of Corollary 3

The result follows from Corollary SA.3, taking $t = \log n$.

□

SA-II.26 Proof of Corollary 4

The result follows from Theorem 2, taking $\delta = n^{-\frac{1}{2}}$ and $t = \log n$.

□

SA-II.27 Proof of Example 1

Define $\mathcal{H} = \{h_{\mathbf{x}} : \mathbf{x} \in \mathcal{X}\}$ where $h_{\mathbf{x}}(\cdot) := b^{-\frac{d}{2}}K(b^{-1}(\mathbf{x}-\cdot))$. Since K is compactly supported and Lipschitz, $\|K\|_\infty < \infty$. Hence $\mathtt{M}_{\mathcal{H}} = b^{-\frac{d}{2}}\|K\|_\infty \lesssim b^{-d/2}$ and $\mathtt{L}_{\mathcal{H}} \leq b^{-\frac{d}{2}-1}\mathtt{L}_{\{K\}} \lesssim b^{-\frac{d}{2}-1}$. Since $\sup_{\mathbf{x}\in\mathcal{X}} \mathrm{Vol}(\mathrm{supp}(h_{\mathbf{x}})) \lesssim$

b^d and each $h_{\mathbf{x}}$ is differentiable, $\mathtt{TV}_{\mathcal{H}} \lesssim \sup_{\mathbf{x}\in\mathcal{X}} \mathrm{Vol}(\mathrm{supp}(h_{\mathbf{x}}))\mathtt{L}_{\mathcal{H}} \lesssim b^{\frac{d}{2}-1}$. To upper bound $\mathtt{K}_{\mathcal{H}}$, consider the following two cases: (i) If $a < b$, then

$$\sup_{\mathbf{x}\in\mathcal{X}} \sup_{\phi\in\mathcal{D}_d(\mathcal{C})} \int h_{\mathbf{x}}(\mathbf{u}) \operatorname{div}(\phi)(\mathbf{u}) d\mathbf{x} / \|\|\phi\|_2\|_\infty \leq \mathtt{L}_{\{h_{\mathbf{x}}\}} \lesssim \mathrm{Vol}(\mathcal{C})\mathtt{L}_{\mathcal{H}} \lesssim b^{-\frac{d}{2}-1} a^d \lesssim b^{-d/2} a^{d-1}.$$

(ii) If $a > b$, then

$$\sup_{\mathbf{x}\in\mathcal{X}} \sup_{\phi\in\mathcal{D}_d(\mathcal{C})} \int h_{\mathbf{x}}(\mathbf{u}) \operatorname{div}(\phi)(\mathbf{u}) d\mathbf{x} / \|\|\phi\|_2\|_\infty \lesssim \sup_{\mathbf{x}\in\mathcal{X}} \mathrm{Vol}(\mathrm{Supp}(h_{\mathbf{x}}))\mathtt{L}_{\mathcal{H}} \lesssim b^d b^{-\frac{d}{2}-1} \lesssim b^{-\frac{d}{2}} b^{d-1} \lesssim b^{-\frac{d}{2}} a^{d-1}.$$

This shows $\mathtt{K}_{\mathcal{H}} \lesssim b^{-\frac{d}{2}}$. Next, by a change of variable,

$$\mathtt{E}_{\mathcal{H}} = \sup_{\mathbf{x}\in\mathcal{X}} \int_{\mathbb{R}^d} b^{-\frac{d}{2}} |K(b^{-1}(\mathbf{x}-\mathbf{u}))| f_X(\mathbf{u}) d\mathbf{u} = \sup_{\mathbf{x}\in\mathcal{X}} \int_{\mathbb{R}^d} b^{-\frac{d}{2}} |K(\mathbf{z})| f_X(\mathbf{x} - h\mathbf{z}) b^d d\mathbf{z} \lesssim b^{d/2}.$$

Now define $g_{\mathbf{x}}(\cdot) = b^{-\frac{d}{2}} \mathtt{M}_{\mathcal{H}}^{-1} K(\cdot)$ for all $\mathbf{x} \in \mathcal{X}$. Then $\mathtt{M}_{\mathcal{H}}^{-1}\mathcal{H} = \{g_{\mathbf{x}}(\frac{\mathbf{x}-\cdot}{b}) : \mathbf{x} \in \mathcal{X}\}$. Then there exists a constant $\mathbf{c}_K$ only depending on $\|K\|_\infty$, $\mathtt{L}_{\{K\}}$ that

$$\sup_{\mathbf{x}\in\mathcal{X}} \|g_{\mathbf{x}}\|_\infty \leq \mathbf{c}_K,$$
$$\sup_{\mathbf{x}\in\mathcal{X}} \sup_{\mathbf{u},\mathbf{v}\in\mathcal{X}} \frac{|g_{\mathbf{x}}(\mathbf{u}) - g_{\mathbf{x}}(\mathbf{v})|}{\|\mathbf{u}-\mathbf{v}\|_\infty} \leq \mathbf{c}_K,$$
$$\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}} \sup_{\mathbf{u}\in\mathcal{X}} \frac{|g_{\mathbf{x}}(\mathbf{u}) - g_{\mathbf{y}}(\mathbf{u})|}{\|\mathbf{x}-\mathbf{y}\|_\infty} \leq \mathbf{c}_K,$$

we can apply Lemma 7 from Cattaneo *et al.* (2024), which is modified upon Lemma 4.1 from Rio (1994), to show that for all $0 < \varepsilon < 1$,

$$N(\varepsilon, \mathtt{M}_{\mathcal{H}}^{-1}\mathcal{H}) \leq \mathbf{c}_K \varepsilon^{-d-1} + 1.$$

Then, by Theorem 1, on a possibly enlarged probability space, $(\xi_n(\mathbf{x}) : \mathbf{x} \in \mathcal{X})$ admits a Gaussian strong approximation with rate function

$$S_n(t) = \mathsf{s}_n \sqrt{t + (d+1)\log n} + \sqrt{\frac{\log n}{nh^d}} (t + (d+1)\log n).$$

To leverage the Lipschitz conditions, observe that

$$\mathtt{l}_{\mathcal{H}} = b^{-\frac{d}{2}} \sup_{\mathbf{x}\in\mathcal{X}} \sup_{\mathbf{u},\mathbf{v}\in\mathcal{X}} \frac{|K(b^{-1}(\mathbf{x}-\mathbf{u})) - K(b^{-1}(\mathbf{x}-\mathbf{v}))|}{\|\mathbf{u}-\mathbf{v}\|_\infty} \lesssim b^{-\frac{d}{2}-1}.$$

The result then follows from Corollary 2. □

SA-II.28 Proof of Example 2

Define a kernel function $k(\cdot,\cdot) : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ by

$$k(\mathbf{u}, \mathbf{x}) = \sqrt{J} \sum_{0 \leq l < J} \mathbb{1}(\mathbf{u} \in \Delta_l) \mathbb{1}(\mathbf{x} \in \Delta_l), \mathbf{u}, \mathbf{x} \in \mathcal{X}.$$

Define $\mathcal{K} = \{k(\cdot, \mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$. Then $\text{Card}(\mathcal{K}) \leq J$ and

$$\begin{aligned}
&\mathtt{M}_{\mathcal{K}} \leq \sqrt{J},\\
&\mathtt{E}_{\mathcal{K}} \leq \max_{0 \leq l < J} \mathbb{P}_X(\Delta_l) \cdot \mathtt{M}_{\mathcal{K}} \leq \rho J^{-1}\sqrt{J} \leq \rho J^{-1/2}.
\end{aligned}$$

Moreover, each function in $\mathcal{K}$ can be written $c\mathbb{1}(\Delta_l)$ for some $l \leq J$, which implies we can take $S = 1$. The result then follows from Theorem 2. □

Lemma SA.13 (Product of VC-classes is a VC-class)**.** *Suppose $\mathcal{F}$ and $\mathcal{S}$ are classes of functions from a measurable space $(\mathcal{X}, \mathcal{B})$ to $\mathbb{R}$ with envelope functions $M_{\mathcal{F}}$ and $M_{\mathcal{S}}$, respectively. Then*

$$\begin{aligned}
\sup_Q N(\mathcal{F} \times \mathcal{S}, M_{\mathcal{F}} M_{\mathcal{S}}, \delta) &\leq \mathtt{N}_{\mathcal{F}}(\delta/2)\mathtt{N}_{\mathcal{S}}(\delta/2), \quad \forall 0 < \delta < 1,\\
J(\mathcal{F} \times \mathcal{S}, M_{\mathcal{F}} M_{\mathcal{S}}, \delta) &\leq \sqrt{2}J(\mathcal{F}, M_{\mathcal{F}}, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{S}, M_{\mathcal{S}}, \delta/\sqrt{2}),
\end{aligned}$$

where $\sup_P$ and $\sup_Q$ are taken over all finite discrete measures on $\mathcal{X}$.

Proof. Let $f, f_1 \in \mathcal{F}$ and $s, s_1 \in \mathcal{S}$. Let $\mathcal{Q}$ be a finite discrete measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

$$\begin{aligned}
&\int |f_1 s_1 - f_2 s_2|^2 d\mathcal{Q} \leq \int |f_1 - f_2|^2 M_S^2 d\mathcal{Q} + \int |s_1 - s_2|^2 M_F^2 d\mathcal{Q}\\
= &\int |f_1 - f_2|^2 d\mathcal{Q}_S \int M_S^2 d\mathcal{Q} + \int |s_1 - s_2|^2 d\mathcal{Q}_F \int M_F^2 d\mathcal{Q},
\end{aligned}$$

where $d\mathcal{Q}_S = M_S^2 d\mathcal{Q} / \int M_S^2 d\mathcal{Q}$ and $d\mathcal{Q}_F = M_F^2 d\mathcal{Q} / \int M_F^2 d\mathcal{Q}$. Take $\mathcal{F}_{\varepsilon\|M_F\|_{\mathcal{Q}_S,2}}$ and $\mathcal{S}_{\varepsilon\|M_S\|_{\mathcal{Q}_F,2}}$ to be $\varepsilon\|M_F\|_{\mathcal{Q},2}$-net of $\mathcal{F}$ and $\varepsilon\|M_S\|_{\mathcal{Q},2}$-net of $\mathcal{S}$ with minimal cardinality. Then for any $f \in \mathcal{F}, s \in \mathcal{S}$, there exists $f_0 \in \mathcal{F}_{\varepsilon\|F\|_{\mathcal{Q}_S,2}}$ and $s_0 \in \mathcal{S}_{\varepsilon\|S\|_{\mathcal{Q}_F,2}}$ such that $\|f - f_0\|_{\mathcal{Q}_S,2}^2 \leq \varepsilon^2\|M_F\|_{\mathcal{Q}_S,2}^2$ and $\|s - s_0\|_{\mathcal{Q}_F,2}^2 \leq \varepsilon^2\|M_S\|_{\mathcal{Q}_F,2}^2$. Hence $\|fs - f_0 s_0\|_{Q,2}^2 \leq 2\varepsilon\|M_F M_S\|_{Q,2}^2$. It follows that

$$\begin{aligned}
J(\mathcal{F} \times \mathcal{S}, M_F M_S, \delta) &\leq \int_0^\delta \sqrt{1 + \log \sup_{\mathcal{Q}} N(\mathcal{F}, \|\cdot\|_{\mathcal{Q},2}, \varepsilon\|M_F\|_{\mathcal{Q},2}/\sqrt{2}) + \log \sup_{\mathcal{Q}} N(\mathcal{M}_{\mathcal{S}}, \|\cdot\|_{\mathcal{Q},2}, \varepsilon\|S\|_{\mathcal{Q},2}/\sqrt{2})} d\varepsilon\\
&\leq \sqrt{2}J(\mathcal{F}, M_F, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{S}, M_S, \delta/\sqrt{2}).
\end{aligned}$$

□

Lemma SA.14 (Covering Number using Covariance Semi-metric)**.** *Assume $\mathcal{F}$ is a class of functions from a measurable space $(\mathcal{X}, \mathcal{B})$ to $\mathbb{R}$ with envelope function $M_{\mathcal{F}}$. Let P be any probability measure on $(\mathcal{X}, \mathcal{B})$. Then for any $0 < \varepsilon < 1$,*

$$N(\mathcal{F}, \|\cdot\|_{P,2}, \varepsilon\|M_{\mathcal{F}}\|_{P,2}) \leq \mathtt{N}_{\mathcal{F}}(\varepsilon).$$

Proof. Let $X_1, X_2, \ldots$ be a sequence of i.i.d random variables with distribution P. Define $\mathcal{Q}_N = \frac{1}{N}\sum_{j=1}^N \delta_{X_j}$. Define $\mathcal{H} = \{(f - g)^2 : f, g \in \mathcal{F}\} \cup \{M_F\}$. Then for all $0 < \varepsilon < 1$,

$$\sup_{\mathcal{Q}} N(\mathcal{H}, \|\cdot\|_{\mathcal{Q},1}, \varepsilon\|M_F^2\|_{\mathcal{Q},1}) \leq \sup_{\mathcal{Q}} N(\mathcal{H}, \|\cdot\|_{\mathcal{Q},1}, \varepsilon\|M_F^2\|_{\mathcal{Q},2}) \leq \sup_{\mathcal{Q}} N(\mathcal{F}, \|\cdot\|_{\mathcal{Q},1}, \varepsilon\|M_F\|_{\mathcal{Q},1})^2.$$

By Theorem 2.4.3 in van der Vaart and Wellner (2013), $\mathcal{H}$ is Glivenko-Cantelli. Let $0 < \varepsilon < 1$ and $\delta > 0$. Then there exists $N \in \mathbb{N}$ and a realization $x_1, \ldots, x_N$ of $X_1, \ldots, X_N$ such that if we denote $P_N = \frac{1}{N}\sum_{i=1}^N \delta_{x_i}$, then for all $f_1, f_2 \in \mathcal{F}$,

$$\begin{aligned}
&\big|\|f_1 - f_2\|_{P,2}^2 - \|f_1 - f_2\|_{P_n,2}^2\big| \leq \delta^2 \varepsilon^2 \|M_F\|_{P,2}^2,\\
&|\|M_F\|_{P,2} - \|M_F\|_{P_n,2}| \leq \delta \|M_F\|_{P,2}.
\end{aligned}$$

Since $P_n \in \mathcal{A}(\mathcal{X})$, there exists $\varepsilon\|M_F\|_{P_n}$-net, $\mathcal{G}$, of $\mathcal{F}$ with minimal cardinality such that for all $f \in \mathcal{F}$, there exists $f_0 \in \mathcal{F}$ such that $\|f - f_0\|_{P_n,2} \leq \varepsilon\|M_F\|_{P_n,2} \leq \varepsilon(\|M_F\|_{P,2} + \delta\|M_F\|_{P,2}) \leq (1+\delta)\varepsilon\|M_F\|_{P,2}$. It follows that for all $f \in \mathcal{F}$, there exists $g \in \mathcal{G}$ such that

$$\|f - g\|_{P,2} \leq \|f - g\|_{P_n,2} + |\|f - g\|_{P,2} - \|f - g\|_{P_n,2}| \leq (1 + 2\delta)\varepsilon\|M_F\|_{P,2},$$

Hence

$$N(\mathcal{F}, \|\cdot\|_{P,2}, \varepsilon\|M_F\|_{P,2}) \leq \sup_{\mathcal{Q}} N(\mathcal{F}, \|\cdot\|_{\mathcal{Q},2}, \varepsilon\|M_F\|_{\mathcal{Q},2}/(1 + 2\delta)).$$

Take $\delta \to 0$ and we get the desired results. □

SA-III Multiplicative-Separable and Residual-Based Empirical Process: Proofs

Assumption SA.1. *Suppose Assumption B holds with $\mathcal{X} = [0,1]^d$. Denote by $\mathbb{P}$ the joint distribution of $(\mathbf{x}_i, y_i)$, $\mathbb{P}_X$ the marginal distribution of $\mathbf{x}_i$, $\mathbb{P}_Y$ the marginal distribution of y_i. Suppose the following two conditions hold.*

(i) *$\mathcal{G}$ is a real-valued pointwise measurable class of functions on $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$ such that $J(\mathcal{G}, \mathtt{M}_\mathcal{G}, 1) < \infty$.*

(ii) *$\mathcal{R}$ be a real-valued pointwise measurable class of functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$ such that $J(\mathcal{R}, M_\mathcal{R}, 1) < \infty$. Furthermore, one of the following holds:*

   (a) *$\mathtt{M}_\mathcal{R} \lesssim 1$ and $\mathtt{pTV}_\mathcal{R} \lesssim 1$, and set $\alpha = 0$, or*

   (b) *$M_\mathcal{R}(y) \lesssim 1 + |y|^\alpha$, $\mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \lesssim 1 + |y|^\alpha$ for all $y \in \mathbb{R}$ and for some $\alpha > 0$, and $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$.*

Assumption SA.2. *Suppose $((\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ are i.i.d. random vectors taking values in $(\mathcal{X} \times \mathbb{R}, \mathcal{B}(\mathcal{X} \times \mathbb{R}))$, $\mathcal{X} \subseteq \mathbb{R}^d$. Denote by $\mathbb{P}$ the joint distribution of $(\mathbf{x}_i, y_i)$, $\mathbb{P}_X$ the marginal distribution of $\mathbf{x}_i$, $\mathbb{P}_Y$ the marginal distribution of y_i. Suppose the following conditions hold.*

(i) *$\mathcal{G}$ is a class of functions on $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$ such that $\mathtt{M}_\mathcal{G} < \infty$ and $\mathcal{G} \subseteq \mathrm{Span}\{\mathbb{1}_{\Delta_l} : 0 \leq l < L\}$, where $\{\Delta_l : 0 \leq l < L\}$ forms a quasi-uniform partition of $\mathcal{X}$ in the sense that*

$$\mathcal{X} \subseteq \sqcup_{0 \leq l < L} \Delta_l \qquad \text{and} \qquad \frac{\max_{0 \leq l < L} \mathbb{P}_X(\Delta_l)}{\min_{0 \leq l < L} \mathbb{P}_X(\Delta_l)} \leq \rho < \infty.$$

   In addition, $J(\mathcal{G}, \mathtt{M}_\mathcal{G}, 1) < \infty$.

(ii) *$\mathcal{R}$ is a real-valued pointwise measurable class of functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$, such that $J(\mathcal{R}, M_\mathcal{R}, 1) < \infty$. Furthermore, one of the following holds:*

(a) $\mathtt{M}_{\mathcal{R}} \lesssim 1$ *and* $\mathtt{pTV}_{\mathcal{R}} \lesssim 1$*, and set* $\alpha = 0$*, or*

(b) $M_{\mathcal{R}}(y) \lesssim 1+|y|^\alpha$, $\mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \lesssim 1+|y|^\alpha$ *for all* $y \in \mathbb{R}$ *and for some* $\alpha > 0$*, and* $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

SA-III.1 Cell Expansions

Definition SA.3 (Cylindered Quasi-Dyadic Expansion of $\mathbb{R}^d$)**.** *Denote by* $\mathbb{P}$ *the joint distribution of* (X, Y)*. Let* $\rho \geq 1$*. A collection of Borel measurable sets in* $\mathbb{R}^{d+1}$*,* $\mathfrak{C}_{M,N}(\mathbb{P}, \rho) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{M+N-j}, 0 \leq j \leq M + N\}$ *is called a* cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$, with depth M for the main subspace $\mathbb{R}^d$ and depth N for the multiplier subspace $\mathbb{R}$, with respect to $\mathbb{P}$, the joint distribution of a random vector (X, Y) taking values in $\mathbb{R}^d \times \mathbb{R}$*, if the following two conditions hold:*

1. *For all* $N \leq j \leq M+N$*,* $0 \leq k < 2^{M+N-j}$*, there exists a set* $\mathcal{X}_{j-N,k} \subseteq \mathbb{R}^d$ *such that* $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathbb{R}$*. Moreover, the class of projected cells onto the main subspace* $\mathbb{R}^d$*,* $\mathtt{p}_X[\mathfrak{C}_{M,N}(\mathbb{P}, \rho)] := \{\mathcal{X}_{l,k} : 0 \leq l \leq M, 0 \leq k < 2^{M-l}\}$*, forms a quasi-dyadic expansion of* $\mathbb{R}^d$ *of depth* M *with respect to* $\mathbb{P}_X$*, the marginal distribution of* X*.*

2. *For all* $0 \leq j < N$*,* $0 \leq k < 2^{M+N-j}$*, take* l, m *to be the unique non-negative integers such that* $k = 2^{N-j}l + m$*, then there exists* $\mathcal{Y}_{l,j,m} \subseteq \mathbb{R}$ *such that* $\mathcal{C}_{j,k} = \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}$*. Moreover, for each* $0 \leq l < 2^M$*,* $\{\mathcal{Y}_{l,j,m} : 0 \leq j \leq N, 0 \leq m < 2^{N-j}\}$ *forms a dyadic expansion of* $\mathbb{R}$ *with respect to the measure* $\mathbb{P}(Y \in \cdot|\mathbf{X} \in \mathcal{X}_{0,l})$*.*

When $\rho = 1$*,* $\mathcal{C}_{M,N}(\mathbb{P}, 1)$ *is called a* cylindered dyadic expansion*.*

Definition SA.4 (Axis-Aligned Quasi-Dyadic Expansion of $\mathbb{R}^d$)**.** *A collection of Borel measurable sets in* $\mathbb{R}^{d+1}$*,* $\mathcal{A}_{M,N}(\mathbb{P}, \rho) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{M+N-j}, 0 \leq j \leq M + N\}$*,* $\rho \geq 1$*, is called an* axis-aligned cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$, with depth M for the main subspace $\mathbb{R}^d$ and depth N for the multiplier subspace $\mathbb{R}$, with respect to $\mathbb{P}$, the joint distribution of (X, Y) taking values in $\mathbb{R}^d \times \mathbb{R}$*, if the following two conditions hold:*

1. $\mathcal{A}_{M,N}(\mathbb{P}, \rho)$ *is a cylindered quasi-dyadic expansion of* $\mathbb{R}^{d+1}$*, with depth* M *for the main subspace* $\mathbb{R}^d$ *and depth* N *for the multiplier subspace* $\mathbb{R}$*, with respect to* $\mathbb{P}$*.*

2. $\mathtt{p}_X[\mathcal{A}_{M,N}(\mathbb{P}, \rho)] := \{\mathcal{X}_{l,k} : 0 \leq l \leq M, 0 \leq k < 2^{M-l}\}$ *forms an axis-aligned quasi-dyadic expansion of* $\mathbb{R}^d$ *of depth* M *with respect to* $\mathbb{P}_X$*, the marginal distribution of* X*.*

When $\rho = 1$*,* $\mathcal{A}_{M,N}(\mathbb{P}, 1)$ *is called an* axis-aligned cylindered dyadic expansion*.*

SA-III.2 Projection onto Piecewise Constant Functions

Due to the multiplicative-separable structure of $g(\mathbf{x}_i)r(y_i)$, we tailor a mapping other than L_2 projection from the space $L_2(\mathbb{R}^{d+1})$ to the space of piecewise constant functions on $\{\mathcal{C}_{0,k} : 0 \leq k < 2^{M+N}\}$, calling it the *product-factorized projection*. This is a technical point that makes the analysis in Lemma SA.19 easier.

First, we define the "projections". For a cylindered quasi-dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}, \rho)$ where $\mathbb{P}$ is the joint distribution of (X, Y), the *product-factorized projection* from $L_2(\mathbb{R}^{d+1})$ to $\mathcal{E}_{M+N} :=$ Span$\{\mathcal{C}_{0,k} = \mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m} : 0 \leq l < 2^M, 0 \leq m < 2^N, k = 2^N l + m\}$ is given by

$$\Pi_1(\mathcal{C}_{M,N}(\mathbb{P}, \rho))[g, r] := \gamma_{M+N,0}(g, r)e_{M+N,0} + \sum_{1\leq j\leq M+N} \sum_{0\leq k<2^{M+N-j}} \widetilde{\gamma}_{j,k}(g, r)\widetilde{e}_{j,k}, \tag{SA-7}$$

where $e_{j,k} = \mathbb{1}(\mathcal{C}_{j,k})$ and $\widetilde{e}_{j,k} = \mathbb{1}(\mathcal{C}_{j-1,2k}) - \mathbb{1}(\mathcal{C}_{j-1,2k+1})$ and

$$\gamma_{j,k}(g,r) = \begin{cases} \mathbb{E}[g(X)r(Y)|X \in \mathcal{X}_{j-N,k}], & \text{if } N \leq j \leq M+N, \\ \mathbb{E}[g(X)|X \in \mathcal{X}_{0,l}] \cdot \mathbb{E}[r(Y)|X \in \mathcal{X}_{0,l}, Y \in \mathcal{Y}_{l,0,m}], & \text{if } j < N, k = 2^{N-j}l + m, \end{cases}$$

and $\widetilde{\gamma}_{j,k}(g,r) = \gamma_{j-1,2k}(g,r) - \gamma_{j-1,2k+1}(g,r)$. We will use Π_1 as a shorthand for $\Pi_1(\mathcal{C}_{M,N}(\mathbb{P},\rho))$. The Haar basis representation on the right hand side of Equation SA-7 recovers the left hand side by adding up layers of more and more local fluctuation. However, at the bottom layers ($1 \leq j \leq N$), the local fluctuation is characterized by a *product-factorized projection* $\mathbb{E}[g(X)|X \in \mathcal{X}_{0,l}] \cdot \mathbb{E}[r(Y)|X \in \mathcal{X}_{0,l}, Y \in \mathcal{Y}_{l,0,m}]$, instead of $\mathbb{E}[g(X)r(Y)|X \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}]$. This makes $\Pi_1(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r]$ in general different from $\Pi_0(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g \cdot r]$.

For the residual empirical process, we define a new projection that adds up the *product-factorized projection* for $g \cdot r$ and the L_2-projection for $g \cdot \theta(\cdot, r)$: For all $(g,r) \in L_2(\mathbb{R}^d) \times L_2(\mathbb{R})$,

$$\Pi_2(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r] := \Pi_1(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r] - \Pi_0(\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)])[g\theta(\cdot,r)], \tag{SA-8}$$

recalling that $\theta(\mathbf{x}, r) = \mathbb{E}[r(Y)|X = \mathbf{x}], \mathbf{x} \in \mathbb{R}^d$. This projection can also be represented in Haar basis as

$$\Pi_2(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r] = \eta_{M+N,0}(g,r)e_{M+N,0} + \sum_{1 \leq j \leq M+N} \sum_{0 \leq k < 2^{M+N-j}} \widetilde{\eta}_{j,k}(g,r)\widetilde{e}_{j,k},$$

where for all $g \in L_2(\mathbb{R}^d)$, $r \in L_2(\mathbb{R})$,

$$\eta(j,k)(g,r) := \begin{cases} 0, & \text{if } N \leq j \leq M+N, \\ \gamma_{j,k}(g,r), & \text{if } j < N, k = 2^{N-j}l + m. \end{cases} \tag{SA-9}$$

We will use Π_2 as a shorthand for $\Pi_2(\mathcal{C}_{M,N}(\mathbb{P},\rho))$.

Now we define the empirical processes indexed by projected functions. By slightly abuse of notations, denote by $(X_n(f) : f \in \mathcal{F})$ the general empirical process based on random sample $((\mathbf{x}_i, y_i) : 1 \leq i \leq n)$, $\mathcal{F} \subseteq L_2(\mathbb{R}^{d+1})$. That is, $X_n(f) := n^{-1/2} \sum_{i=1}^n (f(\mathbf{x}_i, y_i) - \mathbb{E}[f(\mathbf{x}_i, y_i)]), f \in \mathcal{F}$. Then for any $g \in L_2(\mathbb{R}^d)$, $r \in L_2(\mathbb{R})$, we define

$$\begin{aligned} \Pi_1 M_n(g,r) &:= X_n \circ \Pi_1(g,r), \\ \Pi_0 M_n(g,r) &:= X_n \circ \Pi_0[\mathcal{C}_{M,N}(\mathbb{P},\rho)](gr), \\ \Pi_2 R_n(g,r) &:= X_n \circ \Pi_2(g,r), \\ \Pi_0 R_n(g,r) &:= X_n \circ \Pi_0[\mathcal{C}_{M,N}(\mathbb{P},\rho)](gr) - X_n \circ \Pi_0(\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)])[g\theta(\cdot,r)], \end{aligned} \tag{SA-10}$$

where $\Pi_0(\mathcal{C}_{M,N}(\mathbb{P},\rho))$ and $\Pi_0(\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)])$ are the L_2-projections based on cells $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ and $\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)]$, respectively (Equation SA-3).

SA-III.3 Strong Approximation Construction

Lemma SA.15. *Suppose Assumption SA.1 or Assumption SA.2 hold. Suppose $\mathcal{C}_{M,N}(\mathbb{P},\rho), \rho \geq 1$ is a cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$ of depth M in the dimension of $\mathbb{R}^d$ and depth N in the dimension of $\mathbb{R}$ with respect to $\mathbb{P}$. Then, $(\mathcal{G} \times \mathcal{R}) \cup (\mathcal{G} \times \mathcal{V}_{\mathcal{R}}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}) \cup \Pi_2(\mathcal{G} \times \mathcal{R}) \cup \mathcal{E}_{M+N}$ is $\mathbb{P}$-pregaussian.*

The construction essentially follows from the arguments in Section SA-II.3. We start with a Gaussian process indexed by $(\mathcal{G} \times \mathcal{R}) \cup (\mathcal{G} \times \mathcal{V}_{\mathcal{R}}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}) \cup \Pi_2(\mathcal{G} \times \mathcal{R}) \cup \mathcal{E}_{M+N}$ with almost sure continuous sample path, and take conditional quantile transformations of Gaussian process indexed by $\mathbb{1}_{\mathcal{C}_{j,k}}$ to construct counts of $(\mathbf{x}_i, y_i)$'s on the cells $\mathcal{C}_{j,k}$'s. By a Skorohod embedding argument, this Gaussian process can be taken on a possibly enriched probability space. More precisely, we have the following

Lemma SA.16. *Suppose Assumption SA.1 holds. Suppose $\rho = 1$. Then on a possibly enlarged probability space, there exists a Brownian bridge $\mathbb{B}_n$ indexed by $\mathcal{F} = (\mathcal{G} \times \mathcal{R}) \times (\mathcal{G} \times \mathcal{V}_{\mathcal{R}}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}) \cup \Pi_2(\mathcal{G} \times \mathcal{R}) \cup \mathcal{E}_{M+N}$ such that that is mean-zero with almost sure continuous sample paths such that*

$$\mathbb{E}[\mathbb{B}_n(f), \mathbb{B}_n(g)] = \mathbb{C}\mathrm{ov}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n f(\mathbf{x}_i, y_i), \frac{1}{\sqrt{n}}\sum_{i=1}^n g(\mathbf{x}_i, y_i)\right], \quad f, g \in \mathcal{F},$$

and for any finite class of functions $\mathcal{F} \subseteq \mathcal{E}_{M+N}$ and any $x > 0$,

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}}\left|\sum_{i=1}^n f(\mathbf{x}_i, y_i) - \sqrt{n} Z_n(f)\right| \geq 24\sqrt{\|f\|_{\mathcal{E}_{M+N}}^2 x} + 4\sqrt{\mathtt{C}_{\{f\}} x}\right) \leq 2\,\mathrm{Card}(\mathcal{F})\exp(-x),$$

where $\|f\|_{\mathcal{E}_{M+N}}$ and $\mathtt{C}_{\{f\}}$ are defined in Lemma SA.5.

Lemma SA.17. *Suppose Assumption SA.2 holds. Suppose $\rho > 1$. Then on a possibly enlarged probability space, there exists a Brownian bridge $\mathbb{B}_n$ indexed by $\mathcal{F} = (\mathcal{G} \times \mathcal{R}) \times (\mathcal{G} \times \mathcal{V}_{\mathcal{R}}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}) \cup \Pi_2(\mathcal{G} \times \mathcal{R}) \cup \mathcal{E}_{M+N}$ such that $\mathbb{B}_n$ is mean-zero with almost sure continuous sample paths such that*

$$\mathbb{E}[\mathbb{B}_n(f), \mathbb{B}_n(g)] = \mathbb{C}\mathrm{ov}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^n f(\mathbf{x}_i, y_i), \frac{1}{\sqrt{n}}\sum_{i=1}^n g(\mathbf{x}_i, y_i)\right], \quad f, g \in \mathcal{F},$$

and for any finite class of functions $\mathcal{F} \subseteq \mathcal{E}_{M+N}$ and any $x > 0$,

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}}\left|\sum_{i=1}^n f(\mathbf{x}_i, y_i) - \sqrt{n} Z_n(f)\right| \geq C_\rho\sqrt{\|f\|_{\mathcal{E}_{M+N}}^2 x} + C_\rho\sqrt{\mathtt{C}_{\{f\}} x}\right)$$
$$\leq 2\,\mathrm{Card}(\mathcal{F})\exp(-x) + 2^{M+2}\exp\left(-C_\rho n 2^{-M}\right),$$

where C_ρ is a constant that only depends on ρ.

The above two lemmas allow for constructions of Gaussian processes and projected Gaussian processes as counterparts of the empirical processes in Section SA-III.2. In particular, we take $Z_n^M, \Pi_1 Z_n^M$, Z_n^R, $\Pi_2 Z_n^R$ to be the empirical processes indexed by $\mathcal{G} \times \mathcal{R}$ such that for any $g \in \mathcal{G}$, $r \in \mathcal{R}$,

$$Z_n^M(g, r) := \mathbb{B}_n(gr), \quad \Pi_1 Z_n^M(g, r) := \mathbb{B}_n(\Pi_1[g, r]), \quad Z_n^R(g, r) := \mathbb{B}_n(g(r - \theta(\cdot, r))), \quad \Pi_2 Z_n^R(g, r) := \mathbb{B}_n(\Pi_2[g, r]).$$

SA-III.4 Meshing Error

For $0 < \delta \leq 1$, consider the $(\delta \mathtt{M}_{\mathcal{G} \times \mathcal{R}})$-net of $(\mathcal{G} \times \mathcal{R}, e_{\mathbb{P}})$, with cardinality no larger than $\mathtt{N}_{\mathcal{G} \times \mathcal{R}}(\delta)$: Define $\pi_{(\mathcal{G} \times \mathcal{R})_\delta} : \mathcal{G} \times \mathcal{R} \mapsto \mathcal{G} \times \mathcal{R}$ such that $\|\pi_{(\mathcal{G} \times \mathcal{R})_\delta}(h) - h\|_{\mathbb{P},2} \leq \delta \mathtt{M}_{\mathcal{G} \times \mathcal{R}}$ for all $h \in \mathcal{G} \times \mathcal{R}$, where $\mathbb{P}$ is the distribution of $(\mathbf{x}_1, y_1)$ satisfying Assumption B.

Lemma SA.18. *Suppose Assumption SA.1 or Assumption SA.2 hold. For all $t > 0$ and $0 < \delta < 1$,*

$$\mathbb{P}\left[\|M_n - M_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^M \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^M\|_{\mathcal{G}\times\mathcal{R}} > C_1 C_\alpha \mathsf{F}_n(t,\delta)\right] \leq \exp(-t),$$
$$\mathbb{P}\left[\|R_n - R_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^R \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^R\|_{\mathcal{G}\times\mathcal{R}} > C_1 C_\alpha \mathsf{F}_n(t,\delta)\right] \leq \exp(-t),$$

where $C_\alpha = 1 + (2\alpha)^{\frac{\alpha}{2}}$ and

$$\mathtt{F}_n(t,\delta) = J(\delta)\mathtt{M}_\mathcal{G} + \frac{(\log n)^{\alpha/2}\mathtt{M}_\mathcal{G} J^2(\delta)}{\delta^2\sqrt{n}} + \frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}} t + (\log n)^\alpha \frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}} t^\alpha.$$

SA-III.5 Strong Approximation Errors

Lemma SA.19. *Suppose Assumption SA.1 holds. Let $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ be a cylindered dyadic expansion with $\rho = 1$. Suppose $(\mathcal{G}\times\mathcal{R})_\delta$ is a δ-net of $(\mathcal{G}\times\mathcal{R}, e_\mathbb{P})$ with cardinality no greater than $\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)$. Then for all $t > 0$,*

$$\mathbb{P}\left[\|\Pi_1 M_n - \Pi_1 Z_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1 C_\alpha \sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_\mathcal{G}\mathtt{M}_\mathcal{G}}{n} t} + C_1 C_\alpha \sqrt{\frac{\mathtt{C}_{\{(g,r)\}}}{n} t}\right] \leq 2\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta) e^{-t},$$
$$\mathbb{P}\left[\|\Pi_2 R_n - \Pi_2 Z_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1 C_\alpha \sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_\mathcal{G}\mathtt{M}_\mathcal{G}}{n} t} + C_1 C_\alpha \sqrt{\frac{\mathtt{C}_{\{(g,r)\}}}{n} t}\right] \leq 2\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta) e^{-t},$$

where $C_1 > 0$ is a universal constant and $C_\alpha = 1 + (2\alpha)^{\alpha/2}$.

Lemma SA.20. *Suppose Assumption SA.2 holds. Let $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ be a cylindered quasi-dyadic expansion with $\rho > 1$. Suppose $(\mathcal{G}\times\mathcal{R})_\delta$ is a δ-net of $(\mathcal{G}\times\mathcal{R}, e_\mathbb{P})$ with cardinality no greater than $\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)$. Then for all $t > 0$,*

$$\begin{aligned}&\mathbb{P}\left[\|\Pi_1 M_n - \Pi_1 Z_n^M\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1 C_\alpha \sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_\mathcal{G}\mathtt{M}_\mathcal{G}}{n} t} + C_1 C_\alpha \sqrt{\frac{\mathtt{C}_{\{(g,r)\}}}{n} t}\right]\\ &\leq 2\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta) e^{-t} + 2^M \exp\left(-C_\rho n 2^{-M}\right),\\ &\mathbb{P}\left[\|\Pi_2 R_n - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1 C_\alpha \sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_\mathcal{G}\mathtt{M}_\mathcal{G}}{n} t} + C_1 C_\alpha \sqrt{\frac{\mathtt{C}_{\{(g,r)\}}}{n} t}\right]\\ &\leq 2\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta) e^{-t} + 2^M \exp\left(-C_\rho n 2^{-M}\right),\end{aligned}$$

where $C_1 > 0$ is a universal constant and $C_\alpha = 1 + (2\alpha)^{\alpha/2}$.

SA-III.6 Projection Error

The projection error can be decomposed into two parts: One captures the distance from the original function to the L_2 projection, which we call the L_2-projection error, the other captures the distance between Π_1, Π_2 and Π_0, which we call the misspecification error.

SA-III.6.1 Mis-specification Error for M_n-Process

Lemma SA.21. *Suppose Assumption SA.1 or Assumption SA.2 hold. Let $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ with $\rho \geq 1$ be a cylindered quasi-dyadic expansion. Let $\tau > 0$. Define $r_\tau := r\mathbb{1}([-\tau^{\frac{1}{\alpha}}, \tau^{\frac{1}{\alpha}}])$. Then for any $g \in \mathcal{G}, r \in \mathcal{R}$,*

$$\mathbb{E}\left[(\Pi_1 M_n(g,r_\tau) - \Pi_0 M_n(g,r_\tau))^2\right] \leq 2(1+\rho)\tau^2 N^2 \mathtt{V}_\mathcal{G},$$
$$\mathtt{V}_\mathcal{G} := \min\{2\mathtt{M}_\mathcal{G}, \mathtt{L}_\mathcal{G}\|\mathcal{V}\|_\infty\}\left(\sup_{\mathbf{x}} f_X(\mathbf{x})\right)^2 2^M \mathfrak{m}(\mathcal{V})\|\mathcal{V}\|_\infty \mathtt{TV}_\mathcal{G}.$$

SA-III.6.2 L_2-projection Error for M_n-Process

Lemma SA.22. *Suppose Assumption SA.1 or Assumption SA.2 hold. Let $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ with $\rho \geq 1$ be a cylindered quasi-dyadic expansion. Let $\tau > 0$. Define $r_\tau := r\mathbb{1}([-\tau^{\frac{1}{\alpha}}, \tau^{\frac{1}{\alpha}}])$. Then for any $g \in \mathcal{G}, r \in \mathcal{R}$,*

$$\mathbb{E}\left[(\Pi_0 M_n(g,r_\tau) - M_n(g,r_\tau))^2\right] \leq 2\left(2^{-N}\tau^2\mathtt{M}_\mathcal{G}^2 + (1+\rho)\tau^2\mathtt{V}_\mathcal{G}\right).$$

SA-III.6.3 Projection Error for M_n-Process

Combining Lemma SA.21 and SA.22, we can bound the projection error through a truncation argument.

Lemma SA.23. *Suppose Assumption SA.1 or Assumption SA.2 hold. Let $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ with $\rho \geq 1$ be a cylindered quasi-dyadic expansion. Then for all $t > N$,*

$$\mathbb{P}\left[\|M_n - \Pi_1 M_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} > \sqrt{C_{2\alpha}}\sqrt{(1+\rho)N^2\mathtt{V}_\mathcal{G} + 2^{-N}\mathtt{M}_\mathcal{G}^2}\,t^{\alpha+\frac{1}{2}} + C_\alpha\frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}}t^{\alpha+1}\right] \leq 4\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)ne^{-t},$$
$$\mathbb{P}\left[\|Z_n^M - \Pi_1 Z_n^M\|_{(\mathcal{G}\times\mathcal{R})_\delta} > \sqrt{C_{2\alpha}}\sqrt{(1+\rho)N^2\mathtt{V}_\mathcal{G} + C_\alpha 2^{-N}\mathtt{M}_\mathcal{G}^2}\,t^{\frac{1}{2}} + C_\alpha\frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}}t\right] \leq 4\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)ne^{-t},$$

where $C_\alpha = 1 + (2\alpha)^{\frac{\alpha}{2}}$ and $C_{2\alpha} = 1 + (4\alpha)^\alpha$.

SA-III.6.4 Projection Error for R_n-Process

The projection error for R_n-process can be built up upon the error for M_n-process and the observation that

$$\Pi_2 R_n(g,r) - R_n(g,r) = \Big(\Pi_1 M_n(g,r) - M_n(g,r)\Big) - \Big(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})]X_n(g\theta(\cdot,r)) - X_n(g\theta(\cdot,r))\Big),$$
$$\Pi_2 Z_n^R(g,r) - Z_n^R(g,r) = \Big(\Pi_1 Z_n^M(g,r) - Z_n^M(g,r)\Big) - \Big(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})]Z_n^X(g\theta(\cdot,r)) - Z_n^X(g\theta(\cdot,r))\Big),$$

where in both lines, the first bracket is a projection error for an M_n-process that has been studied in Section SA-III.6.3, and the second bracket is a projection error for an X_n-process that has been studied in Section SA-II.5.

Lemma SA.24. *Suppose Assumption SA.1 or Assumption SA.2 hold. Let $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ with $\rho \geq 1$ be a cylindered quasi-dyadic expansion. Then for all $t > N$, with probability at least $1 - 4\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)ne^{-t}$,*

$$\|R_n - \Pi_2 R_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} \lesssim \sqrt{\mathtt{V}_{\mathcal{G}\times\mathcal{V}_\mathcal{R}}}\,t^{\frac{1}{2}} + \sqrt{C_{2\alpha}}\sqrt{(1+\rho)N^2\mathtt{V}_\mathcal{G} + 2^{-N}\mathtt{M}_\mathcal{G}^2}\,t^{\alpha+\frac{1}{2}} + C_\alpha\frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}}t^{\alpha+1},$$
$$\|Z_n^R - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} \lesssim \sqrt{\mathtt{V}_{\mathcal{G}\times\mathcal{V}_\mathcal{R}}}\,t^{\frac{1}{2}} + \sqrt{C_{2\alpha}}\sqrt{(1+\rho)N^2\mathtt{V}_\mathcal{G} + 2^{-N}\mathtt{M}_\mathcal{G}^2}\,t^{\frac{1}{2}} + C_\alpha\frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}}t,$$

where

$$\mathtt{V}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}} := \min\{2\mathtt{M}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}}, \mathtt{L}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}}\|\mathcal{V}\|_\infty\}\left(\sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\right)^2 2^M \mathfrak{m}(\mathcal{V})\|\mathcal{V}\|_\infty \mathtt{TV}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}}.$$

Lemma SA.25 (Covering Number of Conditional Mean)**.** *Suppose (X, Y) is a random variable taking values in $\mathbb{R}^d \times \mathbb{R}$ and $\mathcal{S}$ is a class of measurable functions from $\mathbb{R}$ to $\mathbb{R}$. Consider $\mathcal{V}_{\mathcal{S}} = \{v_s : s \in \mathcal{S}\}$. Then for all $0 < \varepsilon < 1$,*

$$\sup_Q N(\mathcal{V}_{\mathcal{S}}, \|\cdot\|_{Q,2}, \varepsilon\|v_s\|_{Q,2}) \leq \sup_Q N(\mathcal{S}, \|\cdot\|, \varepsilon\|s\|_{Q,2}),$$

where sup *is taken with respect to all finite discrete measures.*

Proof. Let $\mathcal{Q}$ in be a finite discrete measure on $\mathbb{R}^{d_z}$. Let $r, s \in \mathcal{S}$. Define a new probability measure $\widetilde{P}$ on $\mathbb{R}$ by

$$\widetilde{P}(A) = \int \mathbb{E}[\mathbb{1}((\mathbf{z}_i, y_i) \in \mathbb{R}^{d_z} \times A)|\mathbf{z}_i = \mathbf{z}]d\mathcal{Q}(\mathbf{z}), \qquad \forall A \subseteq \mathbb{R}^{d_z}.$$

Then $\int |S| d\widetilde{P} \leq \int_{\mathbb{R}^{d_z}} \mathbb{E}[S(y_i)|\mathbf{z}_i = \mathbf{z}]d\mathcal{Q}(z) < \infty$ since $\sup_{m\in M_S}\|m\|_\infty < \infty$. Hence $\widetilde{P} \in \widetilde{\mathcal{A}}(\mathbb{R})$. Let $r, s \in \mathcal{S}$. Then

$$\int |m_r - m_s|^2 d\mathcal{Q} \leq \int_{\mathbb{R}^{d_z}} \mathbb{E}[|r(y_i) - s(y_i)|^2|\mathbf{z}_i = \mathbf{z}]d\mathcal{Q}(z) = \int |r - s|^2 d\widetilde{P}.$$

Here $\widetilde{P}$ is not necessarily a finite discrete measure, but by similar argument as in Lemma SA.14, there exists $\mathcal{S}_\varepsilon \subseteq \mathcal{S}$ with cardinality no greater than $\sup_{\mathcal{Q}} N(S, \|\cdot\|_{\mathcal{Q},2}, \varepsilon\|S\|_{\mathcal{Q},2})$ such that for any $s \in \mathcal{S}$, there exists $r \in \mathcal{S}_\varepsilon$ such that $\|r - s\|_{\widetilde{P},2} \leq \varepsilon\|S\|_{\widetilde{P},2}$. Hence $\|m_r - m_s\|_{\mathcal{Q},2} \leq \varepsilon\|S\|_{\widetilde{P},2} = \varepsilon\|m_S\|_{\mathcal{Q},2}$. This implies that for any $0 < \varepsilon < 1$,

$$\sup_Q N(M_{\mathcal{S}}, \|\cdot\|_{Q,2}, \varepsilon\|m_S\|_{Q,2}) \leq \sup_Q N(\mathcal{S}, \|\cdot\|, \varepsilon\|S\|_{Q,2}).$$

□

SA-III.7 Proof of Lemma SA.15

By the entropy integral conditions on $\mathcal{G}$ and $\mathcal{R}$ and Lemma SA.13,

$$J(\mathcal{G}\times\mathcal{R}, \mathtt{M}_G M_R, \delta) \leq \sqrt{2}J(\mathcal{G}, \mathtt{M}_G, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{R}, M_R, \delta/\sqrt{2}).$$

Claim 1: There exists $C_\alpha > 0$ such that for all $0 < \delta < 1$,

$$J(\Pi_0(\mathcal{G}\times\mathcal{R}), C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha, \delta) \leq J(\mathcal{G}\times\mathcal{R}, \mathtt{M}_{\mathcal{G}} M_R, \delta).$$

Proof of Claim 1: Under condition (a), $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(Y)|X = \mathbf{x}] \leq 2$ and $|M_R(t)| \leq 1 + |t|^\alpha$ for some constant $\alpha \geq 0$. By Step 2 in Definition SA.3, $\max_{0\leq l<2^{M+N}} \mathbb{E}[\exp(y_i/(N\log 2))|\mathbf{x}_i \in \mathcal{C}_{0,l}] \leq 2$. Hence

$$\max_{0\leq l<2^{M+N}} \sup_{r\in\mathcal{R}} \mathbb{E}[|r(y_i)||(\mathbf{x}_i, y_i) \in \mathcal{C}_{0,l}] \leq 1 + \max_{0\leq l<2^{M+N}} \mathbb{E}[|y_i|^\alpha|\mathbf{x}_i \in \mathcal{C}_{0,l}] \lesssim 1 + (2N\sqrt{\alpha})^\alpha, \tag{SA-11}$$

Hence

$$\|\|\Pi_0(gr)\|_\infty\|_{\mathcal{G}\times\mathcal{R}} \le C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha, \qquad C_\alpha = 1 + (2\sqrt{\alpha})^\alpha. \tag{SA-12}$$

Under condition (b), $\mathtt{M}_{\mathcal{R}} \le 1$. Hence Equation SA-12 holds with $\alpha = 0$. Hence Let Q be a finite discrete measure. Let $f, g \in \mathcal{G} \times R$. Then by definition of Π_0,

$$\|\Pi_0 f - \Pi_0 g\|_{Q,2}^2 \le \sum_{0\le k<2^{M+N}} Q(C_{0,k})(2^{M+N}\int_{C_{0,k}} f - g d\mathbb{P})^2 \le \sum_{0\le k<2^{M+N}} Q(C_{0,k})2^{M+N}\int_{C_{0,k}} (f-g)^2 d\mathbb{P}.$$

Define a measure $\widetilde{Q}$ such that for any $A \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R})$, $\widetilde{Q}(A) = \sum_{0\le k<2^{M+N}} Q(C_{0,k})2^{M+N}\mathbb{P}(A \cap C_{0,k})$, then

$$\|\Pi_0 f - \Pi_0 g\|_{Q,2}^2 \le \|f - g\|_{\widetilde{Q},2}^2.$$

By Lemma SA.14, there exists an $\delta C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha$-net $\mathcal{L}$ of $\mathcal{G}\times\mathcal{R}$ with cardinality no greater than $\sup_Q N(\mathcal{G} \times \mathcal{R}, e_Q, \delta\|\mathtt{M}_{\mathcal{G}} \times M_R\|)$, sup taken over all finite discrete measures on $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$, such that for all $f \in \Pi_0(\mathcal{G}\times\mathcal{R})$, there exists $g \in \mathcal{L}$ such that

$$\|f - g\|_{\widetilde{Q},2}^2 \le \delta^2 \|\mathtt{M}_G M_R\|_{\widetilde{Q},2}^2 \le \delta^2 (C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha)^2.$$

The claim then follows.

Claim 2: There exists $C_\alpha > 0$ such that for all $0 < \delta < 1$,

$$J(\Pi_1(\mathcal{G}\times\mathcal{R}), C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha, \delta) \lesssim J(\mathcal{G}\times\mathcal{R}, \mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}, \delta/3) \text{ and } J(\Pi_2(\mathcal{G}\times\mathcal{R}), C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha, \delta) \lesssim J(\mathcal{G}\times\mathcal{R}, \mathtt{M}_{\mathcal{G}} M_R, \delta/4).$$

Proof of Claim 2: Suppose $\widetilde{\mathbb{P}}$ is a mapping from $\mathcal{B}(\mathbb{R}^{d+1})$ to $[0,1]$ such that

$$\widetilde{\mathbb{P}}(E) = \inf\Bigg\{\sum_{0\le l<2^M}\sum_{0\le m<2^N} \mathbb{E}[\mathbb{1}(X \in A)|X \in \mathcal{X}_{0,l}] \cdot \mathbb{E}[\mathbb{1}(Y \in B)|X \in \mathcal{X}_{0,l}, Y \in \mathcal{Y}_{l,0,m}] : \\ E \subseteq A \times B, A \in \mathcal{B}(\mathbb{R}^d), B \in \mathcal{B}(\mathbb{R})\Bigg\}.$$

It is easy to verify that $\widetilde{\mathbb{P}}$ defines a probability measure on $(\mathbb{R}^d \times \mathbb{R}, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}))$. Recall $\mathcal{C}_{M,N}$ is a collection of cells $\{\mathcal{C}_{j,k} : 0 \le j \le M+N, 0 \le k < 2^{M+N}\}$ where $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathbb{R}$ if $j \ge N$. Take $\mathcal{C}_{M,0} = \{\mathcal{C}_{j,k} : N \le j \le M+N, 0 \le k < 2^{M+N-j}\}$. Let $g \in \mathcal{G}, r \in \mathcal{R}$.

$$\begin{aligned}
\Pi_1[\mathcal{C}_{M,N}(\mathbb{P},\rho)](g,r) &= \Pi_1[\mathcal{C}_{M,0}(\mathbb{P},\rho)](g,r) + \Pi_1[\mathcal{C}_{M,N}(\mathbb{P},\rho)](g,r) - \Pi_1[\mathcal{C}_{M,0}(\mathbb{P},\rho)](g,r) \\
&= \Pi_0[\mathcal{C}_{M,0}(\mathbb{P},\rho)](gr) + \Pi_0[\mathcal{C}_{M,N}(\widetilde{\mathbb{P}},\rho)](gr) - \Pi_0[\mathcal{C}_{M,0}(\widetilde{\mathbb{P}},\rho)](gr), \\
\Pi_2[\mathcal{C}_{M,N}(\mathbb{P},\rho)](g,r) &= \Pi_1[\mathcal{C}_{M,N}(\mathbb{P},\rho)](g,r) - \Pi_0[\mathcal{C}_{M,0}(\mathbb{P},\rho)](g\theta(\cdot,r)).
\end{aligned}$$

Since $\|\Pi_0[\mathcal{C}_{M,N}(\widetilde{\mathbb{P}},\rho)]\|_{\mathcal{G}\times\mathcal{R}} \le C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha$, the previous claim applies not only to $J(\Pi_0[\mathcal{C}_{M,N}(\mathbb{P},\rho)](\mathcal{G}\times\mathcal{R}), C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha, \delta)$ but also to $J(\Pi_0[\mathcal{C}_{M,N}(\widetilde{\mathbb{P}},\rho)](\mathcal{G}\times\mathcal{R}), C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha, \delta)$. Then Claim 2 follows from Claim 1.

Claim 3: There exists $C_\alpha > 0$ such that for all $0 < \delta < 1$,

$$J(\mathcal{G} \times \mathcal{V}_{\mathcal{R}}, C_1 \mathtt{M}_{\mathcal{G}} \sqrt{C_{2\alpha}}, \delta) \leq \sqrt{2} J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2} J(\mathcal{R}, M_R, \delta/\sqrt{2}),$$

C_1 is some absolute constant.

Proof of Claim 3: Let Q be a discrete measure on $\mathbb{R}^d$. Take $\widetilde{Q}$ be the measure on $\mathbb{R}^d \times \mathbb{R}$ such that

$$\widetilde{Q}(E) = \int_{\mathbb{R}^d} \mathbb{E}[\mathbb{1}((\mathbf{x}_1, y_1) \in \mathbb{E}) | \mathbf{x}_1 = \mathbf{x}] dQ(\mathbf{x}), \qquad E \in \mathcal{B}(\mathbb{R}^d).$$

Take $\widetilde{Q}_Y$ to be the marginal of $\widetilde{Q}$ on the last dimension. Then for any $r_1, r_2 \in \mathcal{R}$ such that $\|r_1 - r_2\|_{\widetilde{Q}_Y, 2} \leq \varepsilon \sqrt{C_{2\alpha}}$, we have

$$\begin{aligned}
\|\theta(\cdot, r_1) - \theta(\cdot, r_2)\|_{Q,2}^2 &= \int_{\mathbb{R}^d} |\mathbb{E}[r_1(y_i)|\mathbf{x}_i = \mathbf{x}] - \mathbb{E}[r_2(y_i)|\mathbf{x}_i = \mathbf{x}]|^2 dQ(\mathbf{x}) \\
\leq \int_{\mathbb{R}^d} \mathbb{E}[(r_1(y_i) - r_2(y_i))^2 | \mathbf{x}_i = \mathbf{x}] dQ(\mathbf{x}) &= \int_{\mathbb{R}^d} \int_{\mathbb{R}} (r_1(y) - r_2(y))^2 d\widetilde{Q}(\mathbf{x}, y) \\
= \|r_1 - r_2\|_{\widetilde{Q}_Y, 2}^2 \leq \varepsilon^2 \|M_{\mathcal{R}}\|_{\widetilde{Q}_Y, 2}^2 = \varepsilon^2 \int_{\mathbb{R}^d} \mathbb{E}[M_{\mathcal{R}}(y_i)^2 | \mathbf{x}_i = \mathbf{x}] dQ(\mathbf{x}) &\lesssim \varepsilon^2 C_{2\alpha}.
\end{aligned}$$

It follows that $J(\mathcal{V}_{\mathcal{R}}, C_1 \sqrt{C_{2\alpha}}, \delta) \leq J(\mathcal{R}, M_{\mathcal{R}}, \delta)$, where C_1 some absolute constant. Hence

$$J(\mathcal{G} \times \mathcal{V}_{\mathcal{R}}, C_1 \mathtt{M}_{\mathcal{G}} \sqrt{C_{2\alpha}}, \delta) \leq \sqrt{2} J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2} J(\mathcal{R}, M_R, \delta/\sqrt{2}).$$

Moreover, $\{e_{j,k} : (j,k) \in \mathcal{I})\}$ has cardinality 2^{M+N}. It follows from pointwise separability of $\mathcal{G}$ and $\mathcal{R}$ and Corollary 2.2.9 in van der Vaart and Wellner (2013) that $(\mathcal{G} \times \mathcal{R}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}) \cup \Pi_2(\mathcal{G} \times \mathcal{R}) \cup \mathcal{E}_{M+N}$ is pre-Gaussian. □

SA-III.8 Proof of Lemma SA.16

The result follows from Lemma SA.5 with $(\mathbf{x}_i, y_i)$ replacing $\mathbf{x}_i$. □

SA-III.9 Proof of Lemma SA.17

Define

$$A = \{|\widetilde{U}_{j,k}| \leq c_1 U_{j,k}, \text{ for all } N \leq j \leq M+N, 0 \leq k < 2^{M+N-j}\}.$$

Since in Definition SA.3, $\{\mathcal{Y}_{l,j,m} : 0 \leq j \leq N, 0 \leq m < 2^{N-j}\}$ is a dyadic expansion, we can apply Tusnády's Lemma (Bretagnolle and Massart, 1989, Lemma 4) and Lemma SA.6 to get whenever A holds,

$$\begin{aligned}
&\left| \widetilde{U}_{j,k} - \sqrt{U_{j,k} \frac{\mathbb{P}(\mathcal{C}_{j-1,2k}) \mathbb{P}(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}(\mathcal{C}_{j,k})^2}} \widetilde{\xi}_{j,k} \right| < c_2 \widetilde{\xi}_{j,k}^2 + c_3, \\
&\left| \widetilde{U}_{j,k} \right| \leq 1/c_0 + 2 \sqrt{\frac{\mathbb{P}(\mathcal{C}_{j-1,2k}) \mathbb{P}(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}(\mathcal{C}_{j,k})^2} U_{j,k}} |\widetilde{\xi}_{j,k}|, \quad \forall 1 \leq j \leq M+N, 0 \leq k < 2^{M+N-j}.
\end{aligned}$$

And similarly as in the proof for Lemma SA.7,

$$\mathbb{P}(A^c) \leq 42^M \exp\left(-\min\left\{\frac{c_1^2}{3} \wedge 1\right\}\rho^{-1} n 2^{-M}\right).$$

The rest of the proof follows from Lemma SA.7 by replacing $\mathbf{x}_i$ with $(\mathbf{x}_i, y_i)$. □

SA-III.10 Proof of Lemma SA.18

By Lemma SA.13, for any $0 < \delta < 1$, $\sup_Q N(\mathcal{G} \times \mathcal{R}, \|\cdot\|_{Q,2}, \delta\|\mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}\|_{Q,2}) \leq N(\delta)$ and $J(\delta, \mathcal{G} \times \mathcal{R}, \mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}) \leq J(\delta)$. By definition $\|\pi_{(\mathcal{G}\times\mathcal{R})_\delta} h - h\|_{\mathbb{P},2} \leq \delta\|\mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}\|_{\mathbb{P},2}$, where $\mathbb{P}$ is the joint law for $(\mathbf{x}_i, y_i)$. Take $\mathcal{L} = \{h - \pi_{(\mathcal{G}\times\mathcal{R})_\delta} h : h \in \mathcal{G} \times \mathcal{R}\}$. Take $\mathbb{G}_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [f(\mathbf{x}_i, y_i) - \mathbb{E}[f(\mathbf{x}_i, y_i)]]$. Then, by Theorem 5.2 in Chernozhukov *et al.* (2014),

$$\begin{aligned}\mathbb{E}[\|\mathbb{G}_n\|_{\mathcal{L}}] &\lesssim J(\delta)\mathtt{M}_{\mathcal{G}}\|M_{\mathcal{R}}(y_i)\|_{\mathbb{P},2} + \frac{\mathtt{M}_{\mathcal{G}}\|\max_{1\leq i\leq n} M_{\mathcal{R}}(y_i)\|_{\mathbb{P},2} J^2(\delta)}{\delta^2\sqrt{n}} \\ &\lesssim J(\delta)\mathtt{M}_{\mathcal{G}}(1 + (2\alpha)^{\frac{\alpha}{2}}) + \frac{\mathtt{M}_{\mathcal{G}} J^2(\delta)}{\delta^2\sqrt{n}}(1 + (2\log(n)\alpha)^{\frac{\alpha}{2}}).\end{aligned}$$

Moreover, $\|\max_{1\leq i\leq n}\sup_{g\in\mathcal{G}, r\in\mathcal{R}} |g(\mathbf{x}_i) r(y_i)|\|_{\psi_{\alpha^{-1}}} \lesssim \mathtt{M}_{\mathcal{G}}(\|\max_{1\leq i\leq} y_i\|_{\psi_1})^\alpha \lesssim \mathtt{M}_{\mathcal{G}}(\log n)^\alpha$. Hence, by Theorem 4 in Adamczak (2008), for any $t > 0$, with probability at least $1 - 4\exp(-t)$,

$$\|\mathbb{G}_n\|_{\mathcal{L}} \lesssim J(\delta)\mathtt{M}_{\mathcal{G}}(1 + (2\alpha)^{\frac{\alpha}{2}}) + \frac{\mathtt{M}_{\mathcal{G}} J^2(\delta)}{\delta^2\sqrt{n}}(1 + (2\log(n)\alpha)^{\frac{\alpha}{2}}) + \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t + (\log n)^\alpha \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^\alpha.$$

In particular, $\|\mathbb{G}_n\|_{\mathcal{L}} = \|M_n - M_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}}$. The bound for $\|Z_n^M - Z_n^M \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|$ follows from a standard concentration inequality for Gaussian suprema. The bound for R_n process follows from the fact that if we define $\mathcal{G} \times \overline{R} = \{g(r - \theta(\cdot, r)) : g \in \mathcal{G}, r \in \mathcal{R}\}$, then

$$\sup_Q N(\mathcal{G} \times \overline{R}, \|\cdot\|_{Q,2}, \delta\|\mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}\|_{Q,2}) \leq 2\sup_Q N(\mathcal{G} \times \mathcal{R}, \|\cdot\|_{Q,2}, \delta\|\mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}\|_{Q,2}). \qquad \text{(SA-13)}$$

Now we show the above inequality holds: Let $\mathcal{Q}$ in be a finite discrete measure on $\mathbb{R}^{d_z}$. Let $r, s \in \mathcal{S}$. Define a new probability measure $\widetilde{P}$ on $\mathbb{R}$ by

$$\widetilde{P}(A) = \int \mathbb{E}[\mathbb{1}((\mathbf{x}_i, y_i) \in \mathbb{R}^d \times A) | \mathbf{x}_i = \mathbf{x}] d\mathcal{Q}(\mathbf{x}), \qquad \forall A \subseteq \mathbb{R}^d.$$

Then $\int |S| d\widetilde{P} \leq \int_{\mathbb{R}^d} \mathbb{E}[S(y_i)|\mathbf{x}_i = \mathbf{x}] d\mathcal{Q}(z) < \infty$ since $\sup_{m\in M_S}\|m\|_\infty < \infty$. Hence $\widetilde{P} \in \widetilde{\mathcal{A}}(\mathbb{R})$. Let $r, s \in \mathcal{S}$. Then

$$\int |m_r - m_s|^2 d\mathcal{Q} \leq \int_{\mathbb{R}^{d_z}} \mathbb{E}[|r(y_i) - s(y_i)|^2 | \mathbf{x}_i = \mathbf{x}] d\mathcal{Q}(x) = \int |r - s|^2 d\widetilde{P}.$$

Here $\widetilde{P}$ is not necessarily a finite discrete measure, but by similar argument as in Lemma SA.14, there exists $\mathcal{S}_\varepsilon \subseteq \mathcal{S}$ with cardinality no greater than $\sup_Q N(S, \|\cdot\|_{\mathcal{Q},2}, \varepsilon\|S\|_{\mathcal{Q},2})$ such that for any $s \in \mathcal{S}$, there exists $r \in \mathcal{S}_\varepsilon$ such that $\|r - s\|_{\widetilde{P},2} \leq \varepsilon\|S\|_{\widetilde{P},2}$. Hence $\|m_r - m_s\|_{\mathcal{Q},2} \leq \varepsilon\|S\|_{\widetilde{P},2} = \varepsilon\|m_S\|_{\mathcal{Q},2}$. This implies that for any $0 < \varepsilon < 1$,

$$\sup_Q N(M_{\mathcal{S}}, \|\cdot\|_{Q,2}, \varepsilon \|m_S\|_{Q,2}) \leq \sup_Q N(\mathcal{S}, \|\cdot\|, \varepsilon \|S\|_{Q,2}).$$

□

SA-III.11 Proof of Lemma SA.19

For notational simplicity, we will use $\mathbb{E}[\cdot|\mathcal{X}_{0,l}]$ in short for $\mathbb{E}[\cdot|\mathbf{x}_i \in \mathcal{X}_{0,l}]$, $\mathbb{E}[\cdot|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in short for $\mathbb{E}[\cdot|(\mathbf{x}_i, y_i) \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$. First, we consider the M_n-process.

Layers $N+1 \leq j \leq M+N$**:** For this layers, $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathbb{R}$. By definition of $\widetilde{\gamma}_{j,k}$,

$$\begin{aligned}
\sum_{N<j\leq M+N} \sum_{0\leq k<2^{M+N-j}} |\widetilde{\gamma}_{j,k}(g,r)| &\leq \sum_{N\leq j<M+N} \sum_{0\leq k<2^{M+N-j}} \mathbb{E}\left[|g(\mathbf{x}_i) r(y_i)| |\mathbf{x}_i \in \mathcal{X}_{j-N,k}\right] \\
&\leq \sum_{N\leq j<M+N} \sum_{0\leq k<2^{M+N-j}} \mathbb{E}\left[|g(\mathbf{x}_i) \mathbb{E}\left[r(y_i)|\mathbf{x}_i\right]| |\mathbf{x}_i \in \mathcal{X}_{j-N,k}\right] \\
&\lesssim C_\alpha \sum_{N\leq j<M+N} \sum_{0\leq k<2^{M+N-j}} 2\mathbb{E}\left[|g(\mathbf{x}_i) \mathbb{1}(\mathbf{x}_i \in \mathcal{X}_{j-N,k})|\right] \mathbb{P}\left(\mathbf{x}_i \in \mathcal{X}_{j-N,k}\right)^{-1} \\
&\lesssim C_\alpha \sum_{N\leq j<M+N} \mathtt{E}_{\mathcal{G}} 2^{M+N-j} \lesssim C_\alpha 2^M \mathtt{E}_{\mathcal{G}},
\end{aligned}$$

where in (1) we have used $E[|r(y_i)||\mathbf{x}_i = \mathbf{x}] \lesssim C_\alpha = 1 + (2\alpha)^{\alpha/2}$ for all $\mathbf{x} \in \mathcal{X}$. Moreover, $|\widetilde{\gamma}_{j,k}(g,r)| \lesssim C_\alpha \mathtt{M}_{\mathcal{G}}$ for all $j \in (N, M+N]$, hence

$$\sum_{N\leq j\leq M+N} \sum_{0\leq k<2^{M+N-j}} |\widetilde{\gamma}_{j,k}(g,r)|^2 \lesssim C_\alpha^2 2^M \mathtt{E}_{\mathcal{G}} \mathtt{M}_{\mathcal{G}}.$$

Layers $1 \leq j \leq N$**:** By definition, $\mathcal{C}_{j,k} = \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}$, where $k = 2^{N-j} l + m$, for some unique $l \in [0, 2^M)$ and $m \in [0, 2^{N-j})$. Denote $k = (l,m)$. Fix j and l, sum across m,

$$\sum_{m=0}^{2^{N-j}-1} \left|\widetilde{\gamma}_{j,(l,m)}(g,r)\right| = \sum_{m=0}^{2^{N-j}-1} \left|\mathbb{E}\left[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\right] \left(\mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2m}\right] - \mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2m+1}\right]\right)\right|.$$

Under condition (a), Notice that $|\max(\mathcal{Y}_{l,j-1,0})| \leq \log(\mathbb{E}\left[\exp(r(y_i))|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,0}\right]) \leq \log(2 \cdot 2^N) \leq 2N$, and similarly $|\min(\mathcal{Y}_{l,j-1,2^{N-j}})| \leq 2N$,

$$\begin{aligned}
&\sum_{m=1}^{2^{N-j}-2} \left|\mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2m}\right] - \mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2m+1}\right]\right| \lesssim \mathrm{TV}(r|_{[-2N,2N]}) \lesssim N^\alpha, \\
&\left|\mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,1}\right] - \mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,0}\right]\right| \leq \left(\max_m - \min_m\right) \mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,m}\right] \lesssim C_\alpha N^\alpha, \\
&\left|\mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2^{N-j}-1}\right] - \mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2^{N-j}-2}\right]\right| \lesssim C_\alpha N^\alpha.
\end{aligned}$$

Under condition (b), since $\mathtt{TV}_{\{r\}} \lesssim 1$ and $\mathtt{M}_{\{r\}} \lesssim 1$, the above three inequality still hold. It follows that for all $g \in \mathcal{G}, r \in \mathcal{R}$, fix j, l and sum across m,

$$\sum_{m=0}^{2^{N-j}-1} \left|\widetilde{\gamma}_{j,(l,m)}(g,r)\right| \lesssim C_\alpha N^\alpha \left|\mathbb{E}\left[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\right]\right|.$$

Fix j and sum the above across l,

$$\begin{aligned}\sum_{0 \leq k < 2^{M+N-j}} \left|\widetilde{\gamma}_{j,(l,m)}(g,r)\right| &= \sum_{l=0}^{2^M-1} \sum_{m=0}^{2^{N-j}-1} \left|\widetilde{\gamma}_{j,(l,m)}(g,r)\right| \lesssim C_\alpha N^\alpha \sum_{l=0}^{2^M-1} \mathbb{E}\left[|g(\mathbf{x}_i)\mathbb{1}(\mathcal{X}_{0,l})|\right] \mathbb{P}\left(\mathbf{x}_i \in \mathcal{X}_{0,l}\right)^{-1} \\ &\lesssim C_\alpha N^\alpha 2^M \mathtt{E}_{\mathcal{G}}.\end{aligned}$$

We can now sum across j to get

$$\sum_{j=1}^{N} \sum_{0 \leq k < 2^{M+N-j}} |\widetilde{\gamma}_{j,k}(g,r)| \lesssim C_\alpha N^{\alpha+1} 2^M \mathtt{E}_{\mathcal{G}}.$$

By Equation SA-11, $\sup_{g \in \mathcal{G}, r \in \mathcal{R}} |\widetilde{\gamma}_{j,k}(g,r)| \lesssim C_\alpha N^\alpha \mathtt{M}_{\mathcal{G}}$, and hence

$$\sum_{1 \leq j \leq N} \sum_{0 \leq k < 2^{M+N-j}} |\widetilde{\gamma}_{j,k}(g,r)|^2 \lesssim C_\alpha^2 N^{2\alpha+1} 2^M \mathtt{E}_{\mathcal{G}} \mathtt{M}_{\mathcal{G}}.$$

Strong Approximation for Projected Processes Putting together the previous two parts,

$$\sum_{j=1}^{M+N} \sum_{k=0}^{2^{M+N-j}} \widetilde{\gamma}_{j,k}^2(g,r) \lesssim C_\alpha^2 N^{2\alpha+1} 2^M \mathtt{E}_{\mathcal{G}} \mathtt{M}_{\mathcal{G}}.$$

By Lemma SA.16, we know for any $(g,r) \in \mathcal{G} \times \mathcal{R}$, for any $x > 0$, with probability at least $1 - 2\exp(-x)$,

$$|M_n \circ \Pi_1(g,r) - Z_n \circ \Pi_1(g,r)| \lesssim C_\alpha \sqrt{\frac{N^{2\alpha+1} 2^M \mathtt{E}_{\mathcal{G}} \mathtt{M}_{\mathcal{G}}}{n}} x + C_\alpha \sqrt{\frac{\mathtt{C}_{\{(g,r)\}}}{n}} x,$$

where C_α is a constant that only depends on α. It then follows from the relation between $\gamma_{j,k}$ and $\eta_{j,k}$ which is given in Equation SA-9 that for any $(g,r) \in \mathcal{G} \times \mathcal{R}$, for any $x > 0$, with probability at least $1 - 2\exp(-x)$,

$$|R_n \circ \Pi_2(g,r) - Z_n \circ \Pi_2(g,r)| \lesssim C_\alpha \sqrt{\frac{N^{2\alpha+1} 2^M \mathtt{E}_{\mathcal{G}} \mathtt{M}_{\mathcal{G}}}{n}} x + C_\alpha \sqrt{\frac{\mathtt{C}_{\{(g,r)\}}}{n}} x.$$

□

SA-III.12 Proof of Lemma SA.20

Since $\mathcal{C}_{M,N}$ is a cylindered quasi-dyadic expansion, $\rho^{-1} 2^{-M-N+j} \leq \mathbb{P}(\mathcal{C}_{j,k}) \leq \rho 2^{-M-N+j}$, for all $0 \leq j \leq M+N$, $0 \leq k < 2^{M+N-j}$. Hence following the argument in the proof for Lemma SA.19, for any $g \in \mathcal{G}, r \in \mathcal{R}$,

$$\sum_{j=1}^{M+N} \sum_{k=0}^{2^{M+N-j}} \widetilde{\eta}_{j,k}^2(g,r) \leq \sum_{j=1}^{M+N} \sum_{k=0}^{2^{M+N-j}} \widetilde{\gamma}_{j,k}^2(g,r) \lesssim C_\alpha^2 N^{2\alpha+1} 2^M \mathtt{E}_{\mathcal{G}} \mathtt{M}_{\mathcal{G}}.$$

The result then follows from Lemma SA.16. □

SA-III.13 Proof of Lemma SA.21

Scrutinizing the definition of $\beta_{j,k}$ and $\gamma_{j,k}$ from Sections SA-II.2 and SA-III.2, essentially we are going to show the difference between $\Pi_1 M_n(g, r_\tau)$ and $\Pi_0 M_n(g, r_\tau)$ is driven by the difference between g and $\Pi_0(\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)])g(\mathbf{x}_i)$, the L_2-projection of g onto $\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)]$. Expanding $\Pi_1 M_n(g, r_\tau) - \Pi_0 M_n(g, r_\tau)$ by Haar basis representation,

$$\Pi_1 M_n(g, r_\tau) - \Pi_0 M_n(g, r_\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Delta_i,$$
$$\Delta_i(g, r_\tau) = \sum_{1 \le j \le N} \sum_{0 \le k < 2^{M+N-j}} \left(\widetilde{\gamma}_{j,k}(g, r_\tau) - \widetilde{\beta}_{j,k}(g, r_\tau)\right) \widetilde{e}_{j,k}(\mathbf{x}_i, y_i),$$

where we have used $\widetilde{\gamma}_{j,k}(g, r_\tau) = \widetilde{\beta}_{j,k}(g, r_\tau)$ for $j > N$. Moreover,

$$\mathbb{E}\left[|\Delta_i(g, r_\tau)|\right] \le (1+\rho) \sum_{0 \le j < N} \sum_{0 \le k < 2^{M+N-j}} |\gamma_{j,k}(g, r) - \beta_{j,k}(g, r)| \, \mathbb{P}\left((\mathbf{x}_i, y_i) \in \mathcal{C}_{j,k}\right).$$

Recall in Definition SA.3, $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,l} \times \mathcal{Y}_{l,j,m}$, where $k = 2^{N-j} l + m$, $0 \le l < 2^M$ and $0 \le m < 2^{N-j}$. Since M_R has polynomial growth and r_τ has been truncated,

$$\begin{aligned}
&|\gamma_{j,k}(g, r_\tau) - \beta_{j,k}(g, r_\tau)| = |\mathbb{E}\left[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\right] \cdot \mathbb{E}\left[r_\tau(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}\right] - \mathbb{E}\left[g(\mathbf{x}_i) r_\tau(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}\right]| \\
&= |\mathbb{E}\left[\left(g(\mathbf{x}_i) - \mathbb{E}\left[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\right]\right) r_\tau(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}\right]| \le \tau \left|\mathbb{E}\left[|g(\mathbf{x}_i) - \mathbb{E}\left[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\right]||\mathcal{C}_{j,k}\right]\right|
\end{aligned}$$

Summing across j and k, then by similar argument as in the proof of Lemma SA.9,

$$\begin{aligned}
\mathbb{E}\left[|\Delta_i(g, r_\tau)|\right] &\le (1+\rho)\tau N \mathbb{E}\left[|g(\mathbf{x}_i) - \Pi_0(\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)])g(\mathbf{x}_i)|\right] \\
&\le (1+\rho)\tau N \left(\sup_{\mathbf{x}} f_X(\mathbf{x})\right)^2 2^M \mathfrak{m}(\mathcal{V}) \|\mathcal{V}\|_\infty \mathtt{TV}_{\{g\}}.
\end{aligned}$$

For each fixed j, $\widetilde{e}_{j,k}(\mathbf{x}, y)$ can be non-zero for only one k. Hence, almost surely,

$$\begin{aligned}
|\Delta_i(g, r_\tau)| &= \Big|\sum_{j=1}^N \sum_{0 \le k < 2^{M+N-j}} (\widetilde{\gamma}_{j,k}(g, r_\tau) - \widetilde{\beta}_{j,k}(g, r_\tau)) \widetilde{e}_{j,k}(\mathbf{x}_i, y_i)\Big| \\
&\le \sum_{j=1}^N \max_{0 \le k < 2^{M+N-j}} \left|\widetilde{\gamma}_{j,k}(g, r_\tau) - \widetilde{\beta}_{j,k}(g, r_\tau)\right| \le 2 \sum_{j=0}^{N-1} \max_{0 \le k < 2^{M+N-j}} |\gamma_{j,k}(g, r_\tau) - \beta_{j,k}(g, r_\tau)| \\
&\le 2\tau \sum_{j=0}^{N-1} \max_{0 \le k < 2^{M+N-j}} \left|\mathbb{E}\left[|g(\mathbf{x}_i) - \mathbb{E}\left[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\right]||\mathcal{C}_{j,k}\right]\right| \le 2N\tau \min\{2\mathtt{M}_{\mathcal{G}}, \mathtt{L}_{\mathcal{G}} \|\mathcal{V}\|_\infty\}.
\end{aligned}$$

This shows the results. □

Lemma SA.26. *Suppose g and F are functions from $\mathbb{R}$ to $\mathbb{R}$, where F is bounded and non-decreasing. Suppose T is an interval in $\mathbb{R}$ such that $\inf_{t \in T} g(t) \le 0 \le \sup_{t \in T} g(t)$. Suppose we also have $\mathtt{pTV}_{\{g\},T} :=$*

$\sup_{n\geq 1}\sup_{x_1\leq\cdots\leq x_n\in T}\sum_{i=1}^n |g(x_{i+1})-g(x_i)|<\infty$. *Then*

$$\int_T |g(x)|dF(x) \leq \mathtt{pTV}_{\{g\},T}\int_T 1dF(x).$$

SA-III.14 Proof of Lemma SA.26

The result follows from the observation that for any $x\in T$, $|g(x)|\leq \mathtt{pTV}_{\{g\},T}$. □

SA-III.15 Proof of Lemma SA.22

Denote by $\mathcal{B}$ the σ-algebra generated by $\{\mathbb{1}(\mathcal{C}_{0,k})=\mathbb{1}(\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}): 0\leq k<2^{M+N}, k=2^N l+m\}$. Then

$$\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]-g(\mathbf{x}_i)r_\tau(y_i)=\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]-\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i)+\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i)-g(\mathbf{x}_i)r_\tau(y_i).$$

The first two terms are driven by projection of r_τ on grids $\mathcal{Y}_{l,j,m}$'s, and can be upper bounded through probability measure assigned to each grid (2^{-N}) and total variation of r_τ. We consider the random variable $\mathbb{E}[g(\mathbf{x}_i)\mathbb{1}(g(\mathbf{x}_i)>0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]r_\tau(y_i)$. Take $m_{j,k}^+ := \mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)\mathbb{1}(g(\mathbf{x}_i)>0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]$. Apply Lemma SA.26 with $g(y)=\mathbb{E}[g(\mathbf{x}_i)\mathbb{1}(g(\mathbf{x}_i)>0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]r_\tau(y)-m_{j,k}^+$, $F(y)=\mathbb{P}(y_i\leq y|\mathbf{x}_i=\mathbf{x})$ and interval $T=\mathcal{Y}_{l,j,m}$, to get for each $0\leq l<2^M, 0\leq m<2^N$ and $\mathbf{x}\in\mathcal{X}_{0,l}$,

$$\begin{aligned}&\mathbb{E}\left[|(\mathbb{E}[g(\mathbf{x}_i)\mathbb{1}(g(\mathbf{x}_i)>0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]r_\tau(y_i)-m_{j,k}^+)\mathbb{1}(y_i\in\mathcal{Y}_{l,j,m})|\Big|\mathbf{x}_i=\mathbf{x}\right]\\&\quad\leq\mathbb{P}\left(y_i\in\mathcal{Y}_{l,j,m}|\mathbf{x}_i=\mathbf{x}\right)\mathtt{M}_{\{g\}}\mathtt{TV}_{\{r_\tau|_{\mathcal{Y}_{l,j,m}}\}}.\end{aligned}$$

Similarly, take $m_{j,k}^- := \mathbb{E}[g(\mathbf{x}_i)r(y_i)\mathbb{1}(g(\mathbf{x}_i)<0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]$, and we have for $\mathbf{x}\in\mathcal{X}_{0,l}$,

$$\begin{aligned}&\mathbb{E}\left[|(\mathbb{E}[g(\mathbf{x}_i)\mathbb{1}(g(\mathbf{x}_i)<0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]r_\tau(y_i)-m_{j,k}^-)\mathbb{1}(y_i\in\mathcal{Y}_{l,j,m})|\Big|\mathbf{x}_i=\mathbf{x}\right]\\&\quad\leq\mathbb{P}\left(y_i\in\mathcal{Y}_{l,j,m}|\mathbf{x}_i=\mathbf{x}\right)\mathtt{M}_{\{g\}}\mathtt{TV}_{\{r_\tau|_{\mathcal{Y}_{l,j,m}}\}}.\end{aligned}$$

Combining the two parts and integrate over the event $\mathbf{x}_i\in\mathcal{X}_{0,l}$

$$\begin{aligned}&\mathbb{E}\left[|(\mathbb{E}[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]r(y_i)-m_{0,k})\mathbb{1}(y_i\in\mathcal{Y}_{l,j,m})||\mathbf{x}_i\in\mathcal{X}_{0,l}\right]\\&\quad\leq\mathbb{P}\left(y_i\in\mathcal{Y}_{l,j,m}|\mathbf{x}_i\in\mathcal{X}_{0,l}\right)\mathtt{M}_{\{g\}}\mathtt{TV}_{\{r_\tau|_{\mathcal{Y}_{l,0,m}}\}}\leq 2^{-N}\mathtt{M}_{\{g\}}\mathtt{TV}_{\{r_\tau|_{\mathcal{Y}_{l,0,m}}\}}.\end{aligned}$$

Summing over m, we get for each $0\leq l<2^M$,

$$\mathbb{E}\left[|\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r(y_i)-\mathbb{E}[g(\mathbf{x}_i)r(y_i)|\mathcal{B}]||\mathbf{x}_i\in\mathcal{X}_{0,l}\right]\leq 2^{-N}\mathtt{M}_{\{g\}}\mathtt{TV}_{\{r_\tau\}}.$$

Hence using the polynomial growth of total variation,

$$\mathbb{E}\left[|\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r(y_i)-\mathbb{E}[g(\mathbf{x}_i)r(y_i)|\mathcal{B}]|\right]\leq 2^{-N}\mathtt{M}_{\{g\}}\mathtt{TV}_{\{r_\tau\}}\leq 2^{-N}\mathtt{M}_{\mathcal{G}}\tau.$$

Since $|\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]-\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i)|\leq\mathtt{M}_{\mathcal{G}}\tau$ almost surely,

$$\mathbb{E}\left[(\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r(y_i)-\mathbb{E}[g(\mathbf{x}_i)r(y_i)|\mathcal{B}])^2\right]\leq 2^{-N}\tau^2\mathtt{M}_{\mathcal{G}}^2.$$

The last two terms are essentially driven by the L_2-projection error of g. Denote by $\mathcal{A}$ the σ-algebra generated by $\{\mathbb{1}(\mathcal{X}_{0,l}) : 0 \leq l < 2^M\}$. Then $\mathcal{A} \subseteq \mathcal{B}$. By Jensen's inequality and a similar argument as in the proof of Lemma SA.9,

$$\mathbb{E}\left[(\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i) - g(\mathbf{x}_i)r_\tau(y_i))^2\right] \leq \tau^2\mathbb{E}\left[(g(\mathbf{x}_i) - \mathbb{E}[g(\mathbf{x}_i)|\mathcal{A}])^2\right] \leq (1+\rho)\tau^2\mathtt{V}_{\mathcal{G}}.$$

It then follows that

$$\mathbb{E}\left[\left(\Pi_0 M_n(g, r_\tau) - M_n(g, r_\tau)\right)^2\right] \leq 2\left(2^{-N}\tau^2\mathtt{M}_{\mathcal{G}}^2 + (1+\rho)\tau^2\mathtt{V}_{\mathcal{G}}\right).$$

□

SA-III.16 Proof of Lemma SA.23

We will use a truncation argument for the projection error. First, suppose condition (a) holds. Let $\tau > 0$.

Projection error for truncated processes: By Lemma SA.21, SA.22 and using Bernstein inequality, for all $t > 0$, for each $g \in \mathcal{G}$, $r \in \mathcal{R}$

$$\mathbb{P}\left[|M_n(g, r_\tau) - \Pi_1 M_n(g, r_\tau)| \geq 4\tau\sqrt{(1+\rho)N^2\mathtt{V}_{\mathcal{G}} + 2^{-N}\mathtt{M}_{\mathcal{G}}^2}\sqrt{t} + \frac{4}{3}\tau\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}t\right] \leq 2e^{-t}. \tag{SA-14}$$

Truncation Error: We choose a cutoff τ that satisfies $\tau^{\frac{1}{\alpha}} > \log(2^{N+1})$. Recall Equation SA-11 implies $\max_{0 \leq k < 2^{M+N}} \mathbb{E}\left[|r(y_i)||(\mathbf{x}_i, \mathbf{y}_i) \in C_{0,k}\right] \lesssim C_\alpha N^\alpha$, where $C_\alpha = 1 + (2\alpha)^{\frac{\alpha}{2}}$. The same argument for Equation SA-11 implies $\max_{0 \leq k < 2^{M+N}} \mathbb{E}[(r(y_i))^2|(\mathbf{x}_i, y_i) \in \mathcal{C}_{0,k}] \lesssim 1 + (N\log(2)\sqrt{2\alpha})^{2\alpha} \lesssim C_{2\alpha}N^{2\alpha}$, where $C_{2\alpha} := 1 + (2 \cdot 2\alpha)^{\frac{2\alpha}{2}}$. Hence the following holds almost surely,

$$|\Pi_1 M_n(g, r) - \Pi_1 M_n(g, r_\tau)| \leq \max_{l,m}\left|\mathbb{E}\left[g(\mathbf{x}_i)|\mathcal{X}_{0,l}\right] \cdot \mathbb{E}\left[|r(y_i)|\mathbb{1}(|y_i| \geq \tau^{1/\alpha})\middle|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}\right]\right| \lesssim C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha.$$

Since $\tau^{\frac{1}{\alpha}} > \log(2^{N+1}) > 0.5N$, $\gamma_{0,k} = \beta_{0,k}$ for all k corresponding to $\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}$ for $0 < m < 2^N - 1$, that is, the mismatch only happens at edge cells of y_i, we have

$$\mathbb{E}\left[|\Pi_1 M_n(g, r) - \Pi_1 M_n(g, r_\tau)|^2\right] \lesssim \mathbb{P}\left(\Pi_1 M_n(g, r) - \Pi_1 M_n(g, r_\tau) \neq 0\right) C_{2\alpha}\mathtt{M}_{\mathcal{G}}^2 N^{2\alpha} \leq C_{2\alpha}2^{-N+1}\mathtt{M}_{\mathcal{G}}^2 N^{2\alpha}.$$

Apply Bernstein's inequality for $\Pi_1 M_n(g, r) - \Pi_1 M_n(g, r_\tau)$, for all $t > 0$, with probability at least $1 - 2\exp(-t)$,

$$|\Pi_1 M_n(g, r) - \Pi_1 M_n(g, r_\tau)| \lesssim \sqrt{C_{2\alpha}}2^{-N/2}\mathtt{M}_{\mathcal{G}}N^\alpha\sqrt{t} + C_\alpha\frac{\mathtt{M}_{\mathcal{G}}N^\alpha}{\sqrt{n}}t \leq \sqrt{C_{2\alpha}}2^{-N/2}\mathtt{M}_{\mathcal{G}}N^\alpha\sqrt{t} + C_\alpha\frac{\mathtt{M}_{\mathcal{G}}\tau}{\sqrt{n}}t. \tag{SA-15}$$

Moreover, $\mathbb{V}[M_n(g, r) - M_n(g, r_\tau)] \leq \mathtt{M}_{\mathcal{G}}^2\mathbb{V}[r(y_i) - r_\tau(y_i)] \leq \mathtt{M}_{\mathcal{G}}^2\mathbb{E}[(r(y_i) - r_\tau(y_i))^2] \leq \mathtt{M}_{\mathcal{G}}^2\mathbb{E}[r(y_i)^2\mathbb{1}(|y_i| \geq \tau)] \leq 2^{-N}\mathtt{M}_{\mathcal{G}}^2\max_{0 \leq k < 2^{M+N}}\mathbb{E}[r(y_i)^2|(\mathbf{x}_i, y_i) \in \mathcal{C}_{0,k}] \lesssim C_{2\alpha}\mathtt{M}_{\mathcal{G}}^2 N^{2\alpha}2^{-N}$. By Bernstein inequality and a truncation argument, for all $t > 0$,

$$\begin{aligned}
&\mathbb{P}(\sqrt{n}|M_n(g, r) - M_n(g, r_\tau)| \geq t) \\
&\leq \min_{y > 0} 2\exp\left(-\frac{t^2}{2n\mathbb{V}[M_n(g, r) - M_n(g, r_\tau)] + \frac{2}{3}xy}\right) + 2\mathbb{P}\left(\max_{1 \leq i \leq n}|g(\mathbf{x}_i)(r(y_i) - r_\tau(y_i)| \geq y\right).
\end{aligned}$$

Taking $y = \mathtt{M}_{\mathcal{G}} t^\alpha$, we get for all $t > 0$, with probability at least $1 - 4\exp(-t)$,

$$|M_n(g,r) - M_n(g,r_\tau)| \lesssim \sqrt{C_{2\alpha}} 2^{-N/2} \mathtt{M}_{\mathcal{G}} N^\alpha \sqrt{t} + C_\alpha \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1}. \tag{SA-16}$$

Putting Together: Taking $\tau = t^\alpha > 0.5^\alpha N^\alpha$, we get from Equation SA-14, SA-15 and SA-16 that for all $g \in \mathcal{G}$, $r \in \mathcal{G}$, for all $t > N$, with probability at least $1 - 4n\exp(-t)$,

$$|\Pi_1 M_n(g,r) - M_n(g,r)| \lesssim \sqrt{C_{2\alpha}} \sqrt{(1+\rho) N^2 \mathtt{V}_{\mathcal{G}} + 2^{-N} \mathtt{M}_{\mathcal{G}}^2} t^{\alpha + \frac{1}{2}} + C_\alpha \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1}. \tag{SA-17}$$

The bound for $|\Pi_1 Z_n^M(g,r) - Z_n^M(g,r)|$ follows from the fact that it is a mean-zero Gaussian random variable with variance equal to $\mathbb{V}[\Pi_1 M_n(g,r) - M_n(g,r)]$. The result follows then follows from a union bound over $(g,r) \in (\mathcal{G} \times \mathcal{R})_\delta$.

Now consider the case where condition (b) holds. Condition (b) implies $\mathtt{M}_{\mathcal{R}} \leq 2$. Hence choosing $\tau = 2$, then $M_n(g,r) = M_n(g,r_\tau)$ almost surely for all $g \in \mathcal{G}$, $r \in \mathcal{R}$, that is, there is no truncation error. Hence Equation SA-14 implies Equation SA-17 holds with $\alpha = 0$ and similarly for the Z_n^M counterpart. □

SA-III.17 Proof of Lemma SA.24

By definition of Π_1 and Π_2, by Equation SA-10,

$$\begin{aligned}
\Pi_2 R_n(g,r) - R_n(g,r) &= \Big(\Pi_1 M_n(g,r) - M_n(g,r)\Big) - \Big(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] X_n(g\theta(\cdot,r)) - X_n(g\theta(\cdot,r))\Big),\\
\Pi_2 Z_n^R(g,r) - Z_n^R(g,r) &= \Big(\Pi_1 Z_n^M(g,r) - Z_n^M(g,r)\Big) - \Big(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] Z_n^X(g\theta(\cdot,r)) - Z_n^X(g\theta(\cdot,r))\Big).
\end{aligned}$$

The first two terms on RHS of both lines are bounded from Lemma SA.23. Recall $\mathcal{G} \times \mathcal{V}_{\mathcal{R}} = \{g\theta(\cdot,r) : g \in \mathcal{G}, r \in \mathcal{R}\}$. We know from Lemma SA.9 for all $t > 0$,

$$\begin{aligned}
\mathbb{P}\left(|\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] X_n(g\theta(\cdot,r)) - X_n(g\theta(\cdot,r))| \geq 2\sqrt{\mathtt{V}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}} t} + \frac{4}{3} \cdot \frac{\mathtt{M}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}}}{\sqrt{n}} t\right) &\leq 2\exp(-t),\\
\mathbb{P}\left(|\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] Z_n^X(g\theta(\cdot,r)) - Z_n^X(g\theta(\cdot,r))| \geq 2\sqrt{\mathtt{V}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}} t}\right) &\leq 2\exp(-t).
\end{aligned}$$

Moreover, under condition (a), $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$, hence $\sup_{r\in\mathcal{R}} \sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|r(y_i)||\mathbf{x}_i = \mathbf{x}] \leq 1 + \sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|y_i|^\alpha|\mathbf{x}_i = \mathbf{x}] \lesssim 1 + (\sqrt{\alpha})^{\frac{\alpha}{2}} \leq C_\alpha$ by moment properties of sub-Gaussian random variables. Hence $\mathtt{M}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}} \leq C_\alpha \mathtt{M}_{\mathcal{G}}$. Under condition (b), $\sup_{r\in\mathcal{R}} \|r\|_\infty \leq 2$, hence we also have $\mathtt{M}_{\mathcal{G}\times\mathcal{V}_{\mathcal{R}}} \leq C_\alpha \mathtt{M}_{\mathcal{G}}$. The result then follows from a union bound over $(\mathcal{G}\times\mathcal{R})_\delta$. □

SA-III.18 Proof of Theorem SA.1

We make a reduction via the same Rosenblatt transformation in the proof for Theorem 1. Take $\mathbf{u}_i = \phi_X(\mathbf{x}_i)$ where ϕ_X is defined as in Lemma SA.12. And define $\widetilde{g} = g \circ \phi_X^{-1}$ for each $g \in \mathcal{G}$ and consider $\widetilde{\mathcal{G}} = \{\widetilde{g} : g \in \mathcal{G}\}$. Then for all $g \in \mathcal{G}, r \in \mathcal{R}$,

$$M_n(g,r) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g(\mathbf{x}_i) r(y_i) - \mathbb{E}[g(\mathbf{x}_i) r(y_i)] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{g}(\mathbf{u}_i) r(y_i) - \mathbb{E}[\widetilde{g}(\mathbf{u}_i) r(y_i)] =: \widetilde{M}_n(\widetilde{g}, r).$$

Take $\mathcal{A}_{M,N}(\widetilde{\mathbb{P}},1)$ to be a *axis-aligned cylindered quasi-dyadic expansion of* $\mathbb{R}^{d+1}$*, with depth* M *for the main subspace* $\mathbb{R}^d$ *and depth* N *for the multiplier subspace* $\mathbb{R}$*, with respect to* $\widetilde{\mathbb{P}}$ *the joint distribution of* $(\mathbf{u}_i, y_i)$. Take $\widetilde{Z}_n^M$ to be the mean-zero Gaussian-process in Lemma SA.19 indexed by $\mathcal{G}\times\mathcal{R}$ with the same covariance structure as $\widetilde{M}_n$. Let $(\mathcal{G}\times\mathcal{R})_\delta$ be a $\delta\|\mathtt{M}_{\widetilde{\mathcal{G}}}M_{\mathcal{R}}\|_{\widetilde{\mathbb{P}}}$-net of $\widetilde{\mathcal{G}}\times\mathcal{R}$ with cardinality no greater than $\sup_Q N(\widetilde{\mathcal{G}}\times\mathcal{R}, e_Q, \delta\|\mathtt{M}_{\widetilde{\mathcal{G}}}M_R\|_{\widetilde{\mathbb{P}}})$ where sup is taken over all finite discrete measures on $[0,1]^d\times\mathcal{R}$. By Lemma SA.13, $\sup_Q N(\widetilde{\mathcal{G}}\times\mathcal{R}, e_Q, \delta\|\mathtt{M}_{\widetilde{\mathcal{G}}}M_R\|_{\widetilde{\mathbb{P}}}) \leq N(\delta)$. By Lemma SA.19, the SA error for projected process on δ-net is bounded by: For all $t>0$,

$$\mathbb{P}\Big[\|\Pi_1 M_n - \Pi_1 Z_n^M\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_\alpha\sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}t} + C_\alpha\sqrt{\frac{\mathtt{C}_{\Pi_1(\widetilde{\mathcal{G}}\times\mathcal{R})}}{n}t}\Big] \leq 2N(\delta)e^{-t}.$$

where

$$\mathtt{C}_{\Pi_1(\widetilde{\mathcal{G}}\times\mathcal{R})} = \sup_{f\in\Pi_1(\widetilde{\mathcal{G}}\times\mathcal{R})} \min\left\{\sup_{(j,k)}\left[\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}\widetilde{\beta}^2_{j',k'}(f)\right], \|f\|_\infty^2(M+N)\right\}.$$

Let $f\in\Pi_1(\widetilde{\mathcal{G}}\times\mathcal{R})$. Then there exists $g\in\widetilde{\mathcal{G}}$ and $r\in\mathcal{R}$ such that $f=\Pi_1[g,r]$. Since f is already piecewise-constant, by definition of $\beta_{j,k}$'s and $\gamma_{j,k}$'s, we know $\widetilde{\beta}_{l,m}(f)=\widetilde{\gamma}_{l,m}(g,r)$. Fix (j,k). We consider two cases.

Case 1: $j>N$**.** Then by the design of cell expansions (Section SA-III.1), $\mathcal{C}_{j,k}=\mathcal{X}_{j-N,k}\times\mathbb{R}$. First consider l such that $N\leq j'\leq j$. By definition of $\mathcal{A}_{M,N}(\widetilde{\mathbb{P}},1)$, $\mathcal{U}_j'\subseteq[-2^{-\frac{M+N-j}{d}+2}, 2^{-\frac{M+N-j}{d}+2}]^d$, $\|\mathcal{X}_{j-N,k}\|_\infty\leq 2^{-\frac{M+N-j}{d}+1}$. By definition of $\widetilde{\gamma}_{j',m}$, we have

$$\begin{aligned}\sum_{m:\mathcal{C}_{j',m}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j,k}(\widetilde{g},r)| &\leq 2^{2(M+N-j')}\int_{\mathcal{U}_j'}\int_{\mathcal{X}_{j-N,k}}|\widetilde{g}(\mathbf{x})\theta(\mathbf{x},r)-\widetilde{g}(\mathbf{x}+\mathbf{s})\theta(\mathbf{x}+\mathbf{s},r)|d\mathbf{x}d\mathbf{s}\\ &\leq 2^{2(M+N-j')}\mathtt{K}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}},[0,1]^d}\int_{\mathcal{U}_j'}\|\mathbf{s}\|\|\mathcal{X}_{j-N,k}\|_\infty^{d-1}d\mathbf{s}\\ &\leq 2^{2(M+N-j')}\operatorname{Vol}(\mathcal{U}_j')\|\mathcal{U}_j'\|_\infty\|\mathcal{X}_{j-N,k}\|_\infty^{d-1}\mathtt{K}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}},[0,1]^d}\\ &\leq 2^{\frac{d-1}{d}(j-j')}\mathtt{K}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}},[0,1]^d}.\end{aligned}$$

Then consider j' such that $0\leq j'<N$. Then

$$\begin{aligned}&\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(\widetilde{g},r)|\\ =&\sum_{j':\mathcal{X}_{0,j'}\subseteq\mathcal{X}_{j-N,k}}\sum_{0\leq m<2^{j'}}|\mathbb{E}[\widetilde{g}(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,j'}]|\cdot|\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,j'},y_i\in\mathcal{Y}_{j',j-1,2m}]\\ &\qquad\qquad-\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,j'},y_i\in\mathcal{Y}_{j',j-1,2m+1}]|\\ \leq& C_\alpha\sum_{j':\mathcal{X}_{0,j'}\subseteq\mathcal{X}_{j-N,k}}|\mathbb{E}[\widetilde{g}(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,j'}]|N^\alpha\leq C_\alpha 2^{j-N}\mathtt{M}_{\widetilde{\mathcal{G}}}N^\alpha.\end{aligned}$$

It follows that

$$\begin{aligned}&\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(\widetilde{g},r)|\\&\leq\sum_{N\leq j'<j}(j-j')(j-j'+1)2^{-\frac{j-j'}{d}}\mathtt{K}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}},[0,1]^d}+\sum_{j'<N}(j-j')(j-j'+1)2^{j'-N}\mathtt{M}_{\widetilde{\mathcal{G}}}N^{\alpha}\lesssim\mathtt{K}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}},[0,1]^d}+\mathtt{M}_{\widetilde{\mathcal{G}}}N^{\alpha}.\end{aligned}$$

Case 2: $j\leq N$. Then $\mathcal{C}_{j,k}=\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}$. Hence for any $0\leq j'\leq j$, we have

$$\begin{aligned}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(\widetilde{g},r)|=&|\mathbb{E}[\widetilde{g}(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|\sum_{m':\mathcal{Y}_{l,j',m'}\subseteq\mathcal{Y}_{l,j,m}}|\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l},y_i\in\mathcal{Y}_{l,j-1,2m}]\\&-\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l},y_i\in\mathcal{Y}_{l,j-1,2m+1}]|\\\leq&C_{\alpha}|\mathbb{E}[\widetilde{g}(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|N^{\alpha}\leq C_{\alpha}\mathtt{M}_{\widetilde{\mathcal{G}}}N^{\alpha}.\end{aligned}$$

It follows that

$$\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(\widetilde{g},r)|\leq C_{\alpha}\mathtt{M}_{\widetilde{\mathcal{G}}}N^{\alpha}.$$

Moreover, for all (j,k), we have $\widetilde{\beta}_{j,k}(\widetilde{g},r)\leq C_{\alpha}\mathtt{M}_{\widetilde{\mathcal{G}}}N^{\alpha}$. Now, we bound $\mathtt{K}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}},[0,1]^d}$ in terms of properties of $\widetilde{\mathcal{G}}$ and $\mathcal{V}_{\mathcal{R}}$. Let $\mathcal{C}$ be a cube in $[0,1]^d$ with side length a.

$$\begin{aligned}&\sup_{\varphi\in\mathcal{D}_d(\mathcal{C})}\int g\circ\phi_X^{-1}(\mathbf{u})\theta(\mathbf{x},r)\operatorname{div}(\varphi)(\mathbf{x})d\mathbf{x}/\|\|\varphi\|_2\|_{\infty}\\\leq&\lim_{\varepsilon\downarrow0}\int_{\mathcal{C}}\|\nabla(g_{\varepsilon}\circ\phi_X^{-1}\cdot\theta(\cdot,r))(\mathbf{u})\|_2d\mathbf{u}\\\leq&\lim_{\varepsilon\downarrow0}\int_{\mathcal{C}}\|\nabla(g_{\varepsilon}\circ\phi_X^{-1})(\mathbf{u})\|\theta(\cdot,r)\|_{\infty}+\|g_{\varepsilon}\|_{\infty}\nabla\theta(\mathbf{u},r)\mathbb{1}(\mathbf{u}\in\operatorname{Supp}(g))\|_2d\mathbf{u}\\\leq&\mathtt{K}_{\widetilde{\mathcal{G}}}\mathtt{M}_{\mathcal{V}_{\mathcal{R}}}a^{d-1}+\mathtt{M}_{\widetilde{\mathcal{G}}}\mathtt{L}_{\mathcal{V}_{\mathcal{R}}}a^d\leq(\mathtt{K}_{\widetilde{\mathcal{G}}}\mathtt{M}_{\mathcal{V}_{\mathcal{R}}}+\mathtt{M}_{\widetilde{\mathcal{G}}}\mathtt{L}_{\mathcal{V}_{\mathcal{R}}})a^{d-1}.\end{aligned}$$

Together with Lemma SA.12 for the relation between $\mathtt{K}_{\widetilde{\mathcal{G}}}$ (resp. $\mathtt{M}_{\widetilde{\mathcal{G}}}$) and $\mathtt{K}_{\mathcal{G}}$ (resp. $\mathtt{M}_{\mathcal{G}}$), $\mathtt{K}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}},[0,1]^d}\leq\mathtt{c}_3\mathtt{K}_{\mathcal{G}}\mathtt{M}_{\mathcal{V}_{\mathcal{R}}}+\mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{V}_{\mathcal{R}}}$. Hence

$$\mathtt{C}_{\Pi_1(\widetilde{\mathcal{G}}\times\mathcal{R})}\leq\min\{C_{\alpha}^2(\mathtt{M}_{\mathcal{G}}N^{\alpha})(\mathtt{c}_3\mathtt{K}_{\mathcal{G}}\mathtt{M}_{\mathcal{V}_{\mathcal{R}}}+\mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{V}_{\mathcal{R}}}+\mathtt{M}_{\mathcal{G}}N^{\alpha}),(C_{\alpha}\mathtt{M}_{\mathcal{G}}N^{\alpha})^2(M+N)\}.$$

Since $\mathbf{u}_i\overset{i.i.d}{\sim}\operatorname{Unif}([0,1]^d)$ and the cells $\mathcal{A}_{M,N}(\widetilde{\mathbb{P}},1)$ are obtained via *axis aligned dyadic expansion* and $\mathbf{u}_i$ is uniformly distributed on $[0,1]^d$, we have $\|\mathcal{X}_{0,k}\|_{\infty}\leq2^{-\lfloor M/d\rfloor}$ for all $0\leq k<2^M$. Then by Lemma SA.23 with $\rho=1$, for all $t>N$,

$$\mathbb{P}\Big[\|M_n-\Pi_1M_n\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_{\delta}}>\sqrt{2N^2\mathtt{V}_{\widetilde{\mathcal{G}}}+2^{-N}\mathtt{M}_{\widetilde{\mathcal{G}}}^2}t^{\alpha+\frac12}+\frac{\mathtt{M}_{\widetilde{\mathcal{G}}}}{\sqrt{n}}t^{\alpha+1}\Big]\leq4\mathtt{N}(\delta)ne^{-t},$$
$$\mathbb{P}\Big[\|Z_n^M-\Pi_1Z_n^M\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_{\delta}}>\sqrt{2N^2\mathtt{V}_{\widetilde{\mathcal{G}}}+2^{-N}\mathtt{M}_{\widetilde{\mathcal{G}}}^2}t^{\frac12}+\frac{\mathtt{M}_{\widetilde{\mathcal{G}}}}{\sqrt{n}}t\Big]\leq4\mathtt{N}(\delta)ne^{-t},$$

where

$$\mathtt{V}_{\widetilde{\mathcal{G}}}=\sqrt{d}\min\{2\mathtt{M}_{\widetilde{\mathcal{G}}},\mathtt{L}_{\widetilde{\mathcal{G}}}2^{-\lfloor M/d\rfloor}\}2^{-\lfloor M/d\rfloor}\mathtt{TV}_{\widetilde{\mathcal{G}}}.$$

We find the optimal parameters M^* and N^* by balancing the two terms, choosing either

$$2^{M^*} = \min\left\{\left(\frac{n\mathtt{TV}_{\widetilde{\mathcal{G}}}}{\mathtt{E}_{\widetilde{\mathcal{G}}}}\right)^{\frac{d}{d+1}}, \left(\frac{n\mathtt{L}_{\widetilde{\mathcal{G}}}\mathtt{TV}_{\widetilde{\mathcal{G}}}}{\mathtt{E}_{\widetilde{\mathcal{G}}}\mathtt{M}_{\widetilde{\mathcal{G}}}}\right)^{\frac{d}{d+2}}\right\}, \quad 2^{N^*} = \max\left\{\left(\frac{n\mathtt{M}_{\widetilde{\mathcal{G}}}^{d+1}}{\mathtt{E}_{\widetilde{\mathcal{G}}}\mathtt{TV}_{\widetilde{\mathcal{G}}}^{d}}\right)^{\frac{1}{d+1}}, \left(\frac{n^2\mathtt{M}_{\widetilde{\mathcal{G}}}^{2d+2}}{\mathtt{TV}_{\widetilde{\mathcal{G}}}^{d}\mathtt{L}_{\widetilde{\mathcal{G}}}^{d}\mathtt{E}_{\widetilde{\mathcal{G}}}^{2}}\right)^{\frac{1}{d+2}}\right\}.$$

It follows that for all $t > N_*$, with probability at least $1 - 4nN(\delta)\exp(-t)$,

$$\|M_n - Z_n^M\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta} \leq \sqrt{d}N^* \min\left\{\left(\frac{\mathtt{E}_{\widetilde{\mathcal{G}}}\mathtt{TV}_{\widetilde{\mathcal{G}}}^{d}\mathtt{M}^{d+1}}{n}\right)^{\frac{1}{2(d+1)}}, \left(\frac{\mathtt{E}_{\widetilde{\mathcal{G}}}^2\mathtt{M}_{\widetilde{\mathcal{G}}}^2\mathtt{TV}_{\widetilde{\mathcal{G}}}^{d}\mathtt{L}_{\widetilde{\mathcal{G}}}^{d}}{n^2}\right)^{\frac{1}{2(d+2)}}\right\} t^{\alpha+\frac{1}{2}} + \sqrt{\frac{\mathtt{C}_{\Pi_1(\widetilde{\mathcal{G}}\times\mathcal{R})}}{n}t^{\alpha+1}}.$$

Moreover by Lemma SA.18 we bound fluctuation off-the-net by, for all $t > 0$,

$$\mathbb{P}\left[\|\widetilde{M}_n - \widetilde{M}_n \circ \pi_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta}\|_{\widetilde{\mathcal{G}}\times\mathcal{R}} > C_\alpha \widetilde{\mathsf{F}}_n(t,\delta)\right] \leq \exp(-t),$$
$$\mathbb{P}\left[\|\widetilde{Z}_n^M \circ \pi_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta} - \widetilde{Z}_n^M\|_{\widetilde{\mathcal{G}}\times\mathcal{R}} > C(\mathtt{M}_{\widetilde{\mathcal{G}}\times\mathcal{R}}J(\delta,\widetilde{\mathcal{G}}\times\mathcal{R},\mathtt{M}_{\widetilde{\mathcal{G}}\times\mathcal{R}}) + \delta\mathtt{M}_{\widetilde{\mathcal{G}}\times\mathcal{R}}\sqrt{t})\right] \leq \exp(-t),$$

where

$$\widetilde{\mathsf{F}}_n(t,\delta) := J(\delta,\widetilde{\mathcal{G}}\times\mathcal{R},\mathtt{M}_{\widetilde{\mathcal{G}}}M_{\mathcal{R}})\mathtt{M}_{\widetilde{\mathcal{G}}} + \frac{\log(n)\mathtt{M}_{\widetilde{\mathcal{G}}}J^2(\delta,\widetilde{\mathcal{G}}\times\mathcal{R},\mathtt{M}_{\widetilde{\mathcal{G}}}M_{\mathcal{R}})}{\delta^2\sqrt{n}} + \frac{\mathtt{M}_{\widetilde{\mathcal{G}}}}{\sqrt{n}}t + (\log n)^\alpha\frac{\mathtt{M}_{\widetilde{\mathcal{G}}}}{\sqrt{n}}t^\alpha.$$

The result then follows from the relation between $\mathcal{G}$ quantities and $\widetilde{\mathcal{G}}$ quantities in Lemma SA.12 and the decomposition that

$$\begin{aligned}
\|M_n - Z_n^M\|_{\mathcal{G}\times\mathcal{R}} &= \|\widetilde{M}_n - Z_n^M\|_{\widetilde{\mathcal{G}}\times\mathcal{R}} \\
&\leq \|\widetilde{M}_n - \widetilde{M}_n \circ \pi_{(\widetilde{\mathcal{G}}\times\mathcal{H})_\delta}\|_{\widetilde{\mathcal{G}}\times\mathcal{R}} + \|Z_n^M - Z_n^M \circ \pi_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta}\|_{\widetilde{\mathcal{G}}\times\mathcal{R}} \\
&+ \|\widetilde{M}_n - \Pi_1\widetilde{M}_n\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta} + \|Z_n^M - \Pi_1 Z_n^M\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta} + \|\Pi_1\widetilde{M}_n - \Pi_1 Z_n^M\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta},
\end{aligned}$$

where we have abused the notation to mean the same thing by $Z_n^M(g,r)$ and $Z_n^M(\widetilde{g},r)$. □

SA-III.19 Proof of Theorem SA.2

Suppose $2^M \leq J < 2^{M+1}$. For each $l \in [d]$, we can divide at most 2^M cells into two intervals of equal measure under $\mathbb{P}_X$ such that we get a new partition of $\mathcal{X} = \sqcup_{0\leq j<2^{M+1}}\Delta_l'$ and satisfies

$$\frac{\max_{0\leq l<2^{M+1}}\mathbb{P}_X(\Delta_l')}{\min_{0\leq l<2^{M+1}}\mathbb{P}_X(\Delta_l')} \leq 2\rho.$$

By construction, for each $N \in \mathbb{N}$, there exists an axis-aligned quasi-dyadic expansion $\mathcal{A}_{M+1,N}(\mathbb{P},2\rho) = \{\mathcal{C}_{j,k} : 0 \leq j \leq M+1+N, 0 \leq k < 2^{M+1+N-j}\}$ such that

$$\left\{\mathcal{X}_{0,k} : 0 \leq k < 2^{M+1}\right\} = \left\{\Delta_l' : 0 \leq l < 2^{M+1}\right\},$$

and $\mathcal{G} \subseteq \text{Span}\{\mathbb{1}_{\Delta_j} : 0 \leq j < J\} \subseteq \text{Span}\{\mathcal{X}_{0,k} : 0 \leq k < 2^{M+1}\}$. Hence

$$\Pi_0(g,r) = \Pi_1(g,r) = \sum_{0\leq l<2^{K+1}}\sum_{0\leq m<2^N}\mathbb{1}(\mathcal{X}_{0,l}\times\mathcal{Y}_{j,l,m})g|_{\mathcal{X}_{0,l}}\mathbb{E}[r(y_i)|\mathbf{x}_i \in \mathcal{X}_{0,l}, y_i \in \mathcal{Y}_{j,l,m}]. \quad \text{(SA-18)}$$

Again, consider $(\mathcal{G}\times\mathcal{R})_\delta$ which is a $\delta\|\mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}\|_{\widetilde{\mathbb{P}}}$ of $\mathcal{G}\times\mathcal{R}$ of cardinality no greater than $N(\delta)$. The SA error for projected process on the δ-net is given by Lemma SA.20: For all $t>0$,

$$\begin{aligned}&\mathbb{P}\left[\|\Pi_1 M_n - \Pi_1 Z_n^M\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_\alpha\sqrt{\frac{N^{2\alpha+1}2^{M+1}\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}t} + C_\alpha\sqrt{\frac{\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R})}}{n}t}\right]\\ \leq\quad & 2\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)e^{-t} + 2^M\exp\left(-C_\rho n 2^{-M}\right).\end{aligned}$$

Now we find an upper bound for $\mathtt{C}_{\mathcal{G}\times\mathcal{R}}$. Consider the following two cases.

Case 1: $j\geq N$ Let $g\in\mathcal{G}, r\in\mathcal{R}$. Fix (j,k). Let (j',m') be an index such that $\mathcal{C}_{j',m'}\subseteq\mathcal{C}_{j,k}$. If $N\leq j'\leq M+N$, then by definition of S and the step of splitting each cell into at most two, there exists $l_1,\cdots,l_{2S}\in\{0,\cdots,2^{M+1}-1\}$ with possible duplication such that $g=\sum_{q=1}^{2S} c_q\mathbb{1}(\Delta'_{l_q})$ where $|c_q|\leq\mathtt{M}_{\{g\}}$. Since each Δ'_{l_q} belongs to at most one $\mathcal{X}_{j'-N,k}$, $\widetilde{\gamma}_{j',m'}(\mathbb{1}(\Delta'_{l_q}),r)=0$ if Δ'_{l_q} is not contained in $\mathcal{X}_{j'-N,m'}$ and $|\widetilde{\gamma}_{j',m'}(\mathbb{1}(\Delta'_{l_q}),r)|\leq C_\alpha 2^{-l+1}$ if $\Delta'_{l_q}\subseteq\mathcal{X}_{j'-N,m'}$ where $C_\alpha = 1+(2\sqrt{\alpha})^\alpha$. For j' such that $N\leq j'\leq j$,

$$\sum_{m':\mathcal{C}_{j',m'}\subseteq\mathcal{C}_{j,k}}\left|\widetilde{\gamma}_{j',m'}(g,r)\right|^2 \leq 2S\sum_{q=1}^{2S}\sum_{m':\mathcal{C}_{j',m'}\subseteq\mathcal{C}_{j,k}}\left(c_q\widetilde{\gamma}_{j',m'}(\mathbb{1}(\Delta_{l_q}),r)\right)^2 \leq 2C_\alpha^2 S\sum_{q=1}^{2S}c_q^2 2^{-2l}\leq 4C_\alpha^2S^2\mathtt{M}_{\mathcal{G}}^2 2^{-2l}.$$

For $0\leq j'\leq j$,

$$\begin{aligned}&\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(g,r)|\\ =&\sum_{l:\mathcal{X}_{0,l}\subseteq\mathcal{X}_{j-N,k}}\sum_{0\leq m<2^{j'}}|\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|\cdot|\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l},y_i\in\mathcal{Y}_{l,j-1,2m}]\\ &\qquad\qquad - \mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l},y_i\in\mathcal{Y}_{l,j-1,2m+1}]|\\ \leq& C_\alpha\sum_{l:\mathcal{X}_{0,l}\subseteq\mathcal{X}_{j-N,k}}|\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|N^\alpha\leq C_\alpha 2^{j-N}\mathtt{M}_{\mathcal{G}}N^\alpha.\end{aligned}$$

Since $|\widetilde{\gamma}_{l,m}(g,r)|\lesssim C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha$ for all (l,m), $\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}\widetilde{\gamma}^2_{j',k'}(g,r)\leq C_\alpha^2 2^{j-N}\mathtt{M}_{\mathcal{G}}^2N^{2\alpha}$. Putting together

$$\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}\widetilde{\gamma}^2_{j',k'}(g,r)\lesssim C_\alpha^2S^2\mathtt{M}_{\mathcal{G}}^2 + C_\alpha^2\mathtt{M}_{\mathcal{G}}^2N^{2\alpha}.$$

Case 2: $l<N$ Hence for any $0\leq j'\leq j$, we have

$$\begin{aligned}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(\widetilde{g},r)| =&|\mathbb{E}[\widetilde{g}(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|\sum_{m':\mathcal{Y}_{l,j',m'}\subseteq\mathcal{Y}_{l,j,m}}|\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l},y_i\in\mathcal{Y}_{l,j-1,2m}]\\ &- \mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l},y_i\in\mathcal{Y}_{l,j-1,2m+1}]|\\ \leq& C_\alpha|\mathbb{E}[\widetilde{g}(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|N^\alpha\leq C_\alpha\mathtt{M}_{\widetilde{\mathcal{G}}}N^\alpha.\end{aligned}$$

It follows that

$$\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(\widetilde{g},r)|\leq C_\alpha\mathtt{M}_{\widetilde{\mathcal{G}}}N^\alpha.$$

It follows that

$$\begin{aligned}\mathtt{C}_{\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R})}} &= \sup_{h\in\mathcal{H}} \min\left\{ \sup_{(j,k)} \left[\sum_{l<j} (j-l)(j-l+1)2^{l-j} \sum_{m:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}} \widetilde{\gamma}_{l,m}^2(h) \right], \mathtt{M}^2_{\Pi_1(\mathcal{G}\times\mathcal{R})}(M+N) \right\} \\ &\leq C_\alpha^2 \mathtt{M}_{\mathcal{G}}^2 N^{2\alpha} \min\{M+N, S^2+1\}.\end{aligned}$$

By the characterization of projections in Equation SA-18, we know the mis-specification error is zero, that is, $\Pi_1 M_n(g,r) = \Pi_0 M_n(g,r)$ and $\Pi_1 Z_n^M(g,r) = \Pi_0 M_n(g,r)$. Since g is already piecewise-constant on $\mathcal{X}_{0,l}$'s, the L_2-projection error is solely contributed from r. Consider $\mathcal{B} = \sigma\left(\left\{\mathbb{1}_{\mathcal{C}_{0,k}} : 0 \leq k < 2^{M+N+1}\right\}\right)$. Denote $r_\tau = r|_{[-\tau^{1/\alpha},\tau^{1/\alpha}]}$. Then

$$|\mathbb{E}\left[g(\mathbf{x}_i) r_\tau(y_i)|\mathcal{B}\right] - g(\mathbf{x}_i) r_\tau(y_i)| \leq \mathtt{M}_{\mathcal{G}} \left| r_\tau(y_i) - \mathbb{E}[r_\tau(y_i)|\mathcal{B}]\right|.$$

Then by the same argument as in the proof for Lemma SA.22 and the argument for truncation error in the proof for Lemma SA.23, for all $t > N$,

$$\mathbb{P}\left(\|M_n - \Pi_1 M_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|Z_n^M - \Pi_1 Z_n^M\|_{(\mathcal{G}\times\mathcal{R})_\delta} \geq N\sqrt{2^{-N}\mathtt{M}_{\mathcal{G}}^2} t^{\alpha+\frac{1}{2}} + \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1} \right) \leq 4N(\delta) n e^{-t}. \quad \text{(SA-19)}$$

Then apply Lemma SA.20, we get there exists a mean-zero Gaussian process Z_n^M with the same covariance structure as M_n such that with probability at least $1 - 2\exp(-t) - 2^{M+1}\exp(-C_\rho n 2^{-M-1})$,

$$\|\Pi_1 M_n - \Pi_1 Z_n^M\|_{\mathcal{G}\times\mathcal{R}} \leq C_\rho \min_{\delta\in(0,1)} \left\{ \sqrt{\frac{2^{M+2}\mathtt{M}_{\mathcal{G}}\mathtt{E}_{\mathcal{G}}}{n}} (t + \log N(\delta))^{\alpha+\frac{1}{2}} + \sqrt{\frac{\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R})}}{n}} (t+\log N(\delta))^{\alpha+1} + F_n(t,\delta) \right\}.$$

□

SA-III.20 Proof of Theorem SA.3

We will use the same Rosenblatt transformation as in Theorem SA.1. Taking $\mathbf{u}_i = \phi_X(\mathbf{x}_i)$ and $\widetilde{\mathcal{G}} = \{\widetilde{g} : g \in \mathcal{G}\}$ with $\widetilde{g} = g \circ \phi_X^{-1}$, we have

$$R_n(g,r) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g(\mathbf{x}_i) r(y_i) - g(\mathbf{x}_i)\mathbb{E}[r(y_i)|\mathbf{x}_i] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \widetilde{g}(\mathbf{u}_i) r(y_i) - \widetilde{g}(\mathbf{u}_i)\mathbb{E}[r(y_i)|u_i] =: \widetilde{R}_n(\widetilde{g}, r).$$

Denote by $\widetilde{P}$ the joint distribution of $(\mathbf{u}_i, y_i)$. Take $\mathcal{A}_{M,N}(\widetilde{\mathbb{P}}, 1)$ to be the axis-aligned cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$. Then by Lemma SA.19 and Lemma SA.24, for all $t > N$,

$$\begin{aligned}\mathbb{P}\left[\|\Pi_2 R_n - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_\alpha \sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}} t + C_\alpha \sqrt{\frac{\mathtt{C}_{\Pi_2(\widetilde{\mathcal{G}}\times\mathcal{R})}}{n}} t \right] &\leq 2\mathtt{N}(\delta) e^{-t}, \\ \mathbb{P}\left[\|R_n - \Pi_2 R_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_\alpha \sqrt{2N^2\mathtt{V} + 2^{-N}\mathtt{M}_{\mathcal{G}}^2} t^{\alpha+\frac{1}{2}} + C_\alpha \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1} \right] &\leq 4\mathtt{N}(\delta) n e^{-t}, \\ \mathbb{P}\left[\|Z_n^R - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_\alpha \sqrt{2N^2\mathtt{V} + 2^{-N}\mathtt{M}_{\mathcal{G}}^2} t^{\frac{1}{2}} + C_\alpha \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t \right] &\leq 4\mathtt{N}(\delta) n e^{-t},\end{aligned}$$

where $\mathtt{V} = \sqrt{d} \min\left\{2\mathtt{M}_{\mathcal{G}}, \mathtt{L}2^{-\lfloor M/d \rfloor}\right\} 2^{-\lfloor M/d \rfloor}\mathtt{TV}$, and

$$\mathtt{C}_{\Pi_2(\widetilde{\mathcal{G}}\times\mathcal{R})} = \sup_{f \in \Pi_2(\widetilde{\mathcal{G}}\times\mathcal{R})} \min\left\{\sup_{(j,k)}\left[\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}\widetilde{\beta}^2_{j',k'}(f)\right], \|f\|_\infty^2(M+N)\right\}.$$

Let $f \in \Pi_2(\widetilde{\mathcal{G}}\times\mathcal{R})$. Then there exists $g \in \widetilde{\mathcal{G}}$ and $r \in \mathcal{R}$ such that $f = \Pi_2[g,r]$. Since f is already piecewise-constant, by definition of $\beta_{j,k}$'s and $\eta_{j,k}$'s, we know $\widetilde{\beta}_{l,m}(f) = \widetilde{\eta}_{l,m}(g,r)$. Fix (j,k). We consider two cases.

Case 1: $j > N$. Then by the design of cell expansions (Section SA-III.1), $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k}\times\mathbb{R}$. By definition of $\eta_{l,m}$, for any $N \le j' \le j$, we have $(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}\widetilde{\eta}^2_{j',k'}(g,r) = 0$. Now consider $0 \le j' < N$. Then

$$\begin{aligned}
&\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\eta}_{j',k'}(g,r)| \\
=&\sum_{l:\mathcal{X}_{0,l}\subseteq\mathcal{X}_{j-N,k}}\sum_{0\le m<2^{j'}}|\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|\cdot|\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m}] \\
&\qquad - \mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m+1}]| \\
\le& C_\alpha\sum_{l:\mathcal{X}_{0,l}\subseteq\mathcal{X}_{j-N,k}}|\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|N^\alpha \le C_\alpha 2^{j-N}\mathtt{M}_{\mathcal{G}}N^\alpha.
\end{aligned}$$

It follows that

$$\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\eta}_{j',k'}(g,r)| \le \sum_{j'<j}(j-j')(j-j'+1)2^{j'-N}C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha \lesssim C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha.$$

Case 2: $j \le N$. Then $\mathcal{C}_{j,k} = \mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}$. Hence for any $0 \le j' \le j$, we have

$$\begin{aligned}
\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\eta}_{j',k'}(g,r)| =& |\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|\sum_{m':\mathcal{Y}_{l,j',m'}\subseteq\mathcal{Y}_{l,j,m}}|\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m}] \\
&- \mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m+1}]| \\
\lesssim& C_\alpha|\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|N^\alpha \lesssim C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha.
\end{aligned}$$

It follows that

$$\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\eta}_{j',k'}(g,r)| \lesssim C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha.$$

Moreover, for all (j,k), we have $\widetilde{\beta}_{j,k}(g,r) \lesssim C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha$. Hence $\mathtt{C}_{\Pi_2(\widetilde{\mathcal{G}}\times\mathcal{R})} \lesssim (C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha)^2$. The rest of the proofs follow from choosing optimal M, N and Lemma SA.18 in the same way as in the proof for Theorem SA.1. $\square$

SA-III.21 Proof of Theorem SA.4

Suppose $2^M \leq J < 2^{M+1}$. By the same cell divisions in the proof for Theorem SA.2, there exists a quasi-dyadic expansion $\mathcal{C}_{M+1,N}$ such that

$$\operatorname{Span}\left(\{\mathbb{1}(\Delta_j) : 0 \leq j < J\}\right) \subseteq \operatorname{Span}\left(\left\{\mathbb{1}(\mathcal{X}_{0,l}) : 0 \leq l < 2^{M+1}\right\}\right).$$

By definition, the projection error can be decomposed as

$$R_n(g,r) - \Pi_2 R_n(g,r) = M_n(g,r) - \Pi_1 M_n(g,r) + X_n(g\theta(\cdot,r)) - \Pi_0 X_n(g\theta(\cdot,r)),$$

where Π_0 denotes the L_2-projection from $L_2(\mathbb{R}^d)$ to $\operatorname{Span} = \{\mathbb{1}(\mathcal{X}_{0,l}) : 0 \leq l < 2^{M+1}\}$. Then

$$\begin{aligned}\mathbb{E}\left[(X_n(g\theta(\cdot,r)) - \Pi_0 X_n(g\theta(\cdot,r)))^2\right] &= \sum_{0\leq j<J} \mathbb{P}_X(\Delta_j) g^2|_{\Delta_j} \mathbb{E}\left[(\theta(\mathbf{x}_i,\mathbf{x}) - \Pi_0\theta(\mathbf{x}_i,\mathbf{x}))^2 | \mathbf{x}_i \in \Delta_j\right] \\ &\leq \mathbb{E}[g(\mathbf{x}_i)^2] \max_{0\leq j<J} \|\Delta_j\|_\infty^2 \|\theta(\cdot,r)\|_{\mathrm{Lip}}^2 \leq \mathtt{M}_{\mathcal{G}} \mathtt{E}_{\mathcal{G}} \max_{0\leq j<J} \|\Delta_j\|_\infty^2 \mathtt{L}_{\mathcal{V}_{\mathcal{R}}}^2.\end{aligned}$$

Then $X_n(g\theta(\cdot,r)) - \Pi_0 X_n(g\theta(\cdot,r))$ is bounded through Bernstein inequality and union bound, for all $t > 0$,

$$\mathbb{P}\left(\|X_n(g\theta(\cdot,r)) - \Pi_0 X_n(g\theta(\cdot,r))\|_{(\mathcal{G}\times\mathcal{R})_\delta} \geq \frac{4}{3}\sqrt{\mathtt{M}_{\mathcal{G}}\mathtt{E}_{\mathcal{G}}} \max_{0\leq j<J} \|\Delta_j\|_\infty \mathtt{L}_{\mathcal{V}_{\mathcal{R}}} \sqrt{t} + 2\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t\right) \leq 2\exp(-t).$$

Combining with Lemma SA.20 and Equation SA-19, and the same calculation as in the proof for Theorem SA.3 to get $\mathtt{C}_{\Pi_2(\mathcal{G},\mathcal{R})} \lesssim (C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha)^2$, for all $t > N_*$, with probability at least $1 - 2N(\delta)e^{-t} - 2^M \exp(-C_\rho n 2^{-M})$,

$$\|R_n - Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} \leq \frac{4}{3}\sqrt{\mathtt{M}_{\mathcal{G}}\mathtt{E}_{\mathcal{G}}} \max_{0\leq j<J} \|\Delta_j\|_\infty \mathtt{L}_{\mathcal{V}_{\mathcal{R}}} \sqrt{t} + C_\alpha N_* \sqrt{\frac{J\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}} t^{\alpha+\frac{1}{2}} + C_\alpha \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1},$$

The rest follows from the error for fluctuation off the δ-net given in Lemma SA.18. Notice that the "bias" term $\sqrt{\mathtt{M}_{\mathcal{G}}\mathtt{E}_{\mathcal{G}}} \max_{0\leq j<J} \|\Delta_j\|_\infty \mathtt{L}_{\mathcal{V}_{\mathcal{R}}} \sqrt{t}$ comes from $X_n(g\theta(\cdot,r)) - \Pi_0 X_n(g\theta(\cdot,r))$ in the decomposition.

In the special case that we have a singleton $\mathcal{R} = \{r\}$, we can get rid of the "bias" term by redefining $\varepsilon_i = \operatorname{sign}(r(y_i) - \mathbb{E}[r(y_i)|\mathbf{x}_i])|r(y_i) - \mathbb{E}[r(y_i)|\mathbf{x}_i]|^{1/\alpha}$. Take $\widetilde{r}(u) = \operatorname{sign}(u)|u|^\alpha, u \in \mathbb{R}$. In particular, $\mathbb{E}[\widetilde{r}(\varepsilon_i)|\mathbf{x}_i] = 0$ almost surely. Either r is bounded and we can take $\alpha = 0$, which makes $\widetilde{r}$ also bounded; or $\alpha > 0$ and $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$ and $|r(u)| \lesssim 1 + |u|^\alpha$, which implies $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(\varepsilon_i)|\mathbf{x}_i = \mathbf{x}] \lesssim 2$ and $\widetilde{r}$ has polynomial growth. Then for any $g \in \mathcal{G}$,

$$R_n(g,r) = \frac{1}{\sqrt{n}} \sum_{i=1}^n g(\mathbf{x}_i)\widetilde{r}(\varepsilon_i) - \mathbb{E}[g(\mathbf{x}_i)\widetilde{r}(\varepsilon_i)] =: M_n'(g,\widetilde{r}),$$

where M_n' denotes the empirical process based on random sample $((\mathbf{x}_i,\varepsilon_i) : 1 \leq i \leq n)$. The result then follows from Theorem SA.2. By similar arguments as in the proof of Theorem SA.4,

$$\mathtt{C}_{\Pi_1(\mathcal{G},\{\widetilde{r}\})} = \sup_{f\in\Pi_1(\mathcal{G},\{\widetilde{r}\})} \min\left\{\sup_{(j,k)} \left[\sum_{j'<j} (j-j')(j-j'+1)2^{j'-j} \sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}} \widetilde{\beta}_{j',k'}^2(f)\right], \|f\|_\infty^2 (M+N)\right\},$$

but $\widetilde{\beta}_{j,k}(f)$ vanishes for all $j > N$ and we obtain similarly $\mathtt{C}_{\Pi_1(\mathcal{G},\{\widetilde{r}\})} \lesssim (C_\alpha \mathtt{M}_{\mathcal{G}} N^\alpha)^2$. □

SA-III.22 Proof of Lemma SA.1

Here we concisely flash out the arguments that are standard from empirical process literature.

Convergence rate for each entry of $\widehat{\mathbf{H}}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}}$: Consider $\mathbf{u}_1^\top(\widehat{\mathbf{H}}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}})\mathbf{u}_2$, where $\mathbf{u}_1, \mathbf{u}_2$ are multi-indices such that $|\mathbf{u}_1|, |\mathbf{u}_2| \leq p$. Take $\mathbf{v} = \mathbf{u}_1 + \mathbf{u}_2$. Define

$$g_n(\xi, \mathbf{x}) = \left(\frac{\xi - \mathbf{x}}{h}\right)^{\mathbf{v}} \frac{1}{h^d} K\left(\frac{\xi - \mathbf{x}}{h}\right) \mathbb{1}(\xi \in \mathcal{A}_t), \xi, \mathbf{x} \in \mathcal{X}.$$

Define $\mathcal{F} = \{g_n(\cdot, \mathbf{x}) : \mathbf{x} \in \mathcal{X}\}$. Then $\sup_{\mathbf{x} \in \mathcal{X}} |\mathbf{u}_1^\top(\widehat{\mathbf{H}}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}})\mathbf{u}_2| = \sup_{f \in \mathcal{F}} |\mathbb{E}_n[f(\mathbf{x}_i)] - \mathbb{E}[f(\mathbf{x}_i)]|$. By standard arguments from kernel regression literature, we can show $\mathcal{F}$ forms a VC-type class with exponent d and constant $\operatorname{diam}(\mathcal{X})/b$, $\mathtt{M}_n := \sup_{f \in \mathcal{F}} \sup_{\mathbf{x} \in \mathcal{X}} |f(\mathbf{x})| \lesssim b^{-d}$, $\sigma_n^2 := \sup_{f \in \mathcal{F}} \mathbb{V}[f(\mathbf{x}_i)] \lesssim b^{-d/2}$. By Corollary 5.1 in Chernozhukov *et al.* (2014), we can show $\mathbb{E}[\sup_{f \in \mathcal{F}} |\mathbb{E}_n[f(\mathbf{x}_i)] - \mathbb{E}[f(\mathbf{x}_i)]|] \lesssim (nb^d)^{-1/2}\sqrt{\log n} + (nb^d)^{-1}\log n$. Since $\mathcal{F}$ is separable, we can use Talagrand's inequality (Giné and Nickl, 2016, Theorem 3.3.9) to get for all $t > 0$,

$$\mathbb{P}\Big(\sup_{f \in \mathcal{F}} |\mathbb{E}_n[f(\mathbf{x}_i)] - \mathbb{E}[f(\mathbf{x}_i)]| \geq C_1(nb^d)^{-1/2}\sqrt{t + \log n} + C_1(nb^d)^{-1}(t + \log n)\Big) \leq \exp(-t),$$

where C_1 is a constant not depending on n. This shows $\sup_{\mathbf{x} \in \mathcal{X}} \mathbf{u}_1^\top(\widehat{\mathbf{H}}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}})\mathbf{u}_2 = O((nb^d)^{-1/2}\sqrt{\log n} + (nb^d)^{-1}\log n)$ a.s..

Convergence rate for $\sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\mathbf{H}}_{\mathbf{x}}^{-1} - \mathbf{H}_{\mathbf{x}}^{-1}\|$: Since $\mathbf{H}_{\mathbf{x}}$ and $\widehat{\mathbf{H}}_{\mathbf{x}}$ are finite-dimensional, $\sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\mathbf{H}}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}}\| = O((nb^d)^{-1/2}\sqrt{\log n} + (nb^d)^{-1}\log n)$ a.s.. By Weyl's Theorem, $\sup_{\mathbf{x} \in \mathcal{X}} |\lambda_{\min}(\widehat{\mathbf{H}}_{\mathbf{x}}) - \lambda_{\min}(\mathbf{H}_{\mathbf{x}})| = O((nb^d)^{-1/2}\sqrt{\log n} + (nb^d)^{-1}\log n)$ a.s., which also implies $\inf_{\mathbf{x} \in \mathcal{X}} \lambda_{\min}(\widehat{\mathbf{H}}_{\mathbf{x}}) \gtrsim 1$ a.s.. Hence

$$\sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\mathbf{H}}_{\mathbf{x}}^{-1} - \mathbf{H}_{\mathbf{x}}^{-1}\| \leq \sup_{\mathbf{x} \in \mathcal{X}} \|\widehat{\mathbf{H}}_{\mathbf{x}}^{-1}\| \|\widehat{\mathbf{H}}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}}\| \|\mathbf{H}_{\mathbf{x}}^{-1}\| = O((nb^d)^{-1/2}\sqrt{\log n}), \quad a.s..$$

Convergence rate for $\sup_{\mathbf{x} \in \mathcal{X}} \sup_{r \in \mathcal{R}} \|\mathbf{S}_{\mathbf{x},r}\|$: Consider $\mathbf{v}^\top \mathbf{S}_{\mathbf{x},r}$ where $|\mathbf{v}| \leq p$. Define $\mathcal{H}_1 = \{(\mathbf{z}, y) \mapsto g_n(\mathbf{z}, \mathbf{x})(r(y) - \theta(\mathbf{z}, r)) : \mathbf{x} \in \mathcal{X}, r \in \mathcal{R}_1\}$ and $\mathcal{H}_2 = \{(\mathbf{z}, y) \mapsto g_n(\mathbf{z}, \mathbf{x})(r(y) - \theta(\mathbf{z}, r)) : \mathbf{x} \in \mathcal{X}, r \in \mathcal{R}_2\}$. It is not hard to check both $\mathcal{H}_1$ and $\mathcal{H}_2$ are VC-type classes. By similar arguments as in $\widehat{\mathbf{H}}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}}$, for all $t > 0$,

$$\mathbb{P}\Big(\sup_{h \in \mathcal{H}_2} |\mathbb{E}_n[h(\mathbf{x}_i, y_i)] - \mathbb{E}[h(\mathbf{x}_i, y_i)]| \geq C_2(nb^d)^{-1/2}\sqrt{t + \log n} + C_2(nb^d)^{-1}(t + \log n)\Big) \leq \exp(-t).$$

And if we further assume $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[\exp(y_i)|\mathbf{x}_i = \mathbf{x}] \leq 2$, then for all $t > 0$,

$$\mathbb{P}\Big(\sup_{h \in \mathcal{H}_1} |\mathbb{E}_n[h(\mathbf{x}_i, y_i)] - \mathbb{E}[h(\mathbf{x}_i, y_i)]| \geq C_2(nb^d)^{-1/2}\sqrt{t + \log n} + C_2(nb^d)^{-1}(\log n)(t + \log n)\Big) \leq \exp(-t).$$

Together with finite dimensionality of the vector $\mathbf{S}_{\mathbf{x},r}$,

$$\sup_{\mathbf{x} \in \mathcal{X}} \sup_{r \in \mathcal{R}_1} \|\mathbf{S}_{\mathbf{x},r}\| = O((nb^d)^{-1/2}\sqrt{\log n} + (nb^d)^{-1}(\log n)^2), \quad a.s.,$$
$$\sup_{\mathbf{x} \in \mathcal{X}} \sup_{r \in \mathcal{R}_2} \|\mathbf{S}_{\mathbf{x},r}\| = O((nb^d)^{-1/2}\sqrt{\log n}), \quad a.s.$$

Putting together for Non-Linearity Errors:

$$\sup_{\mathbf{x}\in\mathcal{X}}\sup_{r\in\mathcal{R}_2}|\mathbf{e}_1^\top(\widehat{\mathbf{H}}_{\mathbf{x}}^{-1}-\mathbf{H}_{\mathbf{x}}^{-1})\mathbf{S}_{\mathbf{x},r}| = O((nb^d)^{-1}\log n), \quad a.s.,$$
$$\sup_{\mathbf{x}\in\mathcal{X}}\sup_{r\in\mathcal{R}_1}|\mathbf{e}_1^\top(\widehat{\mathbf{H}}_{\mathbf{x}}^{-1}-\mathbf{H}_{\mathbf{x}}^{-1})\mathbf{S}_{\mathbf{x},r}| = O((nb^d)^{-1}\log n + (nb^d)^{-3/2}(\log n)^{5/2}), \quad a.s..$$

Bias: Take $\mathbf{R}_{\mathbf{x},r} = \mathbb{E}_n\left[\mathbf{r}_p\left(\frac{\mathbf{X}_i-\mathbf{x}}{h}\right)K_h(\mathbf{X}_i-\mathbf{x})\mathfrak{r}_{\mathbf{x}}(\mathbf{X}_i;r)\right]$ where $\mathfrak{r}_{\mathbf{x}}(\xi;r)=\theta(\xi;r)-\sum_{0\leq|\boldsymbol{\nu}|\leq\mathfrak{p}}\frac{\partial_{\boldsymbol{\nu}}\theta(\mathbf{x};r)}{\boldsymbol{\nu}!}(\xi-\mathbf{x})^{\boldsymbol{\nu}}$. Since all $\theta(\cdot;r), r\in\mathcal{R}_\ell$ are $(\mathfrak{p}+1)$-times continuously differentiable with $\sup_{r\in\mathcal{R}_\ell}\sup_{\mathbf{x}\in\mathcal{X}}\max_{|\boldsymbol{\nu}|\leq\mathfrak{p}}|\partial_{\boldsymbol{\nu}}\theta(\mathbf{x};r)|<\infty$, then $\sup_{r\in\mathcal{R}_\ell}\sup_{\mathbf{x}\in\mathcal{X}}|\mathbf{R}_{\mathbf{x},r}| = O(b^{\mathfrak{p}+1})$. We have proved that $\inf_{\mathbf{x}\in\mathcal{X}}\lambda_{\min}(\widehat{\mathbf{H}}_{\mathbf{x}})\gtrsim 1$ a.s.. Hence

$$\sup_{r\in\mathcal{R}_\ell}\sup_{\mathbf{x}\in\mathcal{X}}|\mathbb{E}[\widehat{\theta}(\mathbf{x},r)|\mathbf{x}_1,\cdots,\mathbf{x}_n]-\theta(\mathbf{x},r)| = \sup_{r\in\mathcal{R}_\ell}\sup_{\mathbf{x}\in\mathcal{X}}|\mathbf{e}_1^\top\widehat{\mathbf{H}}_{\mathbf{x}}^{-1}\mathbf{R}_{\mathbf{x},r}| = O(b^{\mathfrak{p}+1}), \qquad a.s., \text{ for } \ell=1,2.$$

SA-III.23 Proof of Lemma SA.2

We use the notation $\mathbb{P}_X(\Delta_l)=\mathbb{P}(\mathbf{x}_i\in\Delta_l)$, and $\widehat{\mathbb{P}}_X(\Delta_l)=n^{-1}\sum_{i=1}^n\mathbb{1}(\mathbf{x}_i\in\Delta_l)$, $0\leq l<L$.

Non-linearity Errors: For $\ell=1,2$, $\mathbf{x}\in\mathcal{X}, r\in\mathcal{R}_\ell$, we have

$$\mathbf{p}(\mathbf{x})^\top(\widehat{\mathbf{J}}^{-1}-\mathbf{J}^{-1})\mathbf{T}_r = \sum_{0\leq l<L}\mathbb{1}(\mathbf{x}\in\Delta_l)(L^{-1}\widehat{\mathbb{P}}_X(\Delta_l)^{-1}-L^{-1}\mathbb{P}_X(\Delta_l)^{-1})\frac{1}{n}\sum_{i=1}^n\frac{\mathbb{1}(\mathbf{x}_i\in\Delta_l)}{L^{-1}}\epsilon_i(r).$$

By maximal inequality for sub-Gaussian random variables (van der Vaart and Wellner, 2013, Lemma 2.2.2), $\max_{0\leq l<L}|L\widehat{\mathbb{P}}_X(\Delta_l)-L\mathbb{P}_X(\Delta_l)| = O(\sqrt{\frac{\log L}{n/L}})$ a.s.. Since $\{\Delta_l: 0\leq l<L\}$ is a quasi-uniform parition on $\mathcal{X}$, $\min_{0\leq l<L}L\mathbb{P}_X(\Delta_l)=\Omega(1)$. Hence

$$\max_{0\leq l<L}|L^{-1}\widehat{\mathbb{P}}_X(\Delta_l)^{-1}-L^{-1}\mathbb{P}_X(\Delta_l)^{-1}| = O(\sqrt{(n/L)^{-1}\log L}), \quad a.s.. \tag{SA-20}$$

Take $\mathcal{H}_\ell=\{(\mathbf{x},y)\mapsto L\mathbb{1}(\mathbf{x}\in\Delta_l)(r(y)-\theta(\mathbf{x},r)): 0\leq l<L, r\in\mathcal{R}_\ell\}$, for $\ell=1,2$. In particular, if we take $\mathcal{G}=\{L\mathbb{1}(\cdot\in\Delta_l): 0\leq l<L\}$, then $\mathcal{G}$ is a VC-type class w.r.p. constant envelope L with constant $\mathtt{c}_{\mathcal{G}}=L$ and exponent $\mathtt{d}_{\mathcal{G}}=1$. In the main text, we explained that both $\mathcal{R}_1$ and $\mathcal{R}_2$ are VC-type class with $\mathtt{c}_{\mathcal{R}_1}=1$, $\mathtt{d}_{\mathcal{R}_1}=1$ and $\mathtt{c}_{\mathcal{R}_2}$ some absolute constant, $\mathtt{d}_{\mathcal{R}_2}=2$. By arguments similar to the proof of Lemma SA-III.10, both $\mathcal{H}_\ell$'s are VC-type class with $\mathtt{c}_{\mathcal{H}_1}=L$, $\mathtt{d}_{\mathcal{H}_1}=1$, $\mathtt{c}_{\mathcal{H}_2}\lesssim L$, $\mathtt{d}_{\mathcal{H}_2}=2$. Since $\sup_{r\in\mathcal{R}_\ell}\max_{0\leq l<L}|\frac{1}{n}\sum_{i=1}^n L\mathbb{1}(\mathbf{x}_i\in\Delta_l)\epsilon_i(r)| = \sup_{h\in\mathcal{H}_\ell}|\mathbb{E}_n[h(\mathbf{x}_i,y_i)]-\mathbb{E}[h(\mathbf{x}_i,y_i)]|$ is the suprema of empirical process, by Corollary 5.1 in Chernozhukov *et al.* (2014),

$$\begin{aligned}
&\sup_{r\in\mathcal{R}_1}\max_{0\leq l<L}\left|\frac{1}{n}\sum_{i=1}^n L\mathbb{1}(\mathbf{x}_i\in\Delta_l)\epsilon_i(r)\right| = O\left(\sqrt{\frac{\log(nL)}{n/L}}+\log(n)\frac{\log(nL)}{n/L}\right) \quad a.s.,\\
&\sup_{r\in\mathcal{R}_2}\max_{0\leq l<L}\left|\frac{1}{n}\sum_{i=1}^n L\mathbb{1}(\mathbf{x}_i\in\Delta_l)\epsilon_i(r)\right| = O\left(\sqrt{\frac{\log(nL)}{n/L}}\right) \quad a.s..
\end{aligned} \tag{SA-21}$$

Putting together Equations SA-20, SA-21, we have

$$\sup_{\mathbf{x}\in\mathcal{X}}\sup_{r\in\mathcal{R}_\ell}\left|\mathbf{p}(\mathbf{x})^\top(\widehat{\mathbf{J}}^{-1}-\mathbf{J}^{-1})\mathbf{T}_r\right| = O\left(\frac{\log(nL)}{n/L}\right)+\mathbb{1}(\ell=1)O\left(\log(n)\left(\frac{\log(nL)}{n/L}\right)^{3/2}\right).$$

Smoothing Bias: Since we have assumed that $\sup_{r \in \mathcal{R}_\ell} \sup_{\mathbf{x},\mathbf{y} \in \mathcal{X}} |\mu(\mathbf{x}, r) - \mu(\mathbf{y}, r)|/\|\mathbf{x}-\mathbf{y}\|_\infty < \infty$, $\ell = 1, 2$,

$$\sup_{\mathbf{x} \in \mathcal{X}} \sup_{r \in \mathcal{R}_l} |\mathbb{E}[\widehat{\mu}(\mathbf{x}, r)|\mathbf{x}_1, \cdots, \mathbf{x}_n] - \mu(\mathbf{x}, r)| = \left| \sum_{0 \le l < L} \mathbb{1}(\mathbf{x} \in \Delta_l) \frac{\sum_{i=1}^n \mathbb{1}(\mathbf{x}_i \in \Delta_l)\mu(\mathbf{x}_i, r)}{\sum_{i=1}^n \mathbb{1}(\mathbf{x}_i \in \Delta_l)} - \mu(\mathbf{x}, r) \right| = O(\max_{0 \le l < L} \|\Delta_l\|_\infty).$$

SA-III.24 Proof of Example SA.1

Recall $\mathcal{G} = \{b^{-d/2}\mathcal{K}_{\mathbf{x}}(\frac{\cdot - \mathbf{x}}{b}) : \mathbf{x} \in \mathcal{X}\}$ with $\mathcal{K}_{\mathbf{x}}(\mathbf{u}) = \mathbf{e}_1^\top \mathbf{H}_{\mathbf{x}}^{-1} \mathbf{p}(\mathbf{u}) K(\mathbf{u})$.

(1) Properties of $\mathcal{G}$

Since $\sup_{\mathbf{x} \in \mathcal{X}} \|\mathbf{H}_{\mathbf{x}}^{-1}\| \lesssim 1$ and K is continuous with compact support, we know

$$\mathtt{M}_{\mathcal{G}} \lesssim b^{-d/2}.$$

By a change of variable, we know

$$\mathtt{E}_{\mathcal{G}} = \sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}\left[\left| b^{-d/2}\mathcal{K}_{\mathbf{x}}\Big(\frac{\mathbf{x}_i - \mathbf{x}}{b}\Big)\right|\right] \lesssim \max_{|\mathbf{v}| \le \mathfrak{p}} b^{d/2} \int \left(\frac{\mathbf{u} - \mathbf{x}}{b}\right)^{\mathbf{v}} \frac{1}{b^d} K\left(\frac{\mathbf{u} - \mathbf{x}}{b}\right) h_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \lesssim b^{d/2}.$$

Moreover, $\sup_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbf{u},\mathbf{u}'} |\mathbf{r}_p\left(\frac{\mathbf{u}-\mathbf{x}}{b}\right) - \mathbf{r}_p(\frac{\mathbf{u}'-\mathbf{x}}{b})|/\|\mathbf{u}-\mathbf{u}'\|_\infty \lesssim b^{-1}$ and $\sup_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbf{u},\mathbf{u}'} |K(\frac{\mathbf{u}-\mathbf{x}}{b}) - K(\frac{\mathbf{u}'-\mathbf{x}}{b})|/\|\mathbf{u}-\mathbf{u}'\|_\infty \lesssim b^{-1}$ It follows that

$$\mathtt{L}_{\mathcal{G}} \lesssim b^{-\frac{d}{2}-1}.$$

Notice that the support of functions in $\mathcal{G}$ has uniformly bounded volume, i.e. $\sup_{g \in \mathcal{G}} \mathrm{Vol}\,(\mathrm{Supp}(g)) \lesssim b^d$. Together with the rate for $\mathtt{L}_{\mathcal{G}}$, we know

$$\mathtt{TV}_{\mathcal{G}} \le \mathtt{L}_{\mathcal{G}} \sup_{g \in \mathcal{G}} \mathrm{Vol}\,(\mathrm{Supp}(g)) \lesssim b^{\frac{d}{2}-1}.$$

Now we will show that $\mathtt{M}_{\mathcal{G}}^{-1}\mathcal{G}$ is a VC-class. We know $\sup_{\mathbf{x},\mathbf{x}' \in \mathcal{X}} \|\mathbf{H}_{\mathbf{x}} - \mathbf{H}_{\mathbf{x}'}\|/\|\mathbf{x} - \mathbf{x}'\|_\infty \lesssim b^{-1}$. Since $\inf_{\mathbf{x} \in \mathcal{X}} \|\mathbf{H}_{\mathbf{x}}\| \gtrsim 1$, we also have $\sup_{\mathbf{x},\mathbf{x}' \in \mathcal{X}} \|\mathbf{H}_{\mathbf{x}}^{-1} - \mathbf{H}_{\mathbf{x}'}^{-1}\|/\|\mathbf{x} - \mathbf{x}'\|_\infty \lesssim b^{-1}$. It follows that

$$\mathtt{L}_{\mathcal{G}} = \sup_{\mathbf{u} \in \mathcal{X}} \sup_{\mathbf{x},\mathbf{x}' \in \mathcal{X}} \left| b^{-d/2}\mathcal{K}_{\mathbf{x}}\Big(\frac{\mathbf{u} - \mathbf{x}}{b}\Big) - b^{-d/2}\mathcal{K}_{\mathbf{x}}\Big(\frac{\mathbf{u} - \mathbf{x}'}{b}\Big)\right| / \|\mathbf{x} - \mathbf{x}'\|_\infty \lesssim b^{-d/2-1}.$$

Consider $h_{\mathbf{x}}(\cdot) = \sqrt{b^d}\mathbf{e}_1^T \mathbf{H}_{\mathbf{x}}^{-1} \mathbf{r}_p(\cdot) K(\cdot)$. Then $b^{-d/2}\mathcal{K}_{\mathbf{x}}(\frac{\cdot - \mathbf{x}}{b}) = h_{\mathbf{x}}\left(\frac{\cdot - \mathbf{x}}{b}\right)$. By the rates of $\mathtt{M}_{\mathcal{G}}$, $\mathtt{L}_{\mathcal{G}}$, $\mathtt{E}_{\mathcal{G}}$, there exists a constant $\mathbf{c}$ only depending on $\|K\|_\infty$, $\mathtt{L}_{\{K\}}$, σ_K, $\overline{f}_X$, $\underline{f}_X$ that

$$\begin{aligned}
&\sup_{\mathbf{x} \in \mathcal{X}} \|h_{\mathbf{x}}\|_\infty \le \mathbf{c}, \\
&\sup_{\mathbf{x} \in \mathcal{X}} \sup_{\mathbf{u},\mathbf{v} \in \mathcal{X}} \frac{|h_{\mathbf{x}}(\mathbf{u}) - h_{\mathbf{x}}(\mathbf{v})|}{\|\mathbf{u} - \mathbf{v}\|_\infty} \le \mathbf{c}, \\
&\sup_{\mathbf{x},\mathbf{y} \in \mathcal{X}} \sup_{\mathbf{u} \in \mathcal{X}} \frac{|h_{\mathbf{x}}(\mathbf{u}) - h_{\mathbf{y}}(\mathbf{u})|}{\|\mathbf{x} - \mathbf{y}\|_\infty} \le \mathbf{c}.
\end{aligned}$$

We can again apply Lemma 7 from Cattaneo *et al.* (2024) to show that, for all $0 < \varepsilon < 1$,

$$N(\mathtt{M}_{\mathcal{G}}^{-1}\mathcal{G}, e_{\mathbb{P}}, \varepsilon) \leq \mathbf{c}\frac{1}{\varepsilon^{d+1}} + 1.$$

(2) Properties of $\widetilde{\mathcal{H}}_1$

Let $g \in \mathcal{G}$. Take $\widetilde{\mathcal{H}}_1^a = \{g \cdot \varphi : g \in \mathcal{G}\}$ and $\widetilde{\mathcal{H}}_1^b = \{g \cdot \theta(\cdot, \mathrm{Id}) : g \in \mathcal{G}\}$. Define $h^a(\mathbf{x}, u) = g(\mathbf{x})\varphi(\mathbf{x}, u)$. Let ι be a real-valued non-negative Lebesgue measurable function on $\mathbb{R}^d$ such that $\int_{\mathbb{R}^d} \iota(\mathbf{u})d\mathbf{u} = 1$. Define $\iota_\varepsilon = \varepsilon^{-d}\iota(\cdot/\varepsilon)$ and $g_\varepsilon = g * \iota_\varepsilon$. Let ξ be a real-valued non-negative Lebesgue measurable function on $\mathbb{R}^{d+1}$ such that $\int_{\mathbb{R}^{d+1}} \xi(\mathbf{u})d\mathbf{u} = 1$. Define $\xi_\varepsilon = \varepsilon^{-d-1}\xi(\cdot/\varepsilon)$ and $\varphi_\varepsilon = \varphi * \xi_\varepsilon$. Then define $h^a_\varepsilon(\mathbf{x}, u) = g_\varepsilon(\mathbf{x})\varphi_\varepsilon(\mathbf{x}, u)$. Then for all $\mathbf{x} \in \mathcal{X}, u \in \mathbb{R}$ and $\varepsilon > 0$,

$$\|\nabla h^a_\varepsilon(\mathbf{x}, u)\|_2 \leq \|\nabla g_\varepsilon(\mathbf{x})\|_2 + \mathtt{M}_{\mathcal{G}}\|\nabla \varphi_\varepsilon(\mathbf{x}, \mathbf{u})\|_2 \mathbb{1}(\mathbf{x} \in \mathrm{Supp}(g_\varepsilon)).$$

Hence by definition of $\mathtt{TV}$ and Dominated Convergence Theorem,

$$\begin{aligned}\mathtt{TV}_{\{h^a\}} \leq& \lim_{\varepsilon \downarrow 0} \int_{\mathcal{X}\times[0,1]} \|\nabla h^a_\varepsilon(\mathbf{x}, u)\|_2 d\mathbf{x}du \leq \lim_{\varepsilon \downarrow 0} \int_{\mathcal{X}} \|\nabla g_\varepsilon(\mathbf{x})\|_2 d\mathbf{x} + \mathtt{M}_{\mathcal{G}} \lim_{\varepsilon \downarrow 0} \int \|\nabla \varphi_\varepsilon(\mathbf{x}, u)\| \mathbb{1}(\mathbf{x} \in \mathrm{Supp}(g_\varepsilon))d\mathbf{x}du \\ \leq& \mathtt{TV}_{\{g\}} + \mathtt{M}_{\mathcal{G}}\, \mathrm{TV}_{\{\varphi\},\mathrm{Supp}(g)\times[0,1]}\,.\end{aligned}$$

Let $\mathcal{C}$ be any cube of side-length a in $\mathbb{R}^{d+1}$. Then

$$\begin{aligned}\mathtt{TV}_{\{h^a\},\mathcal{C}} \leq& \lim_{\varepsilon \downarrow 0} \int_{\mathcal{C}} \|\nabla h^a_\varepsilon(\mathbf{x}, u)\|_2 d\mathbf{x}du \leq \lim_{\varepsilon \downarrow 0} \int_{\mathcal{C}} \|\nabla g_\varepsilon(\mathbf{x})\|_2 d\mathbf{x} + \mathtt{M}_{\mathcal{G}} \lim_{\varepsilon \downarrow 0} \int_{\mathcal{C}} \|\nabla \varphi_\varepsilon(\mathbf{x}, u)\| \mathbb{1}(\mathbf{x} \in \mathrm{Supp}(g_\varepsilon))d\mathbf{x}du \\ \leq& \mathrm{TV}_{\{g\},\mathcal{C}} + \mathtt{M}_{\mathcal{G}}\, \mathrm{TV}_{\{\varphi\},\mathrm{Supp}(g)\times[0,1]\cap\mathcal{C}} \leq \mathtt{M}_{\mathcal{R}}\mathtt{K}_{\mathcal{G}}a^d + \mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{R}}\mathtt{K}_{\{\varphi\}}a^d.\end{aligned}$$

In summary, we have

$$\mathtt{M}_{\widetilde{\mathcal{H}}_1^a} \leq \mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}, \quad \mathtt{TV}_{\widetilde{\mathcal{H}}_1^a} \leq \mathtt{TV}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}} \sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\varphi\},\mathrm{Supp}(g)\times[0,1]}, \quad \mathtt{K}_{\widetilde{\mathcal{H}}_1^a} \leq \mathtt{K}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}}\mathtt{K}_{\{\varphi\}}.$$

Similar argument shows

$$\mathtt{M}_{\widetilde{\mathcal{H}}_1^b} \leq \mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}, \quad \mathtt{TV}_{\widetilde{\mathcal{H}}_1^b} \leq \mathtt{TV}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}} \sup_{g\in\mathcal{G}} \mathtt{TV}_{\mathcal{V}_1,\mathrm{Supp}(g)}, \quad \mathtt{K}_{\widetilde{\mathcal{H}}_1^b} \leq \mathtt{K}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}}\mathtt{K}_{\mathcal{V}_1}.$$

It follows from the assumptions $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\varphi\},\mathrm{supp}(g)\times[0,1]} \lesssim \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g))$ and $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\mathcal{V}_{\mathcal{R}_1},\mathrm{supp}(g)} \lesssim \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g))$ that

$$\mathtt{M}_{\widetilde{\mathcal{H}}_1} \leq \mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}, \quad \mathtt{TV}_{\widetilde{\mathcal{H}}_1} \lesssim \mathtt{TV}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}}\mathfrak{m}(\mathrm{Supp}(g)), \quad \mathtt{K}_{\widetilde{\mathcal{H}}_1} \leq \mathtt{K}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}}\mathtt{K}_{\{\varphi\}} + \mathtt{M}_{\mathcal{G}}\mathtt{K}_{\mathcal{V}_1}.$$

By Lemma SA.13, $\widetilde{\mathcal{H}}_1$ is a VC-type class with constant $c_{\mathcal{G}}c_{\mathcal{R}}2^{\mathtt{d}_{\mathcal{G}}+\mathtt{d}_{\mathcal{R}}}$ and exponent $\mathtt{d}_{\mathcal{G}} + \mathtt{d}_{\mathcal{R}}$ with respect to envelope function $\mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}$.

(3) Properties of $\widetilde{\mathcal{H}}_2$

The main challenge is that $\mathcal{R}_2$ contains non-differentiable indicator. First, we study properties of $\mathcal{G} \times \mathcal{R}_2$. Then by Definition 2,

$$\begin{aligned}
\mathtt{TV}_{\mathcal{G}\times\mathcal{R}_2,[0,1]^{d+1}} &= \sup_{g\in\mathcal{G}}\sup_{y\in\mathbb{R}} \sup_{\substack{\phi\in\mathcal{D}_{d+1}([0,1]^{d+1})\\ \|\|\phi\|_2\|_\infty\le 1}} \int_{[0,1]^d}\int_{[0,1]} g(\mathbf{x})\mathbb{1}(u\le y)\operatorname{div}(\phi)(\mathbf{x},u)dud\mathbf{x} \\
&\le \sup_{g\in\mathcal{G}}\sup_{y\in\mathbb{R}} \sup_{\substack{\phi\in\mathcal{D}_{d}([0,1]^{d})\\ \|\|\phi\|_2\|_\infty\le 1}} \sup_{\substack{\psi\in\mathcal{D}_1([0,1])\\ \|\psi\|_\infty\le 1}} \int_{[0,1]^d}\int_{[0,1]} g(\mathbf{x})\mathbb{1}(u\le y)(\operatorname{div}\phi(\mathbf{x})+\psi'(u))dud\mathbf{x} \\
&= \sup_{g\in\mathcal{G}}\sup_{y\in\mathbb{R}} \sup_{\substack{\phi\in\mathcal{D}_{d}([0,1]^{d})\\ \|\|\phi\|_2\|_\infty\le 1}} \int_{[0,1]^d} g(\mathbf{x})\operatorname{div}\phi(\mathbf{x})d\mathbf{x} + \sup_{\substack{\psi\in\mathcal{D}_1([0,1])\\ \|\psi\|_\infty\le 1}} \int_{[0,1]^d} g(\mathbf{x})d\mathbf{x}(\psi(1)-\psi(0)) \\
\le& \mathtt{TV}_{\mathcal{G},[0,1]^d} + 2\mathtt{E}_{\mathcal{G}}.
\end{aligned}$$

Similar argument as in (2) gives

$$\mathtt{TV}_{\mathcal{G}\times\mathcal{V}_2,[0,1]^d} \le \mathtt{TV}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}}\sup_{g\in\mathcal{G}}\mathtt{TV}_{\mathcal{V}_2,\mathrm{supp}(g)} \lesssim \mathtt{TV}_{\mathcal{G},[0,1]^d} + \mathtt{M}_{\mathcal{G}}\sup_{g\in\mathcal{G}}\mathfrak{m}(\mathrm{supp}(g)).$$

It follows that

$$\mathtt{TV}_{\widetilde{\mathcal{H}}_2} \lesssim \mathtt{TV}_{\mathcal{G},[0,1]^d} + \mathtt{E}_{\mathcal{G}} + \mathtt{M}_{\mathcal{G}}\sup_{g\in\mathcal{G}}\mathfrak{m}(\mathrm{supp}(g)).$$

Consider the change of variable function $T : [0,1]^{d+1} \to \mathbb{R}^{d+1}$ given by $T(\mathbf{x},u) = (\mathbf{x},\varphi(\mathbf{x},u))$. Observe that $\nabla T(\mathbf{x},u)$ is a lower triangular matrix with diagonal $(\mathbf{1},\partial_u\varphi(\mathbf{x},u))$, we have $\|\nabla T(\mathbf{x},u)\|_{\mathrm{op}} = |\partial_u\varphi(\mathbf{x},u)|$, $\det(\nabla T(\mathbf{x},u)) = |\partial_u\varphi(\mathbf{x},u)|$.

$$\begin{aligned}
\mathtt{TV}_{\widetilde{\mathcal{H}}_2} &= \sup_{h\in\mathcal{G}\times\mathcal{R}_2}\lim_{\varepsilon\downarrow 0}\int_{\mathbf{u}\in[0,1]^{d+1}}\|\nabla(h_\varepsilon\circ T)(\mathbf{u})\|d\mathbf{u} = \sup_{h\in\mathcal{G}\times\mathcal{R}_2}\lim_{\varepsilon\downarrow 0}\int_{\mathbf{u}\in[0,1]^{d+1}}\|(\nabla T(\mathbf{u}))^\top\nabla h_\varepsilon(T(\mathbf{u}))\|d\mathbf{u} \\
&= \sup_{h\in\mathcal{G}\times\mathcal{R}_2}\lim_{\varepsilon\downarrow 0}\int_{\mathbf{x}\in T([0,1]^{d+1})}\left\|\left(\nabla T(T^{-1}(\mathbf{x}))\right)^\top\nabla h_\varepsilon(\mathbf{x})\right\|\det\left(\nabla T^{-1}(\mathbf{x})\right)d\mathbf{x} \\
&\le \sup_{h\in\mathcal{G}\times\mathcal{R}_2}\lim_{\varepsilon\downarrow 0}\int_{\mathbf{x}\in T([0,1]^{d+1})}\|\nabla h_\varepsilon(\mathbf{x})\|d\mathbf{x}\|\det(\nabla T)^{-1}\|_\infty\|\|\nabla T\|_{\mathrm{op}}\|_\infty \\
&\le \mathtt{TV}_{\mathcal{G}\times\mathcal{R}_2,[0,1]^{d+1}}\|\det(\nabla T)^{-1}\|_\infty\|\|\nabla T\|_{\mathrm{op}}\|_\infty \le (\mathtt{TV}_{\mathcal{G},[0,1]^d}+2\mathtt{E}_{\mathcal{G}})\|\det(\nabla T)^{-1}\|_\infty\|\|\nabla T\|_{\mathrm{op}}\|_\infty \\
&\lesssim (\mathtt{TV}_{\mathcal{G},[0,1]^d}+\mathtt{E}_{\mathcal{G}}+\mathtt{M}_{\mathcal{G}}\sup_{g\in\mathcal{G}}\mathfrak{m}(\mathrm{supp}(g)))\frac{\max_{(\mathbf{x},u)\in[0,1]^{d+1}}|\partial_u\varphi(\mathbf{x},u)|}{\min_{(\mathbf{x},u)\in[0,1]^{d+1}}|\partial_u\varphi(\mathbf{x},u)|}.
\end{aligned}$$

By Lemma SA.13, $\widetilde{\mathcal{H}}_2$ is a VC-type class with constant $c_{\mathcal{G}}c_{\mathcal{R}}2^{\mathtt{d}_{\mathcal{G}}+\mathtt{d}_{\mathcal{R}}}$ and exponent $\mathtt{d}_{\mathcal{G}}+\mathtt{d}_{\mathcal{R}}$ with respect to envelope function $\mathtt{M}_{\mathcal{G}}$.

(4) Effects of Rosenblatt Transformation

By Lemma SA.12, $\mathtt{TV}_{\mathcal{H}_1} \le \mathtt{TV}_{\widetilde{\mathcal{H}}_1}\overline{f}_Z^2\underline{f}_Z^{-1}$, $\mathtt{TV}_{\mathcal{H}_2} \le \mathtt{TV}_{\widetilde{\mathcal{H}}_2}\overline{f}_Z^2\underline{f}_Z^{-1}$, $\mathtt{M}_{\mathcal{H}_1} = \mathtt{M}_{\widetilde{\mathcal{H}}_1}$, $\mathtt{M}_{\mathcal{H}_2} = \mathtt{M}_{\widetilde{\mathcal{H}}_2}$. Moreover, $\mathcal{H}_1$ and $\mathcal{H}_2$ are VC-type classes with constant $c_{\mathcal{G}}c_{\mathcal{R}}2^{\mathtt{d}_{\mathcal{G}}+\mathtt{d}_{\mathcal{R}}}$ and exponent $\mathtt{d}_{\mathcal{G}}+\mathtt{d}_{\mathcal{R}}$ with respect to envelope functions $\mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}$ and $\mathtt{M}_{\mathcal{G}}$ respectively.

(5) Application of Theorem 1.1 in Rio (1994)

We can now apply Theorem 1.1 in Rio (1994) to get $\{X_n(h) : h \in \mathcal{H}_1\}$ admits a Gaussian strong approximation with rate function

$$C_{d,\varphi}\sqrt{\frac{d\overline{f}_Z^2}{\underline{f}_Z}}\frac{\sqrt{\mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}(\mathtt{TV}_{\mathcal{G}}+\mathtt{M}_{\mathcal{G}}\sup_{g\in\mathcal{G}}\mathtt{TV}_{\{\varphi\},\mathrm{Supp}(g)})}}{n^{\frac{1}{2d+2}}}\sqrt{t+(d+1)\log n}+$$
$$C_{d,\varphi}\sqrt{\frac{\mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}}{n}}\min\left\{\sqrt{\log(n)\mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}},\sqrt{\frac{(2\sqrt{d})^{d-1}\overline{f}_Z^{d+1}}{\underline{f}_Z^d}(\mathtt{K}_{\mathcal{G}}+\mathtt{M}_{\mathcal{G}}\mathtt{K}_{\{\varphi\}}))}\right\}(t+(d+1)\log n),$$

where $C_{d,\varphi,1}$ is a quantity that only depends on d and φ. And $\{X_n(h) : h \in \mathcal{H}_2\}$ admits a Gaussian strong approximation with rate function

$$C_{d,\varphi,2}\sqrt{\frac{d\overline{f}_Z^2}{\underline{f}_Z}}\frac{\sqrt{\mathtt{M}_{\mathcal{G}}\mathtt{TV}'}}{n^{\frac{1}{2d+2}}}\sqrt{t+(d+1))\log n}+C_{d,\varphi,2}\frac{\mathtt{M}_{\mathcal{G}}\mathtt{M}_{\{\varphi\}}}{\sqrt{n}}(t+(d+1)\log n),$$

where $\mathtt{TV}' = (\mathtt{TV}_{\mathcal{G},[0,1]^d} + 2\mathtt{E}_{\mathcal{G}})\|\det(\nabla T)^{-1}\|_\infty\|\|\nabla T\|_{\mathrm{op}}\|_\infty$, and again $C_{d,\varphi,2}$ is a quantity that only depends on d and φ. □

SA-III.25 Proof of Example SA.2

Besides the properties given in the proof of Example SA.1, using product rule we can show $\mathtt{L}_{\widetilde{\mathcal{H}}_1} \lesssim \mathtt{L}_{\mathcal{G}}\mathtt{M}_{\mathcal{R}_1} + \mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{R}_1}\mathtt{L}_{\{\varphi\}} + \mathtt{M}_{\mathcal{G}}\mathtt{L}_{\mathcal{V}_1} \lesssim b^{-d/2-1}$, and by Lemma SA.12, $\mathtt{L}_{\mathcal{H}_1} \lesssim \mathtt{L}_{\widetilde{\mathcal{H}}_1}\overline{f}_Z/\underline{f}_Z$. The result the follows from application of Theorem SA.1. □

SA-III.26 Proof of Example SA.3

The conditions of $\mathcal{G}$ can be verified from Part (1) Properties of $\mathcal{G}$ in Section SA.1. It is easy to check that $\mathcal{R}_1$ satisfies (ii)(b) in Theorem 3 with $\mathtt{c}_{\mathcal{R}_1} = 1$, $\mathtt{d}_{\mathcal{R}_1} = 1$ and $\alpha = 1$. Moreover, $\mathcal{R}_2$ satisfies (ii) (a) in Theorem 3, and we can take $\mathtt{c}_{\mathcal{R}_2}$ to be some absolute constant and $\mathtt{d}_{\mathcal{R}_2} = 2$ by van der Vaart and Wellner (2013, Theorem 2.6.7). The results then follow from Theorem 3.

SA-III.27 Proof of Example 3

In this section, we verify the rates claimed in this section. Recall $\mathcal{G} = \{k_{\mathbf{x}}(\cdot) : \mathbf{x} \in \mathcal{X}\}$ with $k_{\mathbf{x}}(\mathbf{u}) = L^{-1/2}\sum_{0\le l<L}\mathbb{1}(\mathbf{x}\in\Delta_l)\mathbb{1}(\mathbf{u}\in\Delta_l)/\mathbb{P}_X(\Delta_l)$. Since $\{\Delta_l : 0 \le l < L\}$ is a quasi-uniform partition of $\mathcal{X}$, there exists constants $C_1 > 0$ and $C_2 > 0$ not depending on L such that

$$C_1L^{-1} \le \mathbb{P}(\Delta_l) \le C_2L^{-1}, \qquad 0 \le l < L.$$

This gives $\mathtt{M}_{\mathcal{G}} \lesssim L^{1/2}$.

$$\mathtt{E}_{\mathcal{G}} = \sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}\left[\left|\frac{1}{\sqrt{L}}\sum_{0\le l<L}\frac{\mathbb{1}(\mathbf{x}\in\Delta_l)\mathbb{1}(\mathbf{x}_i\in\Delta_l)}{\mathbb{P}_X(\Delta_l)}\right|\right] = \max_{0\le l<L}\mathbb{E}\left[\left|\frac{1}{\sqrt{L}}\frac{\mathbb{1}(\mathbf{x}_i\in\Delta_l)}{\mathbb{P}_X(\Delta_l)}\right|\right] = L^{-1/2}.$$

For any $0 < \varepsilon < 1$, $\mathtt{N}_{\mathcal{G}}(\varepsilon) \leq \mathrm{Card}(\mathcal{G}) = L \leq L\varepsilon^{-1}$. Hence we can take $\mathtt{c}_{\mathcal{G}} = L$ and $\mathtt{d}_{\mathcal{G}} = 1$. $\mathcal{R}_1$ is the singleton of identity function, hence obviously we can take $\mathtt{c}_{\mathcal{R}_1} = 1$ and $\mathtt{d}_{\mathcal{R}_1} = 1$. For $\mathcal{R}_2$, observe that it is a VC sub-graph class of VC-index 2. Hence by van der Vaart and Wellner (2013, Theorem 2.6.7), $\mathcal{R}_2$ is also a VC-type class with $\mathtt{d}_{\mathcal{R}_2} = 2$ and $\mathtt{c}_{\mathcal{R}_2}$ some absolute constant. The claimed results then follow from application of Theorem 4.

SA-III.28 Proof of Theorem 3

We first make a reduction via Rosenblatt transformation. Take $\mathbf{u}_i = \phi_X(\mathbf{x}_i)$ where ϕ_X is defined as in Lemma SA.12. And define $\widetilde{g} = g \circ \phi_X^{-1}$ for each $g \in \mathcal{G}$ and consider $\widetilde{\mathcal{G}} = \{\widetilde{g} : g \in \mathcal{G}\}$. Then

$$R_n(g,r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g(\mathbf{x}_i) r(y_i) - \mathbb{E}[g(\mathbf{x}_i) r(y_i)] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{g}(\mathbf{u}_i) r(y_i) - \mathbb{E}[\widetilde{g}(\mathbf{u}_i) r(y_i)] =: \widetilde{R}_n(\widetilde{g}, r),$$

for all $g \in \mathcal{G}, r \in \mathcal{R}$. Denote by $\widetilde{\mathbb{P}}$ the law of $(\mathbf{u}_i, y_i)$. Consider $\mathcal{A}_{M,N}(\widetilde{\mathbb{P}}, 1)$, the axis-aligned iterative splitting of depth M for the main space $\mathbb{R}^d$ and depth N for the multipler subspace, with respect to $\widetilde{\mathbb{P}}$. Denote $\mathcal{E}_{M+N} := \{\mathbb{1}(\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}) : 0 \leq l < 2^M, 0 \leq m < 2^N\}$ where $\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}$'s are the base level cells given in Definition SA.4. By Lemma SA.12 and Lemma SA.13, it is possible to take a $\delta \mathtt{M}_{\widetilde{\mathcal{G}}} N^{\alpha} = \delta \mathtt{M}_{\mathcal{G}} N^{\alpha}$-net of $\widetilde{\mathcal{G}} \times \mathcal{R}$, $(\widetilde{\mathcal{G}} \times \mathcal{R})_\delta$, with cardinality no greater than $\mathtt{N}_{\mathcal{G} \times \mathcal{R}}(\delta) := \sup_P N(\mathcal{G}, e_P, \delta \mathtt{M}_{\mathcal{G}}/\sqrt{2}) \sup_Q N(\mathcal{R}, e_Q, \delta \|M_{\mathcal{R}}\|_{Q,2}/\sqrt{2})$ where $\sup_P$ is taken over all finite discrete measures on $[0,1]^d$ and $\sup_Q$ is taken over all finite discrete measures on $\mathbb{R}$. By Lemma SA.19, on a possibly enlarged probability space there exists a mean-zero Gaussian process Z_n^R indexed by $\widetilde{\mathcal{G}} \times \mathcal{R} \cup \Pi_2(\widetilde{\mathcal{G}} \times \mathcal{R}) \cup \mathcal{E}_{M+N}$ with almost sure continuous sample path such that

$$\mathbb{E}\left[Z_n^R(g) Z_n^R(f)\right] = \mathbb{E}\left[\widetilde{R}_n(g) \widetilde{R}_n(f)\right], \quad \forall g, f \in \widetilde{\mathcal{G}} \times \mathcal{R} \cup \Pi_2(\widetilde{\mathcal{G}} \times \mathcal{R}) \cup \mathcal{E}_{M+N},$$

and for all $t > 0$,

$$\mathbb{P}\left(\|\Pi_2 \widetilde{R}_n - \Pi_2 Z_n^R\|_{(\widetilde{\mathcal{G}} \times \mathcal{R})_\delta} > C_\alpha \sqrt{\frac{N^{2\alpha+1} 2^M \mathtt{E}_{\widetilde{\mathcal{G}}} \mathtt{M}_{\widetilde{\mathcal{G}}}}{n} t} + C_\alpha \sqrt{\frac{\mathtt{C}_{\Pi_2(\widetilde{\mathcal{G}} \times \mathcal{R})}}{n} t} \right) \leq 2 \mathtt{N}_{\mathcal{G} \times \mathcal{R}}(\delta) e^{-t}, \tag{SA-22}$$

where

$$\mathtt{C}_{\Pi_2(\widetilde{\mathcal{G}} \times \mathcal{R})} = \sup_{f \in \Pi_2(\widetilde{\mathcal{G}} \times \mathcal{R})} \min \left\{ \sup_{(j,k)} \left[\sum_{j' < j} (j - j')(j - j' + 1) 2^{j'-j} \sum_{k' : \mathcal{C}_{j',k'} \subseteq \mathcal{C}_{j,k}} \widetilde{\beta}_{j',k'}^2(f) \right], \|f\|_\infty^2 (M + N) \right\}.$$

Let $f \in \Pi_2(\widetilde{\mathcal{G}} \times \mathcal{R})$. Then there exists $g \in \widetilde{\mathcal{G}}$ and $r \in \mathcal{R}$ such that $f = \Pi_2[g, r]$. Since f is already piecewise-constant, by definition of $\beta_{j,k}$'s and $\eta_{j,k}$'s, we know $\widetilde{\beta}_{l,m}(f) = \widetilde{\eta}_{l,m}(g, r)$. Fix (j,k). We consider two cases. <u>Case 1</u>: $j > N$. Then by the design of cell expansions (Section SA-III.1), $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathbb{R}$. By definition of $\eta_{l,m}$, for any $N \leq j' \leq j$, we have $(j - j')(j - j' + 1) 2^{j'-j} \sum_{k' : \mathcal{C}_{j',k'} \subseteq \mathcal{C}_{j,k}} \widetilde{\eta}_{j',k'}^2(g, r) = 0$. Now consider

$0 \leq j' < N$. Then

$$\begin{aligned}
&\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}} |\widetilde{\eta}_{j',k'}(g,r)| \\
=&\sum_{l:\mathcal{X}_{0,l}\subseteq\mathcal{X}_{j-N,k}} \sum_{0\leq m<2^{j'}} |\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|\cdot|\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m}] \\
&\qquad - \mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m+1}]| \\
\leq& C_\alpha \sum_{l:\mathcal{X}_{0,l}\subseteq\mathcal{X}_{j-N,k}} |\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]| N^\alpha \leq C_\alpha 2^{j-N}\mathtt{M}_{\mathcal{G}} N^\alpha.
\end{aligned}$$

It follows that

$$\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\eta}_{j',k'}(g,r)| \leq \sum_{j'<j}(j-j')(j-j'+1)2^{j'-N}\mathtt{M}_{\mathcal{G}}N^\alpha \lesssim \mathtt{M}_{\mathcal{G}}N^\alpha.$$

<u>Case 2</u>: $j \leq N$. Then $\mathcal{C}_{j,k} = \mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}$. Hence for any $0 \leq j' \leq j$, we have

$$\begin{aligned}
\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}} |\widetilde{\eta}_{j',k'}(g,r)| =& |\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]| \sum_{m':\mathcal{Y}_{l,j',m'}\subseteq\mathcal{Y}_{l,j,m}} |\mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m}] \\
&- \mathbb{E}[r(y_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}, y_i\in\mathcal{Y}_{l,j-1,2m+1}]| \\
\leq& C_\alpha|\mathbb{E}[g(\mathbf{x}_i)|\mathbf{x}_i\in\mathcal{X}_{0,l}]|N^\alpha \leq C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha.
\end{aligned}$$

It follows that

$$\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\eta}_{j',k'}(g,r)| \leq C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha.$$

Moreover, for all (j,k), we have $\widetilde{\beta}_{j,k}(g,r) \leq C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha$. Hence $\mathtt{C}_{\Pi_2(\widetilde{\mathcal{G}}\times\mathcal{R})} \leq (C_\alpha\mathtt{M}_{\mathcal{G}}N^\alpha)^2$. Plug in Equation SA-22, we get for all $t>0$,

$$\mathbb{P}\left(\|\Pi_2\widetilde{R}_n - \Pi_2 Z_n^R\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta} > C_\alpha\sqrt{\frac{N^{2\alpha+1}2^M\mathtt{E}_{\widetilde{\mathcal{G}}}\mathtt{M}_{\widetilde{\mathcal{G}}}}{n}t} + C_\alpha^2\frac{\mathtt{M}_{\mathcal{G}}N^\alpha}{\sqrt{n}}t\right) \leq 2\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)e^{-t}. \tag{SA-23}$$

For projection error, by Lemma SA.24, for all $t > N$, with probability at least $1-8\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta)ne^{-t}$,

$$\|\widetilde{R}_n - \Pi_2\widetilde{R}_n\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta} + \|Z_n^R - \Pi_2 Z_n^R\|_{(\widetilde{\mathcal{G}}\times\mathcal{R})_\delta} \leq C_\alpha\left[\sqrt{\mathtt{V}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}}t^{\frac{1}{2}} + \sqrt{N^2\mathtt{V}_{\mathcal{G}} + 2^{-N}\mathtt{M}_{\mathcal{G}}^2}t^{\alpha+\frac{1}{2}} + \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}t^{\alpha+1}\right], \tag{SA-24}$$

where C_α is a constant that only depends on α and

$$\begin{aligned}
\mathtt{V}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}} &= \min\{2\mathtt{M}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}, \mathtt{L}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}\|\mathcal{V}\|_\infty\}2^M\mathfrak{m}(\mathcal{V})\|\mathcal{V}\|_\infty\mathtt{TV}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}} \lesssim \min\{\mathtt{M}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}, 2^{-\frac{M}{d}}\mathtt{L}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}\}2^{-\frac{M}{d}}\mathtt{TV}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}, \\
\mathtt{V}_{\widetilde{\mathcal{G}}} &= \min\{2\mathtt{M}_{\widetilde{\mathcal{G}}}, \mathtt{L}_{\widetilde{\mathcal{G}}}\|\mathcal{V}\|_\infty\}2^M\mathfrak{m}(\mathcal{V})\|\mathcal{V}\|_\infty\mathtt{TV}_{\widetilde{\mathcal{G}}} \lesssim \min\{\mathtt{M}_{\widetilde{\mathcal{G}}}, 2^{-\frac{M}{d}}\mathtt{L}_{\widetilde{\mathcal{G}}}\}2^{-\frac{M}{d}}\mathtt{TV}_{\widetilde{\mathcal{G}}}.
\end{aligned}$$

Denote $\widetilde{\mathtt{TV}} = \max\{\mathtt{TV}_{\widetilde{\mathcal{G}}}, \mathtt{TV}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}\}$ and $\widetilde{\mathtt{L}} = \max\{\mathtt{L}_{\widetilde{\mathcal{G}}}, \mathtt{L}_{\widetilde{\mathcal{G}}\mathcal{V}_{\mathcal{R}}}\}$. We balance the errors in Equations SA-23 and SA-24 by choosing

$$M = \min\Big\{\Big\lceil \log_2\Big(\frac{n\mathtt{TV}_{\mathcal{G}}}{\mathtt{E}_{\mathcal{G}}}\Big)\Big\rceil, \Big\lceil \log_2\Big(\frac{n\widetilde{\mathtt{TV}}\widetilde{\mathtt{L}}}{\mathtt{E}_{\widetilde{\mathcal{G}}}\mathtt{M}_{\widetilde{\mathcal{G}}}}\Big)\Big\rceil\Big\}, \quad N = \max\Big\{\Big\lceil \log_2\Big(\frac{n\mathtt{M}_{\widetilde{\mathcal{G}}}^{d+1}}{\widetilde{\mathtt{TV}}^d\mathtt{E}_{\widetilde{\mathcal{G}}}}\Big)\Big\rceil, \Big\lceil \log_2\Big(\frac{n^2\mathtt{M}_{\widetilde{\mathcal{G}}}^{2d+2}}{\widetilde{\mathtt{TV}}^d\widetilde{\mathtt{L}}^d\mathtt{E}_{\widetilde{\mathcal{G}}}^2}\Big)\Big\rceil\Big\}.$$

Plug in Equations SA-23 and SA-24, and use the relation between $\widetilde{\mathcal{G}}$ and $\mathcal{G}$ in Lemma SA.12, we have for any $t > 0$, with probability at least $1 - 8\exp(-t)$,

$$\begin{aligned}\|R_n - Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} \leq & \sqrt{d}\min\left\{\left(\frac{\mathtt{c}_1^d\mathtt{E}_{\mathcal{G}}\mathtt{TV}^d\mathtt{M}_{\mathcal{G}}^{d+1}}{n}\right)^{\frac{1}{2(d+1)}}, \left(\frac{\mathtt{c}_1^{\frac{d}{2}}\mathtt{c}_2^{\frac{d}{2}}\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}\mathtt{TV}^{\frac{d}{2}}\mathtt{L}^{\frac{d}{2}}}{n}\right)^{\frac{1}{d+2}}\right\}(t + \mathtt{c}_4\log n + \log \mathtt{N}(\delta))^{\alpha+\frac{3}{2}} \\ & + \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}(t + \mathtt{c}_4\log n + \log\mathtt{N}(\delta))^{\alpha+1}.\end{aligned}$$

The results the follows by the control on meshing error from Lemma SA.18. □

SA-III.29 Proof of Theorem 4

By Lemma SA.13, $N(\delta) \leq \mathtt{c}\delta^{-\mathtt{d}}$. The result follows by plugging in $N(\delta)$ to Lemma SA.4.

References

Adamczak, R. (2008). "A tail inequality for suprema of unbounded empirical processes with applications to Markov chains," *Electronic Journal of Probability*, *13*(34), 1000–1034.

Bretagnolle, J. and Massart, P. (1989). "Hungarian Constructions from the Nonasymptotic Viewpoint," *Annals of Probability*, *17*(1), 239–256.

Brown, L. D., Cai, T. T., and Zhou, H. H. (2010). "Nonparametric regression in exponential families," *Annals of Statistics*, *38*(4), 2005–2046.

Cattaneo, M. D., Chandak, R., Jansson, M., and Ma, X. (2024). "Local Polynomial Conditional Density Estimators," *Bernoulli*.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). "Gaussian approximation of suprema of empirical processes," *Annals of Statistics*, *42*(4), 1564–1597.

De Giorgi, E. (1955). "Nuovi teoremi relativi alle misure $(r-1)$-dimensionali in uno spazio ad r dimensioni," *Ricerche Mat.*, *4*, 95–113.

Dudley, R. M. (2014). *Uniform central limit theorems*, *142*: Cambridge university press.

Giné, E. and Nickl, R. (2016). *Mathematical Foundations of Infinite-dimensional Statistical Models*: Cambridge University Press.

Rio, E. (1994). "Local Invariance Principles and Their Application to Density Estimation," *Probability Theory and Related Fields*, *98*(1), 21–45.

Sakhanenko, A. (1996). "Estimates for the accuracy of coupling in the central limit theorem," *Siberian Mathematical Journal*, *37*(4), 811–823.

van der Vaart, A. and Wellner, J. (2013). *Weak convergence and empirical processes: with applications to statistics*: Springer Science & Business Media.