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

Matias D. Cattaneo† Ruiqi (Rae) Yu‡

November 12, 2024

Abstract

This supplement appendix reports additional theoretical results not discussed in the paper to conserve space, and provides all the technical proofs. Section SA-I introduces additional notation and definitions used in the proofs. Section SA-II studies the general empirical process (Section 3 in the paper). Section SA-III studies the multiplicative-separable empirical process (not discussed in the paper but of independent interest). Section SA-IV studies the residual-based empirical process (Section 4 in the paper). Section SA-V studies the three empirical processes in the context of quasi-uniform Haar basis (Section 5 in the paper).

*Corresponding author: rae.yu@princeton.edu. Support from the National Science Foundation through grants DMS-2210561 and SES-2241575 is gratefully acknowledged.

†Department of Operations Research and Financial Engineering, Princeton University.

‡Department of Operations Research and Financial Engineering, Princeton University.

Contents

SA-I Additional Notation 2
SA-I.1 Additional Main Definitions . . . 2

SA-II General Empirical Process 3
SA-II.1 Preliminary Technical Results . . . 5
SA-II.1.1 Cells Expansions . . . 5
SA-II.1.2 Projection onto Piecewise Constant Functions . . . 6
SA-II.1.3 Strong Approximation Constructions . . . 8
SA-II.1.4 Meshing Error . . . 17
SA-II.1.5 Strong Approximation Errors . . . 18
SA-II.1.6 Projection Error . . . 21
SA-II.2 Surrogate Measure and Normalizing Transformation . . . 22
SA-II.3 General Result: Proof of Theorem 1 . . . 27
SA-II.4 Additional Results . . . 31
SA-II.5 Proofs of Corollaries 1, 2, and 3 . . . 32
SA-II.6 Example 1: Kernel Density Estimation . . . 32
SA-II.6.1 Surrogate Measure and Normalizing Transformation . . . 33
SA-II.6.2 Class $\mathcal{H}$ and Its Corresponding Constants . . . 33

SA-III Multiplicative-Separable Empirical Process 34
SA-III.1 Preliminary Technical Results . . . 35
SA-III.1.1 Cells Expansions . . . 36
SA-III.1.2 Projection onto Piecewise Constant Functions . . . 37
SA-III.1.3 Strong Approximation Construction . . . 38
SA-III.1.4 Meshing Error . . . 42
SA-III.1.5 Strong Approximation Errors . . . 43
SA-III.1.6 Projection Error . . . 46
SA-III.2 General Result . . . 51
SA-III.3 Additional Results . . . 56

SA-IV Residual-Based Empirical Processes 57
SA-IV.1 Preliminary Technical Results . . . 58
SA-IV.1.1 Projection onto Piecewise Constant Functions . . . 58
SA-IV.1.2 Strong Approximation Constructions . . . 59
SA-IV.1.3 Meshing Error . . . 61
SA-IV.1.4 Strong Approximation Errors . . . 62
SA-IV.1.5 Projection Error . . . 64
SA-IV.2 General Result . . . 65
SA-IV.3 Proof of Theorem 2 . . . 68
SA-IV.4 Proof of Corollary 4 . . . 68
SA-IV.5 Example: Local Polynomial Estimators . . . 68

SA-V Quasi-Uniform Haar Basis 76
SA-V.1 General Empirical Process . . . 76
SA-V.1.1 Proof of Theorem 3 . . . 76
SA-V.1.2 Proof of Corollary 5 . . . 78
SA-V.1.3 Example 2: Histogram Density Estimation . . . 78
SA-V.2 Residual-Based (and Multiplicative Separable) Empirical Process . . . 79
SA-V.2.1 Proof of Theorem 4 . . . 84
SA-V.2.2 Proof of Corollary 6 . . . 85
SA-V.3 Example: Haar Partitioning-based Regression . . . 85

SA-I Additional Notation

We introduce additional notation and definitions complementing those given in Section 2 of the paper. See Ambrosio *et al.* (2000), van der Vaart and Wellner (2013), Giné and Nickl (2016), and references therein, for background definitions and more details.

Let $\mathcal{U}, \mathcal{V} \subseteq \mathbb{R}^q$. We define $\mathcal{U} - \mathcal{V} = \{\mathbf{u} - \mathbf{v} : \mathbf{u} \in \mathcal{U}, \mathbf{v} \in \mathcal{V}\}$. We define $\mathcal{U} \triangle \mathcal{V} = (\mathcal{U} \setminus \mathcal{V}) \cup (\mathcal{V} \setminus \mathcal{U})$. Let $\det(\mathbf{A})$ be the determinant of the matrix $\mathbf{A}$. Let $\Phi(z)$ be the distribution function of $\mathsf{Normal}(0,1)$, and $\mathsf{Bern}(p)$ denote the Bernoulli distribution with parameter $p \in (0,1)$. For a real-valued random variable X, the L_p-norm is defined as $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ for $1 \leq p < \infty$. The σ-algebra generated by X is denoted by $\boldsymbol{\sigma}(X)$. For $\alpha > 0$, the ψ_α-norm of X is given by $\|X\|_{\psi_\alpha} = \min\{\lambda > 0 : \mathbb{E}\big[\exp\big((\frac{|X|}{\lambda})^\alpha\big)\big] \leq 2\}$. For $\mathbf{x} \in \mathbb{R}^q$ and $r > 0$, let $B(\mathbf{x}, r)$ denote the Euclidean ball with radius r centered at $\mathbf{x}$. For a matrix $\mathbf{A} \in \mathbb{R}^{q \times q}$, $\|\mathbf{A}\|$ denotes its operator norm. Using standard empirical process notation, $\mathbb{E}_n[f(\mathbf{x}_i)]$ denotes the empirical average $n^{-1}\sum_{i=1}^n [f(\mathbf{x}_i) - \mathbb{E}[f(\mathbf{x}_i)]]$ based on random sample $(\mathbf{x}_i : 1 \leq i \leq n)$. For sequences of real numbers, we write $a_n = \Omega(b_n)$ if there exists some constant C and $N > 0$ such that $n > N$ implies $|a_n| \geq C|b_n|$.

Let $\mathcal{S} \subseteq \mathbb{R}^q$ and Q be a measure on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$. The semi-metric $\mathfrak{d}_Q$ on $L_2(Q)$ is defined by $\mathfrak{d}_Q(f,g) = \big(\|f-g\|_{Q,2}^2 - \big(\int f\,dQ - \int g\,dQ\big)^2\big)^{1/2}$, for $f, g \in L_2(Q)$. For a class $\mathcal{F} \subseteq L_2(Q)$, let $C(\mathcal{F}, \mathfrak{d}_Q)$ denote the class of all continuous functionals on the space $(\mathcal{F}, \mathfrak{d}_Q)$. For $\alpha > 0$, the C^α-norm of a real-valued measurable function f on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ is given by $\|f\|_{C^\alpha} = \max_{|k| \leq \lfloor \alpha \rfloor} \sup_{\mathbf{x} \in \mathcal{S}} |D^k f(\mathbf{x})| + \max_{|k| = \alpha} \sup_{\mathbf{x} \neq \mathbf{y} \in \mathcal{S}} \frac{|D^k f(\mathbf{x}) - D^k f(\mathbf{y})|}{\|\mathbf{x}-\mathbf{y}\|_2^{\alpha - \lfloor \alpha \rfloor}}$. The space $C^\alpha(\mathcal{S})$ denotes the collection of all real-valued measurable functions on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$ with C^α-norm bounded by 1. For real-valued functions f, g on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, the convolution of f and g is the function $f * g$ such that $f * g(x) = \int_{-\infty}^{\infty} f(y) g(x-y) dy, x \in \mathbb{R}$. If $\mathcal{F}$ and $\mathcal{G}$ are two sets of functions from measure space $(\mathcal{U}, \mathcal{B}(\mathcal{U}))$ to $\mathbb{R}$ and $(\mathcal{V}, \mathcal{B}(\mathcal{V}))$ to $\mathbb{R}$, respectively, then $\mathcal{F} \cdot \mathcal{G}$ denotes the class of measurable functions $\{f \cdot g : f \in \mathcal{F}, g \in \mathcal{G}\}$ from $(\mathcal{U} \times \mathcal{V}, \mathcal{B}(\mathcal{U}) \otimes \mathcal{B}(\mathcal{V}))$ to $\mathbb{R}$. For a semi-metric space $(\mathcal{F}, \mathfrak{d})$ of real-valued measurable functions on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, $N_{[\,]}(\varepsilon, \mathcal{F}, \mathfrak{d})$ denotes the bracketing number.

For a probability measure P on $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, a *$P$-Brownian bridge* is a centered Gaussian random function $(W_P(f) : f \in L_2(P))$ with covariance given by $\mathbb{E}[W_P(f) W_P(g)] = P(fg) - P(f)P(g)$ for $f, g \in L_2(P)$. A class $\mathcal{F} \subseteq L_2(P)$ is said to be P-pregaussian if there exists a version of the P-Brownian bridge W_P such that $W_P \in C(\mathcal{F}; \mathfrak{d}_P)$ almost surely.

Finally, we use $a_n \lesssim b_n$ to denote that $a_n = O(b_n)$ with only a universal constant, not a function of the data generating process or related parameters. For $K \in \mathbb{N}$, we repeatedly employ the index sets $\mathcal{I}_K = \{(j,k) \in \mathbb{N} \times \mathbb{N} : 1 \leq j \leq K, 0 \leq k < 2^{K-j}\}$ and $\mathcal{J}_K = \{(j,k) \in \mathbb{N} \times \mathbb{N} : 0 \leq j \leq K, 0 \leq k < 2^{K-j}\}$.

SA-I.1 Additional Main Definitions

Let $\mathcal{F}$ be a class of measurable functions from a probability space $(\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q), \mathbb{P})$ to $\mathbb{R}$. We introduce several additional definitions that capture properties of $\mathcal{F}$, complementing those in Section 2.1.

Definition SA.1. *For a non-empty $\mathcal{C} \subseteq \mathbb{R}^q$, the smoothed uniform total variation of $\mathcal{F}$ over $\mathcal{C}$ is*

$$\mathtt{TV}^*_{\mathcal{F},\mathcal{C}} = \sup_{f \in \mathcal{F}} \inf_{(f_\ell)_{\ell \in \mathbb{N}}} \limsup_{\ell \to \infty} \mathtt{TV}_{\{f_\ell\},\mathcal{C}},$$

where the infimum is taken over all sequences of functions $(f_\ell)_{\ell \in \mathbb{N}}$ such that $f_\ell \to f \in \mathcal{F}$ a.s.-$\mathfrak{m}$ on $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, and f_ℓ is differentiable and bounded by $2\mathtt{M}_{\mathcal{F},\mathcal{C}}$ on $\mathcal{C}$ for all $\ell \geq 1$.

Definition SA.2. *For a non-empty $\mathcal{C} \subseteq \mathbb{R}^q$, the smoothed uniform local total variation of $\mathcal{F}$ over $\mathcal{C}$ is a positive number $\mathtt{K}^*_{\mathcal{F},\mathcal{C}}$ such that for any cube $\mathcal{D} \subseteq \mathbb{R}^q$ with edges of length ℓ parallel to the coordinate axises,*

$$\mathtt{TV}^*_{\mathcal{F},\mathcal{D}\cap\mathcal{C}} \leq \mathtt{K}^*_{\mathcal{F},\mathcal{C}} \ell^{d-1}.$$

Suppose $\mathcal{S}$ is also a class of measurable functions from the probability space $(\mathbb{R}^q, \mathcal{B}(\mathbb{R}^q), \mathbb{P})$ to $\mathbb{R}$. We generalize the definition of the uniform covering number to $\mathcal{F} \times \mathcal{S}$.

Definition SA.3. *For a non-empty $\mathcal{C} \subseteq \mathbb{R}^q$, the uniform covering number of $\mathcal{F} \times \mathcal{S}$ with envelope $M_{\mathcal{F},\mathcal{C}} M_{\mathcal{S},\mathcal{C}}$ over $\mathcal{C}$ is*

$$\mathtt{N}_{\mathcal{F}\times\mathcal{S},\mathcal{C}}(\delta, M_{\mathcal{F},\mathcal{C}} M_{\mathcal{S},\mathcal{C}}) = \sup_{\mu} N(\mathcal{F} \times \mathcal{S}, \lambda_\mu, \delta \| M_{\mathcal{F},\mathcal{C}} M_{\mathcal{S},\mathcal{C}} \|_{\mu,2}), \qquad \delta \in (0, \infty),$$

where the supremum is taken over all finite discrete measures on $(\mathcal{C}, \mathcal{B}(\mathcal{C}))$, and λ_μ is the semi-metric on $\mathcal{F} \times \mathcal{S}$ defined by

$$\lambda_\mu((f_1, g_1), (f_2, g_2))^2 = \int_{\mathcal{C}} (f_1(\mathbf{x}) g_1(\mathbf{x}) - f_2(\mathbf{x}) g_2(\mathbf{x}))^2 \, d\mu(\mathbf{x}).$$

We assume that $M_{\mathcal{F},\mathcal{C}}(\mathbf{u})$ and $M_{\mathcal{S},\mathcal{C}}(\mathbf{u})$ are finite for every $\mathbf{u} \in \mathcal{C}$.

Definition SA.4. *For a non-empty $\mathcal{C} \subseteq \mathbb{R}^q$, the uniform entropy integral of $\mathcal{F} \times \mathcal{S}$ with envelope $M_{\mathcal{F},\mathcal{C}} M_{\mathcal{S},\mathcal{C}}$ over $\mathcal{C}$ is*

$$J_{\mathcal{C}}(\delta, \mathcal{F} \times \mathcal{S}, M_{\mathcal{F},\mathcal{C}} M_{\mathcal{S},\mathcal{C}}) = \int_0^\delta \sqrt{1 + \log \mathtt{N}_{\mathcal{F}\times\mathcal{S},\mathcal{C}}(\varepsilon, M_{\mathcal{F},\mathcal{C}} M_{\mathcal{S},\mathcal{C}})} \, d\varepsilon,$$

where it is assumed that $M_{\mathcal{F},\mathcal{C}}(\mathbf{u}) M_{\mathcal{S},\mathcal{C}}(\mathbf{u})$ is finite for every $\mathbf{u} \in \mathcal{C}$.

SA-II General Empirical Process

Recall that $\mathbf{x}_i \in \mathcal{X} \subseteq \mathbb{R}^d$, $i = 1, \ldots, n$, are i.i.d. random vectors supported on a background probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and the general empirical process is

$$X_n(h) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \big(h(\mathbf{x}_i) - \mathbb{E}[h(\mathbf{x}_i)] \big), \qquad h \in \mathcal{H},$$

where $\mathcal{H}$ is a possibly n-varying class of functions. As briefly explained after Theorem 1 is presented in the paper, its proof relies on the following decomposition:

$$\begin{aligned}
&\|X_n - Z_n^X\|_{\mathcal{H}} \\
&\leq \|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}} + \|X_n - Z_n^X\|_{\mathcal{H}_\delta} + \|Z_n^X \circ \pi_{\mathcal{H}_\delta} - Z_n^X\|_{\mathcal{H}} \\
&\leq \|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}} + \|X_n - \Pi_0 X_n\|_{\mathcal{H}_\delta} + \|\Pi_0 X_n - \Pi_0 Z_n^X\|_{\mathcal{H}_\delta} + \|\Pi_0 Z_n^X - Z_n^X\|_{\mathcal{H}_\delta} + \|Z_n^X \circ \pi_{\mathcal{H}_\delta} - Z_n^X\|_{\mathcal{H}},
\end{aligned}$$

where $\mathcal{H}_\delta$ denotes a discretization (or meshing) of $\mathcal{H}$ (i.e., δ-net of $\mathcal{H}$), and the terms $\|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}}$ and $\|Z_n^X \circ \pi_{\mathcal{H}_\delta} - Z_n^X\|_{\mathcal{H}}$ capture the fluctuations (or oscillations) of X_n and Z_n^X relative to the meshing for each of the stochastic processes. These terms are handled using standard arguments for empirical processes. Then, following Rio (1994), the term $\|X_n - Z_n^X\|_{\mathcal{H}_\delta}$ is further decomposed into three terms: $\|\Pi_0 X_n - \Pi_0 Z_n^X\|_{\mathcal{H}_\delta}$ and $\|\Pi_0 Z_n^X - Z_n^X\|_{\mathcal{H}_\delta}$ represent a mean square projection onto a Haar function space, where $\Pi_0 X_n(h) = X_n \circ \Pi_0 h$ with Π_0 the L_2 projection onto piecewise constant functions on a carefully chosen partition of $\mathcal{X}$, while the final term $\|\Pi_0 X_n - \Pi_0 Z_n^X\|_{\mathcal{H}_\delta}$ captures the coupling between the projected empirical process and the projected Gaussian process (on a δ-net of $\mathcal{H}$, after the L_2 projection).

The proof of Theorem 1 first constructs the Gaussian process $(Z_n^X(h) : h \in \mathcal{H})$ on a possibly enlarged probability space supporting the empirical process $(X_n(h) : h \in \mathcal{H})$, and then bounds each of the five error terms described above. The proof is given in Section SA-II.3, and it exploits the existence of a surrogate measure and normalizing transformation (Section SA-II.2), along with a collection preliminary technical results (Section SA-II.1) that may be of independent interest. More specifically, our preliminary technical results are organized as follows:

- Section SA-II.1.1 introduces a class of recursive quasi-dyadic cells expansion of $\mathcal{X}$, which we employ to generalize prior dyadic cell results in the literature.
- Section SA-II.1.2 introduces the L_2 projection onto piecewise constant functions, which can be written as a linear combination of the Haar basis based on the cells. As a consequence, the empirical process indexed by L_2-projected functions can be written as linear combinations of counts of i.i.d. data.
- Section SA-II.1.3 constructs the Gaussian process $(Z_n^X(h) : h \in \mathcal{H})$. Since the constant approximation within each recursive partitioning cell generates counts based on i.i.d. data, the construction boils down to coupling binomial random variables with Gaussian random variables. The celebrated Tusnády's inequality couples $\mathsf{Bin}(n, \frac{1}{2})$ with $\mathsf{Normal}(\frac{n}{2}, \frac{n}{4})$, and gives an almost sure bound on the coupling error. In particular, the Gaussian random variable is given by a quantile transformation of the binomial random variable. Building on the quantile transformation idea, our Lemma SA.4 studies the coupling between $\mathsf{Bin}(n,p)$ and $\mathsf{Normal}(np, np(1-p))$, with the error bound given on a high probability set. Due to the dyadic correlation structure, a conditional quantile transformation is used to generated the Binomial–Gaussian pairs down the dyadic cells. Since the constructed Gaussian random variables have a joint distribution that coincides with the Brownian bridge integrated on cells, the Skorohod embedding lemma (Dudley, 2014, Lemma 3.35) is then used to construct the Brownian bridge $(Z_n^X(h) : h \in \mathcal{H})$ on a possibly enriched probability space supporting the data distribution.
- Section SA-II.1.4 handles the meshing errors $\|X_n - X_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}}$ and $\|Z_n^X \circ \pi_{\mathcal{H}_\delta} - Z_n^X\|_{\mathcal{H}}$ using standard empirical process results, which give the contribution $\mathsf{F}(\delta)$ emerging from Talagrand's inequality (Giné and Nickl, 2016, Theorem 3.3.9) combined with a standard maximal inequality (Chernozhukov *et al.*, 2014, Theorem 5.2). This allows us to focus on the error on the δ-net to study $\|X_n - Z_n^X\|_{\mathcal{H}_\delta}$.
- Section SA-II.1.5 handles the strong approximation error $\|\Pi_0 X_n - \Pi_0 Z_n^X\|_{\mathcal{H}_\delta}$. Building on the Tusnády's Lemma, Rio (1994, Theorem 2.1) established a remarkable coupling result for bounded functions L_2-projected on a dyadic cells expansion of $\mathcal{X}$. Our Lemma SA.7 builds on his powerful ideas, and establishes an analogous result for the case of Lipschitz functions L_2-projected on dyadic cells expansions of $\mathcal{X}$, thereby obtaining a tighter coupling error. A limitation of these results is that they only apply to a dyadic cell expansion due to the specifics of Tusnády's Lemma. Leveraging the coupling between $\mathsf{Bin}(n,p)$ and $\mathsf{Normal}(np, np(1-p))$, our Lemma SA.8 established a coupling result for bounded functions L_2-projected on a quasi-dyadic cells, although the result is restricted to a high probability event.

- Section SA-II.1.6 handles the L_2-projection errors $\|X_n - \Pi_0 X_n\|_{\mathcal{H}_\delta}$ and $\|\Pi_0 Z_n^X - Z_n^X\|_{\mathcal{H}_\delta}$ using Bernstein inequality, and taking into account explicitly the potential Lipschitz structure of the functions as well as the generic cell structure.

Section SA-II.2 introduces a reduction argument via the surrogate measure and the normalizing transformation in order to apply the preliminary technical results from Section SA-II.1 to prove Theorem 1. Specifically, the surrogate measure and normalizing transformation reduce the problem to the case where $\mathbf{x}_i \sim \mathsf{Uniform}([0,1]^d)$. Section SA-II.3 gives the proof of Theorem 1. Section SA-II.4 presents additional results of independent interest, which are used in Section SA-II.5 to prove the results discussed in Section 3.2 of the paper. Finally, Section SA-II.6 provides technical details underlying Example 1 in the paper.

SA-II.1 Preliminary Technical Results

This section presents preliminary technical results that are used to prove Theorem 1. Whenever possible, these results are presented at a higher level of generality, and therefore may be of independent theoretical interest. Throughout this section, we employ the following assumption.

Assumption SA.1. *Suppose $(\mathbf{x}_i : 1 \leq i \leq n)$ are i.i.d. random vectors taking values in $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with common law $\mathbb{P}_X$ supported on $\mathcal{X} \subseteq \mathbb{R}^d$, and the following conditions hold.*

(i) *$\mathcal{H}$ is a real-valued pointwise measurable class of functions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.*

(ii) *$\mathtt{M}_{\mathcal{H},\mathcal{X}} < \infty$ and $J_{\mathcal{X}}(1, \mathcal{H}, \mathtt{M}_{\mathcal{H},\mathcal{X}}) < \infty$.*

Compared to the assumptions in Theorem 1, this assumption does not require the existence of a surrogate measure and normalizing transformation. It will be applied in the analysis of each term in the error decomposition, where we work with the $\mathbb{P}_X$ distribution. Section SA-II.2 illustrates how the normalizing transformation enables the use of the surrogate measure $\mathbb{Q}_{\mathcal{H}}$, providing greater flexibility in the data generating process. This reduction through the normalizing transformation is a crucial step in the proof of Theorem 1 (Section SA-II.3).

SA-II.1.1 Cells Expansions

We introduce two definitions of quasi-dyadic cells expansions. Recall that $\mathcal{I}_K = \{(j,k) \in \mathbb{N} \times \mathbb{N} : 1 \leq j \leq K, 0 \leq k < 2^{K-j}\}$ and $\mathcal{J}_K = \{(j,k) \in \mathbb{N} \times \mathbb{N} : 0 \leq j \leq K, 0 \leq k < 2^{K-j}\}$.

Definition SA.5 (Quasi-Dyadic Expansion)**.** *A collection of Borel measurable sets in $\mathbb{R}^d$, $\mathcal{C}_K(\mathbb{P}, \rho) = \{\mathcal{C}_{j,k} : (j,k) \in \mathcal{J}_K\}$, is called a quasi-dyadic expansion of depth K with respect to probability measure $\mathbb{P}$ if the following three conditions hold:*

(i) $\mathbb{P}(\mathcal{C}_{K,0}) = 1$.

(ii) *$\mathcal{C}_{j,k} = \mathcal{C}_{j-1,2k} \sqcup \mathcal{C}_{j-1,2k+1}$, for all $(j,k) \in \mathcal{J}_K$.*

(iii) $\max_{0 \leq k < 2^K} \mathbb{P}(\mathcal{C}_{0,k}) / \min_{0 \leq k < 2^K} \mathbb{P}(\mathcal{C}_{0,k}) \leq \rho < \infty$.

When $\rho = 1$, $\mathcal{C}_K(\mathbb{P}, 1)$ is called a dyadic expansion 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 $(j,k) \in \mathcal{I}_K$, 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.5. Furthermore, $\mathbb{P}(\mathcal{C}_{j-1,2k}) = \mathbb{P}(\mathcal{C}_{j-1,2k+1}) = \frac{1}{2}\mathbb{P}(\mathcal{C}_{j,k})$ in the special case $\rho = 1$, 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.6 (Axis-Aligned Quasi-Dyadic Expansion)**.** *A collection of Borel measurable sets in* $\mathbb{R}^d$*,* $\mathcal{A}_K(\mathbb{P},\rho) = \{\mathcal{C}_{j,k} : (j,k) \in \mathcal{J}_K\}$*, is an axis-aligned quasi-dyadic expansion of depth* K *with respect to probability measure* $\mathbb{P}$ *if it can be constructed via the following procedure:*

(i) Initialization $(q = 0)$*: Take* $\mathcal{C}_{K-q,0} = \mathrm{Supp}(\mathbb{P})$*.*

(ii) 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* $\mathcal{C}_{K-q,2k} = \mathcal{C}_{K-q+1,k} \cap \{\mathbf{x} \in \mathbb{R}^d : \mathbf{e}_s^\top \mathbf{x} \leq c_{K-q+1,k}\}$ *and* $\mathcal{C}_{K-q,2k+1} = \mathcal{C}_{K-q+1,k} \cap \{\mathbf{x} \in \mathbb{R}^d : \mathbf{e}_s^\top \mathbf{x} > c_{K-q+1,k}\}$ *where* $c_{K-q+1,k}$ *is a number chosen so 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 the collection* $(\mathcal{C}_{0,k} : 0 \leq k < 2^K)$ *has been constructed.*

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

SA-II.1.2 Projection onto Piecewise Constant Functions

For a quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P},\rho)$, the span of the Haar basis based on the terminal cells is

$$\mathcal{E}_K = \mathrm{Span}\{\mathbb{1}_{\mathcal{C}_{0,k}} : 0 \leq k < 2^K\}.$$

For $h \in L_2(\mathbb{P})$, the mean square projection of h onto $\mathcal{E}_K$ is

$$\Pi_0(\mathcal{C}_K(\mathbb{P},\rho))[h] = \sum_{0 \leq k < 2^K} \frac{\mathbb{1}_{\mathcal{C}_{0,k}}}{\mathbb{P}(\mathcal{C}_{0,k})} \int_{\mathcal{C}_{0,k}} h(\mathbf{u}) d\mathbb{P}(\mathbf{u}).$$

Because $\Pi_0(\mathcal{C}_K(\mathbb{P},\rho))[h]$ is a linear combination of Haar functions, we obtain the following orthogonal decomposition.

Lemma SA.1. *For a quasi-dyadic expansion* $\mathcal{C}_K(\mathbb{P},\rho)$ *and any* $h \in L_2(\mathbb{P})$*, the mean square projection* $\Pi_0(\mathcal{C}_K(\mathbb{P},\rho))[h]$ *satisfies*

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

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 all $(j,k) \in \mathcal{I}_K = \{(j,k) \in \mathbb{N} \times \mathbb{N} : 1 \leq j \leq K, 0 \leq k < 2^{K-j}\}$.

Proof of Lemma SA.1. First, we show that $\{e_{K,0}\} \cup \{\widetilde{e}_{j,k} : (j,k) \in \mathcal{I}_K\}$ is an orthogonal basis. For each $(j,k) \in \mathcal{I}_K$,

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

where $\langle \cdot, \cdot \rangle$ denotes the inner product on $L_2(\mathbb{P})$ given by $\langle f, g \rangle = \int_{\mathbb{R}^d} f(\mathbf{u}) g(\mathbf{u}) d\mathbb{P}(\mathbf{u})$, $f, g \in L_2(\mathbb{P})$. Let $(j_1,k_1), (j_2,k_2) \in \mathcal{I}_K$ such that $(j_1,k_1) \neq (j_2,k_2)$. We show $\langle e_{j_1,k_1}, e_{j_2,k_2} \rangle = 0$ by considering two cases.

- *Case 1*: $j_1 = j_2$ and $k_1 \neq k_2$, 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*: $j_1 \neq j_2$ and, without loss of generality, we assume $j_1 < j_2$. By (1) in Definition SA.5, 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.5 again, either $\mathcal{C}_{j_1,k_1} \subseteq \mathcal{C}_{j_2-1,2k_2}$ or $\mathcal{C}_{j_1,k_1} \subseteq \mathcal{C}_{j_2-1,2k_2+1}$. Assume, without loss of generality, that $\mathcal{C}_{j_1,k_1} \subseteq \mathcal{C}_{j_2-1,2k_2}$. Then, for any $(j_1,k_1), (j_2,k_2) \in \mathcal{I}_K$,

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

Thus, $\{e_{K,0}\} \cup \{\widetilde{e}_{j,k} : (j,k) \in \mathcal{I}_K\}$ is an orthogonal basis for $\mathcal{E}_K$, and the L_2 projection for all $h \in L_2(\mathbb{P})$ is

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

For all $(j,k) \in \mathcal{I}_K$, the coefficients are given by

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

Moreover,

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

This concludes the proof. $\square$

To save notation, we will write Π_0 for $\Pi_0(\mathcal{C}_K(\mathbb{P},\rho))$ whenever the underlying cells expansion is clear from the context. For a class of functions $\mathcal{H}$ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P})$ such that $\mathcal{H} \subseteq L_2(\mathbb{P})$, denote $\Pi_0\mathcal{H} = \{\Pi_0 h : h \in \mathcal{H}\}$.

SA-II.1.3 Strong Approximation Constructions

This section employs the notations and conventions introduced in Sections SA-II.1.1 and SA-II.1.2. Unless explicitly stated otherwise, we assume a quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P}_X,\rho)$ is given. Let $(\widetilde{\xi}_{j,k} : (j,k) \in \mathcal{I}_K)$ be i.i.d. standard Gaussian random variables. Let $F_{(j,k),m}$ be the cumulative distribution function of $(S_{j,k} - mp_{j,k})/\sqrt{mp_{j,k}(1-p_{j,k})}$, where $S_{j,k}$ is a $\mathsf{Bin}(m, p_{j,k})$ random variable with $p_{j,k} = \mathbb{P}_X(\mathcal{C}_{j-1,2k})/\mathbb{P}_X(\mathcal{C}_{j,k})$, and $G_{(j,k),m}(t) = \inf\{x : F_{(j,k),m}(x) \geq t\}$.

We define the collection of random variables $(U_{j,k} : (j,k) \in \mathcal{J}_K)$ and $(\widetilde{U}_{j,k} : (j,k) \in \mathcal{I}_K)$ via the following iterative scheme:

1. *Initialization* $(j = K)$: $U_{K,0} = n$.

2. *Iteration* $(j = K, K-1, \dots, 1)$: For each $1 \leq j \leq K$, and given $(U_{l,k} : j < l \leq K, 0 \leq k < 2^{K-l})$, solve for $(U_{j,k} : 0 \leq k < 2^{K-j})$ such that

$$\begin{aligned} \widetilde{U}_{j,k} &= \sqrt{U_{j,k}p_{j,k}(1-p_{j,k})}\, G_{(j,k),U_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k}), \\ \widetilde{U}_{j,k} &= (1-p_{j,k})U_{j-1,2k} - p_{j,k}U_{j-1,2k+1} = U_{j-1,2k} - p_{j,k}U_{j,k}, \\ U_{j-1,2k} &+ U_{j-1,2k+1} = U_{j,k}, \end{aligned} \tag{SA-1}$$

where $0 \leq k < 2^{K-j}$. Continue till $(U_{0,k} : 0 \leq k < 2^K)$ are defined.

Then, $(U_{j,k} : (j,k) \in \mathcal{J}_K)$ has the same joint distribution as $(\sum_{i=1}^n e_{j,k}(\mathbf{x}_i) : (j,k) \in \mathcal{J}_K)$ from Lemma SA.1. By the Vorob'ev–Berkes–Philipp theorem (Dudley, 2014, Theorem 1.31), $(\widetilde{\xi}_{j,k} : (j,k) \in \mathcal{I}_K)$ 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 $(j,k) \in \mathcal{J}_K$. We will show that the $\widetilde{\xi}_{j,k}$'s can be given as a Brownian bridge indexed by $\widetilde{e}_{j,k}$'s from Lemma SA.1. Recall the definitions given in Section SA-II.1.2.

Lemma SA.2. *Suppose Assumption SA.1 holds, and a quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P}_X,\rho)$ is given. Then, $\mathcal{H} \cup \Pi_0\mathcal{H} \subseteq L_2(\mathbb{P}_X)$ and is $\mathbb{P}_X$-pregaussian.*

Proof of Lemma SA.2. To simplify notation, the parameters of $\mathcal{H}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{C}$ is omitted. Since $\mathtt{M}_{\mathcal{H}} < \infty$, $\mathcal{H} \cup \Pi_0\mathcal{H} \subseteq L_2(\mathbb{P}_X)$. Definition of Π_0 from Section SA-II.1.2 implies that $\mathtt{M}_{\mathcal{H}}$ is an envelope for $\Pi_0\mathcal{H}$.

Claim: For all $0 < \delta < 1$, $J(\Pi_0\mathcal{H}, \mathtt{M}_{\mathcal{H}}, \delta) \leq J(\mathcal{H}, \mathtt{M}_{\mathcal{H}}, \delta)$.

Proof of Claim: Let Q be a finite discrete measure on $\mathcal{X}$. Let $f, g \in \mathcal{H}$. Then, by the definition of Π_0 and Jensen's inequality,

$$\|\Pi_0 f - \Pi_0 g\|_{Q,2}^2 \leq \sum_{0 \leq k < 2^K} Q(\mathcal{C}_{0,k}) 2^K \int_{\mathcal{C}_{0,k}} (f-g)^2 d\mathbb{P}_X.$$

Define a measure $\widetilde{Q}$ such that for any $A \in \mathcal{B}(\mathbb{R}^d)$, $\widetilde{Q}(A) = \sum_{0 \leq k < 2^K} Q(\mathcal{C}_{0,k}) 2^K \mathbb{P}_X(A \cap \mathcal{C}_{0,k})$, then

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

Take $\mathcal{L}$ to be a $\delta \mathtt{M}_{\mathcal{H}}$-net of $\mathcal{H}$ over $\mathcal{X}$ with respect to $\|\cdot\|_{\widetilde{Q},2}$ with cardinality no greater than $\mathtt{N}_{\mathcal{H}}(\delta, \mathtt{M}_{\mathcal{H}})$. Let $\Pi_0 f$ be in an arbitrary function in $\Pi_0\mathcal{H}$, there exists $g \in \mathcal{L}$ such that $\|\Pi_0 f - \Pi_0 g\|_{Q,2}^2 \le \|f-g\|_{\widetilde{Q},2}^2 \le \delta^2 \mathtt{M}_{\mathcal{H}}^2$. The claim then follows.

It follows from the claim and (ii) from Assumption SA.1 that $J(1, \mathcal{H} \cup \Pi_0\mathcal{H}, \mathtt{M}_{\mathcal{H}}) < \infty$. By Dominated Convergence Theorem, $\lim_{\delta \downarrow 0} J(\delta, \mathcal{H} \cup \Pi_0\mathcal{H}) < \infty$. Since $\mathtt{M}_{\mathcal{H}} < \infty$, $\mathcal{H} \cup \Pi_0\mathcal{H}$ is totally bounded with respect to $\|\cdot\|_{\mathbb{P}_X,2}$. By separability of $\mathcal{H}$ and van der Vaart and Wellner (2013, Corollary 2.2.9), $\mathcal{H} \cup \Pi_0\mathcal{H}$ is $\mathbb{P}_X$-pregaussian. □

Under the conditions of Lemma SA.2, take $(Z_n^X(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H})$ to be a $\mathbb{P}_X$-Brownian bridge such that $Z_n^X(\cdot) \in C(\mathcal{H} \cup \Pi_0\mathcal{H}, \mathfrak{d}_{\mathbb{P}_X})$ almost surely. Since $(Z_n^X(\widetilde{e}_{j,k}) : (j,k) \in \mathcal{I}_K)$ are independent random variables with distribution $\mathsf{Normal}(0, \mathbb{P}_X(\mathcal{C}_{j-1,2k})\mathbb{P}_X(\mathcal{C}_{j-1,2k+1})\mathbb{P}_X(\mathcal{C}_{j,k})^{-1})$ for $(j,k) \in \mathcal{I}_K$, by Skorohod Embedding lemma (Dudley, 2014, Lemma 3.35), on a possibly enlarged probability space, the Brownian bridge $(Z_n^X(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H})$ can be constructed such that it satisfies

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

for all $(j,k) \in \mathcal{I}_K$. Moreover, for all $g \in \Pi_0\mathcal{H}$,

$$\sqrt{n} X_n(g) = \sum_{j=1}^{K} \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(g) \widetilde{U}_{j,k} \qquad \text{and} \qquad \sqrt{n} Z_n^X(g) = \sum_{j=1}^{K} \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(g) \widetilde{V}_{j,k},$$

where $\widetilde{V}_{j,k} = \sqrt{n} Z_n^X(\widetilde{e}_{j,k})$ for all $(j,k) \in \mathcal{I}_K$. The difference between $X_n(g)$ and $Z_n^X(g)$, for all $g \in \Pi_0\mathcal{H}$, will rely on the coefficient $(\widetilde{\beta}_{j,k}(g) : (j,k) \in \mathcal{I}_K, g \in \Pi_0\mathcal{H})$ 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 Rio (1994, Theorem 2.1) is stated for i.i.d. $\mathsf{Uniform}([0,1])$ random variables, the underlying process only depends through the counts of the random variables taking values in each interval of the form $[k2^{-j}, (k+1)2^{-j})$ for $(j,k) \in \mathcal{J}_K$, which have the same distribution as the counts $(\sum_{i=1}^n \mathbb{1}(\mathbf{x}_i \in \mathcal{C}_{j,k}) : (j,k) \in \mathcal{J}_K)$. Therefore, we have the following corollary of Rio (1994, Theorem 2.1) under Assumption SA.1. Recall the definitions given in Section SA-II.1.2.

Lemma SA.3. *Suppose Assumption SA.1 holds, a dyadic expansion $\mathtt{C}_K(\mathbb{P}_X, 1)$ is given, and $(Z_n^X(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H})$ is the Gaussian process constructed as in (SA-2) on a possibly enlarged probability space. Then, for any $g \in \Pi_0\mathcal{H}$ and any $t > 0$,*

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

where

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

using the definitions in Lemma SA.1, and

$$\mathtt{C}_{\mathcal{F},K} = \sup_{f \in \mathcal{F}} \min\left\{ \sup_{(j,k) \in \mathcal{I}_K} \left[\sum_{1 \le l < j} (j-l)(j-l+1) 2^{l-j} \sum_{0 \le m < 2^{K-l} : \mathcal{C}_{l,m} \subseteq \mathcal{C}_{j,k}} \widetilde{\beta}_{l,m}^2(f) \right] + \mathtt{M}_{\{f\},\mathcal{C}_{K,0}}^2, K\mathtt{M}_{\{f\},\mathcal{C}_{K,0}}^2 \right\},$$

for any $\mathcal{F} \subseteq \mathcal{H} \cup \Pi_0\mathcal{H}$.

Proof of Lemma SA.3. Let $(w_i : 1 \leq i \leq n)$ be i.i.d. $\mathsf{Uniform}([0,1])$, and $I_{j,k} = [k2^{-j}, (k+1)2^{-j})$ for $(j,k) \in \mathcal{J}_K$. 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 $t \in [0,1]$. Take

$$\begin{aligned} v_{j,k} &= \sqrt{n}\int_0^1 \mathbb{1}(t \in I_{j,k}) dB(t), \qquad (j,k) \in \mathcal{J}_K, \\ \widetilde{v}_{j,k} &= v_{j-1,2k} - v_{j-1,2k+1}, \qquad (j,k) \in \mathcal{I}_K. \end{aligned}$$

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

1. *Initialization*: $u_{K,0} = n$.
2. *Iteration*: For each $0 \leq j \leq K-1$, and given $(u_{l,k} : 0 \leq k < 2^{K-l}, j < l \leq K)$, then solve for $(u_{j,k} : 0 \leq k < 2^{K-j})$ such that

$$\begin{aligned} \widetilde{u}_{j,k} &= \frac{1}{2}\sqrt{u_{j,k}} G_{u_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k}), \\ \widetilde{u}_{j,k} &= \frac{1}{2}u_{j-1,2k} - \frac{1}{2}u_{j-1,2k+1} = u_{j-1,2k} - \frac{1}{2}u_{j,k}, \\ u_{j-1,2k} + u_{j-1,2k+1} &= u_{j,k}, \end{aligned}$$

for $0 \leq k < 2^{K-j}$. Continue till $(u_{0,k} : 0 \leq k < 2^K)$ are defined.

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 $(B(t) : 0 \leq t \leq 1)$ such that $u_{j,k} = \sum_{i=1}^n \mathbb{1}(w_i \in I_{j,k})$ for all $(j,k) \in \mathcal{J}_K$, almost surely.

Observe $\{(\widetilde{u}_{j,k}, \widetilde{v}_{j,k}) : (j,k) \in \mathcal{I}_K\}$ and $\{(\widetilde{U}_{j,k}, \widetilde{V}_{j,k}) : (j,k) \in \mathcal{I}_K\}$ have the same joint distribution, and

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

for all $g \in \Pi_0\mathcal{H}$. Thus the distribution of the process $\{(X_n(g), Z_n^X(g)) : g \in \Pi_0\mathcal{H}\}$ is the same as distribution of

$$\big((x_n(g), z_n(g)) : g \in \Pi_0\mathcal{H}\big) = \Bigg(\Bigg(\frac{1}{\sqrt{n}}\sum_{j=1}^K \sum_{0 \leq k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\widetilde{u}_{j,k}, \frac{1}{\sqrt{n}}\sum_{j=1}^K \sum_{0 \leq k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\widetilde{v}_{j,k}\Bigg) : g \in \Pi_0\mathcal{H}\Bigg),$$

We can then apply Rio (1994, Theorem 2.1) on $((x_n(g), z_n(g)) : g \in \Pi_0\mathcal{H})$ and use its equi-distribution as $((X_n(g), Z_n^X(g)) : g \in \Pi_0\mathcal{H})$ to get for any $\mathbf{p} = (p_1, \cdots, p_K)$ with positive components such that $\sum_{i=1}^K p_i \leq 1$, if we take $q_i = (2^i p_i)^{-1}$ and

$$M(\mathbf{p}, g) = 4 \sup_{(j,k) \in \mathcal{I}_K} \Bigg[\sum_{1 \leq l < j} q_{j-l} \sum_{0 \leq m < 2^{K-l} : \mathcal{C}_{l,m} \subseteq \mathcal{C}_{j,k}} \widetilde{\beta}_{l,m}^2(g)\Bigg], \qquad g \in \Pi_0\mathcal{H},$$

then for any $t > 0$ and $g \in \Pi_0 \mathcal{H}$,

$$\mathbb{P}\left(\sqrt{n}\big|X_n(g) - Z_n^X(g)\big| \geq (\sqrt{M(\mathbf{p}, g)} + \mathtt{M}_{\{g\},\mathcal{C}_{K,0}})t + \left(\Big(\sum_{i=1}^{K} q_i/2\Big)^{1/2} + 3\right)\|g\|_{\mathcal{E}_K}\sqrt{t}\right) \leq 2\exp(-t).$$

Following Rio (1994, Section 3), we choose either $p_i = \frac{1}{2}\big(\frac{1}{K} + \frac{1}{i(i+1)}\big)$ to get

$$M(\mathbf{p}, g) \leq 8K\mathtt{M}_{\{g\},\mathcal{C}_{K,0}} \qquad \text{and} \qquad \sum_{i=1}^{K} \frac{q_i}{2} < 8,$$

or $p_i = \frac{1}{i(i+1)}$ to get

$$M(\mathbf{p}, g) \leq \sup_{(j,k)\in\mathcal{I}_K}\left[\sum_{1\leq l<j}(j-l)(j-l+1)2^{l-j}\sum_{0\leq m<2^{K-l}:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}}\widetilde{\beta}^2_{l,m}(g)\right] \text{ and } \sum_{i=1}^{K}\frac{q_i}{2} < 4.$$

The conclusion then follows. □

Lemma SA.3 relies on a coupling of $\mathsf{Bin}(m, 1/2)$ random variables with Gaussian random variables. A weaker coupling also holds for $\mathsf{Bin}(m, p)$ with the error term only depending on how far away p is bounded away from 0 and 1, as the following lemma establishes.

Lemma SA.4. *Suppose $X \sim \mathsf{Bin}(m, p)$ where $0 < \underline{p} < p < \overline{p} < 1$. Then, there exists a a random variable $Z \sim \mathsf{Normal}(0, 1)$, and constants $c_0, c_1, c_2, c_3 > 0$ only depending on $\underline{p}$ and $\overline{p}$, such that whenever the event $A = \{|X - mp| \leq c_1 m\}$ occurs and $c_0\sqrt{m} \geq 1$, we have*

$$\Big|X - mp - \sqrt{mp(1-p)}Z\Big| \leq c_2 Z^2 + c_3 \qquad \textit{and} \qquad |X - mp| \leq \frac{1}{c_0} + 2\sqrt{mp(1-p)}|Z|.$$

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

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

and $c_1 = 15c_0\sqrt{\underline{p}(1-\overline{p})}$, $c_2 = 1/(15c_0)$, $c_3 = 1/c_0$, and then set

$$Z = \Phi^{-1} \circ F\Big((X - mp)/\sqrt{mp(1-p)}\Big).$$

That is, Z can be taken via the quantile transformation based on $F(x) = \mathbb{P}(X - mp < \sqrt{mp(1-p)}x)$.

Proof of Lemma SA.4. Let $(X_j : 1 \leq j \leq m)$ be i.i.d. $\mathsf{Bern}(p)$ with $0 < \underline{p} < p < \overline{p} < 1$. Take $\xi_j = (X_j - p)/\sqrt{mp(1-p)}$ and $S_m = \sum_{j=1}^m \xi_j$. Then, for any $a \in \mathbb{R}$,

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

Take $c_0 > 0$ such that

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

Then, for any $m \in \mathbb{N}$ and $\lambda = c_0\sqrt{m}$, we have $60\lambda L(2\lambda) \leq 1$. Sakhanenko (1996, Lemma 2) implies that, whenever $c_0\sqrt{m} \geq 1$ and the event $\{|S_m| < c_0\sqrt{m}\}$ occurs,

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

Moreover, Z can be taken such that $Z = \Phi^{-1} \circ F(S_m)$.

We then proceed as in the proof for Lemma 2 in Brown *et al.* (2010), where they show for each $0 < p < 1$, the coupling exits with c_0 to c_3 not depending on m, though they did not give explicit dependency of c_0 to c_3 on p. Take c_1 such that $c_1/(60c_0) < 1/2$. In particular, we can take $c_1 = 15c_0$. Then, on the event $\{|S_m| < c_1\sqrt{m}\}$,

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

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

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

Since $X = \sum_{j=1}^m X_j \sim \mathsf{Bin}(m,p)$, whenever the event $\left\{|X - mp| < c_1 m\sqrt{\underline{p}(1-\overline{p})}\right\}$ occurs and $c_0\sqrt{m} \geq 1$,

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

Moreover, $|S_m| \leq \frac{2}{c_0\sqrt{m}} + 2|Z|$ implies $|X - mp| \leq \frac{1}{c_0} + 2\sqrt{mp(1-p)}|Z|$. □

This generalization of Tusnády's Lemma enables the following strong approximation for the case of a quasi-dyadic cells expansion.

Lemma SA.5. *Suppose Assumption SA.1 holds, a quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P}_X, \rho)$ is given with $\rho > 1$, and $(Z_n^X(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H})$ is the Gaussian process constructed as in (SA-2) on a possibly enlarged probability space. Then, for any $g \in \Pi_0\mathcal{H}$ and for any $t > 0$,*

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

where c_ρ is a constant that only depends on ρ, and $\|g\|^2_{\mathcal{E}_K}$ and $\mathsf{C}_{\{g\},K}$ are defined in Lemma SA.3.

Proof of Lemma SA.5. We adopt the coupling method from Section 2 of Rio (1994), extending it to accommodate quasi-dyadic cells. Instead of applying the well-known Tusnády inequality as in Rio (1994), which states that for $X \sim \mathsf{Bin}(m, \frac{1}{2})$, there exists $Z \sim \mathsf{Normal}(0,1)$ such that almost surely:

$$\left|X - \frac{m}{2} - \left(\frac{\sqrt{m}}{2}\right)Z\right| \leq 1 + \frac{Z^2}{8}, \quad \text{and} \quad |X - \frac{m}{2}| \leq 1 + \frac{\sqrt{m}}{2}|Z|,$$

we rely on Lemma SA.4, which allows for coupling in the case of $\mathsf{Bin}(m,p)$ with $p \neq \frac{1}{2}$, though restricted to a high-probability set. The proof proceeds in two parts: Part 1 establishes an upper bound for the small-probability event where the coupling inequalities from Lemma SA.4 fail to hold; Part 2 decomposes the error $X_n(g) - Z_n^X(g)$ into the coupling errors corresponding to each pair of cells $(\mathcal{C}_{j-1,2k}, \mathcal{C}_{j-1,2k+1})$, following the strategy in Rio (1994), while accounting for the restriction to the high-probability set.

Part 1: Strong Approximation Set-up. By the construction at Equation (2), condition on $U_{j,k}$, $\widetilde{U}_{j,k}$ has the same distribution as $2\mathsf{Bin}(U_{j,k}, p_{j,k}) - U_{j,k}$, and the conditional quantile transformation relation $\widetilde{U}_{j,k} = \sqrt{U_{j,k} p_{j,k}(1-p_{j,k})} G_{(j,k),U_{j,k}} \circ \Phi(\widetilde{\xi}_{j,k})$ holds. This allows for application of Lemma SA.4. Let $\overline{p} = \rho$, $\underline{p} = \rho^{-1}$, c_0 to be the positive solution of

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

$c_1 = 15 c_0 \sqrt{\underline{p}(1-\overline{p})}$, $c_2 = 1/(15 c_0)$, and $c_3 = 1/c_0$. Consider the small probability set $\mathcal{A}$ where the coupling inequalities from Lemma SA.4 are not guaranteed to hold,

$$\mathcal{A} = \left\{ |\widetilde{U}_{j,k}| \leq c_1 U_{j,k} : (j,k) \in \mathcal{I}_K \right\},$$

and notice that we can always take $c_1 \leq 1$ because $|\widetilde{U}_{j,k}| \leq U_{j,k}$ almost surely. Using Lemma SA.4 conditional on $U_{j,k}$, whenever $\mathcal{A}$ occurs,

$$\begin{aligned} \left| \widetilde{U}_{j,k} - \sqrt{U_{j,k} p_{j,k}(1-p_{j,k})} \widetilde{\xi}_{j,k} \right| &< c_2 \widetilde{\xi}_{j,k}^2 + c_3, \\ \left| \widetilde{U}_{j,k} \right| &\leq 1/c_0 + 2\sqrt{p_{j,k}(1-p_{j,k})} |\widetilde{\xi}_{j,k}|, \end{aligned} \tag{SA-3}$$

for all $(j,k) \in \mathcal{I}_K$.

To bound $\mathbb{P}(\mathcal{A}^c)$, first notice that by Chernoff's inequality for Binomial distribution, $\mathbb{P}(U_{j,k} \leq \mathbb{E}[U_{j,k}]/2) \leq \exp(-\mathbb{E}[U_{j,k}]/8)$ for all $(j,k) \in \mathcal{I}_K$, and $\mathbb{P}(U_{j,k} \leq 2^{-1}\rho^{-1} n 2^{j-K}) \leq \exp(-8^{-1}\rho^{-1} n 2^{j-K})$ for all $(j,k) \in \mathcal{I}_K$ because $\rho^{-1} n 2^{j-K} \leq \mathbb{E}[U_{j,k}] \leq \rho n 2^{j-K}$. Furthermore, using Hoeffding's inequality and the fact that $\widetilde{U}_{j,k} = U_{j-1,2k} - p_{j,k} U_{j,k} = U_{j-1,2k} - \mathbb{E}[U_{j-1,2k} | U_{j,k}]$,

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

Putting these together, and using the union bound,

$$\begin{aligned} \mathbb{P}(\mathcal{A}^c) &\leq \sum_{(j,k) \in \mathcal{I}_K} \mathbb{P}(|\widetilde{U}_{j,k}| > c_1 U_{j,k}) \\ &\leq \sum_{(j,k) \in \mathcal{I}_K} \mathbb{P}\left(U_{j,k} \leq \frac{1}{2}\rho^{-1} n 2^{-K+j} \right) + \mathbb{P}\left(|\widetilde{U}_{j,k}| \geq c_1 U_{j,k} \middle| U_{j,k} \geq \frac{1}{2}\rho^{-1} n 2^{-K+j} \right) \\ &\leq \sum_{j=1}^{K} \sum_{0 \leq k < 2^{K-j}} \left\{ \exp(-8^{-1}\rho^{-1} n 2^{j-K}) + 2\exp(-c_1^2 \rho^{-1} n 2^{j-K}/3) \right\} \\ &\leq 3 \cdot 2^K \exp(-\min\{c_1^2/3, 1/8\} \rho^{-1} n 2^{-K}). \end{aligned} \tag{SA-4}$$

Part 2: Bounding Strong Approximation Error. We show that the proof of Theorem 2.1 in Rio (1994) still goes through for an approximate dyadic scheme. In other words, we show that the approximate dyadic scheme gives essentially the same Gaussian coupling rates as the dyadic scheme (Section SA-II.1.1). We employ the same notation as in Rio (1994), and for $g \in \Pi_0\mathcal{H}$, define

$$\Delta(g) = (X - Z)(g), \qquad X(g) = \sum_{j=1}^{K} \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\widetilde{U}_{j,k}, \qquad Z(g) = \sum_{j=1}^{K} \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\widetilde{V}_{j,k},$$

$$\Delta_1(g) = (X - Y)(g), \qquad Y(g) = \sum_{j=1}^{K} \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\sqrt{U_{j,k}\widetilde{p}_{j,k}(1-\widetilde{p}_{j,k})}\widetilde{\xi}_{j,k},$$

$$\Delta_2(g) = (Y - Z)(g)$$

It suffices to verify the following two claims.

Claim 1: $\mathbb{E}[\exp(t\Delta_1(g))\mathbb{1}(\mathcal{A})] \le \prod_{j=1}^{K}\prod_{0 \le k < 2^{K-j}} \mathbb{E}[\cosh(t\widetilde{\beta}_{j,k}(g)(2 + \widetilde{\xi}_{j,k}^2/4))]$ for all $g \in \Pi_0\mathcal{H}$. Then, it follows from the proof of Lemma 2.2 in Rio (1994) that

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

for all $|t| < 1$.

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

$$\Delta_3(g) = \sum_{j=1}^{K} \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\widetilde{\xi}_{j,k} \left(1 + \sum_{l=j}^{K} \sum_{0 \le q < 2^{K-l}} 2^{-|j-l|/2}\left|\widetilde{\xi}_{l,q}\right|\mathbb{1}(\mathcal{C}_{l,q} \supseteq \mathcal{C}_{j,k})\right),$$

for all $g \in \Pi_0\mathcal{H}$, and c_ρ a constant that only depends on ρ.

Proof of Claim 1: Let $\mathcal{F}_j = \boldsymbol{\sigma}(\{\widetilde{\xi}_{l,k} : j < l \le K, 0 \le k < 2^{K-l}\})$ for all $1 \le j < K$. In particular, $\boldsymbol{\sigma}(\{U_{l,k} : j \le l \le K, 0 \le k < 2^{K-l}\}) \subseteq \mathcal{F}_j$. Then, by Equation SA-3, for all $t \in \mathbb{R}$,

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

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

$$S_j(t) = \exp\left(t \sum_{0 \le k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\left(\widetilde{U}_{j,k} - \sqrt{U_{j,k}\widetilde{p}_{j,k}(1-\widetilde{p}_{j,k})}\widetilde{\xi}_{j,k}\right)\right),$$

so that $\mathbb{E}[\exp(t\Delta_1)\mathbb{1}(\mathcal{A})] = \mathbb{E}[\prod_{j=1}^K S_j(t)\mathbb{1}(\mathcal{A})]$, and

$$T_j(t) = \prod_{0 \le k < 2^{K-j}} \cosh\left(t\widetilde{\beta}_{j,k}(g)(c_2\widetilde{\xi}_{j,k}^2 + c_3)\right),$$

so that $\prod_{j=1}^K \prod_{0\le k<2^{K-j}} \mathbb{E}[\cosh(t\widetilde{\beta}_{j,k}(2+\widetilde{\xi}_{j,k}^2/4))] = \mathbb{E}[\prod_{j=1}^K T_j(t)]$. By Equation SA-3, for all $1 \le j \le K$,

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

It follows that

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

where in the last line, we have used independence of $(\widetilde{\xi}_{j,k} : 1 \le j \le K, 0 \le k < 2^{K-j})$. Without loss of generality, we assume that $c_\rho \sup_{\mathbf{x}\in\mathcal{C}_{K,0}} |g(\mathbf{x})| \le 1$. Since we know $(\widetilde{\xi}_{j,k}, 1 \le j \le K, 0 \le k < 2^{K-j})$ are i.i.d. standard Gaussian, the same upper bound established in Rio (1994) for the right hand side of the last display holds: for all $g \in \Pi_0\mathcal{H}$, $|t| < 1$,

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

which concludes the verification of the first claim.

<u>Proof of Claim 2</u>: Denote $q_{j,k} = \mathbb{P}_X(\mathcal{C}_{j,k})$ for $(j,k) \in \mathcal{J}_K$. By Equation (2), for any $g \in \Pi_0\mathcal{H}$, we have

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

We will use the same strategy as in Rio (1994) adapted to the quasi-dyadic case. Fix $0 \le l < 2^{K-j}$ and $0 \le j \le K$, and let k_l be the unique integer in $[0, 2^{K-l})$ such that $\mathcal{C}_{l,k_l} \supseteq \mathcal{C}_{j,k}$. Then,

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

By Equation SA-3, when the event $\mathcal{A}$ holds,

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

For the first summand,

$$\sum_{l=j}^{K-1}\sqrt{\frac{q_{j,k}}{q_{l+1,k_{l+1}}}}2\sqrt{\frac{q_{l+1,2k_l+1}}{q_{l,k_l}}}|\widetilde{\xi}_{l,k_l}| = \sum_{l=j}^{K-1}\sqrt{\prod_{j<s\leq l}p_{s,k_s}}2\sqrt{p_{l,k_l}}|\widetilde{\xi}_{l,k_l}| \leq c_\rho\sum_{l=j}^{K-1}2^{-(l-j)/2}|\widetilde{\xi}_{l,k_l}|.$$

For the second summand, we separate it into two terms as in Rio (1994). For $\mathbb{1}(\widetilde{U}_{l,k_l} \leq 0)$, we have

$$\sum_{l=j}^{K-1}\sqrt{\frac{q_{j,k}}{q_{l+1,k_{l+1}}}}\frac{\min\{c_0^{-1}, -\widetilde{U}_{l,k_l}\}}{\sqrt{\frac{q_{l+1,k_{l+1}}}{q_{l,k_l}}U_{l,k_l}} + \sqrt{U_{l+1,k_{l+1}}}}\mathbb{1}(\widetilde{U}_{l,k_l} \leq 0)$$
$$= \sum_{l=j}^{K-1}\sqrt{\frac{q_{j,k}}{q_{l+1,k_{l+1}}}}\frac{\min\{c_0^{-1}, -\widetilde{U}_{l,k_l}\}}{\sqrt{U_{l+1,k_{l+1}} - \widetilde{U}_{l,k_l}} + \sqrt{U_{l+1,k_{l+1}}}}\mathbb{1}(\widetilde{U}_{l,k_l} \leq 0) \leq c_\rho,$$

since $\sup_{0\leq x\leq u}\min\{c_0^{-1}, x\}/(\sqrt{u} + \sqrt{u+x}) \lesssim 1$. For $\mathbb{1}(\widetilde{U}_{l,k_l} > 0)$, we have

$$\sum_{l=j}^{K-1}\sqrt{\frac{q_{j,k}}{q_{l+1,k_{l+1}}}}\frac{\min\{c_0^{-1}, \widetilde{U}_{l,k_l}\}}{\sqrt{\frac{q_{l+1,k_{l+1}}}{q_{l,k_l}}U_{l,k_l}} + \sqrt{U_{l+1,k_{l+1}}}}\mathbb{1}(\widetilde{U}_{l,k_l} > 0)$$
$$\leq \sum_{l=j}^{K-1}\sqrt{\frac{q_{j,k}}{q_{l+1,k_{l+1}}}}\left(\sqrt{U_{l+1,k_{l+1}}} - \sqrt{\frac{q_{l+1,k_{l+1}}}{q_{l,k_l}}U_{l,k_l}}\right)\mathbb{1}\left(\frac{q_{l+1,k_{l+1}}}{q_{l,k_l}}U_{l,k_l} \leq U_{l+1,k_{l+1}} \leq \frac{q_{l+1,k_{l+1}}}{q_{l,k_l}}U_{l,k_l} + c_0^{-1}\right)$$
$$\leq \sum_{l=j}^{K-1}\sqrt{\frac{q_{j,k}}{q_{l+1,k_{l+1}}}}\sqrt{c_0^{-1}} = \sum_{l=j}^{K-1}\sqrt{\prod_{j<s\leq l}p_{s,k_s}}\sqrt{c_0^{-1}} \leq c_\rho.$$

It follows that when the event $\mathcal{A}$ holds,

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

Using an induction argument, for all $g \in \Pi_0\mathcal{H}$, $t > 0$,

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

For any random variable W, define $\gamma_W(t) = \log(\mathbb{E}[\exp(tc_\rho W)])$ for all $t > 0$, and $h_W(u) = \sup_{t>0}(tu -$

$\gamma_W(u)$). Combining Equation (SA-5), for any $g \in \Pi_0\mathcal{H}$, $t > 0$,

$$\mathbb{P}(\Delta_1(g) \geq t \text{ and } \mathcal{A}) \leq \inf_{u>0} \mathbb{P}(\exp(\Delta_1(g)u) \geq \exp(tu) \text{ and } \mathcal{A}) \leq \inf_{u>0} \exp(-tu)\mathbb{E}[\exp(\Delta_1(g)u)\mathbb{1}(\mathcal{A})]$$
$$\leq \exp\left(-h_{\Delta_1(g)}(t)\right) = \exp\left(-\sup_{u>0}\left(tu + \frac{83}{3}c_\rho^2\|g\|_{\mathcal{E}_K}^2 \log(1-u^2/16)\right)\right),$$

hence for any $t > 0$,

$$\mathbb{P}(|\Delta_1(g)| \geq C c_\rho \|g\|_{\mathcal{E}_K}\sqrt{t} + Ct \text{ and } \mathcal{A}) = \mathbb{P}(\Delta_1(g) \geq h^{-1}_{\Delta_1(g)}(t) \text{ and } \mathcal{A}) \leq 2\exp(-t). \tag{SA-7}$$

By Equation (SA-6), for any $t > 0$,

$$\mathbb{P}(\Delta_2(g) \geq t \text{ and } \mathcal{A}) \leq \inf_{u>0} \exp(-tu)\mathbb{E}[\exp(\Delta_2(g)u)\mathbb{1}(\mathcal{A})] \leq \exp\left(-h_{\Delta_3(g)}(t)\right). \tag{SA-8}$$

Since $\Delta_3(g)$ only depends on $((\widetilde{\xi}_{j,k}, \widetilde{\beta}_{j,k}(g)) : (j,k) \in \mathcal{I}_K)$, the rest of the proof follows from Lemma 2.4 in Rio (1994). In particular, define

$$\Delta_4(g) = \sum_{j=1}^{K} \sum_{0 \leq k < 2^{K-j}} \widetilde{\beta}_{j,k}(g)\widetilde{\xi}_{j,k}, \qquad \Delta_5(g) = \Delta_3(g) - \Delta_4(g),$$

then identifying that $\Delta_4(g)$ is Gaussian and applying Rio (1994, Lemma 2.4) with two choices of p_i-sequence, $p_i = \frac{1}{2}(\frac{1}{K} + \frac{1}{i(i+1)})$ and $p_i = \frac{1}{i(i+1)}$ separately on $\Delta_5(g)$, we get for any $t > 0$, and $g \in \Pi_0\mathcal{H}$,

$$\mathbb{P}\left(|\Delta_2(g)| \geq c_\rho\|g\|_{\mathcal{E}_K}\sqrt{t} + c_\rho\sqrt{\mathtt{C}_{\{g\},K}}t \text{ and } \mathcal{A}\right) \leq \mathbb{P}\left(|\Delta_3(g)| \geq c_\rho\|g\|_{\mathcal{E}_K}\sqrt{t} + c_\rho\sqrt{\mathtt{C}_{\{g\},K}t}\right) \leq 2\exp(-t).$$

Combining Equation (SA-4), (SA-7) and (SA-8), we get the stated result. □

SA-II.1.4 Meshing Error

For $0 < \delta \leq 1$, consider the $(\delta\mathtt{M}_{\mathcal{H},\mathcal{X}})$-net of $(\mathcal{H}, \|\cdot\|_{\mathbb{P}_X,2})$ over $\mathcal{X}$, $\mathcal{H}_\delta$, with cardinality no larger than $\mathtt{N}_{\mathcal{H},\mathcal{X}}(\delta, \mathtt{M}_{\mathcal{H},\mathcal{X}})$. Define $\pi_{\mathcal{H}_\delta} : \mathcal{H} \mapsto \mathcal{H}$ such that $\|\pi_{\mathcal{H}_\delta}(h) - h\|_{\mathbb{P}_X,2} \leq \delta\mathtt{M}_{\mathcal{H},\mathcal{X}}$ for all $h \in \mathcal{H}$. To simplify notation, in this section the parameters of $\mathcal{H}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{C}$ is omitted whenever there is no confusion.

Lemma SA.6. *Suppose Assumption SA.1 holds, a quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P}_X, \rho)$ is given, $(Z_n^X(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H})$ is the Gaussian process constructed as in (SA-2) on a possibly enlarged probability space, and $\mathcal{H}_\delta$ is chosen in Section SA-II.1.4. Then, for all $t > 0$ and $0 < \delta < 1$,*

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

Proof of Lemma SA.6. Take $\mathcal{L} = \{h - \pi_{\mathcal{H}_\delta}(h) : h \in \mathcal{H}\}$ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$. Then, $\sup_{l \in \mathcal{L}}\|l\|_{\mathbb{P}_X,2} \leq \delta\mathtt{M}_{\mathcal{H}}$ and, for all $0 < \varepsilon < \delta$,

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

Hence $J(u, \mathcal{L}, \mathtt{M}_{\mathcal{H}}) \leq 2J(u, \mathcal{H}, \mathtt{M}_{\mathcal{H}})$ for all $0 < u < \delta$. By Chernozhukov *et al.* (2014, Theorem 5.2), we have

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

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

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

By van der Vaart and Wellner (2013, Corollary 2.2.9),

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

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

$$\mathbb{P}\left(\|Z_n - Z_n \circ \pi_{\mathcal{H}_\delta}\|_{\mathcal{H}} \gtrsim J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}})\mathtt{M}_{\mathcal{H}} + \delta \mathtt{M}_{\mathcal{H}} \sqrt{t}\right) \leq \exp(-t),$$

which concludes the proof. □

SA-II.1.5 Strong Approximation Errors

To simplify notation, in this section the parameters of $\mathcal{H}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{C}$ is omitted whenever there is no confusion. The next lemma controls the strong approximation error for projected processes.

Lemma SA.7. *Suppose Assumption SA.1 holds, a dyadic expansion $\mathfrak{C}_K(\mathbb{P}_X, 1)$ is given, $(Z_n^X(h) : h \in \mathcal{H} \cup \mathcal{E}_{K,\mathtt{M}_{\mathcal{H}}})$ is the Gaussian process constructed as in (SA-2) on a possibly enlarged probability space, and $\mathcal{H}_\delta$ is chosen as in Section SA-II.1.4. For each $1 \leq j \leq K$, define the j-th level difference set*

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

Then, for all $t > 0$,

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

where

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

and $\mathtt{C}_{\mathcal{H}_\delta, K}$ is defined in Lemma SA.3. In the above display, f_X denotes the Lebesgue density of $\mathbb{P}_X$: if it does not exist, the term $\sqrt{d} \sup_{\mathbf{x} \in \mathcal{X}} f_X^2(\mathbf{x}) 2^{2(K-j)} \|\mathcal{U}_j\|_\infty \mathfrak{m}(\mathcal{U}_j) \mathtt{TV}^*_{\mathcal{H}_\delta}$ is taken to be infinity.

Proof of Lemma SA.7. We employ the same strategy as in the proof of Theorem 1.1 from Rio (1994), noting that incorporating the Lipschitz condition can lead to a tighter bound for strong approximation error.

A very first bound that we can obtain for is

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

If we further assume $\mathbb{P}_X$ admits a Lebesgue density f_X, then an analysis based on total variation of h can be done as follows. For each $1\le j\le K$, there exists unique integers $j_1,\dots,j_d$ such that $0\le j_1\le\dots\le j_d\le j_1+1$ and $\sum_{i=1}^d j_i = j$. In particular, there exists a unique $l=l(j)\in\{1,2,\dots,d\}$ such that either $l\le d-1$ and $j_l<j_{l+1}$ or $l=d$ and $j_d<j_1+1$.

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

Since we have assumed f is bounded from above on $\mathcal{X}$ and hence on $\mathcal{C}_{K,0}$, and $\mathcal{C}_{j-1,2k+1}-\{\mathbf{x}\}\subseteq\mathcal{C}_{j-1,2k+1}-\mathcal{C}_{j-1,2k}$,

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

and therefore

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

where $\mathcal{U}_j = \cup_{0\le k<2^{K-j}}(\mathcal{C}_{j-1,2k+1}-\mathcal{C}_{j-1,2k})$. Let $(h_\ell)_{\ell\in\mathbb{N}}$ be any sequence of real-valued functions on $(\mathcal{X},\mathcal{B}(\mathcal{X}))$ such that $h_\ell\to h$ $\mathfrak{m}$-almost surely, and are bounded by $2\mathtt{M}_{\mathcal{H}}$ on $\mathcal{X}$. Since we assumed $\mathtt{M}_{\mathcal{H}}<\infty$, and h_ℓ and h are bounded by $2\mathtt{M}_{\mathcal{H}}$ with $\int_{\mathbb{R}^d} 2\mathtt{M}_{\mathcal{H}} f_X(\mathbf{x})d\mathbf{x}\le 2\mathtt{M}_{\mathcal{H}}<\infty$, Dominated Convergence Theorem implies for any $\mathbf{x}\in\mathcal{U}_j$,

$$\begin{aligned}
\int_{\sqcup_{0\le k<2^{K-j}}\mathcal{C}_{j-1,2k}} |h(\mathbf{x})-h(\mathbf{x}+\mathbf{s})|f_X(\mathbf{x})d\mathbf{x} &= \lim_{\ell\to\infty}\int_{\sqcup_{0\le k<2^{K-j}}\mathcal{C}_{j-1,2k}} |h_\ell(\mathbf{x})-h_\ell(\mathbf{x}+\mathbf{s})|f_X(\mathbf{x})d\mathbf{x}\\
&= \lim_{\ell\to\infty}\int_{\sqcup_{0\le k<2^{K-j}}\mathcal{C}_{j-1,2k}}\int_0^{\|\mathbf{s}\|} \|\nabla h_\ell(\mathbf{x}+t\mathbf{s}/\|\mathbf{s}\|)\|f_X(\mathbf{x})dtd\mathbf{x}\\
&= \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\|\mathbf{s}\|\limsup_{\ell\to\infty}\mathtt{TV}^*_{\{h_\ell\}}.
\end{aligned}$$

Since the above inequality holds for all sequences $(h_\ell)_{\ell\in\mathbb{N}}$ such that $h_\ell\to h$ $\mathfrak{m}$-almost surely, and are bounded by $2\mathtt{M}_{\mathcal{H}}$ on $\mathcal{X}$, Definition SA.1 implies

$$\int_{\sqcup_{0\leq k<2^{K-j}}\mathcal{C}_{j-1,2k}} |h(\mathbf{x})-h(\mathbf{x}+\mathbf{s})| f_X(\mathbf{x}) d\mathbf{x} \leq \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x}) \|\mathbf{s}\| \mathtt{TV}^*_{\{h\}}$$
$$\leq \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x}) \sqrt{d} \|\mathcal{U}_j\|_\infty \mathtt{TV}^*_{\{h\}}.$$

It follows that

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

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

$$\sup_{h\in\mathcal{H}_\delta} \|h\|^2_{\mathcal{E}_K} = \sup_{h\in\mathcal{H}_\delta} \sum_{j=1}^K \sum_{0\leq k<2^{K-j}} |\widetilde{\beta}_{j,k}(h)|^2 \leq \sup_{h\in\mathcal{H}_\delta} \sum_{j=1}^K \min\{\mathtt{M}_{\mathcal{H}_\delta}, \|\mathcal{U}_j\|_\infty \mathtt{L}_{\mathcal{H}_\delta}\} \sum_{0\leq k<2^{K-j}} |\widetilde{\beta}_{j,k}(h)| \leq \mathcal{R}_K(\mathcal{H}_\delta),$$

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

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

Applying Lemma SA.3, for any $h\in\mathcal{H}_\delta$, for any $t>0$, with probability at least $1-2\exp(-t)$,

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

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

Lemma SA.8. *Suppose Assumption SA.1 holds, a quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P}_X,\rho)$ is given with $\rho>1$, $(Z_n^X(h): h\in\mathcal{H}\cup\Pi_0\mathcal{H})$ is the Gaussian process constructed at Equation (SA-2) on a possibly enlarged probability space, and $\mathcal{H}_\delta$ is chosen in Section SA-II.1.4. Then, for all $t>0$,*

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

where C_ρ is a constant only depending on ρ, $\mathcal{R}_K(\mathcal{H}_\delta)$ is defined in Lemma SA.7, and $\mathtt{C}_{\mathcal{H}_\delta,K}$ is defined in Lemma SA.3.

Proof of Lemma SA.8. This follows from Lemma SA.5 and the fact that

$$\sup_{h\in\mathcal{H}_\delta} \|h\|^2_{\mathcal{E}_K} \leq \mathcal{R}_K(\mathcal{H}_\delta), \qquad h\in\mathcal{H}_\delta,$$

from the proof of Lemma SA.7. □

SA-II.1.6 Projection Error

To simplify notation, in this section the parameters of $\mathcal{H}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{C}$ is omitted whenever there is no confusion. The following lemma controls the mean square projection onto piecewise constant functions.

Lemma SA.9. *Suppose Assumption SA.1 holds, a dyadic expansion $\mathfrak{C}_K(\mathbb{P}_X, 1)$ is given, $(Z_n^X(h) : h \in \mathcal{H} \cup \Pi_0\mathcal{H})$ is the Gaussian process constructed as in (SA-2) on a possibly enlarged probability space, and $\mathcal{H}_\delta$ is chosen in Section SA-II.1.4. In addition, assume $\mathbb{P}_X$ admits a Lebesgue density f_X supported on $\mathcal{X} \subseteq \mathbb{R}^d$. Define quasi-dyadic variation set $\mathcal{V} = \cup_{0 \le k < 2^K}(\mathcal{C}_{0,k} - \mathcal{C}_{0,k})$. Then, for all $t > 0$,*

$$\mathbb{P}\Big[\|X_n - X_n \circ \Pi_0\|_{\mathcal{H}_\delta} > \sqrt{4\mathtt{V}_{\mathcal{H}_\delta} t} + \frac{4\mathtt{B}_{\mathcal{H}_\delta}}{3\sqrt{n}} t\Big] \le 2\mathtt{N}_{\mathcal{H}_\delta}(\delta, \mathtt{M}_{\mathcal{H}_\delta}) e^{-t},$$
$$\mathbb{P}\Big[\|Z_n^X - Z_n^X \circ \Pi_0\|_{\mathcal{H}_\delta} > \sqrt{4\mathtt{V}_{\mathcal{H}_\delta} t}\Big] \le 2\mathtt{N}_{\mathcal{H}_\delta}(\delta, \mathtt{M}_{\mathcal{H}_\delta}) e^{-t},$$

where

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

In particular, if $\mathbb{P}_X = \mathsf{Uniform}([0,1]^d)$ and $\mathfrak{C}_K(\mathbb{P}_X, 1) = \mathcal{A}_K(\mathbb{P}_X, 1)$, then for all $t > 0$,

$$\mathbb{P}\Big[\|X_n - X_n \circ \Pi_0\|_{\mathcal{H}_\delta} > \sqrt{4d \min\{2\mathtt{M}_{\mathcal{H}_\delta}, \mathtt{L}_{\mathcal{H}_\delta} 2^{-K}\} 2^{-K} \mathtt{TV}^*_{\mathcal{H}_\delta} t} + \frac{4\min\{2\mathtt{M}_{\mathcal{H}_\delta}, \mathtt{L}_{\mathcal{H}_\delta} 2^{-K}\}}{3\sqrt{n}} t\Big] \le 2\mathtt{N}_{\mathcal{H}_\delta}(\delta, \mathtt{M}_{\mathcal{H}_\delta}) e^{-t},$$
$$\mathbb{P}\Big[\|Z_n^X - Z_n^X \circ \Pi_0\|_{\mathcal{H}_\delta} > \sqrt{4d \min\{2\mathtt{M}_{\mathcal{H}_\delta}, \mathtt{L}_{\mathcal{H}_\delta} 2^{-K}\} 2^{-K} \mathtt{TV}^*_{\mathcal{H}_\delta} t}\Big] \le 2\mathtt{N}_{\mathcal{H}_\delta}(\delta, \mathtt{M}_{\mathcal{H}_\delta}) e^{-t}.$$

Proof of Lemma SA.9. Let $h \in \mathcal{H}$. Then, $|h(\mathbf{x}_i) - \Pi_0 h(\mathbf{x}_i)| \le \min\{2\mathtt{M}_{\mathcal{H}_\delta}, \mathtt{L}_{\mathcal{H}_\delta}\|\mathcal{V}\|_\infty\} = \mathtt{B}_{\mathcal{H}_\delta}$,

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

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

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

Let $(h_\ell)_{\ell \in \mathbb{N}}$ be any sequence of real-valued functions on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ such that $h_\ell \to h$ $\mathfrak{m}$-almost surely, and are bounded by $2\mathtt{M}_\mathcal{H}$ on $\mathcal{X}$. Since we assumed $\mathtt{M}_\mathcal{H} < \infty$, and h_ℓ and h are bounded by $2\mathtt{M}_\mathcal{H}$, by Dominated

Convergence Theorem we have that

$$\begin{aligned}\int_{\mathcal{C}_{K,0}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| \, f_X(\mathbf{x}) d\mathbf{x} &= \lim_{\ell\to\infty} \int_{\mathcal{C}_{K,0}} |h_\ell(\mathbf{x}) - h_\ell(\mathbf{x}+\mathbf{s})| f_X(\mathbf{x}) d\mathbf{x} \\ &\leq \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x}) \cdot \lim_{\ell\to\infty} \int_{\mathcal{X}} \int_0^{\|\mathbf{s}\|} \|\nabla h_\ell(\mathbf{x} + t\mathbf{s}/\|\mathbf{s}\|)\| dt d\mathbf{x} \\ &\leq \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x}) \cdot \int_0^{\|\mathbf{s}\|} \lim_{\ell\to\infty} \int_{\mathcal{X}} \|\nabla h_\ell(\mathbf{x} + t\mathbf{s}/\|\mathbf{s}\|)\| d\mathbf{x} dt \\ &\leq \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x}) \cdot \|\mathbf{s}\| \limsup_{\ell\to\infty} \mathtt{TV}_{\{h_\ell\}}.\end{aligned}$$

Since this holds for any sequence $(h_\ell)_{\ell\in\mathbb{N}}$ $h_\ell \to h$ $\mathfrak{m}$-almost surely, and are bounded by $2\mathtt{M}_{\mathcal{H}}$ on $\mathcal{X}$, hence $\int_{\mathcal{X}} |h(\mathbf{x}) - h(\mathbf{x}+\mathbf{s})| \, d\mathbf{x} \leq \|\mathbf{s}\| \mathtt{TV}^*_{\{h\}}$. It follows that

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

and

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

for all $h \in \mathcal{H}_\delta$. Then, by Bernstein inequality, for any $t > 0$,

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

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 $t \leq 2\sqrt{\mathtt{V}_{\mathcal{H}_\delta}}\sqrt{u} + \frac{4}{3}\frac{\mathtt{B}_{\mathcal{H}_\delta}}{\sqrt{n}} u$. 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 $\mathbb{V}[X_n(h) - X_n(\Pi_0 h)]$ and a union bound argument. □

SA-II.2 Surrogate Measure and Normalizing Transformation

This section studies the properties of the surrogate measure $\mathbb{Q}_{\mathcal{H}}$ and normalizing transformation $\phi_{\mathcal{H}}$ introduced in condition (ii) of Theorem 1. The following lemma characterizes the connections between the original and the transformed parameters of $\mathcal{H}$ (Definitions 4 to 12) when deploying $\mathbb{Q}_{\mathcal{H}}$ and $\phi_{\mathcal{H}}$.

Lemma SA.10. *Suppose following conditions hold.*

(i) *$\mathcal{H}$ is a real-valued pointwise measurable class of functions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.*

(ii) *There exists a surrogate measure $\mathbb{Q}_{\mathcal{H}}$ for $\mathbb{P}_X$ with respect to $\mathcal{H}$ such that $\mathbb{Q}_{\mathcal{H}} = \mathfrak{m} \circ \phi_{\mathcal{H}}$, where the normalizing transformation $\phi_{\mathcal{H}} : \mathcal{Q}_{\mathcal{H}} \mapsto [0,1]^d$ is a diffeomorphism.*

Let $\widetilde{\mathcal{H}} = \{h \circ \phi_{\mathcal{H}}^{-1} : h \in \mathcal{H}\}$. *Then,*

$$\begin{aligned}
\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d} &= \mathtt{M}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}, \qquad \mathtt{E}_{\widetilde{\mathcal{H}},[0,1]^d} = \mathtt{E}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}, \\
\mathtt{N}_{\widetilde{\mathcal{H}},[0,1]^d}(\varepsilon, \mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}) &= \mathtt{N}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}(\varepsilon, \mathtt{M}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}), \qquad \varepsilon \in (0,1), \\
\mathtt{L}_{\widetilde{\mathcal{H}},[0,1]^d} &\leq \mathtt{c}_2 \mathtt{L}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}, \qquad \mathtt{c}_2 = \sup_{\mathbf{x} \in \mathcal{Q}_{\mathcal{H}}} \frac{1}{\sigma_d(\nabla \phi_{\mathcal{H}}(\mathbf{x}))}, \\
\mathtt{TV}^*_{\widetilde{\mathcal{H}},[0,1]^d} &\leq d^{-1} \mathtt{c}_1 \mathtt{TV}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}, \qquad \mathtt{c}_1 = d \sup_{\mathbf{x} \in \mathcal{Q}_{\mathcal{H}}} \prod_{j=1}^{d-1} \sigma_j(\nabla \phi_{\mathcal{H}}(\mathbf{x})), \\
\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d} &\leq d^{-1/2} \mathtt{c}_3 \mathtt{K}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}, \qquad \mathtt{c}_3 = 2^{d-1} d^{d/2-1} \mathtt{c}_1 \mathtt{c}_2^{d-1}.
\end{aligned}$$

Proof of Lemma SA.10. The first three identities are self-evident. Consider next the relation between $\mathtt{L}_{\widetilde{\mathcal{H}},[0,1]^d}$ and $\mathtt{L}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}$: for any $h \in \mathcal{H}$, using a change of variables and the differentiability of $\phi_{\mathcal{H}}$,

$$\begin{aligned}
\mathtt{L}_{\{h \circ \phi_{\mathcal{H}}^{-1}\},[0,1]^d} &= \sup_{\mathbf{u},\mathbf{u}' \in [0,1]^d} \frac{|h \circ \phi_{\mathcal{H}}^{-1}(\mathbf{u}) - h \circ \phi_{\mathcal{H}}^{-1}(\mathbf{u}')|}{\|\mathbf{u} - \mathbf{u}'\|} \\
&\leq \sup_{\mathbf{x},\mathbf{x}' \in \mathcal{Q}_{\mathcal{H}}} \frac{|h(\mathbf{x}) - h(\mathbf{x}')|}{\|\mathbf{x} - \mathbf{x}'\|} \frac{\|\mathbf{x} - \mathbf{x}'\|}{\|\phi_{\mathcal{H}}(\mathbf{x}) - \phi_{\mathcal{H}}(\mathbf{x}')\|} \\
&\leq \mathtt{L}_{\{h\},\mathcal{Q}_{\mathcal{H}}} \sup_{\mathbf{u},\mathbf{u}' \in [0,1]^d} \frac{|\phi_{\mathcal{H}}^{-1}(\mathbf{u}) - \phi_{\mathcal{H}}^{-1}(\mathbf{u}')|}{\|\mathbf{u} - \mathbf{u}'\|} \\
&\leq \mathtt{L}_{\{h\},\mathcal{Q}_{\mathcal{H}}} \sup_{\mathbf{z} \in [0,1]^d} \sigma_1(\nabla \phi_{\mathcal{H}}^{-1}(\mathbf{z})) \\
&= \mathtt{L}_{\{h\},\mathcal{Q}_{\mathcal{H}}} \sup_{\mathbf{x} \in \mathcal{Q}_{\mathcal{H}}} \sigma_d(\nabla \phi_{\mathcal{H}}(\mathbf{x}))^{-1},
\end{aligned}$$

and the result follows.

Now consider the relation between $\mathtt{TV}_{\widetilde{\mathcal{H}},[0,1]^d}$ and $\mathtt{TV}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}$. First suppose all functions in $\mathcal{H}$ are differentiable, an integration by parts based on the definition of uniform total variation (Definition 5) and a change of variables calculation gives

$$\begin{aligned}
\mathtt{TV}_{\{h \circ \phi_{\mathcal{H}}^{-1}\},[0,1]^d} &= \sup_{\varphi \in \mathcal{D}_d([0,1]^d)} \int_{[0,1]^d} h \circ \phi_{\mathcal{H}}^{-1}(\mathbf{x}) \operatorname{div}(\varphi)(\mathbf{x}) d\mathbf{x} / \| \|\varphi\|_2 \|_\infty \\
&= \int_{\mathbf{u} \in [0,1]^d} \|\nabla (h \circ \phi_{\mathcal{H}}^{-1})(\mathbf{u})\| d\mathbf{u} \\
&= \int_{\mathbf{u} \in [0,1]^d} \|\nabla \phi_{\mathcal{H}}^{-1}(\mathbf{u})^\top \nabla h(\phi_{\mathcal{H}}^{-1}(\mathbf{u}))\| d\mathbf{u} \\
&= \int_{\mathcal{Q}_{\mathcal{H}}} \|\nabla \phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))^\top \nabla h(\mathbf{x})\| \cdot |\det(\nabla \phi_{\mathcal{H}}(\mathbf{x}))| d\mathbf{x} \\
&\leq \int_{\mathcal{Q}_{\mathcal{H}}} \|\nabla h(\mathbf{x})\| d\mathbf{x} \sup_{\mathbf{x} \in \mathcal{Q}_{\mathcal{H}}} |\det(\nabla \phi_{\mathcal{H}}(\mathbf{x}))| \cdot \|\nabla \phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\| \\
&\leq \mathtt{c}_1 \mathtt{TV}_{\{h\},\mathcal{Q}_{\mathcal{H}}},
\end{aligned}$$

where in the last line we have used $|\det(\nabla \phi_{\mathcal{H}}(\mathbf{x}))| = \prod_{j=1}^d \sigma_j(\nabla \phi_{\mathcal{H}}(\mathbf{x}))$, and since $\phi_{\mathcal{H}}$ is a diffeomorphism, $\|\nabla \phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\| = \sigma_1(\nabla \phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))) = \sigma_d(\nabla \phi_{\mathcal{H}}(\mathbf{x}))^{-1}$. Now consider $\mathcal{H}$ which contains possibly non-differentiable functions. Take $\psi : \mathbb{R}^d \to \mathbb{R}$ to be any smooth function with compact support such that $\int_{\mathbb{R}^d} \psi(\mathbf{z}) d\mathbf{z} = 1$, and take $\psi_\varepsilon(\cdot) = \varepsilon^{-d} \psi(\cdot/\varepsilon)$. For each $\ell \in \mathbb{N}$, define $h_\ell = h * \psi_{\varepsilon_\ell}$, where $(\varepsilon_\ell)_{\ell \in \mathbb{N}}$ is a sequence of non-increasing real positive numbers converging to zero with ε_1 small enough. Then

$$\mathtt{TV}^*_{\widetilde{\mathcal{H}},[0,1]^d} \leq \sup_{h\in\mathcal{H}} \limsup_{\ell\to\infty} \mathtt{TV}_{\{h_\ell\circ\phi_{\mathcal{H}}^{-1}\},[0,1]^d} = \sup_{h\in\mathcal{H}} \limsup_{\ell\to\infty} \mathsf{c}_1 \mathtt{TV}_{\{h_\ell\},\mathcal{Q}_{\mathcal{H}}} \leq \mathsf{c}_1 \mathtt{TV}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}},$$

where the first inequality is due to $(h_\ell)_{\ell\in\mathbb{N}}$ being a particular sequence satisfying Definition SA.1, the second inequality from Lemma SA.10, the third inequality due to $\mathtt{TV}_{\{h*\psi\},\mathcal{Q}_{\mathcal{H}}} \leq \mathtt{TV}_{\{h\},\mathcal{Q}_{\mathcal{H}}}$ for any smooth ψ.

Moreover, let $\mathcal{C} \subseteq \mathbb{R}^d$ be a cube with edges of length a parallel to the coordinate axises. Then, $\phi_{\mathcal{H}}^{-1}(\mathcal{C})$ is contained in another cube $\mathcal{C}'$ with edges of length at most $2\sqrt{d}\sup_{\mathbf{x}\in[0,1]^d}\|\nabla\phi_{\mathcal{H}}^{-1}(\mathbf{x})\|\mathsf{a}$. Again, we first assume that each $h \in \mathcal{H}$ is differentiable. Using a change of variables for the total variation calculation and the definition of $\mathtt{K}_{\{h\},\mathcal{Q}_{\mathcal{H}}}$ (Definition 5), for any $h \in \mathcal{H}$,

$$\begin{aligned}
\mathtt{TV}_{\{h\circ\phi_{\mathcal{H}}^{-1}\},\mathcal{C}} &= \int_{\mathcal{C}} \|\nabla(h\circ\phi_{\mathcal{H}}^{-1})(\mathbf{u})\| d\mathbf{u} \\
&\leq \int_{\mathcal{C}'} \|\nabla(h\circ\phi_{\mathcal{H}}^{-1})(\phi_{\mathcal{H}}(\mathbf{x}))\| \det(\nabla\phi_{\mathcal{H}}(\mathbf{x})) d\mathbf{x} \\
&\leq \int_{\mathcal{C}'} \|\nabla h(\mathbf{x})\| d\mathbf{x} \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} |\det(\nabla\phi_{\mathcal{H}}(\mathbf{x}))| \|\nabla\phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\| \\
&\leq \mathtt{K}_{\{h\},\mathcal{Q}_{\mathcal{H}}} \Big(2\sqrt{d} \sup_{\mathbf{x}\in[0,1]^d} \|\nabla\phi_{\mathcal{H}}^{-1}(\mathbf{x})\|\mathsf{a}\Big)^{d-1} \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} |\det(\nabla\phi_{\mathcal{H}}(\mathbf{x}))| \|\nabla\phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\| \\
&= d^{-1}(2\sqrt{d})^{d-1}\mathsf{c}_1\mathsf{c}_2^{d-1}\mathtt{K}_{\{h\},\mathcal{Q}_{\mathcal{H}}}\mathsf{a}^{d-1} \\
&= d^{-1/2}\mathsf{c}_3\mathtt{K}_{\{h\},\mathcal{Q}_{\mathcal{H}}}\mathsf{a}^{d-1},
\end{aligned}$$

which implies

$$\mathtt{K}_{\{\widetilde{h}\},[0,1]^d} \leq d^{-1/2}\mathsf{c}_3\mathtt{K}_{\{h\},\mathcal{Q}_{\mathcal{H}}}.$$

By similar smoothing arguments as for the $\mathtt{TV}$ terms, we can also show that $\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d} \leq d^{-1/2}\mathsf{c}_3\mathtt{K}_{\mathcal{H},[0,1]^d}$ even when $\mathcal{H}$ contains possibly non-differentiable functions. □

Lemma SA.11. *We recap the statements in Section 3.1 and present their proofs.*

- ***Case 1: Uniform on Rectangle.*** *Suppose that $\mathbf{x}_i \sim \mathsf{Uniform}(\mathcal{X})$ with $\mathcal{X} = \times_{l=1}^d[\mathsf{a}_l, \mathsf{b}_l]$, where $-\infty < \mathsf{a}_l < \mathsf{b}_l < \infty$, $l = 1, 2, \ldots, d$. Setting $\mathbb{Q}_{\mathcal{H}} = \mathbb{P}_X$, a valid normalizing transformation is $\phi_{\mathcal{H}}(x_1, \cdots, x_d) = ((\mathsf{b}_1 - \mathsf{a}_1)^{-1}(x_1 - \mathsf{a}_1), \cdots, (\mathsf{b}_d - \mathsf{a}_d)^{-1}(x_d - \mathsf{a}_d))$, which verifies assumption (ii) in Theorem 1. In this case, $\mathsf{c}_1 = d\max_{1\leq l\leq d}|\mathsf{b}_l - \mathsf{a}_l|\prod_{l=1}^d|\mathsf{b}_l - \mathsf{a}_l|^{-1}$, $\mathsf{c}_2 = \max_{1\leq l\leq d}|\mathsf{b}_l - \mathsf{a}_l|$ and $\mathsf{c}_3 = 2^{d-1}d^{d/2}\max_{1\leq l\leq d}|\mathsf{b}_l - \mathsf{a}_l|^d\prod_{l=1}^d|\mathsf{b}_l - \mathsf{a}_l|^{-1}$.*

- ***Case 2: Rectangular $\mathcal{Q}_{\mathcal{H}}$.*** *Suppose that $\mathbb{Q}_{\mathcal{H}}$ admits a Lebesgue density f_Q supported on $\mathcal{Q}_{\mathcal{H}} = \times_{l=1}^d[\mathsf{a}_l, \mathsf{b}_l]$, $-\infty \leq \mathsf{a}_l < \mathsf{b}_l \leq \infty$. Then, the Rosenblatt transformation $\phi_{\mathcal{H}} = T_{\mathbb{Q}_{\mathcal{H}}}$ is a normalizing transformation, and we obtain*

$$\begin{aligned}
\mathsf{c}_1 &= d\sup_{\mathbf{u}\in\mathcal{Q}_{\mathcal{H}}} \frac{f_Q(\mathbf{u})}{\min\{f_{Q,1}(u_1), f_{Q,2|1}(u_2|u_1), \cdots, f_{Q,d|-d}(u_d|u_1, \cdots, u_{d-1})\}}, \\
\mathsf{c}_2 &= \sup_{\mathbf{u}\in\mathcal{Q}_{\mathcal{H}}} \frac{1}{\min\{f_{Q,1}(u_1), f_{Q,2|1}(u_2|u_1), \cdots, f_{Q,d|-d}(u_d|u_1, \cdots, u_{d-1})\}},
\end{aligned}$$

and $\mathsf{c}_3 = 2^{d-1} d^{d/2-1} \mathsf{c}_1 \mathsf{c}_2^{d-1}$.

This case covers several examples of interest, which give primitive conditions for assumption (ii) in Theorem 1:

(a) Suppose $\mathcal{Q}_{\mathcal{H}} = \times_{l=1}^{d} [\mathsf{a}_l, \mathsf{b}_l]$ *is bounded. Then,*

$$\mathsf{c}_1 \leq d \frac{\overline{f}_Q^2}{\underline{f}_Q} \overline{\mathcal{Q}}_{\mathcal{H}} \qquad \textit{and} \qquad \mathsf{c}_2 \leq \frac{\overline{f}_Q}{\underline{f}_Q} \overline{\mathcal{Q}}_{\mathcal{H}}.$$

(b) Suppose $\mathcal{Q}_{\mathcal{H}} = \times_{l=1}^{d} [\mathsf{a}_l, \mathsf{b}_l]$ *is unbounded. To fix ideas, let* $\mathbf{x}_i \sim \mathsf{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. *Then, setting* $\mathbb{Q}_{\mathcal{H}} = \mathbb{P}_X$ *and* $\phi_{\mathcal{H}} = T_{\mathbb{P}_X}$ *also gives a valid normalizing transformation, with*

$$\begin{aligned} \mathsf{c}_1 &\leq d \sup_{\mathbf{x} \in \mathcal{Q}_{\mathcal{H}}} \max\{f_{X,1}(x_1), f_{X,2|1}(x_2|x_1), \cdots, f_{X,d|-d}(x_d|x_{-d})\}^{d-1} \\ &\leq d \min_{1 \leq k \leq d} \{\boldsymbol{\Sigma}_{k,k} - \boldsymbol{\Sigma}_{k,1:k-1} \boldsymbol{\Sigma}_{1:k-1,1:k-1}^{-1} \boldsymbol{\Sigma}_{1:k-1,k}\}^{-(d-1)/2} \end{aligned}$$

bounded, but c_2 *(and hence* c_3*) unbounded.*

- ***Case 3: Non-Rectangular*** $\mathcal{Q}_{\mathcal{H}}$. *Suppose that* $\mathbb{Q}_{\mathcal{H}}$ *admits a Lebesgue density* f_Q *supported on* $\mathcal{Q}_{\mathcal{H}}$, *and there exists a diffeomorphism* $\chi : \mathcal{Q}_{\mathcal{H}} \mapsto [0,1]^d$. *Setting* $\phi_{\mathcal{H}} = T_{\mathbb{Q}_{\mathcal{H}} \circ \chi^{-1}} \circ \chi$ *gives a valid normalizing transformation, with*

$$\mathsf{c}_1 \leq d \frac{\overline{f}_Q^2}{\underline{f}_Q} \mathsf{S}_\chi \qquad \textit{and} \qquad \mathsf{c}_2 \leq \frac{\overline{f}_Q}{\underline{f}_Q} \mathsf{S}_\chi,$$

where $\mathsf{S}_\chi = \frac{\sup_{\mathbf{x} \in [0,1]^d} |\det(\nabla \chi^{-1}(\mathbf{x}))|}{\inf_{\mathbf{x} \in [0,1]^d} |\det(\nabla \chi^{-1}(\mathbf{x}))|} \| \|\nabla \chi^{-1}\|_2 \|_\infty$.

Proof of Lemma SA.11. We consider the three cases separately.

Case 1: Uniform on Rectangle. For every $\mathbf{x} \in \mathcal{Q}_{\mathcal{H}}$, the singular values of $\nabla \phi_{\mathcal{H}}(\mathbf{x})$ are $(\mathsf{b}_1 - \mathsf{a}_1)^{-1}, \cdots, (\mathsf{b}_d - \mathsf{a}_d)^{-1}$. The values of c_1 and c_2 (and hence c_3) then follow.

Case 2: Rectangular $\mathcal{Q}_{\mathcal{H}}$. We start with a proof for a general result for $\mathsf{c}_1, \mathsf{c}_2, \mathsf{c}_3$, and then prove upper bounds for (a) and (b).

1. The General Case. Since $\mathbb{Q}$ has a Lebesgue density f_Q,

$$\nabla T_{\mathbb{Q}}(\mathbf{x}) = \begin{bmatrix} f_{Q,1}(x_1) & 0 & \cdots & 0 \\ * & f_{Q,2|1}(x_2|x_1) & \cdots & 0 \\ * & * & \vdots & 0 \\ * & * & \cdots & f_{Q,d|1,\cdots,d-1}(x_d|x_1, \cdots, x_{d-1}) \end{bmatrix}, \qquad \mathbf{x} \in \mathcal{Q}_{\mathcal{H}}.$$

Because the singular values of $\nabla \phi_{\mathcal{H}}(\mathbf{x}) = \nabla T_{\mathbb{Q}}(\mathbf{x})$ are the values on the diagonal,

$$\mathsf{c}_1 = d \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} \frac{f_Q(\mathbf{x})}{\min\{f_{Q,1}(x_1), f_{Q,2|1}(x_2|x_1), \cdots, f_{Q,d|-d}(x_d|x_{-d})\}},$$
$$\mathsf{c}_2 = \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} \max\{f_{Q,1}(x_1)^{-1}, f_{Q,2|1}(x_2|x_1)^{-1}, \cdots, f_{Q,d|-d}(x_d|x_{-d})^{-1}\}.$$

2. Case (a): $\mathcal{Q}_{\mathcal{H}} = \times_{l=1}^{d}[\mathsf{a}_l, \mathsf{b}_l]$ is bounded. Since we assumed the existence of an $\mathcal{Q}_{\mathcal{H}}$ that is compact and $\inf_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} f_Q(\mathbf{x}) > 0$, integrating on the rectangle gives

$$\frac{\underline{f}_Q}{\overline{f}_Q}\frac{1}{\overline{L}} \le f_{Q,k|1,\cdots,k-1}(x_k|x_1,\cdots,x_{k-1}) = \frac{\int_{\prod_{l=k+1}^{d}[a_l,b_l]} f_Q(x_1,\cdots,x_k,\mathbf{z})d\mathbf{z}}{\int_{\prod_{l=k}^{d}[a_l,b_l]} f_Q(x_1,\cdots,x_{k-1},\mathbf{u})d\mathbf{u}} \le \frac{\overline{f}_Q}{\underline{f}_Q}\frac{1}{\underline{L}},$$

where $\overline{L} = \max_{1\le l\le d}(b_l - a_l)$ and $\underline{L} = \min_{1\le l\le d}(b_l - a_l)$. Plugging in the generic bounds for c_1 and c_2,

$$\mathsf{c}_1 \le d\left(\frac{\overline{f}_Q}{\underline{f}_Q}\max_{1\le k\le d}\frac{1}{b_k - a_k}\right)^{d-1} \qquad \text{and} \qquad \mathsf{c}_2 \le \frac{\overline{f}_Q}{\underline{f}_Q}\max_{1\le k\le d}|b_k - a_k|.$$

3. Case (b): $\mathbf{x}_i \sim \mathsf{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The bound on c_1 follows from properties of the conditional distribution of multivariate Gaussian distribution. Since $\inf_{\mathbf{x}\in\mathbb{R}^k} f_{k|1,\cdots,k-1}(x_k|x_1,\cdots,x_{k-1}) = 0$ for $1 \le k \le d$, c_2 (and hence c_3) are unbounded.

Case 3: Non-rectangular $\mathcal{Q}_{\mathcal{H}}$. Since both $T_{\mathbb{Q}_{\mathcal{H}}}$ and χ are diffeomorphisms, we can use chain rule to get,

$$\begin{aligned}
\mathsf{c}_1 &= \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} \prod_{j=1}^{d-1} \sigma_j(\nabla\phi_{\mathcal{H}}(\mathbf{x})) \\
&= \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} \det(\nabla\phi_{\mathcal{H}}(\mathbf{x}))\|\nabla\phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\| \\
&\le \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} \det(\nabla T_{\mathbb{Q}_{\mathcal{H}}}(\chi(\mathbf{x})))\det(\nabla\chi(\mathbf{x}))\|\nabla\chi^{-1}(\chi(\mathbf{x}))\|_2\|\nabla T_{\mathbb{Q}_{\mathcal{H}}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\|_2.
\end{aligned}$$

Take $\mathbf{w}_i = \chi(\mathbf{x}_i)$, and denote by f_W the density of $\mathbf{w}_i$. Then

$$\nabla T_{\mathbb{Q}_{\mathcal{H}}}(x_1,\cdots,x_d) = \begin{bmatrix} f_{W_1}(x_1) & 0 & \cdots & 0 \\ * & f_{W_2|W_1}(x_2|x_1) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ * & * & \cdots & f_{W_d|W_1,\cdots,W_{d-1}}(x_d|x_1,\cdots,x_{d-1}) \end{bmatrix},$$

where $*$ denotes values that won't affect determinant or operator norm of the matrix $\nabla T_{\mathbb{Q}_{\mathcal{H}}}$. Hence,

$$\det(\nabla T_{\mathbb{Q}_{\mathcal{H}}}(\chi(\mathbf{x}))) = f_W(\chi(\mathbf{x})) = f_X(\mathbf{x})|\det(\nabla\chi(\mathbf{x}))|^{-1}$$

and

$$\|\nabla T_{\mathbb{Q}_{\mathcal{H}}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\|_2 = \sigma_d(\nabla T_{\mathbb{Q}_{\mathcal{H}}}(\chi(\mathbf{x})))^{-1} \leq \frac{\sup_{\mathbf{w}\in[0,1]^d} f_W(\mathbf{w})}{\inf_{\mathbf{w}\in[0,1]^d} f_W(\mathbf{w})}$$
$$\leq \frac{\sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} f_X(\mathbf{x})}{\inf_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} f_X(\mathbf{x})} \cdot \frac{\sup_{\mathbf{x}\in[0,1]^d} |\det(\nabla\chi^{-1}(\mathbf{x}))|}{\inf_{\mathbf{x}\in[0,1]^d} |\det(\nabla\chi^{-1}(\mathbf{x}))|}.$$

Putting together, we have

$$\mathsf{c}_1 \leq \frac{\overline{f}_X^2}{\underline{f}_X}\mathsf{S}_\chi,$$

with $\overline{f}_X = \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} f_X(\mathbf{x})$, $\underline{f}_X = \inf_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} f_X(\mathbf{x})$, $\mathsf{S}_\chi = \frac{\sup_{\mathbf{x}\in[0,1]^d} |\det(\nabla\chi^{-1}(\mathbf{x}))|}{\inf_{\mathbf{x}\in[0,1]^d} |\det(\nabla\chi^{-1}(\mathbf{x}))|}\|\|\nabla\chi^{-1}\|_2\|_\infty$. Also,

$$\mathsf{c}_2 = \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} \|\nabla\phi_{\mathcal{H}}^{-1}(\phi_{\mathcal{H}}(\mathbf{x}))\|_2 \leq \sup_{\mathbf{u}\in[0,1]^d} \|\nabla\chi^{-1}(\mathbf{u})\|_2 \sup_{\mathbf{u}\in[0,1]^d} \|\nabla T_{\mathbb{Q}_{\mathcal{H}}}^{-1}(\mathbf{u})\|_2$$
$$\leq \sup_{\mathbf{u}\in[0,1]^d} \|\nabla\chi^{-1}(\mathbf{u})\|_2 \frac{\sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} f_X(\mathbf{x})}{\inf_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}} f_X(\mathbf{x})} \cdot \frac{\sup_{\mathbf{x}\in[0,1]^d} |\det(\nabla\chi^{-1}(\mathbf{x}))|}{\inf_{\mathbf{x}\in[0,1]^d} |\det(\nabla\chi^{-1}(\mathbf{x}))|} \leq \frac{\overline{f}_X}{\underline{f}_X}\mathsf{S}_\chi.$$

This completes the proof. □

SA-II.3 General Result: Proof of Theorem 1

First, we make a reduction through the surrogate measure and normalizing transformation. We want to show that under assumption (ii) in Theorem 1, the empirical process $(X_n(h) : h \in \mathcal{H})$ can be written as an empirical process based on i.i.d $\mathsf{Uniform}([0,1]^d)$ random variables. Let $\mathcal{Z}_{\mathcal{H}} = \mathcal{X} \cap \operatorname{Supp}(\mathcal{H})$. Since $\mathbb{Q}_{\mathcal{H}} = \mathfrak{m} \circ \phi_{\mathcal{H}}$ by Assumption (ii) in Theorem 1, and $\mathbb{Q}_{\mathcal{H}}|_{\mathcal{Z}_{\mathcal{H}}} = \mathbb{P}_X|_{\mathcal{Z}_{\mathcal{H}}}$,

$$\mathbb{P}_X|_{\mathcal{Z}_{\mathcal{H}}} = \mathfrak{m} \circ \phi_{\mathcal{H}}|_{\mathcal{Z}_{\mathcal{H}}}.$$

To define the $\mathsf{Uniform}([0,1]^d)$ random variables on the probability space that $\mathbf{x}_i$'s live in, we define a joint probability measure $\mathbb{O}$ on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^{2d}))$ such that for all $A \in \mathcal{B}(\mathbb{R}^{2d})$:

$$\mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}} \times \mathcal{Z}_{\mathcal{H}})) = \mathbb{P}_X(\Pi_{1:d}(A \cap \{(\mathbf{x},\mathbf{x}) : \mathbf{x} \in \mathcal{Z}_{\mathcal{H}}\})),$$
$$\mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}} \times \mathcal{Z}_{\mathcal{H}}^c)) = \mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}}^c \times \mathcal{Z}_{\mathcal{H}})) = 0,$$
$$\mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}}^c \times \mathcal{Z}_{\mathcal{H}}^c)) = \int_{\mathcal{Z}_{\mathcal{H}}^c \cap \Pi_{d+1:2d}(A)} \frac{\mathbb{P}_X(A^{\mathbf{u}} \cap \mathcal{Z}_{\mathcal{H}}^c)}{\mathbb{P}_X(\mathcal{Z}_{\mathcal{H}}^c)} d(\mathfrak{m} \circ \phi_{\mathcal{H}})(\mathbf{u}),$$

where $\Pi_{1:d}(A) = \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x},\mathbf{u}) \in A \text{ for some } \mathbf{u} \in \mathbb{R}^d\}$, $\Pi_{d+1:2d}(A) = \{\mathbf{u} \in \mathbb{R}^d : (\mathbf{x},\mathbf{u}) \in A \text{ for some } \mathbf{x} \in \mathbb{R}^d\}$, and $A^{\mathbf{u}} = \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x},\mathbf{u}) \in A\}$. See Figure 1 for a graphical illustration.

Then we can check that (i) the marginals of $\mathbb{O}$ are $\mathbb{P}_X$ and $\mathfrak{m} \circ \phi_{\mathcal{H}}$, respectively; (ii) $\mathbb{O}|_{\mathcal{Z}_{\mathcal{H}} \times \mathbb{R}^d \cup \mathbb{R}^d \times \mathcal{Z}_{\mathcal{H}}}$ is supported on $\{(\mathbf{x},\mathbf{x}) : \mathbf{x} \in \mathcal{Z}_{\mathcal{H}}\}$. By Skorohod embedding (Dudley, 2014, Lemma 3.35), on a possibly enlarged probability space, there exists a $\mathbf{u}_i, 1 \leq i \leq n$ i.i.d. Uniform$([0,1]^d)$ such that $(\mathbf{x}_i, \phi_{\mathcal{H}}^{-1}(\mathbf{u}_i))$ has joint law $\mathbb{O}$. In particular, if $\mathbf{x}_i \in \mathcal{Z}_{\mathcal{H}}$, then $\mathbf{x}_i = \phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)$; if $\mathbf{x}_i \in \mathcal{Z}_{\mathcal{H}}^c$, then $\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i) \in \mathcal{Z}_{\mathcal{H}}^c$, and since $\mathcal{Q}_{\mathcal{H}} \subseteq \mathcal{X} \cup (\cap_{h\in\mathcal{H}} \operatorname{Supp}(h)^c)$, $\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i) \in \cap_{h\in\mathcal{H}} \operatorname{Supp}(h)^c$.

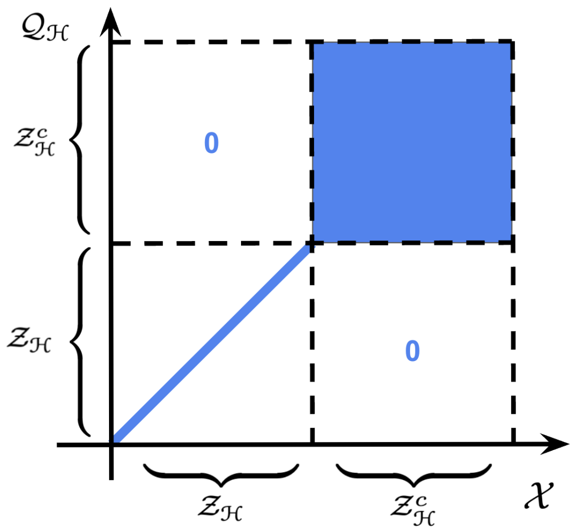

Figure 1: Illustration of $\mathbb{O}$. $\mathbb{O}$ concentrates on $\{(\mathbf{x},\mathbf{x}) : \mathbf{x} \in \mathcal{Z}_{\mathcal{H}}\}$ in $\mathcal{Z}_{\mathcal{H}} \times \mathcal{Z}_{\mathcal{H}}$, agrees with the zero measure on $\mathcal{Z}_{\mathcal{H}} \times \mathcal{Z}_{\mathcal{H}}^c$ and $\mathcal{Z}_{\mathcal{H}}^c \times \mathcal{Z}_{\mathcal{H}}$, and agrees with the product measure of $\mathbb{P}_X \otimes (\mathfrak{m} \circ \phi_{\mathcal{H}})$ on $\mathcal{Z}_{\mathcal{H}}^c \times \mathcal{Z}_{\mathcal{H}}^c$.

Thus, we take $\widetilde{h} = h \circ \phi_{\mathcal{H}}^{-1}$, and consider the new class of functions $\widetilde{\mathcal{H}} = \{\widetilde{h} : h \in \mathcal{H}\}$. For any $h \in \mathcal{H}$,

$$\begin{aligned}
X_n(h) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[h(\mathbf{x}_i) - \mathbb{E}[h(\mathbf{x}_i)]\big] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[h(\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)) - \mathbb{E}[h(\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i))]\big] \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big[\widetilde{h}(\mathbf{u}_i) - \mathbb{E}[\widetilde{h}(\mathbf{u}_i)]\big],
\end{aligned}$$

where the second equality follows because $\mathbf{x}_i = \phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)$ on the event $\{\mathbf{x}_i \in \mathcal{Z}_{\mathcal{H}}\}$, and $h(\mathbf{x}_i) = h(\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)) = 0$ (a.s.) on the event $\{\mathbf{x}_i \in \mathcal{Z}_{\mathcal{H}}^c\}$. Hence, we work with an equivalent empirical process

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

In particular, $\mathbf{u}_i$ has the uniform distribution $\mathbb{P}_U$ which has a Lebesgue density f_U that is bounded from above and below on its support $[0,1]^d$,

$$(X_n(h) : h \in \mathcal{H}) = (\widetilde{X}_n(\widetilde{h}) : \widetilde{h} \in \widetilde{\mathcal{H}}) \qquad \text{almost surely,}$$

and Assumption SA.1 is satisfied with the random sample $(\mathbf{u}_i : 1 \leq i \leq n)$ with $\mathbf{u}_i \sim \mathbb{P}_U$ and the class of functions $\widetilde{\mathcal{H}}$. We thus consider $\mathcal{A}_K(\mathbb{P}_U, 1)$, an axis aligned dyadic expansion of depth K with respect to probability measure $\mathbb{P}_U = \mathsf{Uniform}([0,1]^d)$. Suppose $\mathcal{E}_K$ $(\mathtt{M}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}} = \mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d})$ and $\Pi_0 = \Pi_0[\mathcal{A}_K(\mathbb{P}_U, 1)]$ are defined based on $\mathcal{A}_K(\mathbb{P}_U, 1)$ as in Section SA-II.1.2. By Lemma SA.2 and Lemma SA.10, $\widetilde{\mathcal{H}} \cup \Pi_0 \widetilde{\mathcal{H}}$ is $\mathbb{P}_U$-pregaussian, hence by the same construction given in Section SA-II.1.3, on a possibly enlarged probability space, there exists a mean-zero Gaussian process $\widetilde{Z}_n^X$ indexed by $\widetilde{\mathcal{H}} \cup \Pi_0 \widetilde{\mathcal{H}}$ such that with almost sure continuous sample path such that

$$\mathbb{E}[\widetilde{Z}_n^X(g)\widetilde{Z}_n^X(f)] = \mathbb{E}[\widetilde{X}_n(g)\widetilde{X}_n(f)], \qquad \forall g, f \in \widetilde{\mathcal{H}} \cup \Pi_0\widetilde{\mathcal{H}},$$

and $U_{j,k} = \sum_{i=1}^n e_{j,k}(\mathbf{u}_i)$ for all $(j,k) \in \mathcal{J}_K$. Let $\widetilde{\mathcal{H}}_\delta$ be a $\delta\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d} = \delta\mathtt{M}_{\mathcal{H},\mathcal{Q}_{\mathcal{H}}}$-net of $\widetilde{\mathcal{H}}$ with cardinality no greater than $\mathtt{N}_{\widetilde{\mathcal{H}},[0,1]^d}(\delta, \mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d})$.

The proof proceeds by bounding each of the terms in the decomposition

$$\|\widetilde{X}_n - \widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}} \leq \underbrace{\|\widetilde{X}_n - \widetilde{X}_n \circ \pi_{\widetilde{\mathcal{H}}_\delta}\|_{\widetilde{\mathcal{H}}} + \|\widetilde{Z}_n^X \circ \pi_{\widetilde{\mathcal{H}}_\delta} - \widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}}}_{\text{meshing error}} + \underbrace{\|\widetilde{X}_n - \widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}_\delta}}_{\text{error on net}},$$

$$\|\widetilde{X}_n - \widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}_\delta} \leq \underbrace{\|\Pi_0\widetilde{X}_n - \Pi_0\widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}_\delta}}_{\text{approximation error}} + \underbrace{\|\widetilde{X}_n - \Pi_0\widetilde{X}_n\|_{\widetilde{\mathcal{H}}_\delta} + \|\Pi_0\widetilde{Z}_n^X - \widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}_\delta}}_{\text{projection error}},$$

and then balancing their contributions.

Given the cells $\mathcal{A}_K(\mathbb{P}_U, 1)$, we have $\mathcal{U}_j \subseteq [-2^{-\frac{K-j}{d}+1}, 2^{-\frac{K-j}{d}+1}]^d$. Then, by Lemma SA.7, for all $t > 0$,

$$\mathbb{P}\left[\|\widetilde{X}_n \circ \Pi_0 - \widetilde{Z}_n^X \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta} > 48\sqrt{\frac{\mathcal{R}_K(\widetilde{\mathcal{H}}_\delta)}{n}t} + 4\sqrt{\frac{\mathtt{C}_{\widetilde{\mathcal{H}}_\delta,K}}{n}t}\right] \leq 2\mathtt{N}_{\widetilde{\mathcal{H}},[0,1]^d}(\delta, \mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d})e^{-t},$$

where

$$\mathcal{R}_K(\widetilde{\mathcal{H}}_\delta) \leq \begin{cases} \min\{\mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\mathtt{M}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}, \mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\mathtt{L}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\}, & \text{if } d = 1, \\ \min\{2^K\mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\mathtt{M}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}, K\mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\mathtt{L}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\}, & \text{if } d = 2, \\ \min\{2^{K(d-1)}\mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\mathtt{M}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}, 2^{K(d-2)}\mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\mathtt{L}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}\} & \text{if } d \geq 3. \end{cases}$$

Now we calculate the $\mathtt{C}_{\widetilde{\mathcal{H}}_\delta,K}$ term. Let $\widetilde{h} \in \widetilde{\mathcal{H}}$ and take $(\widetilde{h}_\ell)_{\ell\in\mathbb{N}}$ be any sequence of real-valued functions on $([0,1]^d, \mathcal{B}([0,1]^d))$ such that $\widetilde{h}_\ell \to \widetilde{h}$ $\mathfrak{m}$-almost surely, and are bounded by $2\mathtt{M}_{\mathcal{H}}$ on $\mathcal{X}$. Moreover, by Dominated Convergence Theorem, the definition of $\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d}$ and similar arguments as in the proof of Lemma SA.7, for each $(j,k) \in \mathcal{I}_K$,

$$\begin{aligned} \sum_{m:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}} \left|\widetilde{\beta}_{j,k}(\widetilde{h})\right| &\leq 2^{2(K-l)} \int_{\mathcal{U}_l}\int_{\mathcal{C}_{j,k}} |\widetilde{h}(\mathbf{x}) - \widetilde{h}(\mathbf{x}+\mathbf{s})| d\mathbf{x} d\mathbf{s} \\ &= \lim_{\ell\to\infty} 2^{2(K-l)} \int_{\mathcal{U}_l}\int_{\mathcal{C}_{j,k}} |\widetilde{h}_\ell(\mathbf{x}) - \widetilde{h}_\ell(\mathbf{x}+\mathbf{s})| d\mathbf{x} d\mathbf{s} \\ &\leq \limsup_{\ell\to\infty} 2^{2(K-l)} \int_{\mathcal{U}_l} \|\mathbf{s}\| \mathtt{TV}_{\{\widetilde{h}_\ell\},\mathcal{C}_{j,k}} d\mathbf{s}. \end{aligned}$$

Since the above inequality holds for all sequences $(h_\ell)_{\ell\in\mathbb{N}}$ such that $\widetilde{h}_\ell \to \widetilde{h}$ $\mathfrak{m}$-almost surely, and are bounded by $2\mathtt{M}_{\mathcal{H}}$ on $[0,1]^d$, Definitions SA.1 and SA.2 implies

$$\begin{aligned}
\sum_{m:\mathcal{C}_{l,m}\subseteq\mathcal{C}_{j,k}} \left|\widetilde{\beta}_{j,k}(\widetilde{h})\right| &\leq 2^{2(K-l)}\int_{\mathcal{U}_l}\|\mathbf{s}\|\mathtt{TV}^*_{\widetilde{\mathcal{H}},\mathcal{C}_{j,k}}d\mathbf{s} \\
&\leq 2^{2(K-l)}\int_{\mathcal{U}_l}\|\mathbf{s}\|\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d}\|\mathcal{C}_{j,k}\|_\infty^{d-1}d\mathbf{s} \\
&\leq \sqrt{d}2^{2(K-l)}\,\mathfrak{m}(\mathcal{U}_l)\|\mathcal{U}_l\|_\infty\|\mathcal{C}_{j,k}\|_\infty^{d-1}\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d} \\
&\leq \sqrt{d}2^{\frac{d-1}{d}(j-l)}\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d}.
\end{aligned}$$

It follows from the definition of $\mathtt{C}_{\widetilde{\mathcal{H}},K}$ in Lemma SA.3 that

$$\mathtt{C}_{\widetilde{\mathcal{H}},K} \leq \min\{\sqrt{K\mathtt{M}^2_{\widetilde{\mathcal{H}},[0,1]^d}}, \sqrt{\sqrt{d}\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d} + \mathtt{M}^2_{\widetilde{\mathcal{H}},[0,1]^d}}\}.$$

For projection error, by Lemma SA.9, for all $t > 0$, with probability at least $1 - 2\mathtt{N}_{\widetilde{\mathcal{H}},[0,1]^d}(\delta, \mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d})e^{-t}$,

$$\begin{aligned}
\|\widetilde{X}_n - \widetilde{X}_n \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta} &\leq \sqrt{4d\min\{2\mathtt{M}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}, \mathtt{L}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}2^{-K}\}2^{-K}\mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}t} \\
&\quad + \frac{4\min\{2\mathtt{M}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}, \mathtt{L}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}2^{-K}\}}{3\sqrt{n}}t, \\
\|\widetilde{Z}_n^X - \widetilde{Z}_n^X \circ \Pi_0\|_{\widetilde{\mathcal{H}}_\delta} &\leq \sqrt{4d\min\{2\mathtt{M}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}, \mathtt{L}_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}2^{-K}\}2^{-K}\mathtt{TV}^*_{\widetilde{\mathcal{H}}_\delta,[0,1]^d}t}.
\end{aligned}$$

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 $1 - 2\exp(-t)$,

$$\|\widetilde{X}_n - \widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}_\delta} \leq \widetilde{\mathsf{A}}_n(t,\delta),$$

where

$$\begin{aligned}
\widetilde{\mathsf{A}}_n(t,\delta) =& \min\left\{\mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}}, \mathtt{l}_{n,d}\sqrt{\mathtt{L}_{\widetilde{\mathcal{H}},[0,1]^d}}\right\}\sqrt{d\mathtt{TV}^*_{\widetilde{\mathcal{H}},[0,1]^d}}\sqrt{(t + \log\widetilde{\mathtt{N}}_{\widetilde{\mathcal{H}},[0,1]^d}(\delta,\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}))} \\
&+ \sqrt{\frac{\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}}, \sqrt{\sqrt{d}\mathtt{K}^*_{\widetilde{\mathcal{H}},[0,1]^d} + \mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}}\}(t + \log\widetilde{\mathtt{N}}_{\widetilde{\mathcal{H}},[0,1]^d}(\delta,\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d})).
\end{aligned}$$

By Lemma SA.6 we bound the meshing error by, for all $t > 0$,

$$\begin{aligned}
\mathbb{P}\left[\|\widetilde{X}_n - \widetilde{X}_n\circ\pi_{\widetilde{\mathcal{H}}_\delta}\|_{\widetilde{\mathcal{H}}} > C\mathsf{F}_n(t,\delta)\right] &\leq \exp(-t), \\
\mathbb{P}\left[\|\widetilde{Z}_n^X\circ\pi_{\widetilde{\mathcal{H}}_\delta} - \widetilde{Z}_n^X\|_{\widetilde{\mathcal{H}}} > C(\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}J(\delta,\widetilde{\mathcal{H}},\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}) + \delta\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}\sqrt{t})\right] &\leq \exp(-t),
\end{aligned}$$

where

$$\mathsf{F}_n(t,\delta) = J(\delta,\widetilde{\mathcal{H}},\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d})\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d} + \frac{\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}J^2(\delta,\widetilde{\mathcal{H}},\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d})}{\delta^2\sqrt{n}} + \delta\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}\sqrt{t} + \frac{\mathtt{M}_{\widetilde{\mathcal{H}},[0,1]^d}}{\sqrt{n}}t.$$

Take the Gaussian process $(Z_n(h) : h \in \mathcal{H})$ such that, almost surely, $Z_n(h) = \widetilde{Z}_n(\widetilde{h})$ for all $h \in \mathcal{H}$. The result then follows from the decomposition that

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

and Lemma SA.10 to establish the relationships between the parameters of $\mathcal{H}$ over $\mathcal{Q}_{\mathcal{H}}$ and those of $\widetilde{\mathcal{H}}$ over $[0,1]^d$. □

SA-II.4 Additional Results

In what follows, we drop the dependence on $\mathcal{C} = \mathcal{Q}_{\mathcal{F}}$ for all quantities in Definitions 4-12. That is, to save notation, we set $\mathtt{TV}_{\mathcal{F}} = \mathtt{TV}_{\mathcal{F},\mathcal{Q}_{\mathcal{F}}}$, $\mathtt{K}_{\mathcal{F}} = \mathtt{K}_{\mathcal{F},\mathcal{Q}_{\mathcal{F}}}$, $\mathtt{M}_{\mathcal{F},\mathcal{X}} = \mathtt{M}_{\mathcal{F},\mathcal{Q}_{\mathcal{F}}}$, $M_{\mathcal{F},\mathcal{X}}(\mathbf{u}) = M_{\mathcal{F},\mathcal{Q}_{\mathcal{F}}}(\mathbf{u})$, $\mathtt{L}_{\mathcal{F}} = \mathtt{L}_{\mathcal{F},\mathcal{Q}_{\mathcal{F}}}$, and so on, whenever there is no confusion.

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

$$\mathsf{S}_n(t) = \mathsf{m}_{n,d}\sqrt{\mathtt{c}_1\mathtt{M}_{\mathcal{H}}\mathtt{TV}_{\mathcal{H}}}\sqrt{t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n)} + \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{\mathtt{c}_3\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n))$$

in Theorem 1.

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

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

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. □

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

$$\begin{aligned}\mathsf{S}_n(t) =\ & \min\left\{\mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\mathcal{H}}}, \mathsf{l}_{n,d}\sqrt{\mathtt{c}_2\mathtt{L}_{\mathcal{H}}}\right\}\sqrt{\mathtt{TV}_{\mathcal{H}}}\sqrt{t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n)} \\ & + \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{\mathtt{c}_3\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + \mathtt{d}_{\mathcal{H}}\log(\mathtt{c}_{\mathcal{H}}n))\end{aligned}$$

in Theorem 1.

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. □

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

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

Proof of Corollary SA.3. Under the polynomial entropy condition, $\log \mathtt{N}_{\mathcal{H}}(\delta) \leq \mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}}$, $J(\delta, \mathcal{H}, \mathtt{M}_{\mathcal{H}}) \leq \sqrt{\mathtt{a}_{\mathcal{H}}}(2 - \mathtt{b}_{\mathcal{H}})^{-1}\delta^{-\mathtt{b}_{\mathcal{H}}/2+1}$,

$$\mathsf{A}_n(t,\delta) \leq \min\{\mathsf{m}_{n,d}\sqrt{\mathtt{M}_{\mathcal{H}}}, \mathsf{l}_{n,d}\sqrt{\mathtt{c}_2\mathtt{L}_{\mathcal{H}}}\}\sqrt{\mathtt{TV}_{\mathcal{H}}(t + \mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}})}$$
$$+ \sqrt{\frac{\mathtt{M}_{\mathcal{H}}}{n}}\min\{\sqrt{\log n}\sqrt{\mathtt{M}_{\mathcal{H}}}, \sqrt{\mathtt{c}_3\mathtt{K}_{\mathcal{H}} + \mathtt{M}_{\mathcal{H}}}\}(t + \mathtt{a}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}}),$$
$$\mathsf{F}_n(t,\delta) \leq \mathtt{a}_{\mathcal{H}}(2 - \mathtt{b}_{\mathcal{H}})^{-2}\left(\mathtt{M}_{\mathcal{H}}\delta^{-\mathtt{b}_{\mathcal{H}}/2+1} + \frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}\delta^{-\mathtt{b}_{\mathcal{H}}} + \delta\mathtt{M}_{\mathcal{H}}\sqrt{t} + \frac{\mathtt{M}_{\mathcal{H}}}{\sqrt{n}}t\right).$$

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

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

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

SA-II.5 Proofs of Corollaries 1, 2, and 3

Proof of Corollary 1. Take $t = C\log n$ with $C > 1$ in Corollary SA.1. □

Proof of Corollary 2. Take $t = C\log n$ with $C > 1$ in Corollary SA.2. □

Proof of Corollary 3. Take $t = C\log n$ with $C > 1$ in Corollary SA.3. □

SA-II.6 Example 1: Kernel Density Estimation

To simplify notation, in this section the parameters of $\mathcal{H}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Q}_{\mathcal{H}}$, and the index $\mathcal{C}$ is omitted whenever there is no confusion.

SA-II.6.1 Surrogate Measure and Normalizing Transformation

We show that the two sets of primitive conditions discussed in the paper imply condition (ii) in Theorem 1.

First, consider the case $\mathcal{X} = \times_{l=1}^{d}[\mathsf{a}_l, \mathsf{b}_l]$, $-\infty \leq \mathsf{a}_l < \mathsf{b}_l \leq \infty$ and $\mathcal{W}$ is arbitrary. Observe that $\mathbb{Q}_{\mathcal{H}} = \mathbb{P}_X$ is always a valid surrogate measure for $\mathbb{P}_X$ with respect to $\mathcal{H}$, according to Definition 2. The conclusion then follows from Case 1 from Section 3.1 with $f_Q = f_X$.

Second, consider the case when $\mathcal{X}$ may be unbounded. We present a general construction, and then specialize it to the example discussed in the paper. Suppose we can find $\mathcal{Q}_{\mathcal{H}}$ diffeomorphic to $[0,1]^d$ such that $\mathcal{X} \cap \operatorname{Supp}(\mathcal{H}) \subseteq \mathcal{Q}_{\mathcal{H}} \subseteq \mathcal{X} \cup \operatorname{Supp}(\mathcal{H})^c$, with $\mathbb{P}_X(\mathcal{X} \cap \operatorname{Supp}(\mathcal{H})) < 1$ and $\mathfrak{m}(\mathcal{Q}_{\mathcal{H}} \setminus (\mathcal{X} \cap \operatorname{Supp}(\mathcal{H}))) > 0$. Setting $\mathbb{Q}_{\mathcal{H}}$ to be the probability measure with Lebesgue density f_Q such that

$$f_Q(\mathbf{x}) = \begin{cases} f_X(\mathbf{x}), & \text{if } \mathbf{x} \in \mathcal{X} \cap \operatorname{Supp}(\mathcal{H}), \\ (1 - \mathbb{P}_X(\mathcal{X} \cap \operatorname{Supp}(\mathcal{H})))/\,\mathfrak{m}(\mathcal{Q}_{\mathcal{H}} \setminus (\mathcal{X} \cap \operatorname{Supp}(\mathcal{H}))), & \text{if } \mathbf{x} \in \mathcal{Q}_{\mathcal{H}} \setminus (\mathcal{X} \cap \operatorname{Supp}(\mathcal{H})), \\ 0, & \text{otherwise.} \end{cases}$$

then $\mathbb{Q}_{\mathcal{H}}$ is a surrogate measure of $\mathbb{P}_X$ with respect to $\mathcal{H}$. Suppose χ is a diffeomorphism from $\mathcal{Q}_{\mathcal{H}}$ to $[0,1]^d$. Since we assumed $\mathcal{X} \cap \operatorname{Supp}(\mathcal{H}) \subseteq \mathcal{Q}_{\mathcal{H}} \subseteq \mathcal{X} \cup \operatorname{Supp}(\mathcal{H})^c$, with $\mathbb{P}_X(\mathcal{X} \cap \operatorname{Supp}(\mathcal{H})) < 1$ and $\mathfrak{m}(\mathcal{Q}_{\mathcal{H}} \setminus (\mathcal{X} \cap \operatorname{Supp}(\mathcal{H}))) > 0$, we can check that (1) f_Q is supported and positive on $\mathcal{Q}_{\mathcal{H}}$, (2) f_Q agrees with f_X on $\mathcal{X} \cap \operatorname{Supp}(\mathcal{H})$. Then Case 2 in Section 3.1 implies $\phi_{\mathcal{H}} = T_{\mathbb{Q}_{\mathcal{H}} \circ \chi^{-1}} \circ \chi$ is a valid normalizing transformation, and condition (ii) in Theorem 1 holds. Suppose $0 < \inf_{\mathbf{x} \in \mathcal{X} \cap \operatorname{Supp}(\mathcal{H})} f_X(\mathbf{x}) < \sup_{\mathbf{x} \in \mathcal{X} \cap \operatorname{Supp}(\mathcal{H})} f_X(\mathbf{x}) < \infty$ and $\frac{\sup_{\mathbf{x} \in [0,1]^d} |\det(\nabla \chi^{-1}(\mathbf{x}))|}{\inf_{\mathbf{x} \in [0,1]^d} |\det(\nabla \chi^{-1}(\mathbf{x}))|} \| \| \nabla \chi^{-1} \|_2 \|_\infty < \infty$, then we have $\mathsf{c}_1 = O(1)$ and $\mathsf{c}_2 = O(1)$ (and hence $\mathsf{c}_3 = O(1)$).

For a concrete example, consider the case $\mathcal{X} = \mathbb{R}_+^d$, $\mathcal{W} = \times_{l=1}^{d}[\mathsf{a}_l, \mathsf{b}_l]$, $0 \leq \mathsf{a}_l < \mathsf{b}_l < \infty$, and $\mathcal{K} = [-1,1]^d$. Observe that $\operatorname{Supp}(\mathcal{H}) \cap \mathcal{X} = \times_{l=1}^{d}[(\mathsf{a}_l - b)_+, \mathsf{b}_l + b] = \times_{l=1}^{d}[\bar{\mathsf{a}}_l, \bar{\mathsf{b}}_l]$. Since $\mathcal{X} = \mathbb{R}_+^d$, $\mathbb{P}_X(\times_{l=1}^{d}[\bar{\mathsf{a}}_l, \bar{\mathsf{b}}_l]) < 1$. Moreover, we can check that $\mathcal{X} \cap \operatorname{Supp}(\mathcal{H}) \subseteq \mathcal{Q}_{\mathcal{H}} \subseteq \mathcal{X} \cup \operatorname{Supp}(\mathcal{H})^c$ and $\mathbb{Q}_{\mathcal{H}}$ agrees with $\mathbb{P}_X$ on $\mathcal{X} \cap \operatorname{Supp}(\mathcal{H})$. The rest then follows from the general construction above.

SA-II.6.2 Class $\mathcal{H}$ and Its Corresponding Constants

Let $\mathcal{H} = \{h_{\mathbf{w}} : \mathbf{w} \in \mathcal{W}\}$ with $h_{\mathbf{w}}(\cdot) = b^{-d/2} K(b^{-1}(\mathbf{w} - \cdot))$. Since K is compactly supported and Lipschitz, $\mathtt{M}_{\{K\}} < \infty$. Hence, $\mathtt{M}_{\mathcal{H}} = b^{-d/2} \mathtt{M}_{\{K\}} \leq C_K b^{-d/2}$ and $\mathtt{L}_{\mathcal{H}} \leq b^{-\frac{d}{2}-1} \mathtt{L}_{\{K\}} \leq C_K b^{-d/2-1}$, where C_K is a constant that only depends on the kernel function K. Since $\sup_{\mathbf{w} \in \mathcal{W}} \mathfrak{m}(\operatorname{Supp}(h_{\mathbf{w}})) \leq C_K b^d$ and each $h_{\mathbf{w}}$ is differentiable,

$$\mathtt{TV}_{\mathcal{H}} = \sup_{\mathbf{w} \in \mathcal{W}} \int \| \nabla h_{\mathbf{w}}(\mathbf{u}) \| d\mathbf{u} \leq \sup_{\mathbf{w} \in \mathcal{W}} \mathfrak{m}(\operatorname{Supp}(h_{\mathbf{w}})) \mathtt{L}_{\mathcal{H}} \leq C_K b^{d/2-1}.$$

To upper bound $\mathtt{K}_{\mathcal{H}}$, let $\mathcal{D} \subseteq \mathcal{Q}_{\mathcal{H}}$ be a cube with edges of length a parallel to the coordinate axises. Consider the following two cases: (i) if $\mathsf{a} < b$, then $\mathtt{TV}_{\mathcal{H},\mathcal{D}} \leq C_K b^{-d/2-1} \mathsf{a}^d \leq C_K b^{-d/2} \mathsf{a}^{d-1}$; (ii) if $\mathsf{a} > b$, then $\mathtt{TV}_{\mathcal{H},\mathcal{D}} \leq C_K \sup_{\mathbf{w} \in \mathcal{W}} \mathfrak{m}(\operatorname{Supp}(h_{\mathbf{w}})) \mathtt{L}_{\mathcal{H}} \leq C_K b^d b^{-d/2-1} \leq C_K b^{-d/2} b^{d-1} \leq C_K b^{-d/2} \mathsf{a}^{d-1}$. This shows

$$\mathtt{K}_{\mathcal{H}} \leq C_K b^{-d/2}.$$

Next, by a change of variables,

$$\mathtt{E}_{\mathcal{H}} = \sup_{\mathbf{w} \in \mathcal{W}} \int b^{-\frac{d}{2}} |K(b^{-1}(\mathbf{w} - \mathbf{u}))| f_X(\mathbf{u}) d\mathbf{u} = \sup_{\mathbf{w} \in \mathcal{W}} \int b^{-\frac{d}{2}} |K(\mathbf{z})| f_X(\mathbf{w} - h\mathbf{z}) b^d d\mathbf{z} \leq C_K b^{d/2}.$$

Finally, we check that $\mathcal{H}$ is a VC-type class. We will apply Lemma 7 from Cattaneo *et al.* (2024) on the class $\mathtt{M}_{\mathcal{H}}^{-1}\mathcal{H}$. To check the conditions in this lemma, define $g_{\mathbf{w}}(\cdot) = b^{-\frac{d}{2}}\mathtt{M}_{\mathcal{H}}^{-1}K(\cdot)$ for all $\mathbf{w} \in \mathcal{W}$. Note that $g_{\mathbf{w}}$ is the same function for all $\mathbf{w} \in \mathcal{W}$ in this setting (but, more generally, our results allow for functions varying with the evaluation point such as in the case of boundary adaptive kernels). Then $\mathtt{M}_{\mathcal{H}}^{-1}\mathcal{H} = \{g_{\mathbf{w}}(\frac{\mathbf{w}-\cdot}{b}) : \mathbf{w} \in \mathcal{W}\}$, and there exists a constant c_K, only depending on $\mathtt{M}_{\{K\}}$ and $\mathtt{L}_{\{K\}}$, such that

$$\sup_{\mathbf{w}\in\mathcal{W}} \|g_{\mathbf{w}}\|_\infty \leq c_K, \qquad \sup_{\mathbf{w}\in\mathcal{W}} \sup_{\mathbf{u},\mathbf{v}\in\mathcal{Q}_{\mathcal{H}}} \frac{|g_{\mathbf{w}}(\mathbf{u}) - g_{\mathbf{w}}(\mathbf{v})|}{\|\mathbf{u}-\mathbf{v}\|_\infty} \leq c_K, \qquad \sup_{\mathbf{w},\mathbf{w}'\in\mathcal{W}} \sup_{\mathbf{u}\in\mathcal{Q}_{\mathcal{H}}} \frac{|g_{\mathbf{w}}(\mathbf{u}) - g_{\mathbf{w}'}(\mathbf{u})|}{\|\mathbf{w}-\mathbf{w}'\|_\infty} \leq c_K.$$

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 $0 < \varepsilon < 1$, $\mathtt{N}_{\mathtt{M}_{\mathcal{H}}^{-1}\mathcal{H}}(\varepsilon, 1) \leq c_K\varepsilon^{-d-1} + 1$, and hence

$$\mathtt{N}_{\mathcal{H}}(\varepsilon, \mathtt{M}_{\mathcal{H}}) \leq c_K\varepsilon^{-2d-2} + 1,$$

The conclusions on uniform Gaussian strong approximation rates then follow from Corollaries 1–3.

SA-III Multiplicative-Separable Empirical Process

Let $\mathbf{z}_i = (\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^d \times \mathbb{R}$, $i = 1, \ldots, n$, be i.i.d. random vectors supported on a background probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The multiplicative-separable empirical process is

$$G_n(g, r) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \big(g(\mathbf{x}_i)r(y_i) - \mathbb{E}[g(\mathbf{x}_i)r(y_i)]\big), \qquad g \in \mathcal{G}, r \in \mathcal{R},$$

where $\mathcal{G}$ and $\mathcal{R}$ are possibly n-varying classes of functions. Notably, if we take $\mathcal{H} = \mathcal{G}\cdot\mathcal{R} = \{g\cdot r : g \in \mathcal{G}, r \in \mathcal{R}\}$, then the above process can also be written as a generic empirical process based on $(\mathbf{z}_i : 1 \leq i \leq n)$ because

$$X_n(h) = X_n(g \cdot r) = \frac{1}{\sqrt{n}} \sum_{i=1}^n ((g \cdot r)(\mathbf{z}_i) - \mathbb{E}[(g \cdot r)(\mathbf{z}_i)]), \qquad h = g \cdot r \in \mathcal{H} = \mathcal{G} \cdot \mathcal{R}.$$

Hence, the same decomposition for the X_n process also applies for the G_n process:

$$\begin{aligned}
\|G_n - Z_n^G\|_{\mathcal{G}\times\mathcal{R}} &\leq \|G_n - Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^G \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^G\|_{\mathcal{G}\times\mathcal{R}} \\
&\leq \|\Pi_1 Z_n^G - Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|\Pi_1 G_n - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} \\
&\quad + \|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^G \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^G\|_{\mathcal{G}\times\mathcal{R}},
\end{aligned}$$

where $(\mathcal{G}\times\mathcal{R})_\delta$ denotes a discretization (or meshing) of $\mathcal{G}\times\mathcal{R}$ (i.e., δ-net of $\mathcal{G}\times\mathcal{R}$), and the terms $\|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}}$ and $\|Z_n^G \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^G\|_{\mathcal{G}\times\mathcal{R}}$ capture the fluctuations (or oscillations) of G_n and Z_n^G relative to the meshing for each of the stochastic processes. $\|\Pi_1 G_n - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and $\|\Pi_1 Z_n^G - Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ represent projections onto a Haar function space, where $\Pi_1 G_n(h) = G_n \circ \Pi_1 h$. The operator Π_1 is a projection onto piecewise constant functions that respects the multiplicative structure of the G_n process. The final term $\|\Pi_1 G_n - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ captures the coupling between the empirical process and the Gaussian process (on a δ-net of $\mathcal{G}\times\mathcal{R}$, after the projection Π_1).

A general result under uniform entropy integral conditions is presented in Section SA-III.2, and a corollary under a VC-type condition is presented in Section SA-III.3. The proofs exploit the existence of a surrogate measure and normalizing transformation of $\mathcal{G}$ with respect to $\mathbb{P}_X$, the law of $\mathbf{x}_1$, as developed in Section SA-II.2. The preliminary technical results differ from those in Section SA-II by explicitly leveraging the multiplicative structure of the empirical process, and are organized as follows.

- Section SA-III.1.1 introduces the class of *cylindered quasi-dyadic cell expansions* based on $\mathbb{P}_Z$, which can be viewed as a special case of the *quasi-dyadic cell expansions* from Definition SA.5 that leverages the multiplicative structure. This cell expansion is tailored to the multiplicative structure, with the upper layers corresponding to splits in the $\mathbf{x}_i$-direction and the lower layers handling divisions along the y_i-direction.

- Section SA-III.1.2 introduces an alternative to the L_2 projection onto piecewise constant functions on the chosen cells: the *product-factorized projection*, Π_1. This projection exploits the multiplicative structure of G_n, allowing the empirical process to treat $\mathbf{x}_i$ and y_i as independent in layers where cells divide along $\mathcal{Y}$, thereby isolating contributions from $\mathcal{G}$ and $\mathcal{R}$. To analyze the projection errors $\|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})\delta}$ and $\|Z_n^G - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})\delta}$, we also define the L_2 projection onto piecewise constant functions on the chosen cells, Π_0.

- Section SA-III.1.3 constructs the Gaussian process $(Z_n^G(g,r) : (g,r) \in \mathcal{G}\times\mathcal{R})$. These constructions are essentially the same as those in Section SA-II.1.3, relying on coupling binomial random variables with Gaussian random variables.

- Section SA-III.1.4 handles the meshing errors $\|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}}$ and $\|Z_n^G \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^G\|_{\mathcal{G}\times\mathcal{R}}$ using standard empirical process results, which give the contribution $\mathsf{F}(\delta)$ emerging from Talagrand's inequality (Giné and Nickl, 2016, Theorem 3.3.9) combined with a standard maximal inequality (Chernozhukov *et al.*, 2014, Theorem 5.2). This allows us to focus on the error on the δ-net to simply study $\|G_n - Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$.

- Section SA-III.1.5 addresses the strong approximation error $\|\Pi_1 G_n - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$. The multiplicative structure of G_n and the pre-factorization of coefficients in $\Pi_1 G_n$ and $\Pi_1 Z_n^G$ enable a new bound on the strong approximation error for the empirical process indexed by piecewise constant functions. Specifically, we establish a bound on $\mathbb{E}[\|\Pi_1 G_n - \Pi_1 Z_n^G\|^2_{(\mathcal{G}\times\mathcal{R})_\delta}]$ that is polynomial in the number of splits along the y_i-direction and exponential in the number of splits along the $\mathbf{x}_i$-direction. This is a key step in achieving a Gaussian strong approximation rate that treats splits along the y_i-dimension as residual contributions.

- Section SA-III.1.6 addresses the projection errors $\|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and $\|Z_n^G - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$. We begin by comparing the two projections, bounding the differences $\|\Pi_1 G_n - \Pi_0 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and $\|\Pi_1 Z_n^G - \Pi_0 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$. Next, we control the L_2 projection errors $\|G_n - \Pi_0 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and $\|Z_n^G - \Pi_0 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ using Bernstein inequality and similar arguments as in Section SA-II.1.6.

SA-III.1 Preliminary Technical Results

This section presents preliminary technical results that are used to prove Theorem SA.1. Whenever possible, these results are presented at a higher level of generality, and therefore may be of independent theoretical interest. Throughout this section, we employ the following assumption.

Assumption SA.2. *Suppose* $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ *are i.i.d. random vectors taking values in* $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$*, where* $(\mathbf{x}_1, y_1)$ *has joint distribution* $\mathbb{P}_Z$*. Suppose* $\mathbf{x}_1$ *has distribution* $\mathbb{P}_X$ *supported on* $\mathcal{X} \subseteq \mathbb{R}^d$*,* y_1 *has distribution* $\mathbb{P}_Y$ *supported on* $\mathcal{Y} \subseteq \mathbb{R}$*, and the following conditions hold.*

(i) $\mathcal{G}$ *is a real-valued pointwise measurable class of functions on* $(\mathcal{X}, \mathcal{B}(\mathcal{X}), \mathbb{P}_X)$.

(ii) $\mathtt{M}_{\mathcal{G},\mathcal{X}} < \infty$ *and* $J_{\mathcal{X}}(1, \mathcal{G}, \mathtt{M}_{\mathcal{G},\mathcal{X}}) < \infty$.

(iii) $\mathcal{R}$ *is a real-valued pointwise measurable class of functions on* $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}), \mathbb{P}_Y)$.

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

(v) $J_{\mathcal{Y}}(\mathcal{R}, M_{\mathcal{R},\mathcal{Y}}, 1) < \infty$.

Compared to the assumptions in Theorem 2, this assumption does not require the existence of a surrogate measure or a normalizing transformation. It will be applied in the analysis of terms in the error decomposition, where we work with the distribution $\mathbb{P}_Z$, and an extra condition on the existence of Lebesgue density of $\mathbb{P}_X$ is assumed whenever necessary (Section SA-III.1.6). The surrogate measure and the normalizing transformation will be used in the proof of Theorem SA.1 with the help of Section SA-II.2, providing greater flexibility in the data generating process.

SA-III.1.1 Cells Expansions

Definition SA.7 (Cylindered Quasi-Dyadic Expansion of $\mathbb{R}^d$)**.** *Let* $\mathbb{P}$ *denote the joint distribution of* $(\mathbf{X}, Y)$, *a random vector taking values in* $(\mathbb{R}^d \times \mathbb{R}, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}))$, *and let* $\mathbb{P}_X$ *be the marginal distribution of* $\mathbf{X}$. *For a given* $\rho \geq 1$, *a collection of Borel measurable sets in* $\mathbb{R}^{d+1}$, $\mathscr{C}_{M,N}(\mathbb{P}, \rho) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{M+N-j}, 0 \leq j \leq M + N\}$, *is called a cylindered quasi-dyadic expansion of* $\mathbb{R}^{d+1}$ *of depth* M *for the main subspace* $\mathbb{R}^d$ *and depth* N *for the multiplier subspace* $\mathbb{R}$ *with respect to* $\mathbb{P}$ *if the following 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* $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathcal{Y}_{*,N,0}$, *with* $\mathcal{Y}_{*,N,0}$ *a subset of* $\mathbb{R}$ *and* $\mathbb{P}(\mathcal{C}_{M+N,0}) = 1$. *The collection* $\mathscr{C}_M(\mathbb{P}_X, \rho) = \{\mathcal{X}_{l,k} : 0 \leq l \leq M, 0 \leq k < 2^{M-l}\}$ *forms a quasi-dyadic expansion of depth* M *with respect to* $\mathbb{P}_X$.

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

When $\rho = 1$, $\mathscr{C}_{M,N}(\mathbb{P}, 1)$ *is called a* cylindered dyadic expansion. *For notational simplicity, we write* $\mathtt{p}_X[\mathscr{C}_{M,N}(\mathbb{P}, \rho)] = \{\mathcal{X}_{l,k} : 0 \leq l \leq M, 0 \leq k < 2^{M-l}\}$.

Definition SA.8 (Axis-Aligned Quasi-Dyadic Expansion of $\mathbb{R}^d$)**.** *Let* $\mathbb{P}$ *denote the joint distribution of* $(\mathbf{X}, Y)$, *a random vector taking values in* $(\mathbb{R}^d \times \mathbb{R}, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}))$, *and let* $\mathbb{P}_X$ *be the marginal distribution of* $\mathbf{X}$. *For a given* $\rho \geq 1$, *a collection of Borel measurable sets in* $\mathbb{R}^{d+1}$, $\mathcal{A}_{M,N}(\mathbb{P}, \rho) = \{\mathcal{C}_{j,k} : 0 \leq k < 2^{M+N-j}, 0 \leq j \leq M + N\}$, *is called an axis-aligned cylindered quasi-dyadic expansion of* $\mathbb{R}^{d+1}$ *of depth* M *in the main subspace* $\mathbb{R}^d$ *and depth* N *in the multiplier subspace* $\mathbb{R}$ *with respect to* $\mathbb{P}$ *if the following conditions hold:*

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

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

SA-III.1.2 Projection onto Piecewise Constant Functions

Consider a cylindered quasi-dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P},\rho)$ where $\mathbb{P}$ is the joint distribution of a random vector $(\mathbf{X},Y)$ taking values in $(\mathbb{R}^d\times\mathbb{R},\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{B}(\mathbb{R}))$. Define the span of the Haar basis over the terminal cells as described in Section SA-II.1.2, specifically

$$\mathcal{E}_{M+N}=\operatorname{Span}\{\mathbb{1}_{\mathcal{C}_{0,k}}:0\leq k<2^{M+N}\}.$$

For $h\in L_2(\mathbb{P})$, recall that the mean square projection of h onto $\mathcal{E}_{M+N}$ is given by

$$\Pi_0(\mathcal{C}_{M,N}(\mathbb{P},\rho))[h]=\sum_{0\leq k<2^{M+N}}\frac{\mathbb{1}_{\mathcal{C}_{0,k}}}{\mathbb{P}(\mathcal{C}_{0,k})}\int_{\mathcal{C}_{0,k}}h(\mathbf{u})\,d\mathbb{P}(\mathbf{u}).$$

and the β-coefficients are defined by

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

Then we still have

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

where

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

for all $(j,k)\in\mathcal{I}_{M+N}=\{(j,k)\in\mathbb{N}\times\mathbb{N}:1\leq j\leq M+N,0\leq k<2^{M+N-j}\}$. We refer to $\Pi_0(\mathcal{C}_{M,N}(\mathbb{P},\rho))$ as Π_0 for simplicity.

To address the separable structure of $g(\mathbf{X})r(Y)$, we define the *product-factorized projection* from $L_2(\mathbb{P})$ to $\mathcal{E}_{M+N}=\operatorname{Span}\{\mathcal{C}_{0,k}=\mathcal{X}_{0,l}\times\mathcal{Y}_{l,0,m}:0\leq l<2^M,0\leq m<2^N,k=2^Nl+m\}$, defined as

$$\Pi_1(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r]=\gamma_{M+N,0}(g,r)e_{M+N,0}+\sum_{1\leq j\leq M+N}\sum_{0\leq k<2^{M+N-j}}\widetilde{\gamma}_{j,k}(g,r)\widetilde{e}_{j,k},\tag{SA-9}$$

and

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

and $\widetilde{\gamma}_{j,k}(g,r)=\gamma_{j-1,2k}(g,r)-\gamma_{j-1,2k+1}(g,r)$. We refer to $\Pi_1(\mathcal{C}_{M,N}(\mathbb{P},\rho))$ as Π_1 for simplicity.

The Haar basis representation in Equation (SA-9) decomposes the function into layers of increasingly localized fluctuations. However, at lower layers ($1\leq j\leq N$), the local fluctuation is characterized by the *product-factorized projection* $\mathbb{E}[g(\mathbf{X})|\mathbf{X}\in\mathcal{X}_{0,l}]\cdot\mathbb{E}[r(Y)|\mathbf{X}\in\mathcal{X}_{0,l},Y\in\mathcal{Y}_{l,0,m}]$, rather than $\mathbb{E}[g(\mathbf{X})r(Y)|\mathbf{X}\in\mathcal{X}_{0,l}\times\mathcal{Y}_{l,0,m}]$. This distinction makes $\Pi_1(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r]$ generally different from $\Pi_0(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g\cdot r]$.

Now, we define the empirical processes indexed by projected functions. For any real valued functions g

on $\mathbb{R}^d$ and r on $\mathbb{R}$ such that $\int_{\mathbb{R}^d}\int_{\mathbb{R}} g(\mathbf{x})^2 \mathbb{P}(dyd\mathbf{x}) < \infty$ and $\int_{\mathbb{R}^d}\int_{\mathbb{R}} r(y)^2 \mathbb{P}(dyd\mathbf{x}) < \infty$, we define

$$\begin{aligned}\Pi_1 G_n(g,r) &= X_n \circ \Pi_1[\mathcal{C}_{M,N}(\mathbb{P},\rho)](g,r),\\ \Pi_0 G_n(g,r) &= X_n \circ \Pi_0[\mathcal{C}_{M,N}(\mathbb{P},\rho)](gr),\end{aligned} \tag{SA-10}$$

recalling $(X_n(f) : f \in \mathcal{F})$ is the empirical process based on a random sample $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \le i \le n)$ with

$$X_n(f) = n^{-1/2}\sum_{i=1}^{n}(f(\mathbf{x}_i, y_i) - \mathbb{E}[f(\mathbf{x}_i, y_i)]).$$

SA-III.1.3 Strong Approximation Construction

In this section, we construct the Gaussian process Z_n^G (along with some auxiliary Gaussian processes) on a possibly enlarged probability space to couple with the empirical process G_n.

Lemma SA.12. *Suppose Assumption SA.2 holds, and a cylindered quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P}_Z,\rho)$ is given. Then, $(\mathcal{G}\cdot\mathcal{R}) \cup \Pi_0(\mathcal{G}\times\mathcal{R}) \cup \Pi_1(\mathcal{G}\times\mathcal{R})$ is $\mathbb{P}_Z$-pregaussian.*

Proof of Lemma SA.12. By the entropy integral conditions on $\mathcal{G}$ and $\mathcal{R}$ and Definitions 10 and SA.4,

$$\begin{aligned}J_{\mathcal{X}\times\mathcal{Y}}(\mathcal{G}\cdot\mathcal{R}, \mathtt{M}_{\mathcal{G},\mathcal{X}}M_{\mathcal{R},\mathcal{Y}},\delta) &= J_{\mathcal{X}\times\mathcal{Y}}(\mathcal{G}\times\mathcal{R}, \mathtt{M}_{\mathcal{G},\mathcal{X}}M_{\mathcal{R},\mathcal{Y}},\delta)\\ &\le \sqrt{2}J_{\mathcal{X}\times\mathcal{Y}}(\overline{\mathcal{G}}, \mathtt{M}_{\mathcal{G},\mathcal{X}},\delta/\sqrt{2}) + \sqrt{2}J_{\mathcal{X}\times\mathcal{Y}}(\overline{\mathcal{R}}, M_{\mathcal{R},\mathcal{Y}},\delta/\sqrt{2})\\ &\le \sqrt{2}J_{\mathcal{X}}(\mathcal{G}, \mathtt{M}_{\mathcal{G},\mathcal{X}},\delta/\sqrt{2}) + \sqrt{2}J_{\mathcal{Y}}(\mathcal{R}, M_{\mathcal{R},\mathcal{Y}},\delta/\sqrt{2})\end{aligned}$$

where $\overline{\mathcal{G}} = \{(\mathbf{x},y) \in \mathcal{X}\times\mathcal{Y} \mapsto g(\mathbf{x}) : g \in \mathcal{G}\}$ and $\overline{\mathcal{R}} = \{(\mathbf{x},y) \in \mathcal{X}\times\mathcal{Y} \mapsto r(y) : r \in \mathcal{R}\}$.

<u>Claim 1</u>: For all $0 < \delta < 1$,

$$J_{\mathcal{X}\times\mathcal{Y}}(\Pi_0(\mathcal{G}\times\mathcal{R}), c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G},\mathcal{X}}N^\alpha,\delta) \le J_{\mathcal{X}\times\mathcal{Y}}(\mathcal{G}\times\mathcal{R}, \mathtt{M}_{\mathcal{G},\mathcal{X}}M_{\mathcal{R},\mathcal{Y}},\delta),$$

where $c_{\mathtt{v},\alpha} = \mathtt{v}\max\{1 + (2\alpha)^{\frac{\alpha}{2}}, 1 + (4\alpha)^\alpha\}$.
<u>Proof of Claim 1</u>: We consider the two cases of whether $\alpha > 0$ in Assumption SA.2 (iv) separately.

If $\alpha > 0$, by Step 2 in Definition SA.7, $\max_{0\le l<2^{M+N}} \mathbb{E}[\exp(y_i/(N\log 2))|(\mathbf{x}_i, y_i) \in \mathcal{C}_{0,l}] \le 2$. Hence

$$\begin{aligned}\max_{0\le l<2^{M+N}}\sup_{r\in\mathcal{R}} \mathbb{E}[|r(y_i)||(\mathbf{x}_i,y_i)\in\mathcal{C}_{0,l}] &\le \mathtt{v}(1 + \max_{0\le l<2^{M+N}} \mathbb{E}[|y_i|^\alpha|(\mathbf{x}_i,y_i)\in\mathcal{C}_{0,l}])\\ &\le \mathtt{v}(1 + (2N\sqrt{\alpha})^\alpha).\end{aligned} \tag{SA-11}$$

Definition of Π_0 then implies

$$\sup_{g\in\mathcal{G}}\sup_{r\in\mathcal{R}}\sup_{(\mathbf{x},y)\in\mathcal{C}_{M+N,0}} |\Pi_0(gr)(\mathbf{x},y)| \le c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G},\mathcal{X}}N^\alpha. \tag{SA-12}$$

Moreover, if $\alpha = 0$, Assumption SA.2 (iv) implies $\mathtt{M}_{\mathcal{R},\mathcal{Y}} \le 1$, Equations (SA-11), (SA-12) hold with $\alpha = 0$.

Let Q be a finite discrete measure on $\mathcal{X} \times \mathcal{Y}$. Definition of Π_0 and Jensen's inequality implies

$$\begin{aligned}\|\Pi_0 f - \Pi_0 g\|_{Q,2}^2 &\leq \sum_{0 \leq k < 2^{M+N}} Q(C_{0,k})(2^{M+N} \int_{C_{0,k}} f - g d\mathbb{P}_Z)^2 \\ &\leq \sum_{0 \leq k < 2^{M+N}} Q(C_{0,k}) 2^{M+N} \int_{C_{0,k}} (f-g)^2 d\mathbb{P}_Z, \qquad \forall f, g \in \mathcal{G} \cdot \mathcal{R}.\end{aligned}$$

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

$$\|\Pi_0 f - \Pi_0 g\|_{Q,2}^2 \leq \|f - g\|_{\widetilde{Q},2}^2, \qquad \forall f, g \in \mathcal{G} \cdot \mathcal{R}.$$

Lemma SA.15 implies that there exists an $\delta c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha$-net $\mathcal{L}$ of $\mathcal{G} \times \mathcal{R}$ with cardinality no greater than $\mathtt{N}_{\mathcal{G} \times \mathcal{R}, \mathcal{X} \times \mathcal{Y}}(\delta, \|\mathtt{M}_{\mathcal{G},\mathcal{X}} M_{\mathcal{R},\mathcal{Y}}\|_{\widetilde{Q},2})$ such that for all $f \in \Pi_0(\mathcal{G} \times \mathcal{R})$, there exists $g \in \mathcal{L}$ such that

$$\|f - g\|_{\widetilde{Q},2}^2 \leq \delta^2 \|\mathtt{M}_{\mathcal{G},\mathcal{X}} M_{\mathcal{R},\mathcal{Y}}\|_{\widetilde{Q},2}^2 \leq \delta^2 (c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha)^2.$$

The claim then follows.

Claim 2: For all $0 < \delta < 1$,

$$J_{\mathcal{X} \times \mathcal{Y}}(\Pi_1(\mathcal{G} \times \mathcal{R}), c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha, \delta) \lesssim J_{\mathcal{X} \times \mathcal{Y}}(\mathcal{G} \times \mathcal{R}, \mathtt{M}_{\mathcal{G},\mathcal{X}} M_{\mathcal{R},\mathcal{Y}}, \delta/3).$$

Proof of Claim 2: Definition SA.5 and the definition of product factorized projection imply that for the upper layers with $N \leq j \leq M + N$,

$$\gamma_{j,k}(g, r) = \mathbb{E}[g(\mathbf{x}_i) r(y_i) | \mathbf{x}_i \in \mathcal{X}_{j-N,k}] = \mathbb{E}[g(\mathbf{x}_i) r(y_i) | (\mathbf{x}_i, y_i) \in \mathcal{C}_{j-N,k}],$$

that is, the coefficients coincide with those from the mean square projection. Take $\mathcal{C}_{M,0} = \{\mathcal{C}_{j,k} : N \leq j \leq M + N, 0 \leq k < 2^{M+N-j}\}$ to be the collection of all upper layer cells down to the N-th layer, then

$$\Pi_1[\mathcal{C}_{M,0}(\mathbb{P}_Z, \rho)](g, r) = \Pi_0[\mathcal{C}_{M,0}(\mathbb{P}_Z, \rho)](gr), \qquad g \in \mathcal{G}, r \in \mathcal{R}.$$

For the lower layers $0 \leq j < N$, suppose $\widetilde{\mathbb{P}}_Z$ is a mapping from $\mathcal{B}(\mathbb{R}^{d+1})$ to $[0, 1]$ such that

$$\begin{aligned}\widetilde{\mathbb{P}}_Z(E) = \inf \Bigg\{ &\sum_{\ell=1}^{\mathfrak{L}} \sum_{0 \leq l < 2^M} \sum_{0 \leq m < 2^N} \mathbb{E}[\mathbb{1}(\mathbf{x}_i \in A_\ell) | \mathbf{x}_i \in \mathcal{X}_{0,l}] \cdot \mathbb{E}[\mathbb{1}(y_i \in B_\ell) | \mathbf{x}_i \in \mathcal{X}_{0,l}, y_i \in \mathcal{Y}_{l,0,m}] : \\ &E \subseteq \sqcup_{\ell=1}^{\mathfrak{L}} A_\ell \times B_\ell \text{ with } A_\ell \times B_\ell, 1 \leq l \leq \mathfrak{L} \in \mathbb{N}, \text{ disjoint rectangles in } \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}) \Bigg\},\end{aligned}$$

with $E \in \mathcal{B}(\mathbb{R}^{d+1})$. It follows that $\widetilde{\mathbb{P}}_Z$ defines a probability measure on $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$, and

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

where $\mathbb{E}_{\widetilde{\mathbb{P}}_Z}$ means the expectation is taken with $(\mathbf{x}_i, y_i)$ following the law of $\widetilde{\mathbb{P}}_Z$ instead of $\mathbb{P}_Z$. This implies

$$\Pi_1[\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)](g,r) - \Pi_1[\mathcal{C}_{M,0}(\mathbb{P}_Z, \rho)](g,r) = \Pi_0[\mathcal{C}_{M,N}(\widetilde{\mathbb{P}}_Z, \rho)](gr) - \Pi_0[\mathcal{C}_{M,0}(\widetilde{\mathbb{P}}_Z, \rho)](gr), \quad g \in \mathcal{G}, r \in \mathcal{R}.$$

We can then express the Π_1 projection of (g,r) as three L_2 projections as follows:

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

Since $\|\Pi_0[\mathcal{C}_{M,N}(\widetilde{\mathbb{P}}_Z, \rho)]\|_{\mathcal{G} \times \mathcal{R}} \le c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha$, Claim 1 applies to all of the three terms:

$$\begin{aligned}
&J_{\mathcal{X} \times \mathcal{Y}}(\Pi_0[\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)](\mathcal{G} \times \mathcal{R}), c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha, \delta) + J_{\mathcal{X} \times \mathcal{Y}}(\Pi_0[\mathcal{C}_{M,N}(\widetilde{\mathbb{P}}_Z, \rho)](\mathcal{G} \times \mathcal{R}), c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha, \delta) \\
&\quad + J_{\mathcal{X} \times \mathcal{Y}}(\Pi_0[\mathcal{C}_{M,0}(\widetilde{\mathbb{P}}_Z, \rho)](\mathcal{G} \times \mathcal{R}), c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha, \delta) \lesssim J_{\mathcal{X} \times \mathcal{Y}}(\mathcal{G} \times \mathcal{R}, \mathtt{M}_{\mathcal{G},\mathcal{X}} M_{\mathcal{R},\mathcal{Y}}, \delta).
\end{aligned}$$

Then Claim 2 follows from Claim 1.

Putting together,

$$\begin{aligned}
&J_{\mathcal{X} \times \mathcal{Y}}((\mathcal{G} \times \mathcal{R}) \cup \Pi_0(\mathcal{G} \times \mathcal{R}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}), \mathtt{M}_{\mathcal{G},\mathcal{X}} M_{\mathcal{R},\mathcal{Y}} + c_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G},\mathcal{X}} N^\alpha, 1) \\
&\lesssim J_{\mathcal{X}}(\mathcal{G}, \mathtt{M}_{\mathcal{G},\mathcal{X}}, 1) + J_{\mathcal{Y}}(\mathcal{R}, M_{\mathcal{R},\mathcal{Y}}, 1) < \infty,
\end{aligned}$$

and the conclusion follows from separability of $\mathcal{G}$ and $\mathcal{R}$, and van der Vaart and Wellner (2013, Corollary 2.2.9). □

The construction of the Gaussian process essentially follows from the arguments in Section SA-II.1.3 with $\mathbf{z}_i$'s replacing $\mathbf{x}_i$'s. (Recall that $\mathbf{z}_i = (\mathbf{x}_i, y_i)$ in this section.) We start with a Gaussian process indexed by $(\mathcal{G} \cdot \mathcal{R}) \cup \Pi_0(\mathcal{G} \times \mathcal{R}) \cup \Pi_1(\mathcal{G} \times \mathcal{R})$ with almost sure continuous sample paths, 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 result.

Lemma SA.13. *Suppose Assumption SA.2 holds and a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, 1)$ is given. Then on a possibly enlarged probability space, there exists a $\mathbb{P}_Z$-Brownian bridge B_n indexed by $\mathcal{F} = (\mathcal{G} \cdot \mathcal{R}) \cup \Pi_0(\mathcal{G} \times \mathcal{R}) \cup \Pi_1(\mathcal{G} \times \mathcal{R})$ with almost sure continuous trajectories on $(\mathcal{F}, \mathfrak{d}_{\mathbb{P}_Z})$ such that for any $f \in \mathcal{F}$ and any*

$x > 0$,

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

where for both $\|f\|_{\mathcal{E}_{M+N}}^2$ *and* $\mathtt{C}_{\{f\},M+N}$ *are defined in Lemma SA.3.*

Proof of Lemma SA.13. The result follows from Lemma SA.12 and Lemma SA.3 with $(\mathbf{x}_i, y_i)$ replacing $\mathbf{x}_i$. □

Lemma SA.14. *Suppose Assumption SA.2 holds and a cylindered quasi-dyadic expansion* $\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)$ *with* $\rho > 1$ *is given. Then on a possibly enlarged probability space, there exists a* $\mathbb{P}_Z$*-Brownian bridge* B_n *indexed by* $\mathcal{F} = (\mathcal{G} \cdot \mathcal{R}) \cup \Pi_0(\mathcal{G} \times \mathcal{R}) \cup \Pi_1(\mathcal{G} \times \mathcal{R})$ *with almost sure continuous trajectories on* $(\mathcal{F}, \mathfrak{d}_{\mathbb{P}_Z})$ *such that for any* $f \in \mathcal{F}$ *and* $x > 0$,

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

where C_ρ *is a constant that only depends on* ρ.

Proof of Lemma SA.14. Replacing $\mathbf{x}_i$ by $\mathbf{z}_i = (\mathbf{x}_i, y_i)$ in Section SA-II.1.3 (and with the help of the pregaussian lemma SA.12), suppose we constructed as therein on a possibly enlarged probability space the i.i.d standard Gaussian random variables $(\widetilde{\xi}_{j,k} : (j,k) \in \mathcal{I}_{M+N})$ and the Binomial counts $(U_{j,k} : (j,k) \in \mathcal{J}_{M+N}) = (\sum_{i=1}^n e_{j,k}(\mathbf{z}_i) : (j,k) \in \mathcal{J}_{M+N})$. Again, we take $\widetilde{U}_{j,k} = U_{j-1,2k} - U_{j-1,2k+1}$ for $(j,k) \in \mathcal{I}_{M+N}$. By Definition SA.7, the upper layer cells ($N \leq j \leq M+N$) may not be dyadic with respect to $\mathbb{P}_Z$, but the lower layer cells ($0 \leq j < N$) are. Tusnády's Lemma (Bretagnolle and Massart, 1989, Lemma 4) and Lemma SA.4 then imply whenever the event $\mathcal{A}$ holds, with

$$\mathcal{A} = \{|\widetilde{U}_{j,k}| \leq c_{1,\rho} U_{j,k}, \text{ for all } N \leq j \leq M+N, 0 \leq k < 2^{M+N-j}\},$$

we know the following relations hold almost surely in $\mathbb{P}_Z$,

$$\left|\widetilde{U}_{j,k} - \sqrt{U_{j,k}\frac{\mathbb{P}_Z(\mathcal{C}_{j-1,2k})\mathbb{P}_Z(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}_Z(\mathcal{C}_{j,k})^2}}\widetilde{\xi}_{j,k}\right| < c_{2,\rho}\widetilde{\xi}_{j,k}^2 + c_{3,\rho},$$
$$\left|\widetilde{U}_{j,k}\right| \leq 1/c_{0,\rho} + 2\sqrt{\frac{\mathbb{P}_Z(\mathcal{C}_{j-1,2k})\mathbb{P}_Z(\mathcal{C}_{j-1,2k+1})}{\mathbb{P}_Z(\mathcal{C}_{j,k})^2}U_{j,k}}|\widetilde{\xi}_{j,k}|,$$

for all $(j,k) \in \mathcal{I}_{M+N}$, and where $c_{0,\rho}, c_{1,\rho}, c_{2,\rho}, c_{3,\rho}$ are constants that only depends on ρ. By similar argument as in the proof for Lemma SA.5, $\mathbb{P}(\mathcal{A}^c) \leq 3 \cdot 2^M \exp(-\min\{c_{1,\rho}^2/3, 1/8\}\rho^{-1} n 2^{-M})$. The rest of the proof follows from Lemma SA.5 by replacing $\mathbf{x}_i$ with $(\mathbf{x}_i, y_i)$. □

The above two lemmas allow for constructions of Gaussian processes and projected Gaussian processes as counterparts of the empirical processes in Section SA-II.1.3. In particular, we take $Z_n^G, \Pi_0 Z_n^G, \Pi_1 Z_n^G$ to be the empirical processes indexed by $\mathcal{G} \times \mathcal{R}$ such that

$$Z_n^G(g, r) = B_n(g \cdot r), \qquad (g, r) \in \mathcal{G} \times \mathcal{R}. \tag{SA-13}$$

We also define the following ancillary processes for analysis:

$$\Pi_0 Z_n^G(g,r) = B_n(\Pi_0[g \cdot r]), \qquad \Pi_1 Z_n^G(g,r) = B_n(\Pi_1[g,r]), \qquad (g,r) \in \mathcal{G} \times \mathcal{R}. \tag{SA-14}$$

In particular, $(Z_n^G(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ has almost sure continuous trajectories in $(\mathcal{G} \times \mathcal{R}, \mathfrak{d}_{\mathbb{P}_Z})$

The following ancillary lemma for uniform covering number and uniform entropy integrals is used in the proof of Lemma SA.12.

Lemma SA.15 (Covering Number using Covariance Semi-metric)**.** *Assume $\mathcal{F}$ is a class of functions from a measurable space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ to $\mathbb{R}$ with envelope function $M_{\mathcal{F},\mathcal{X}}$. Let P be a probability measure on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. Then, for any $0 < \varepsilon < 1$,*

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

Proof of Lemma SA.15. The proof essentially follows from the arguments for (van der Vaart and Wellner, 2013, Theorem 2.5.2), but we present here for completeness. Define $\mathcal{H} = \{(f-g)^2 : f, g \in \mathcal{F}\} \cup \{M_{\mathcal{F},\mathcal{X}}\}$. Then, for all $0 < \varepsilon < 1$,

$$\sup_Q N(\mathcal{H}, \|\cdot\|_{Q,1}, \varepsilon \|M_{\mathcal{F},\mathcal{X}}^2\|_{Q,1}) \leq \sup_Q N(\mathcal{H}, \|\cdot\|_{Q,1}, \varepsilon \|M_{\mathcal{F},\mathcal{X}}^2\|_{Q,2}) \leq \sup_Q N(\mathcal{F}, \|\cdot\|_{Q,1}, \varepsilon \|M_{\mathcal{F},\mathcal{X}}\|_{Q,1})^2,$$

where the supremums are all taken over finite discrete measures on $\mathcal{X}$. By Theorem 2.4.3 in van der Vaart and Wellner (2013), $\mathcal{H}$ is Glivenko-Cantelli. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with distribution P. Define $Q_N = \frac{1}{N}\sum_{j=1}^N \delta_{X_j}$. 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 $f_1, f_2 \in \mathcal{F}$,

$$\begin{aligned}
&\left| \|f_1 - f_2\|_{P,2}^2 - \|f_1 - f_2\|_{P_N,2}^2 \right| \leq \delta^2 \varepsilon^2 \|M_{\mathcal{F},\mathcal{X}}\|_{P,2}^2, \\
&\left| \|M_{\mathcal{F},\mathcal{X}}\|_{P,2} - \|M_{\mathcal{F},\mathcal{X}}\|_{P_N,2} \right| \leq \delta \|M_{\mathcal{F},\mathcal{X}}\|_{P,2}.
\end{aligned}$$

Since P_N is a finite discrete measure on $\mathcal{X}$, there exists an $\varepsilon\|M_{\mathcal{F},\mathcal{X}}\|_{P_N}$-net, $\mathcal{G}$, of $\mathcal{F}$ with minimal cardinality such that for all $f \in \mathcal{F}$, there exists $f_0 \in \mathcal{G}$ such that $\|f - f_0\|_{P_N,2} \leq \varepsilon \|M_{\mathcal{F},\mathcal{X}}\|_{P_N,2} \leq \varepsilon(\|M_{\mathcal{F},\mathcal{X}}\|_{P,2} + \delta\|M_{\mathcal{F},\mathcal{X}}\|_{P,2}) \leq (1+\delta)\varepsilon\|M_{\mathcal{F},\mathcal{X}}\|_{P,2}$. It follows that for all $f \in \mathcal{F}$, there exists $g \in \mathcal{G}$ such that

$$\|f - g\|_{P,2} \leq \|f - g\|_{P_N,2} + \left| \|f-g\|_{P,2} - \|f-g\|_{P_N,2} \right| \leq (1+2\delta)\varepsilon \|M_{\mathcal{F},\mathcal{X}}\|_{P,2},$$

Hence, $N(\mathcal{F}, \|\cdot\|_{P,2}, \varepsilon\|M_{\mathcal{F},\mathcal{X}}\|_{P,2}) \leq \mathtt{N}_{\mathcal{F},\mathcal{X}}(\varepsilon/(1+2\delta), \mathtt{M}_{\mathcal{F},\mathcal{X}})$. Take $\delta \to 0$ to obtain the desired result. □

SA-III.1.4 Meshing Error

To simplify notation, the parameters of $\mathcal{G}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{X}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{X} \times \mathcal{Y}$, and the index $\mathcal{X} \times \mathcal{Y}$ is omitted where there is no ambiguity. We also define, for $\delta \in (0, 1]$,

$$\mathtt{N}(\delta) = \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}}) \mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}})$$

and

$$J(\delta) = \sqrt{2}J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}).$$

For $0 < \delta \leq 1$, consider a $\delta \mathtt{M}_{\mathcal{G}} \|M_{\mathcal{R}}\|_{\mathbb{P}_Y,2}$-net of $(\mathcal{G} \times \mathcal{R}, \|\cdot\|_{\mathbb{P}_Z,2})$, denoted by $(\mathcal{G} \times \mathcal{R})_\delta$, with cardinality at most $\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta, \mathtt{M}_{\mathcal{G}}\|M_{\mathcal{R}}\|_{\mathbb{P}_Y,2})$. Define the projection onto the δ-net as a mapping $\pi_{(\mathcal{G}\times\mathcal{R})_\delta} : \mathcal{G} \times \mathcal{R} \to \mathcal{G} \times \mathcal{R}$ such that $\|\pi_{(\mathcal{G}\times\mathcal{R})_\delta}(g, r) - gr\|_{\mathbb{P}_Z,2} \leq \delta \mathtt{M}_{\mathcal{G}}\|M_{\mathcal{R}}\|_{\mathbb{P}_Y,2}$ for all $g \in \mathcal{G}$ and $r \in \mathcal{R}$.

Lemma SA.16. *Suppose Assumption SA.2 holds, a cylindered quasi-dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)$ is given, $(Z_n^G(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ is the Gaussian process constructed as in Equation (SA-13) on a possibly enlarged probability space, and $(\mathcal{G} \times \mathcal{R})_\delta$ is chosen in Section SA-III.1.4. For all $t > 0$ and $0 < \delta < 1$,*

$$\mathbb{P}\big[\|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^G \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^G\|_{\mathcal{G}\times\mathcal{R}} > C_1 c_{\mathtt{v},\alpha} \mathsf{F}_n^G(t, \delta)\big] \leq 8\exp(-t),$$

where $c_{\mathtt{v},\alpha} = \mathtt{v}(1 + (2\alpha)^{\frac{\alpha}{2}})$ and

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

Proof of Lemma SA.16. By standard empirical process arguments, we can show for any $0 < \delta < 1$, $\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta, \mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}) \leq \mathtt{N}(\delta)$ and $J(\delta, \mathcal{G} \times \mathcal{R}, \mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}) \leq J(\delta)$. By definition of $\pi_{(\mathcal{G}\times\mathcal{R})_\delta}$, $\|\pi_{(\mathcal{G}\times\mathcal{R})_\delta} h - h\|_{\mathbb{P}_Z,2} \leq \delta\|\mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}\|_{\mathbb{P}_Z,2} = \delta \mathtt{M}_{\mathcal{G}}\|M_{\mathcal{R}}\|_{\mathbb{P}_Y,2}$. Take $\mathcal{L} = \{h - \pi_{(\mathcal{G}\times\mathcal{R})_\delta} h : h \in \mathcal{G} \times \mathcal{R}\}$. Then, by Theorem 5.2 in Chernozhukov *et al.* (2014),

$$\begin{aligned}
\mathbb{E}[\|X_n\|_{\mathcal{L}}] &\lesssim J(\delta)\mathtt{M}_{\mathcal{G}}\|M_{\mathcal{R}}(y_i)\|_2 + \frac{\mathtt{M}_{\mathcal{G}}\|\max_{1\leq i\leq n} M_{\mathcal{R}}(y_i)\|_2 J^2(\delta)}{\delta^2\sqrt{n}} \\
&\lesssim c_{\mathtt{v},\alpha} J(\delta)\mathtt{M}_{\mathcal{G}} + c_{\mathtt{v},\alpha}(\log n)^{\alpha/2}\frac{\mathtt{M}_{\mathcal{G}} J^2(\delta)}{\delta^2\sqrt{n}}.
\end{aligned}$$

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

$$\|X_n\|_{\mathcal{L}} \lesssim c_{\mathtt{v},\alpha} J(\delta)\mathtt{M}_{\mathcal{G}} + c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}} J^2(\delta)}{\delta^2\sqrt{n}} + c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}t + c_{\mathtt{v},\alpha}(\log n)^\alpha \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}t^\alpha.$$

In particular, $\|X_n\|_{\mathcal{L}} = \|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}}$. The bound for $\|Z_n^G - Z_n^G \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|$ follows from a standard concentration inequality for Gaussian suprema. □

SA-III.1.5 Strong Approximation Errors

To simplify notation, the parameters of $\mathcal{G}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{X}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{X} \times \mathcal{Y}$, and the index $\mathcal{X} \times \mathcal{Y}$ is omitted where there is no ambiguity. Recall we also define, for $\delta \in (0, 1]$,

$$\mathtt{N}(\delta) = \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}})\mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}})$$

and

$$J(\delta) = \sqrt{2}J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}).$$

Lemma SA.17. *Suppose Assumption SA.2 holds, a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, 1)$ is given, $(\Pi_1 Z_n^G(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ is the Gaussian process constructed as in Equation (SA-14) on a possibly enlarged probability space, and $(\mathcal{G} \times \mathcal{R})_\delta$ is chosen in Section SA-III.1.4. Then for all $t > 0$,*

$$\mathbb{P}\Big[\|\Pi_1 G_n - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1 c_{\mathtt{v},\alpha}\sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}t} + C_1 c_{\mathtt{v},\alpha}\sqrt{\frac{\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R})_\delta, M+N}}{n}t}\Big] \leq 2\mathtt{N}(\delta)e^{-t},$$

where $C_1 > 0$ is a universal constant.

Proof of Lemma SA.17. To simplify notation, we will use $\mathbb{E}[\cdot|\mathcal{X}_{0,l}]$ in short for $\mathbb{E}[\cdot|\mathbf{x}_i \in \mathcal{X}_{0,l}]$, and $\mathbb{E}[\cdot|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in short for $\mathbb{E}[\cdot|(\mathbf{x}_i, y_i) \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$.

Layers $N + 1 \leq j \leq M + N$: For these layers, $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathcal{Y}_{*,N,0}$. By definition of $\widetilde{\gamma}_{j,k}$,

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

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

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

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,

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

<u>*Case 1*</u>: $\alpha > 0$ in (iv) from Assumption SA.2. Then $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$, and Markov's inequality implies $\min\{|y| : y \in \mathcal{Y}_{l,0,0})\} \leq \log(\mathbb{E}[\exp(|y_i|)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,0}]) \leq \log(2 \cdot 2^N) \leq 2N$, and similarly $\min\{|y| : y \in \mathcal{Y}_{l,0,2^N-1}\} \leq 2N$. Hence the middle cells satisfy $\mathcal{Y}_{l,j,m} \subseteq [-2N, 2N]$ for all $0 \leq j < N$, $1 \leq m \leq 2^{N-j} - 2$, and

$$\sum_{m=1}^{2^{N-j}-2} \big|\mathbb{E}[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2m}] - \mathbb{E}[r(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j-1,2m+1}]\big| \leq \mathtt{pTV}_{r|_{[-2N,2N]}} \leq c_{\mathtt{v},\alpha}N^\alpha,$$

and for the left-most cells,

$$\begin{aligned}&\left|\mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j-1,1}\right]-\mathbb{E}\left[r(y_i)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j-1,0}\right]\right|\\&\quad\leq\max_{0\leq m<2^{N-j+1}}\mathbb{E}[r(y_i)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j-1,m}]-\min_{0\leq m<2^{N-j+1}}\mathbb{E}[r(y_i)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j-1,m}]\leq 2c_{\mathtt{v},\alpha}N^{\alpha},\end{aligned}$$

and similarly for the right-most cells,

$$\left|\mathbb{E}[r(y_i)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j-1,2^{N-j}-1}]-\mathbb{E}[r(y_i)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j-1,2^{N-j}-2}]\right|\leq 2c_{\mathtt{v},\alpha}N^{\alpha}.$$

Case 2: $\alpha=0$ in (iv) from Assumption SA.2, since $\mathtt{pTV}_{\{r\}}\leq 2\mathtt{v}$ and $\mathtt{M}_{\{r\}}\leq 2\mathtt{v}$ for all $r\in\mathcal{R}$, 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,

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

Fix j and sum the above across l,

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

We can now sum across j to get

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

By Equation (SA-11), $\sup_{g\in\mathcal{G},r\in\mathcal{R}}\max_{(j,k)\in\mathcal{I}_{M+N}}|\widetilde{\gamma}_{j,k}(g,r)|\leq 2c_{\mathtt{v},\alpha}N^{\alpha}\mathtt{M}_{\mathcal{G}}$, and hence

$$\sum_{1\leq j\leq N}\sum_{0\leq k<2^{M+N-j}}|\widetilde{\gamma}_{j,k}(g,r)|^2\leq 4c_{\mathtt{v},\alpha}^2N^{2\alpha+1}2^M\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}},\qquad g\in\mathcal{G},r\in\mathcal{R}.$$

Putting Together: Putting together the previous two parts,

$$\sum_{j=1}^{M+N}\sum_{k=0}^{2^{M+N-j}}\widetilde{\gamma}_{j,k}^2(g,r)\leq 6c_{\mathtt{v},\alpha}^2N^{2\alpha+1}2^M\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}},\qquad g\in\mathcal{G},r\in\mathcal{R}.$$

By Lemma SA.13, we know for any $(g,r)\in\mathcal{G}\times\mathcal{R}$, for any $x>0$, with probability at least $1-2\exp(-x)$,

$$|G_n\circ\Pi_1(g,r)-\Pi_1Z_n^G(g,r)|\lesssim c_{\mathtt{v},\alpha}\sqrt{\frac{N^{2\alpha+1}2^M\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}x}+c_{\mathtt{v},\alpha}\sqrt{\frac{\mathtt{C}_{\Pi_1\{(g,r)\},M+N}}{n}x},$$

and the proof is complete. □

Lemma SA.18. *Suppose Assumption SA.2 holds, a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z,\rho)$ is given with*

$\rho > 1$, $(\Pi_1 Z_n^G(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ *is the Gaussian process constructed as in Equation* (SA-14) *on a possibly enlarged probability space, and* $(\mathcal{G}\times\mathcal{R})_\delta$ *is chosen in Section SA-III.1.4. Then for all* $t > 0$*,*

$$\mathbb{P}\Big[\|\Pi_1 G_n - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1 C_\rho c_{\mathtt{v},\alpha}\sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}t} + C_1 C_\rho c_{\mathtt{v},\alpha}\sqrt{\frac{\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R})_\delta, M+N}}{n}t}\Big]$$
$$\leq 2\mathtt{N}(\delta)e^{-t} + 2^M \exp\left(-C_\rho n 2^{-M}\right),$$

where $C_1 > 0$ *is a universal constant,* $c_{\mathtt{v},\alpha} = \mathtt{v}(1+(2\alpha)^{\alpha/2})$ *and* C_ρ *is a constant that only depends on* ρ*.*

Proof of Lemma SA.18. Since $\mathcal{C}_{M,N}$ is a cylindered quasi-dyadic expansion, $\rho^{-1}2^{-M-N+j} \leq \mathbb{P}_Z(\mathcal{C}_{j,k}) \leq \rho 2^{-M-N+j}$, for all $0 \leq j \leq M+N$, $0 \leq k < 2^{M+N-j}$. The same argument for Lemma SA.17 implies

$$\sum_{j=1}^{M+N}\sum_{k=0}^{2^{M+N-j}} \widetilde{\gamma}_{j,k}^2(g,r) \leq c_\rho c_{\mathtt{v},\alpha}^2 N^{2\alpha+1}2^M \mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}, \qquad g \in \mathcal{G}, r \in \mathcal{R},$$

where c_ρ is a constant that only depends on ρ. The result then follows from Lemma SA.14. □

SA-III.1.6 Projection Error

To simplify notation, the parameters of $\mathcal{G}$ (Definitions 4 to 12, SA.1, SA.2) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{X}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G}\times\mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{X}\times\mathcal{Y}$, and the index $\mathcal{X}\times\mathcal{Y}$ is omitted where there is no ambiguity. Recall we also define, for $\delta \in (0,1]$,

$$\mathtt{N}(\delta) = \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}})\mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}})$$

and

$$J(\delta) = \sqrt{2}J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}).$$

To analyze the projection error, we employ the decomposition

$$\Pi_1 G_n(g,r) - G_n(g,r) = \big(\Pi_0 G_n(g,r) - G_n(g,r)\big) + \big(\Pi_1 G_n(g,r) - \Pi_0 G_n(g,r)\big),$$

where $\Pi_0 G_n(g,r) - G_n(g,r)$ represents the L_2 projection error, and $\Pi_1 G_n(g,r) - \Pi_0 G_n(g,r)$ denotes the mis-specification error. Specifically, the L_2 projection error captures the minimum loss incurred by projecting onto the class of piecewise constant functions over the cells $\mathcal{E}_{M+N}$. In contrast, the mis-specification error reflects the additional loss introduced when shifting from the L_2 projection to the product-factorized projection.

First we bound the mis-specification error.

Lemma SA.19. *Suppose Assumption SA.2 holds, a cylindered dyadic expansion* $\mathcal{C}_{M,N}(\mathbb{P}_Z, 1)$ *is given,* $(\Pi_0 Z_n^G(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ *and* $(\Pi_1 Z_n^G(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ *are the Gaussian processes constructed as in Equation* (SA-14) *on a possibly enlarged probability space, and* $(\mathcal{G}\times\mathcal{R})_\delta$ *is chosen in Section SA-III.1.4. Suppose* $\mathbb{P}_X$ *admits a Lebesgue density* f_X *supported on* $\mathcal{X} \subseteq \mathbb{R}^d$*. Let* $\tau > 0$*. Define* $r_\tau = r\mathbb{1}([-\tau^{\frac{1}{\alpha}}, \tau^{\frac{1}{\alpha}}])$*.*

Then, for any $g \in \mathcal{G}, r \in \mathcal{R}$,

$$\mathbb{E}\left[\left(\Pi_1 G_n(g, r_\tau) - \Pi_0 G_n(g, r_\tau)\right)^2\right] = \mathbb{E}\left[\left(\Pi_1 Z_n^G(g, r_\tau) - \Pi_0 Z_n^G(g, r_\tau)\right)^2\right] \leq 4\mathtt{v}^2(1+\tau)^2 N^2 \mathtt{V}_{\mathcal{G}},$$

where

$$\mathtt{V}_{\mathcal{G}} = \min\{2\mathtt{M}_{\mathcal{G}}, \mathtt{L}_{\mathcal{G}}\|\mathcal{V}_M\|_\infty\}\Big(\sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\Big)^2 2^M \mathfrak{m}(\mathcal{V}_M)\|\mathcal{V}_M\|_\infty \mathtt{TV}^*_{\mathcal{G}},$$

and, as in Section SA-II.1.6, $\mathcal{V}_M = \cup_{0 \leq l < 2^M}(\mathcal{X}_{0,l} - \mathcal{X}_{0,l})$ is the upper level quasi-dyadic variation set.

Proof of Lemma SA.19. To simplify notation, we will use $\mathbb{E}[\cdot|\mathcal{X}_{0,l}]$ in short for $\mathbb{E}[\cdot|\mathbf{x}_i \in \mathcal{X}_{0,l}]$, and $\mathbb{E}[\cdot|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in short for $\mathbb{E}[\cdot|(\mathbf{x}_i, y_i) \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in this proof.

Expanding $\Pi_1 G_n(g, r_\tau) - \Pi_0 G_n(g, r_\tau)$ by Haar basis representation,

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

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

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

Recall in Definition SA.7, $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,l} \times \mathcal{Y}_{l,j,m}$, where $k = 2^{N-j}l + m$, $0 \leq l < 2^M$ and $0 \leq m < 2^{N-j}$. Definitions of $\gamma_{j,k}$ and $\beta_{j,k}$ from Section SA-III.1.2 imply

$$\begin{aligned}
|\gamma_{j,k}(g, r_\tau) - \beta_{j,k}(g, r_\tau)| &= |\mathbb{E}[g(\mathbf{x}_i)|\mathcal{X}_{0,l}] \cdot \mathbb{E}[r_\tau(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}] - \mathbb{E}[g(\mathbf{x}_i) r_\tau(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]| \\
&= |\mathbb{E}[(g(\mathbf{x}_i) - \mathbb{E}[g(\mathbf{x}_i)|\mathcal{X}_{0,l}]) \, r_\tau(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]| \\
&\leq \mathtt{v}(1+\tau)\mathbb{E}[|g(\mathbf{x}_i) - \mathbb{E}[g(\mathbf{x}_i)|\mathcal{X}_{0,l}]||\mathcal{C}_{j,k}],
\end{aligned}$$

where the first line is simply the definitions of $\gamma_{j,k}$ and $\beta_{j,k}$; the second line is because $\boldsymbol{\sigma}(\mathbb{1}(\mathbf{x}_i \in \mathcal{X}_{0,l})) \subseteq \boldsymbol{\sigma}(\mathbb{1}((\mathbf{x}_i, y_i) \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}))$; and the third line is because Assumption SA.2 (iv) implies $\sup_{y \in \mathbb{R}} |r_\tau(y)| \leq \mathtt{v}(1+\tau)$ for all $r \in \mathcal{R}$. Summing across j and k, then by similar argument as in the proof of Lemma SA.9,

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

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

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

This shows the results. □

Next we bound the L_2 projection error.

Lemma SA.20. *Suppose Assumption SA.2 holds, a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z,1)$ is given, $(Z_n^G(g,r):g\in\mathcal{G},r\in\mathcal{R})$ and $(\Pi_0 Z_n^G(g,r):g\in\mathcal{G},r\in\mathcal{R})$ are the Gaussian processes constructed as in Equations (SA-13) and (SA-14) on a possibly enlarged probability space, and $(\mathcal{G}\times\mathcal{R})_\delta$ is chosen in Section SA-III.1.4. Suppose $\mathbb{P}_X$ admits a Lebesgue density f_X supported on $\mathcal{X}\subseteq\mathbb{R}^d$. Let $\tau>0$. Define $r_\tau=r\mathbb{1}([-\tau^{\frac{1}{\alpha}},\tau^{\frac{1}{\alpha}}])$. Then for any $g\in\mathcal{G},r\in\mathcal{R}$,*

$$\mathbb{E}\left[\left(\Pi_0 Z_n^G(g,r_\tau)-Z_n^G(g,r_\tau)\right)^2\right]=\mathbb{E}\left[\left(\Pi_0 G_n(g,r_\tau)-G_n(g,r_\tau)\right)^2\right]\le 4\mathtt{v}^2(1+\tau)^2\left(2^{-N}\mathtt{M}_{\mathcal{G}}^2+\mathtt{V}_{\mathcal{G}}\right),$$

where $\mathtt{V}_{\mathcal{G}}$ is defined in Lemma SA.19.

Proof of Lemma SA.20. To simplify notation, we will use $\mathbb{E}[\cdot|\mathcal{X}_{0,l}]$ in short for $\mathbb{E}[\cdot|\mathbf{x}_i\in\mathcal{X}_{0,l}]$, and $\mathbb{E}[\cdot|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]$ in short for $\mathbb{E}[\cdot|(\mathbf{x}_i,y_i)\in\mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}]$ in this proof.

Let $\mathcal{B}=\boldsymbol{\sigma}(\{\mathbb{1}((\mathbf{x}_i,y_i)\in\mathcal{C}_{0,k}):0\le k<2^{M+N}\})$ be the σ-algebra generated by $\{\mathbb{1}((\mathbf{x}_i,y_i)\in\mathcal{C}_{0,k}):0\le k<2^{M+N}\}$. Then the difference between the L_2 projection and the original can be expressed as

$$\begin{aligned}
\Pi_0(g\cdot r_\tau)(\mathbf{x}_i,y_i)-g(\mathbf{x}_i)r_\tau(y_i)&=\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]-g(\mathbf{x}_i)r_\tau(y_i)\\
&=\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]-\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i)+\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i)-g(\mathbf{x}_i)r_\tau(y_i).
\end{aligned}$$

By Definition SA.7, each cell $\mathcal{C}_{0,k}$ is of the form of a product, that is,

$$\mathcal{C}_{0,k}=\mathcal{X}_{0,l}\times\mathcal{Y}_{l,0,m}\text{ with }k=2^N l+m,$$

where $0\le k<2^{M+N}$, $0\le l<2^M$ and $0\le m<2^N$.

The first two terms $\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]-\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i)$ in the decomposition are driven by projection of r_τ on grids $\mathcal{Y}_{l,0,m}$'s, and can be upper bounded through probability measure assigned to each grid (2^{-N}) and total variation of r_τ. We consider the positive and negative parts separately: Consider the function

$$q_{l,m}^+(y)=\mathbb{E}[g(\mathbf{x}_i)\mathbb{1}(g(\mathbf{x}_i)\ge 0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,0,m}]r_\tau(y)-\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)\mathbb{1}(g(\mathbf{x}_i)\ge 0)|\mathcal{X}_{0,l}\times\mathcal{Y}_{l,0,m}],\quad y\in\mathcal{Y}_{l,0,m}.$$

Either $q^+_{l,m}$ is constantly zero on $\mathcal{Y}_{l,0,m}$ or $q^+_{l,m}$ takes both positive and negative values on $\mathcal{Y}_{l,0,m}$. Under either case, we have $|q^+_{l,m}(y)| \le \mathtt{pTV}_{\{q^+_{l,m}\},\mathcal{Y}_{l,0,m}}$ for all $y \in \mathcal{Y}_{l,0,m}$. Hence

$$\begin{aligned}
\mathbb{E}[|q^+_{l,m}(y_i)|\mathbb{1}(y_i \in \mathcal{Y}_{l,0,m})|\mathbf{x}_i = \mathbf{x}] &= \int_{\mathcal{Y}_{l,0,m}} |q^+_{l,m}(y)| d\mathbb{P}(y_i \le y|\mathbf{x}_i = \mathbf{x}) \\
&\le \mathbb{P}(y_i \in \mathcal{Y}_{l,0,m}|\mathbf{x}_i = \mathbf{x})\mathtt{pTV}_{\{q^+_{l,m}\},\mathcal{Y}_{l,0,m}} \\
&\le \mathbb{P}(y_i \in \mathcal{Y}_{l,0,m}|\mathbf{x}_i = \mathbf{x})\mathtt{M}_{\{g\}}\mathtt{pTV}_{\{r_\tau\},\mathcal{Y}_{l,0,m}}, \qquad \mathbf{x} \in \mathcal{X}_{0,l}.
\end{aligned}$$

Similarly,

$$q^-_{l,m}(y) = \mathbb{E}[g(\mathbf{x}_i)\mathbb{1}(g(\mathbf{x}_i) < 0)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}]r_\tau(y) - \mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)\mathbb{1}(g(\mathbf{x}_i) < 0)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}], \quad y \in \mathcal{Y}_{l,0,m},$$

and we have

$$\mathbb{E}[|q^-_{l,m}(y_i)|\mathbb{1}(y_i \in \mathcal{Y}_{l,0,m})|\mathbf{x}_i = \mathbf{x}] \le \mathbb{P}(y_i \in \mathcal{Y}_{l,0,m}|\mathbf{x}_i = \mathbf{x})\mathtt{M}_{\{g\}}\mathtt{pTV}_{\{r_\tau\},\mathcal{Y}_{l,0,m}}, \qquad \mathbf{x} \in \mathcal{X}_{0,l}.$$

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

$$\begin{aligned}
&\mathbb{E}\Big[\big|\mathbb{E}[g(\mathbf{x}_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}]r_\tau(y_i) - \mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}]\big|\mathbb{1}(y_i \in \mathcal{Y}_{l,0,m})\Big|\mathbf{x}_i \in \mathcal{X}_{0,l}\Big] \\
&\quad \le 2\mathbb{P}(y_i \in \mathcal{Y}_{l,0,m}|\mathbf{x}_i \in \mathcal{X}_{0,l})\mathtt{M}_{\{g\}}\mathtt{pTV}_{\{r_\tau\},\mathcal{Y}_{l,0,m}} \le 2 \cdot 2^{-N}\mathtt{M}_{\{g\}}\mathtt{pTV}_{\{r_\tau\},\mathcal{Y}_{l,0,m}}.
\end{aligned}$$

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

$$\mathbb{E}\big[|\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i) - \mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]|\big|\mathbf{x}_i \in \mathcal{X}_{0,l}\big] \le 2 \cdot 2^{-N}\mathtt{M}_{\{g\}}\mathtt{pTV}_{\{r_\tau\},\mathcal{Y}_{*,N,0}}.$$

Hence, using the polynomial growth of total variation,

$$\mathbb{E}\big[|\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i) - \mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}]|\big] \le 2^{-N}\mathtt{M}_{\{g\}}\mathtt{pTV}_{\{r_\tau\},\mathcal{Y}_{*,N,0}} \le 2 \cdot 2^{-N}\mathtt{M}_{\mathcal{G}}\mathtt{v}(1+\tau).$$

Since $\big|\mathbb{E}[g(\mathbf{x}_i)r_\tau(y_i)|\mathcal{B}] - \mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i)\big| \le 2\mathtt{M}_{\mathcal{G}}\mathtt{v}(1+\tau)$ almost surely,

$$\mathbb{E}\big[(\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r(y_i) - \mathbb{E}[g(\mathbf{x}_i)r(y_i)|\mathcal{B}])^2\big] \le 4 \cdot 2^{-N}\mathtt{v}^2(1+\tau)^2\mathtt{M}_{\mathcal{G}}^2.$$

Now we look at the last two terms $\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i) - g(\mathbf{x}_i)r_\tau(y_i)$, which are essentially driven by the L_2-projection error of g. Denote by $\mathcal{A} = \boldsymbol{\sigma}(\{\mathbb{1}(\mathbf{x}_i \in \mathcal{X}_{0,l}) : 0 \le l < 2^M\})$ the σ-algebra generated by $\{\mathbb{1}(\mathbf{x}_i \in \mathcal{X}_{0,l}) : 0 \le l < 2^M\}$. Then $\mathcal{A} \subseteq \mathcal{B}$. By Jensen's inequality and a similar argument as in the proof of Lemma SA.9,

$$\mathbb{E}\big[(\mathbb{E}[g(\mathbf{x}_i)|\mathcal{B}]r_\tau(y_i) - g(\mathbf{x}_i)r_\tau(y_i))^2\big] \le 4\mathtt{v}^2(1+\tau)^2\mathbb{E}\big[(g(\mathbf{x}_i) - \mathbb{E}[g(\mathbf{x}_i)|\mathcal{A}])^2\big] \le 4\mathtt{v}^2(1+\tau)^2\mathtt{V}_{\mathcal{G}}.$$

It then follows that $\mathbb{E}[(\Pi_0 G_n(g,r_\tau) - G_n(g,r_\tau))^2] \le 4\mathtt{v}^2(1+\tau)^2\left(2^{-N}\mathtt{M}_{\mathcal{G}}^2 + \mathtt{V}_{\mathcal{G}}\right)$. □

Using a truncation argument and the previous two lemmas, we get the bound on Π_1-projection error with tail control.

Lemma SA.21. *Suppose Assumption SA.2 holds, a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, 1)$ is given,*

$(Z_n^G(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ *and* $(\Pi_1 Z_n^G(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ *are the Gaussian processes constructed as in Equations* (SA-13) *and* (SA-14) *on a possibly enlarged probability space, and* $(\mathcal{G} \times \mathcal{R})_\delta$ *is chosen in Section SA-III.1.4. Suppose* $\mathbb{P}_X$ *admits a Lebesgue density* f_X *supported on* $\mathcal{X} \subseteq \mathbb{R}^d$*. Then for all* $t > N$,

$$\mathbb{P}\Big[\|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1\sqrt{c_{\mathtt{v},2\alpha}}\sqrt{N^2 \mathtt{V}_{\mathcal{G}} + 2^{-N}\mathtt{M}_{\mathcal{G}}^2}\, t^{\alpha+\frac{1}{2}} + C_1 c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1}\Big] \leq 4\mathtt{N}(\delta) n e^{-t},$$

$$\mathbb{P}\Big[\|Z_n^G - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1\sqrt{c_{\mathtt{v},2\alpha}}\sqrt{N^2 \mathtt{V}_{\mathcal{G}} + 2^{-N}\mathtt{M}_{\mathcal{G}}^2}\, t^{\frac{1}{2}} + C_1 c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t\Big] \leq 4\mathtt{N}(\delta) n e^{-t},$$

where $c_{\mathtt{v},\alpha} = \mathtt{v}(1 + (2\alpha)^{\frac{\alpha}{2}})$, $c_{\mathtt{v},2\alpha} = \mathtt{v}^2(1 + (4\alpha)^\alpha)$*, and* C_1 *is a universal constant.*

Proof of Lemma SA.21. To simplify notation, we will use $\mathbb{E}[\cdot|\mathcal{X}_{0,l}]$ in short for $\mathbb{E}[\cdot|\mathbf{x}_i \in \mathcal{X}_{0,l}]$, and $\mathbb{E}[\cdot|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in short for $\mathbb{E}[\cdot|(\mathbf{x}_i, y_i) \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in this proof. We will use a truncation argument and consider the cases of whether $\alpha > 0$ in (iv) of Assumption SA.2 separately.

First, suppose $\alpha > 0$ in (iv) of Assumption SA.2. Let $\tau > 0$ such that $\tau^{\frac{1}{\alpha}} > \log(2^{N+1})$.

Projection error for truncated processes: By Lemmas SA.19 and SA.20, and using Bernstein inequality, for all $t > 0$, for each $g \in \mathcal{G}$, $r \in \mathcal{R}$,

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

Truncation Error: Recall Equation (SA-11) implies $\max_{0 \leq k < 2^{M+N}} \mathbb{E}[|r(y_i)||(\mathbf{x}_i, \mathbf{y}_i) \in \mathcal{C}_{0,k}] \leq c_{\mathtt{v},\alpha} N^\alpha$. The same argument implies $\max_{0 \leq k < 2^{M+N}} \mathbb{E}[r(y_i)^2|(\mathbf{x}_i, y_i) \in \mathcal{C}_{0,k}] \leq \mathtt{v}^2(1 + (N\log(2)\sqrt{2\alpha})^{2\alpha}) \leq c_{\mathtt{v},2\alpha} N^{2\alpha}$. Hence the following holds almost surely,

$$\begin{aligned}
|\Pi_1 G_n(g,r) - \Pi_1 G_n(g, r_\tau)| &\leq \max_{0 \leq l < 2^M} \max_{0 \leq m < 2^N} \Big|\mathbb{E}[g(\mathbf{x}_i)|\mathcal{X}_{0,l}]\mathbb{E}\big[|r(y_i)|\mathbb{1}(|y_i| \geq \tau^{1/\alpha})\big|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,0,m}\big]\Big| \\
&\leq c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G}} N^\alpha.
\end{aligned}$$

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

$$\begin{aligned}
\mathbb{E}\Big[\big|\Pi_1 G_n(g,r) - \Pi_1 G_n(g, r_\tau)\big|^2\Big] &\leq \mathbb{P}\big(\Pi_1 G_n(g,r) - \Pi_1 G_n(g,r_\tau) \neq 0\big) c_{\mathtt{v},2\alpha}\mathtt{M}_{\mathcal{G}}^2 N^{2\alpha} \\
&\leq c_{\mathtt{v},2\alpha} 2^{-N+1}\mathtt{M}_{\mathcal{G}}^2 N^{2\alpha}.
\end{aligned}$$

Using Bernstein's inequality, for all $t > 0$, with probability at least $1 - 2\exp(-t)$,

$$\begin{aligned}
|\Pi_1 G_n(g,r) - \Pi_1 G_n(g,r_\tau)| &\lesssim \sqrt{c_{\mathtt{v},2\alpha}} 2^{-N/2}\mathtt{M}_{\mathcal{G}} N^\alpha \sqrt{t} + c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}} N^\alpha}{\sqrt{n}} t \\
&\lesssim \sqrt{c_{\mathtt{v},2\alpha}} 2^{-N/2}\mathtt{M}_{\mathcal{G}} N^\alpha \sqrt{t} + c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}}\tau}{\sqrt{n}} t.
\end{aligned}$$

Moreover, using $\mathbb{P}(|y_i| \geq \tau) \leq 2 \cdot 2^{-N}$, we have

$$\begin{aligned}
\mathbb{E}[(G_n(g,r) - G_n(g,r_\tau))^2] &\leq \mathtt{M}_{\mathcal{G}}^2\mathbb{E}[(r(y_i) - r_\tau(y_i))^2] \leq \mathtt{M}_{\mathcal{G}}^2\mathbb{E}[r(y_i)^2\mathbb{1}(|y_i| \geq \tau)] \\
&\leq 2 \cdot 2^{-N}\mathtt{M}_{\mathcal{G}}^2 \max_{0 \leq k < 2^{M+N}} \mathbb{E}[r(y_i)^2|(\mathbf{x}_i, y_i) \in \mathcal{C}_{0,k}] \leq 2c_{\mathtt{v},2\alpha}\mathtt{M}_{\mathcal{G}}^2 N^{2\alpha} 2^{-N}.
\end{aligned}$$

By Bernstein inequality and a truncation argument, for all $t > 0$,

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

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

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

Putting Together: Taking $\tau = t^\alpha > 0.5^\alpha N^\alpha$, we get from the previous bounds on $G_n(g,r_\tau) - \Pi_1 G_n(g,r_\tau)$, $\Pi_1 G_n(g,r) - \Pi_1 G_n(g,r_\tau)$, and $G_n(g,r) - G_n(g,r_\tau)$ that for all $g \in \mathcal{G}$, $r \in \mathcal{G}$, for all $t > N$, with probability at least $1 - 4n\exp(-t)$,

$$|\Pi_1 G_n(g,r) - G_n(g,r)| \lesssim \sqrt{c_{\mathtt{v},2\alpha}} \sqrt{N^2 \mathtt{V}_\mathcal{G} + 2^{-N} \mathtt{M}_\mathcal{G}^2} t^{\alpha+\frac{1}{2}} + c_{\mathtt{v},\alpha} \frac{\mathtt{M}_\mathcal{G}}{\sqrt{n}} t^{\alpha+1}. \tag{SA-15}$$

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

Next, suppose $\alpha = 0$ in (iv) of Assumption SA.2. This implies $\mathtt{M}_\mathcal{R} \leq 2\mathtt{v}$. Hence choosing $\tau = 2\mathtt{v}$, then $G_n(g,r) = G_n(g,r_\tau)$ almost surely for all $g \in \mathcal{G}$, $r \in \mathcal{R}$, that is, there is no truncation error. Hence the bound on $G_n(g,r_\tau) - \Pi_1 G_n(g,r_\tau)$ implies Equation (SA-15) holds with $\alpha = 0$ and similarly for the Z_n^G counterpart. □

SA-III.2 General Result

This section presents the main result for the G_n-process. To simplify notation, the parameters of $\mathcal{G}$ and $\mathcal{G} \cdot \mathcal{V}_\mathcal{R}$ (Definitions 4 to 12, SA.1, SA.2) are taken with $\mathcal{C} = \mathcal{Q}_\mathcal{G}$, and the index $\mathcal{Q}_\mathcal{G}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{Q}_\mathcal{G} \times \mathcal{Y}$, and the index $\mathcal{Q}_\mathcal{G} \times \mathcal{Y}$ is omitted where there is no ambiguity.

Theorem SA.1. *Suppose $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ are i.i.d. random vectors taking values in $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$ with common law $\mathbb{P}_Z$, where $\mathbf{x}_i$ has distribution $\mathbb{P}_X$ supported on $\mathcal{X} \subseteq \mathbb{R}^d$, y_i has distribution $\mathbb{P}_Y$ supported on $\mathcal{Y} \subseteq \mathbb{R}$, and the following conditions hold.*

(i) *$\mathcal{G}$ is a real-valued pointwise measurable class of functions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.*

(ii) *There exists a surrogate measure $\mathbb{Q}_\mathcal{G}$ for $\mathbb{P}_X$ with respect to $\mathcal{G}$ such that $\mathbb{Q}_\mathcal{G} = \mathfrak{m} \circ \phi_\mathcal{G}$, where the normalizing transformation $\phi_\mathcal{G} : \mathcal{Q}_\mathcal{G} \mapsto [0,1]^d$ is a diffeomorphism.*

(iii) *$\mathtt{M}_\mathcal{G} < \infty$ and $J(\mathcal{G}, \mathtt{M}_\mathcal{G}, 1) < \infty$.*

(iv) *$\mathcal{R}$ is a real-valued pointwise measurable class of functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$.*

(v) $J(\mathcal{R}, M_{\mathcal{R}}, 1) < \infty$, *where* $M_{\mathcal{R}}(y) + \mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \leq \mathtt{v}(1 + |y|^{\alpha})$ *for all* $y \in \mathcal{Y}$, *for some* $\mathtt{v} > 0$, *and for some* $\alpha \geq 0$. *Furthermore, if* $\alpha > 0$, *then* $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

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

- $\mathbb{E}[G_n(g_1,r_1)G_n(g_2,r_2)] = \mathbb{E}[Z_n^G(g_1,r_1)Z_n^G(g_2,r_2)]$ *for all* $(g_1,r_1),(g_2,r_2) \in \mathcal{G} \times \mathcal{R}$, *and*
- $\mathbb{P}\big[\|G_n - Z_n^G\|_{\mathcal{G}\times\mathcal{R}} > C_1 c_{\mathtt{v},\alpha}\mathsf{T}_n^G(t)\big] \leq C_2 e^{-t}$ *for all* $t > 0$,

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

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

with

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

where

$$\mathtt{c}_1 = d\sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}}\prod_{j=1}^{d-1}\sigma_j(\nabla\phi_{\mathcal{H}}(\mathbf{x})), \qquad \mathtt{c}_2 = \sup_{\mathbf{x}\in\mathcal{Q}_{\mathcal{H}}}\frac{1}{\sigma_d(\nabla\phi_{\mathcal{H}}(\mathbf{x}))}, \qquad \mathtt{c}_3 = d^{-1/2}(2\sqrt{d})^{d-1}\mathtt{c}_1\mathtt{c}_2^{d-1},$$

and

$$\begin{aligned}
\mathcal{V}_{\mathcal{R}} &= \{\theta(\cdot, r) : r \in \mathcal{R}\}, \\
\mathtt{N}(\delta) &= \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}})\mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}}), \qquad \delta \in (0,1], \\
J(\delta) &= \sqrt{2}J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}), \qquad \delta \in (0,1], \\
M^* &= \Big\lfloor \log_2\min\Big\{\Big(\frac{\mathtt{c}_1 n\mathtt{TV}_{\mathcal{G}}}{\mathtt{E}_{\mathcal{G}}}\Big)^{\frac{d}{d+1}}, \Big(\frac{\mathtt{c}_1\mathtt{c}_2 n\mathtt{L}_{\mathcal{G}}\mathtt{TV}_{\mathcal{G}}}{\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}\Big)^{\frac{d}{d+2}}\Big\}\Big\rfloor, \\
N^* &= \Big\lceil \log_2\max\Big\{\Big(\frac{n\mathtt{M}_{\mathcal{G}}^{d+1}}{\mathtt{c}_1^d\mathtt{E}_{\mathcal{G}}\mathtt{TV}_{\mathcal{G}}^d}\Big)^{\frac{1}{d+1}}, \Big(\frac{n^2\mathtt{M}_{\mathcal{G}}^{2d+2}}{\mathtt{c}_1^d\mathtt{c}_2^d\mathtt{TV}_{\mathcal{G}}^d\mathtt{L}_{\mathcal{G}}^d\mathtt{E}_{\mathcal{G}}^2}\Big)^{\frac{1}{d+2}}\Big\}\Big\rceil.
\end{aligned}$$

Proof of Theorem SA.1. To simplify notation, we will use $\mathbb{E}[\cdot|\mathcal{X}_{0,l}]$ in short for $\mathbb{E}[\cdot|\mathbf{x}_i \in \mathcal{X}_{0,l}]$, and $\mathbb{E}[\cdot|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in short for $\mathbb{E}[\cdot|(\mathbf{x}_i, y_i) \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in this proof.

First, we make a reduction via the surrogate measure and normalizing transformation. Since $\mathrm{Supp}(\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}) \subseteq \mathrm{Supp}(\mathcal{G})$, we know $\mathcal{Q}_{\mathcal{G}}$ is also a surrogate measure for $\mathbb{P}_X$ with respect to $\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}$, and $\phi_{\mathcal{G}}$ remains a valid normalizing transformation. Let $\mathcal{Z}_{\mathcal{G}} = \mathcal{X} \cap \mathrm{Supp}(\mathcal{G})$. Since $\mathbb{Q}_{\mathcal{G}} = \mathfrak{m} \circ \phi_{\mathcal{G}}$ by assumption (ii) in Theorem 1, and $\mathbb{Q}_{\mathcal{G}}|_{\mathcal{Z}_{\mathcal{G}}} = \mathbb{P}_X|_{\mathcal{Z}_{\mathcal{G}}}$,

$$\mathbb{P}_X|_{\mathcal{Z}_{\mathcal{G}}} = \mathfrak{m} \circ \phi_{\mathcal{G}}|_{\mathcal{Z}_{\mathcal{G}}}.$$

To define the $\mathsf{Uniform}([0,1]^d)$ random variables on the probability space that $(\mathbf{x}_i, y_i)$'s live in, we define a joint probability measure $\mathbb{O}$ on $(\mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^{2d+1}))$ such that for all $A \in \mathcal{B}(\mathbb{R}^{2d+1})$:

$$\begin{aligned}
&\mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}} \times \mathbb{R} \times \mathcal{Z}_{\mathcal{H}})) = \mathbb{P}_Z(\Pi_{1:d+1}(A \cap \{(\mathbf{x}, y, \mathbf{x}) : \mathbf{x} \in \mathcal{Z}_{\mathcal{H}}, y \in \mathbb{R}\})),\\
&\mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}} \times \mathbb{R} \times \mathcal{Z}_{\mathcal{H}}^c)) = \mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}}^c \times \mathbb{R} \times \mathcal{Z}_{\mathcal{H}})) = 0,\\
&\mathbb{O}(A \cap (\mathcal{Z}_{\mathcal{H}}^c \times \mathbb{R} \times \mathcal{Z}_{\mathcal{H}}^c)) = \int_{\mathcal{Z}_{\mathcal{H}}^c \cap \Pi_{d+2:2d+1}(A)} \frac{\mathbb{P}_Z(A^{\mathbf{u}} \cap (\mathcal{Z}_{\mathcal{H}}^c \times \mathbb{R}))}{\mathbb{P}_Z(\mathcal{Z}_{\mathcal{H}}^c \times \mathbb{R})} d(\mathfrak{m} \circ \phi_{\mathcal{H}})(\mathbf{u}),
\end{aligned}$$

where $\Pi_{1:d+1}(A) = \{\mathbf{z} \in \mathbb{R}^{d+1} : (\mathbf{z}, \mathbf{u}) \in A \text{ for some } \mathbf{u} \in \mathbb{R}^d\}$, $\Pi_{d+2:2d+1}(A) = \{\mathbf{u} \in \mathbb{R}^d : (\mathbf{z}, \mathbf{u}) \in A \text{ for some } \mathbf{z} \in \mathbb{R}^{d+1}\}$, and $A^{\mathbf{u}} = \{\mathbf{z} \in \mathbb{R}^{d+1} : (\mathbf{z}, \mathbf{u}) \in A\}$.

Then we can check that (i) the marginals of $\mathbb{O}$ are $\mathbb{P}_Z$ and $\mathfrak{m} \circ \phi_{\mathcal{H}}$, respectively; (ii) $\mathbb{O}|_{\mathcal{Z}_{\mathcal{H}} \times \mathbb{R} \times \mathbb{R}^d \cup \mathbb{R}^d \times \mathbb{R} \times \mathcal{Z}_{\mathcal{H}}}$ is supported on $\{(\mathbf{x}, y, \mathbf{x}) : \mathbf{x} \in \mathcal{Z}_{\mathcal{H}}, y \in \mathbb{R}\}$. By Skorohod embedding (Dudley, 2014, Lemma 3.35), on a possibly enlarged probability space, there exists a $\mathbf{u}_i, 1 \le i \le n$ i.i.d. Uniform$([0,1]^d)$ such that $(\mathbf{z}_i = (\mathbf{x}_i, y_i), \phi_{\mathcal{H}}^{-1}(\mathbf{u}_i))$ has joint law $\mathbb{O}$. In particular, if $\mathbf{x}_i \in \mathcal{Z}_{\mathcal{H}}$, then $\mathbf{x}_i = \phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)$; if $\mathbf{x}_i \in \mathcal{Z}_{\mathcal{H}}^c$, then $\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i) \in \mathcal{Z}_{\mathcal{H}}^c$, and since $\mathcal{Q}_{\mathcal{H}} \subseteq \mathcal{X} \cup (\cap_{h \in \mathcal{H}} \mathrm{Supp}(h)^c)$, $\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i) \in \cap_{h \in \mathcal{H}} \mathrm{Supp}(h)^c$. In particular, $\sup_{\mathbf{u} \in [0,1]^d} \mathbb{E}[\exp(|y_i|) | \mathbf{u}_i = \mathbf{u}] \le 2$.

By the same argument as in the proof for Theorem 1, assumption (ii) implies that on a possibly enriched probability space, there exists $(\mathbf{u}_i : 1 \le i \le n)$ i.i.d distributed with law $\mathbb{P}_U = \mathsf{Uniform}([0,1]^d)$, and

$$g(\mathbf{x}_i) = g(\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)), \qquad \forall g \in \mathcal{G}, 1 \le i \le n.$$

Define $\widetilde{G}_n$ to be the empirical process based on $((\mathbf{u}_i, y_i) : 1 \le i \le n)$, and

$$\widetilde{G}_n(f, s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big[f(\mathbf{u}_i) s(y_i) - \mathbb{E}[f(\mathbf{u}_i) s(y_i)] \Big],$$

and take $\widetilde{\mathcal{G}} = \{g \circ \phi_{\mathcal{H}}^{-1} : g \in \mathcal{G}\}$, then

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

The relation between constants for $\widetilde{\mathcal{G}}$ and constants for $\mathcal{G}$ can be deduced from Lemma SA.10. Hence, without loss of generality, we assume $(\mathbf{x}_i : 1 \le i \le n)$ are i.i.d under common law $\mathbb{P}_X = \mathsf{Uniform}([0,1]^d)$ distributed and $\mathcal{X} = [0,1]^d$.

Take $\mathcal{A}_{M,N}(\mathbb{P}_Z, 1)$ to be an axis-aligned cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$, of depth M for the main subspace $\mathbb{R}^d$ and depth N for the multiplier subspace $\mathbb{R}$ with respect to $\mathbb{P}_Z$. Take $(Z_n^G(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ and $(\Pi_1 Z_n^G(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ to be the mean-zero Gaussian processes constructed as in Equations (SA-13) and (SA-14). Let $(\mathcal{G} \times \mathcal{R})_\delta$ be a $\delta \|\mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}\|_{\mathbb{P}_Z}$-net of $\mathcal{G} \times \mathcal{R}$ with cardinality no greater than $\mathtt{N}_{\mathcal{G} \times \mathcal{R}}(\delta, \mathtt{M}_{\mathcal{G}} M_{\mathcal{R}})$. By standard empirical process argument, $\mathtt{N}_{\mathcal{G} \times \mathcal{R}}(\delta, \mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}) \le \mathtt{N}(\delta)$. By Lemma SA.16, the meshing error can be bounded by: For all $t > 0$,

$$\mathbb{P}\big[\|G_n - G_n \circ \pi_{(\mathcal{G} \times \mathcal{R})_\delta}\|_{\mathcal{G} \times \mathcal{R}} + \|Z_n^G - Z_n^G \circ \pi_{(\mathcal{G} \times \mathcal{R})_\delta}\|_{\mathcal{G} \times \mathcal{R}} > C_1 c_{\mathtt{v},\alpha} \mathsf{F}_n^G(t, \delta) \big] \le 8 \exp(-t),$$

where C_1 is a universal constant and $c_{\mathtt{v},\alpha} = \mathtt{v}(1 + (2\alpha)^{\frac{\alpha}{2}})$. Lemma SA.17 implies that the strong approximation error for the projected process on δ-net is bounded by: For all $t > 0$,

$$\mathbb{P}\Big[\|\Pi_1 G_n - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_1 c_{\mathtt{v},\alpha}\sqrt{\frac{N^{2\alpha+1}2^M \mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}t} + C_1 c_{\mathtt{v},\alpha}\sqrt{\frac{\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R}),M+N}}{n}t}\Big] \leq 2\mathtt{N}(\delta)e^{-t}.$$

where

$$\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R}),M+N} = \sup_{f\in\Pi_1(\mathcal{G}\times\mathcal{R})} \min\Bigg\{ \sup_{(j,k)\in\mathcal{I}_{M+N}} \Bigg[\sum_{j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}\widetilde{\beta}^2_{j',k'}(f)\Bigg],$$
$$\sup_{\mathbf{z}\in\mathcal{C}_{M+N,0}} f(\mathbf{z})^2(M+N)\Bigg\}.$$

Now we upper bound the left hand side of the minimum. Let $f \in \Pi_1(\mathcal{G}\times\mathcal{R})$. Then there exists $g \in \mathcal{G}$ 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 $(j,k) \in \mathcal{I}_{M+N}$. We consider two cases.

Case 1: $j > N$. By Definition SA.7, $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathcal{Y}_{*,N,0}$. By definition of $\mathcal{A}_{M,N}(\mathbb{P}_Z, 1)$ and the assumption that $\mathbf{x}_i$'s are $\mathsf{Uniform}([0,1]^d)$ distributed, $\|\mathcal{X}_{j-N,k}\|_\infty \leq 2^{-\frac{M+N-j}{d}+1}$.

Consider j' such that $N \leq j' \leq j$. By definition of $\mathcal{A}_{M,N}(\mathbb{P}_Z, 1)$ and the assumption that $\mathbf{x}_i$'s are $\mathsf{Uniform}([0,1]^d)$ distributed, the j'-th level difference set $\mathcal{U}_{j'} = \cup_{0\leq k<2^{M+N-j'}}(\mathcal{C}_{j'-1,2k+1} - \mathcal{C}_{j'-1,2k})$ is contained in $[-2^{-\frac{M+N-j'}{d}+2}, 2^{-\frac{M+N-j'}{d}+2}]^d$. Let $g \in \mathcal{G}$, $r \in \mathcal{R}$. By definition of $\widetilde{\gamma}_{j',m}$ and similar arguments to those in the proof of Lemma SA.7,

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

Next, consider j' such that $0 \leq j' < N$, we know

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

It follows that

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

Case 2: $j \leq N$**.** Then $\mathcal{C}_{j,k} = \mathcal{X}_{0,l}\times\mathcal{Y}_{l,j,m}$ with $k = 2^{N-j}l+m$, and $\mathcal{C}_{j',k'} = \mathcal{X}_{0,l'}\times\mathcal{Y}_{l',j',m'}$ with $k' = 2^{N-j'}l'+m'$. In particular, $\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}$ implies $l'=l$ and $\mathcal{Y}_{l',j',m'}\subseteq\mathcal{Y}_{l,j,m}$. By a similar argument to the proof in Lemma SA.17 (Layers $1\leq j\leq N$), for any $0\leq j'\leq j$,

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

Using the elementary inequality that $x(x+1)\leq 30\cdot 2^{x/4}$ for $x>0$, we can get

$$\sum_{1\leq j'<j}(j-j')(j-j'+1)2^{j'-j}\sum_{k':\mathcal{C}_{j',k'}\subseteq\mathcal{C}_{j,k}}|\widetilde{\gamma}_{j',k'}(g,r)| \leq 60c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G}}N^{\alpha}.$$

Moreover, for all (j,k), we have $\widetilde{\beta}_{j,k}(g,r)\leq c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G}}N^{\alpha}$. This implies that

$$\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R}),M+N} \lesssim c^2_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G}}N^{\alpha}\min\{\mathtt{K}^*_{\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}}}+\mathtt{M}_{\mathcal{G}}N^{\alpha}, \mathtt{M}_{\mathcal{G}}N^{\alpha}(M+N)\}.$$

Since $\mathbf{x}_i \overset{i.i.d.}{\sim} \mathsf{Uniform}([0,1]^d)$ and the cells $\mathcal{A}_{M,N}(\mathbb{P}_Z,1)$ are obtained via *axis aligned dyadic expansion*, we have $\|\mathcal{X}_{0,k}\|_\infty \leq 2^{-\lfloor M/d\rfloor}$ for all $0\leq k<2^M$. Then by Lemma SA.21, for all $t>N$,

$$\begin{aligned}
\mathbb{P}\Big[\|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} \gtrsim \sqrt{c_{\mathtt{v},2\alpha}}\sqrt{N^2\mathtt{V}_{\mathcal{G}} + 2^{-N}\mathtt{M}^2_{\mathcal{G}}}t^{\alpha+\frac{1}{2}} + c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}t^{\alpha+1}\Big] \leq 4\mathtt{N}(\delta)ne^{-t}, \\
\mathbb{P}\Big[\|Z^G_n - \Pi_1 Z^G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} \gtrsim \sqrt{c_{\mathtt{v},2\alpha}}\sqrt{N^2\mathtt{V}_{\mathcal{G}} + 2^{-N}\mathtt{M}^2_{\mathcal{G}}}t^{\frac{1}{2}} + c_{\mathtt{v},\alpha}\frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}}t\Big] \leq 4\mathtt{N}(\delta)ne^{-t},
\end{aligned}$$

where $c_{\mathtt{v},\alpha} = \mathtt{v}(1+(2\alpha)^{\frac{\alpha}{2}})$ and $c_{\mathtt{v},2\alpha} = \mathtt{v}^2(1+(4\alpha)^{\alpha})$, and

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

We find the optimal parameters M^* and N^* by balancing the term $\sqrt{\frac{2^M\mathtt{E}_{\mathcal{G}}\mathtt{M}_{\mathcal{G}}}{n}}$ from the bound on $\|\Pi_1 G_n - \Pi_1 Z^G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and the term $\mathtt{V}_{\mathcal{G}}$ from the bounds on $\|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and $\|Z_n - \Pi_1 Z^G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$, choosing

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

It follows that for all $t > N_*$, with probability at least $1 - 4n\mathtt{N}(\delta)\exp(-t)$,

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

The result then follows from the decomposition that

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

and Lemma SA.10 for the reduction to the case of $\mathsf{Uniform}([0,1]^d)$ distributed $\mathbf{x}_i$'s. □

SA-III.3 Additional Results

This section presents the additional result for the G_n-process under VC-type entropy conditions. To simplify notation, the parameters of $\mathcal{G}$ and $\mathcal{G}\cdot\mathcal{V}_\mathcal{R}$ (Definitions 4 to 12, SA.1, SA.2) are taken with $\mathcal{C} = \mathcal{Q}_\mathcal{G}$, and the index $\mathcal{Q}_\mathcal{G}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G}\times\mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{Q}_\mathcal{G}\times\mathcal{Y}$, and the index $\mathcal{Q}_\mathcal{G}\times\mathcal{Y}$ is omitted where there is no ambiguity.

Corollary SA.4 (VC-Type Lipschitz Functions)**.** *Suppose the conditions of Theorem SA.1 and the following additional conditions hold.*

(i) $\mathcal{G}$ *is a VC-type class with respect to envelope* $\mathtt{M}_\mathcal{G}$ *with constant* $\mathtt{c}_\mathcal{G} \geq e$ *and exponent* $\mathtt{d}_\mathcal{G} \geq 1$ *over* $\mathcal{Q}_\mathcal{G}$.

(ii) $\mathcal{R}$ *is a VC-type class with respect to envelope* $M_\mathcal{R}$ *with constant* $\mathtt{c}_\mathcal{R} \geq e$ *and exponent* $\mathtt{d}_\mathcal{R} \geq 1$ *over* $\mathcal{Y}$.

(iii) There exists a constant $\mathtt{k}$ *such that* $|\log_2 \mathtt{E}_\mathcal{G}| + |\log_2 \mathtt{TV}| + |\log_2 \mathtt{M}_\mathcal{G}| \leq \mathtt{k}\log_2 n$*, where we take* $\mathtt{TV} = \max\{\mathtt{TV}_\mathcal{G}, \mathtt{TV}_{\mathcal{G}\cdot\mathcal{V}_\mathcal{R}}\}$.

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

- $\mathbb{E}[G_n(g_1,r_1)G_n(g_2,r_2)] = \mathbb{E}[Z_n^G(g_1,r_1)Z_n^G(g_2,r_2)]$ *for all* $(g_1,r_1),(g_2,r_2) \in \mathcal{G}\times\mathcal{R}$*, and*
- $\mathbb{P}\left[\|G_n - Z_n^G\|_{\mathcal{G}\times\mathcal{R}} > C_1 c_{\mathtt{v},\alpha}\mathsf{T}_n^G(t)\right] \leq C_2 e^{-t}$ *for all* $t > 0$,

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

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

with $\mathtt{c} = \mathtt{c}_\mathcal{G}\mathtt{c}_\mathcal{R}$, $\mathtt{d} = \mathtt{d}_\mathcal{G} + \mathtt{d}_\mathcal{R}$.

Proof of Corollary SA.4. The proof follows by Theorem SA.1 with $\delta = n^{-1/2}$, and

$$\mathtt{N}(n^{-1/2}) = \mathtt{N}_{\mathcal{G}}(1/\sqrt{2n}, \mathtt{M}_{\mathcal{G}})\mathtt{N}_{\mathcal{R}}(1/\sqrt{2n}, M_{\mathcal{R}}) \leq \mathtt{c}_{\mathcal{G}}\mathtt{c}_{\mathcal{R}}(2\sqrt{n})^{\mathtt{d}_{\mathcal{G}}+\mathtt{d}_{\mathcal{R}}} = \mathtt{c}(2\sqrt{n})^{\mathtt{d}},$$

and

$$\begin{aligned} J(n^{-1/2}) &= \sqrt{2}J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, 1/\sqrt{2n}) + \sqrt{2}J(\mathcal{R}, M_{\mathcal{R}}, 1/\sqrt{2n}) \\ &\leq 3n^{-1/2}\sqrt{\mathtt{d}_{\mathcal{G}} \log(\mathtt{c}_{\mathcal{G}}\sqrt{n})} + 3\delta\sqrt{\mathtt{d}_{\mathcal{R}} \log(\mathtt{c}_{\mathcal{R}}\sqrt{n})} \\ &\leq 3\delta\sqrt{(\mathtt{d}_{\mathcal{G}} + \mathtt{d}_{\mathcal{R}}) \log(\mathtt{c}_{\mathcal{G}}\mathtt{c}_{\mathcal{R}} n)} \leq 3\delta\sqrt{\mathtt{d} \log(\mathtt{c} n)}. \end{aligned}$$

The conclusion follows. □

SA-IV Residual-Based Empirical Processes

Recall that $\mathbf{z}_i = (\mathbf{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^d \times \mathbb{R}$, $i = 1, \ldots, n$, are i.i.d. random vectors supported on a background probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and the *residual-based empirical process* is

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

In particular, $(R_n(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ can be seen as a combination of the two empirical processes studied in the previous sections: for $r \in \mathcal{R}$ and $\mathbf{x} \in \mathcal{X}$,

$$R_n(g, r) = G_n(g, r) - X_n(g\,\theta(\cdot, r)), \qquad \theta(\mathbf{x}, r) = \mathbb{E}[r(y_i)|\mathbf{x}_i = \mathbf{x}],$$

where

$$\begin{aligned} G_n(g, r) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[g(\mathbf{x}_i)r(y_i) - \mathbb{E}[g(\mathbf{x}_i)r(y_i)]\Big], \\ X_n(g\,\theta(\cdot, r)) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[g(\mathbf{x}_i)\theta(\mathbf{x}_i, r) - \mathbb{E}[g(\mathbf{x}_i)\theta(\mathbf{x}_i, r)]\Big]. \end{aligned}$$

Results for the X_n process (Section SA-II) and for the G_n process (Section SA-III) will be used to handle the terms above. The same error decomposition as in Sections SA-II and SA-III also applies here:

$$\begin{aligned} \|R_n - Z_n^R\|_{\mathcal{G}\times\mathcal{R}} &\leq \|R_n - Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|R_n - R_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^R \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^R\|_{\mathcal{G}\times\mathcal{R}} \\ &\leq \|\Pi_2 Z_n^R - Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|R_n - \Pi_2 R_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|\Pi_2 R_n - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} \\ &\quad + \|R_n - R_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^R \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^R\|_{\mathcal{G}\times\mathcal{R}}, \end{aligned}$$

where $(\mathcal{G}\times\mathcal{R})_\delta$ denotes a discretization (or meshing) of $\mathcal{G}\times\mathcal{R}$ (i.e., δ-net of $\mathcal{G}\times\mathcal{R}$), and the terms $\|R_n - R_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}}$ and $\|Z_n^R \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^R\|_{\mathcal{G}\times\mathcal{R}}$ capture the fluctuations (or oscillations) of R_n and Z_n^R relative to the meshing for each of the stochastic processes. $\|\Pi_2 R_n - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and $\|\Pi_2 Z_n^R - Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ represent projections onto a Haar function space, where $\Pi_2 R_n(h) = R_n \circ \Pi_2 h$. The operator Π_2 is a projection onto piecewise constant functions that respects the multiplicative structure of the R_n process. The final term

$\|\Pi_2 R_n - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ captures the coupling between the empirical process and the Gaussian process (on a δ-net of $\mathcal{G}\times\mathcal{R}$, after the projection Π_2).

The general result under uniform entropy integral conditions is presented in Section SA-IV.2. Theorem 2 and Corollary 4 then follow from that general result. The proofs leverage the existence of a surrogate measure and a normalizing transformation of $\mathcal{G}$ with respect to $\mathbb{P}_X$, the distribution of $\mathbf{x}_1$, as developed in Section SA-II.2. We will use the same class of cylindered quasi-dyadic cell expansions as in Section SA-III.1.1, which explicitly exploits the multiplicative structure of R_n. Bounds for each term in the error decomposition are provided in Section SA-IV.1, which boils down to handle the extra $X_n(g\,\theta(\cdot, r))$ term compared to the results in Section SA-III.1 and is organized as follows:

- Section SA-IV.1.1 introduces the *conditional mean adjusted product-factorized projection* that combines the *product-factorized projection* for the $G_n(g, r)$ part and the L_2 projection for the $X_n(g\,\theta(\cdot, r))$ part.

- Section SA-IV.1.2 constructs the Gaussian process $(Z_n^R(g, r) : (g, r) \in \mathcal{G}\times\mathcal{R})$. The construction is essentially the same as those in Section SA-II.1.3, relying on coupling binomial random variables with Gaussian random variables.

- Section SA-IV.1.3 handles the meshing errors $\|R_n - R_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}}$ and $\|Z_n^R \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^R\|_{\mathcal{G}\times\mathcal{R}}$ using standard empirical process results.

- Section SA-IV.1.4 addresses the strong approximation error $\|\Pi_2 R_n - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta}$. With the help of the relation between Π_1 and Π_2, we can reuse results from Section SA-III.1.5.

- Section SA-IV.1.5 addresses the projection errors $\|R_n - \Pi_2 R_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ and $\|Z_n^R - \Pi_1 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta}$. We use the results from Section SA-III.1.6 for $\|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta}$, and deal with $\|X_n(g\theta(\cdot, r)) - \Pi_0 X_n(g\theta(\cdot, r))\|_{(\mathcal{G}\times\mathcal{R})_\delta}$ using results from Section SA-II.1.6.

SA-IV.1 Preliminary Technical Results

This section presents preliminary technical results that are used to prove Theorem SA.1. Whenever possible, these results are presented at a higher level of generality, and therefore may be of independent theoretical interest. Throughout this section, we assume the same set of conditions (Assumption SA.2) on data generate process as in Section SA-III.1.

Compared to the assumptions in Theorem 2, this assumption does not require the existence of a surrogate measure or a normalizing transformation. It will be applied in the analysis of terms in the error decomposition, where we work with the $\mathbb{P}_Z$ distribution and extra condition on the existence of Lebesgue density of $\mathbb{P}_X$ is assumed whenever necessary (Section SA-IV.1.5). The surrogate measure and the normalizing transformation will be used in the proof of Theorem SA.1 with the help of Section SA-II.2, providing greater flexibility in the data generating process.

SA-IV.1.1 Projection onto Piecewise Constant Functions

For the residual empirical process, we tailor a projection to piecewise constant functions on the quasi-dyadic cells that differs from the mean square projection from Section SA-II.1.2 and the product-factorized projection from Section SA-III.1.2. Given a cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$, $\mathcal{C}_{M,N}(\mathbb{P}, \rho)$ with $\mathbb{P}$ the law of random vector $(\mathbf{X}, Y) \in \mathbb{R}^d \times \mathbb{R}$, and recall the definition of $\mathcal{E}_{M+N}$ from Section SA-II.1.2, for any real valued functions g on $\mathbb{R}^d$ and r on $\mathbb{R}$ such that $\int_{\mathbb{R}^d}\int_{\mathbb{R}} g(\mathbf{x})^2\mathbb{P}(dyd\mathbf{x}) < \infty$ and $\int_{\mathbb{R}^d}\int_{\mathbb{R}} r(y)^2\mathbb{P}(dyd\mathbf{x}) < \infty$, the *conditional mean adjusted product-factorized projection* of g and r is defined as

$$\Pi_2(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r] = \Pi_1(\mathcal{C}_{M,N}(\mathbb{P},\rho))[g,r] - \Pi_0(\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)])[g\,\theta(\cdot,r)], \qquad \text{(SA-16)}$$

where $\theta(\mathbf{x},r) = \mathbb{E}[r(Y)|\mathbf{X}=\mathbf{x}]$ for $r \in \mathcal{R}$ and $\mathbf{x} \in \mathcal{X}$, and $\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)] = \{\mathcal{X}_{l,k} : 0 \le l \le M, 0 \le k < 2^{M-l}\}$ as defined in Definition SA.7. We denote the collection of conditional mean functions based on $\mathcal{R}$ by $\mathcal{V}_{\mathcal{R}} = \{\theta(\cdot,r) : r \in \mathcal{R}\}$.

This projection can also be represented using the Haar basis as

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

with

$$\eta_{j,k}(g,r) = \begin{cases} 0, & \text{if } N \le j \le M+N, \\ \gamma_{j,k}(g,r), & \text{if } j < N. \end{cases} \qquad \text{(SA-17)}$$

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

Next, we define the empirical processes indexed by these projected functions. With a slight abuse of notation, let $(X_n(f) : f \in \mathcal{F})$ be the empirical process based on a random sample $((\mathbf{x}_i, y_i) : 1 \le i \le n)$, where $\mathcal{F}$ is a class of real-valued functions on $\mathbb{R}^{d+1}$. Specifically, $X_n(f) = n^{-1/2}\sum_{i=1}^n (f(\mathbf{x}_i,y_i) - \mathbb{E}[f(\mathbf{x}_i,y_i)])$ for $f \in \mathcal{F}$. For any real valued functions g on $\mathbb{R}^d$ and r on $\mathbb{R}$ such that $\int_{\mathbb{R}^d}\int_{\mathbb{R}} g(\mathbf{x})^2\mathbb{P}(dyd\mathbf{x}) < \infty$ and $\int_{\mathbb{R}^d}\int_{\mathbb{R}} r(y)^2\mathbb{P}(dyd\mathbf{x}) < \infty$, we define

$$\begin{aligned} \Pi_2 R_n(g,r) &= X_n \circ \Pi_2(g,r), \\ \Pi_0 R_n(g,r) &= X_n \circ \Pi_0[\mathcal{C}_{M,N}(\mathbb{P},\rho)](gr) - X_n \circ \Pi_0(\mathtt{p}_X[\mathcal{C}_{M,N}(\mathbb{P},\rho)])[g\,\theta(\cdot,r)]. \end{aligned} \qquad \text{(SA-18)}$$

SA-IV.1.2 Strong Approximation Constructions

Lemma SA.22. *Suppose Assumption SA.2 holds, and a cylindered quasi-dyadic expansion $\mathcal{C}_K(\mathbb{P}_Z,\rho)$ is given. Then, $(\mathcal{G}\cdot\mathcal{R}) \cup (\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}}) \cup \Pi_1(\mathcal{G}\times\mathcal{R}) \cup \Pi_2(\mathcal{G}\times\mathcal{R}) \cup \Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})](\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}})$ is $\mathbb{P}_Z$-pregaussian.*

Proof. Recall we have shown in the proof of Lemma SA.12 that for all $0 < \delta < 1$,

$$\begin{aligned} J_{\mathcal{X}\times\mathcal{Y}}(\mathcal{G}\cdot\mathcal{R}, \mathtt{M}_{\mathcal{G},\mathcal{X}}M_{\mathcal{R},\mathcal{Y}}, \delta) &\lesssim \sqrt{2}J_{\mathcal{X}}(\mathcal{G}, \mathtt{M}_{\mathcal{G},\mathcal{X}}, \delta/\sqrt{2}) + \sqrt{2}J_{\mathcal{Y}}(\mathcal{R}, M_{\mathcal{R},\mathcal{Y}}, \delta/\sqrt{2}), \\ J_{\mathcal{X}\times\mathcal{Y}}(\Pi_1(\mathcal{G}\times\mathcal{R}), c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G},\mathcal{X}}N^{\alpha}, \delta) &\lesssim \sqrt{2}J_{\mathcal{X}}(\mathcal{G}, \mathtt{M}_{\mathcal{G},\mathcal{X}}, \delta/(3\sqrt{2})) + \sqrt{2}J_{\mathcal{Y}}(\mathcal{R}, M_{\mathcal{R},\mathcal{Y}}, \delta/(3\sqrt{2})), \end{aligned}$$

where $c_{\mathtt{v},\alpha} = \mathtt{v}(1+(2\alpha)^{\frac{\alpha}{2}})$. Lemma SA.25 implies $J_{\mathcal{X}}(\mathcal{V}_{\mathcal{R}}, \theta(\cdot, M_{\mathcal{R},\mathcal{Y}}), \delta) \le J_{\mathcal{Y}}(\mathcal{R}, M_{\mathcal{R},\mathcal{Y}}, \delta)$. Since Assumption SA.2 (iv) implies $\sup_{\mathbf{x}\in\mathcal{X}} \theta(\cdot, M_{\mathcal{R},\mathcal{Y}}) \le c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G}}$, we know for all $0 < \delta < 1$,

$$\begin{aligned} J_{\mathcal{X}}(\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}}, c_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G}}, \delta) &\le \sqrt{2}J_{\mathcal{X}}(\mathcal{G}, \mathtt{M}_{\mathcal{G},\mathcal{X}}, \delta/\sqrt{2}) + \sqrt{2}J_{\mathcal{X}}(\mathcal{V}_{\mathcal{R}}, \theta(\cdot, M_{\mathcal{R},\mathcal{Y}}), \delta/\sqrt{2}) \\ &\le \sqrt{2}J_{\mathcal{X}}(\mathcal{G}, \mathtt{M}_{\mathcal{G},\mathcal{X}}, \delta/\sqrt{2}) + \sqrt{2}J_{\mathcal{Y}}(\mathcal{R}, M_{\mathcal{R},\mathcal{Y}}, \delta/\sqrt{2}). \end{aligned}$$

The same argument for Lemma SA.12 implies that for all $0 < \delta < 1$,

$$J_{\mathcal{X}}(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})](\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}), C_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G},\mathcal{X}}N^{\alpha}, \delta) \leq J_{\mathcal{X}}(\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}, C_{\mathtt{v},\alpha}\mathtt{M}_{\mathcal{G},\mathcal{X}}N^{\alpha}, \delta).$$

Moreover Lemma SA.15 implies $\Pi_2(\mathcal{G} \times \mathcal{R}) \subseteq \Pi_1(\mathcal{G} \times \mathcal{R}) + \Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})](\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}})$. It follows from pointwise separability of $\mathcal{G}$ and $\mathcal{R}$ and Corollary 2.2.9 in van der Vaart and Wellner (2013) that $(\mathcal{G} \cdot \mathcal{R}) \cup (\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}) \cup \Pi_2(\mathcal{G} \times \mathcal{R}) \cup \Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})](\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}})$ is $\mathbb{P}_Z$-pregaussian. □

Lemma SA.23. *Suppose Assumption SA.2 holds and a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, 1)$ is given. Then on a possibly enlarged probability space, there exists a $\mathbb{P}_Z$-Brownian bridge B_n indexed by $\mathcal{F} = (\mathcal{G} \cdot \mathcal{R}) \cup \Pi_0(\mathcal{G} \times \mathcal{R}) \cup \Pi_1(\mathcal{G} \times \mathcal{R})$ with almost sure continuous trajectories on $(\mathcal{F}, \mathfrak{d}_{\mathbb{P}_Z})$ such that for any $f \in \mathcal{F}$ and any $x > 0$,*

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

where for both $\|f\|_{\mathcal{E}_{M+N}}^2$ and $\mathtt{C}_{\{f\},M+N}$ are defined in Lemma SA.3.

Proof of Lemma SA.23. The result follows from Lemma SA.22 and the same argument as Lemma SA.13. □

Lemma SA.24. *Suppose Assumption SA.2 holds and a cylindered quasi-dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)$ with $\rho > 1$ is given. Then on a possibly enlarged probability space, there exists a Brownian bridge B_n indexed by $\mathcal{F} = (\mathcal{G} \cdot \mathcal{R}) \cup (\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}) \cup \Pi_1(\mathcal{G} \times \mathcal{R}) \cup \Pi_2(\mathcal{G} \times \mathcal{R}) \cup \Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})](\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}})$ with almost sure continuous trajectories on $(\mathcal{F}, \mathfrak{d}_{\mathbb{P}_Z})$ such that for any $f \in \mathcal{F}$ and any $x > 0$,*

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

where C_ρ is a constant that only depends on ρ.

Proof of Lemma SA.24. The result follows from Lemma SA.22 and the same argument as Lemma SA.14. □

The above two lemmas enable the construction of Gaussian processes and their projected counterparts as analogs to the empirical processes defined in Section SA-II.1.3 and Section SA-III.1.3. In particular, we define Z_n^R and $\Pi_2 Z_n^R$ as Gaussian processes indexed by $\mathcal{G} \times \mathcal{R}$ such that, for any $g \in \mathcal{G}$ and $r \in \mathcal{R}$,

$$\begin{aligned} Z_n^R(g, r) &= B_n(g(r - \theta(\cdot, r))), \\ \Pi_2 Z_n^R(g, r) &= B_n(\Pi_2[g, r]). \end{aligned} \tag{SA-19}$$

We also define the following ancillary processes for analysis:

$$\begin{aligned} Z_n^G(g, r) &= B_n(gr), \qquad \Pi_1 Z_n^G(g, r) = B_n(\Pi_1[g, r]), \\ Z_n^X(g\,\theta(\cdot, r)) &= B_n(g\,\theta(\cdot, r)), \qquad \Pi_0 Z_n^X(g\,\theta(\cdot, r)) = B_n(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})][g\,\theta(\cdot, r)]). \end{aligned} \tag{SA-20}$$

Since for any $g_1, g_2 \in \mathcal{G}$, $r_1, r_2 \in \mathcal{R}$,

$$\mathfrak{d}_{\mathbb{P}_Z}(g_1(r_1 - \theta(\cdot, r_1)), g_2(r_2 - \theta(\cdot, r_2))) \leq 2\mathfrak{d}_{\mathbb{P}_Z}(g_1 r_1, g_2 r_2),$$

and B_n has almost sure continuous sample trajectories on $(\mathcal{G} \cdot \mathcal{R}, \mathfrak{d}_{\mathbb{P}_Z})$, Equation (SA-19) also implies $(Z_n^R(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ has almost sure continuous sample trajectories on $(\mathcal{G} \times \mathcal{R}, \mathfrak{d}_{\mathbb{P}_Z})$.

The following ancillary lemma on uniform covering number of the class of conditional means is used for the proof of Lemma SA.22.

Lemma SA.25. *Suppose $\mathcal{S}$ is a class of functions from a measurable space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ to $\mathbb{R}$, where $\mathcal{Y} \subseteq \mathbb{R}$, with envelope function $M_{\mathcal{S},\mathcal{Y}}$. Let $\mathcal{V}_{\mathcal{S}}$ be the class of conditional means $\{\theta(\cdot, s) : s \in \mathcal{S}\}$ with $\theta(\mathbf{x}, s) = \mathbb{E}[s(y_i)|\mathbf{x}_i = \mathbf{x}]$ for $\mathbf{x} \in \mathcal{X}$. Then*

$$\mathtt{N}_{\mathcal{V}_{\mathcal{S}},\mathcal{X}}(\delta, \theta(\cdot, M_{\mathcal{S},\mathcal{Y}})) \leq \mathtt{N}_{\mathcal{S},\mathcal{Y}}(\delta, M_{\mathcal{S},\mathcal{Y}}).$$

Proof of Lemma SA.25. Let $\mathcal{Q}$ be a finite discrete measure on $\mathbb{R}^d$, and 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 $\int |\theta(\cdot, M_{\mathcal{S},\mathcal{Y}})| \, d\widetilde{P} \leq \int_{\mathbb{R}^d} \mathbb{E}[M_{\mathcal{S},\mathcal{Y}}(y_i)|\mathbf{x}_i = \mathbf{x}] \, d\mathcal{Q}(z) < \infty$, since $\sup_{m \in \mathcal{V}_S} \|m\|_\infty < \infty$.

For $r, s \in \mathcal{S}$, we have

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

Here, $\widetilde{P}$ is not necessarily finite or discrete, but by a similar argument as in Lemma SA.15, there exists a subset $\mathcal{S}_\varepsilon \subseteq \mathcal{S}$ with cardinality no greater than $\mathtt{N}_{\mathcal{S},\mathcal{Y}}(\delta, M_{\mathcal{S},\mathcal{Y}})$, such that for any $s \in \mathcal{S}$, there exists $r \in \mathcal{S}_\varepsilon$ with $\|r - s\|_{\widetilde{P},2} \leq \varepsilon \|\theta(\cdot, M_{\mathcal{S},\mathcal{Y}})\|_{\widetilde{P},2}$. Hence, $\|m_r - m_s\|_{\mathcal{Q},2} \leq \varepsilon \|\theta(\cdot, M_{\mathcal{S},\mathcal{Y}})\|_{\widetilde{P},2} = \varepsilon \|\theta(\cdot, M_{\mathcal{S},\mathcal{Y}})\|_{\mathcal{Q},2}$. The conclusion then follows. □

SA-IV.1.3 Meshing Error

To simplify notation, the parameters of $\mathcal{G}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{X}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{X} \times \mathcal{Y}$, and the index $\mathcal{X} \times \mathcal{Y}$ is omitted where there is no ambiguity. We also define

$$J(\delta) = \sqrt{2} J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2} J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}), \qquad \delta \in (0, 1],$$
$$\mathtt{N}(\delta) = \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}}) \mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}}), \qquad \delta \in (0, 1].$$

For $0 < \delta \leq 1$, consider a $\delta \mathtt{M}_{\mathcal{G}} \|M_{\mathcal{R}}\|_{\mathbb{P}_Y,2}$-net of $(\mathcal{G} \times \mathcal{R}, \|\cdot\|_{\mathbb{P}_Z,2})$, denoted by $(\mathcal{G} \times \mathcal{R})_\delta$, with cardinality at most $\mathtt{N}_{\mathcal{G} \times \mathcal{R}}(\delta, \mathtt{M}_{\mathcal{G}} \|M_{\mathcal{R}}\|_{\mathbb{P}_Y,2})$. Define the projection onto the δ-net as a mapping $\pi_{(\mathcal{G} \times \mathcal{R})_\delta} : \mathcal{G} \times \mathcal{R} \to \mathcal{G} \times \mathcal{R}$ such that $\|\pi_{(\mathcal{G} \times \mathcal{R})_\delta}(g, r) - gr\|_{\mathbb{P}_Z,2} \leq \delta \mathtt{M}_{\mathcal{G}} \|M_{\mathcal{R}}\|_{\mathbb{P}_Y,2}$ for all $g \in \mathcal{G}$ and $r \in \mathcal{R}$.

Lemma SA.26. *Suppose Assumption SA.2 holds, a cylindered quasi-dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)$ is given, $(Z_n^R(g, r) : g \in \mathcal{G}, r \in \mathcal{R})$ is the Gaussian process constructed as in (SA-19) on a possibly enlarged probability space, and $(\mathcal{G} \times \mathcal{R})_\delta$ is chosen in Section SA-III.1.4. For all $t > 0$ and $0 < \delta < 1$,*

$$\mathbb{P}\big[\|R_n - R_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^R \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^R\|_{\mathcal{G}\times\mathcal{R}} > C_1 c_{\mathtt{v},\alpha} \mathtt{F}_n^R(t, \delta)\big] \leq \exp(-t),$$

where $c_{\mathtt{v},\alpha} = \mathtt{v}(1 + (2\alpha)^{\frac{\alpha}{2}})$ and

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

Proof of Lemma SA.26. Recall for any $g \in \mathcal{G}$, $r \in \mathcal{R}$,

$$R_n(g, r) = G_n(g, r) + X_n[\mathtt{p}_X(\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho))](g\,\theta(\cdot, r)).$$

Lemma SA.16 implies that for any $t > 0$ and $0 < \delta < 1$,

$$\mathbb{P}\big[\|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}} + \|Z_n^G \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta} - Z_n^G\|_{\mathcal{G}\times\mathcal{R}} > C_1 c_{\mathtt{v},\alpha} \mathtt{F}_n^R(t, \delta)\big] \leq \exp(-t).$$

For $g \in \mathcal{G}$, $r \in \mathcal{R}$, and take $(g_0, r_0) = \pi_{(\mathcal{G}\times\mathcal{R})_\delta}$, Jensen's inequality implies

$$\begin{aligned}
&\|X_n(g\,\theta(\cdot, r)) - X_n(g_0\,\theta(\cdot, r_0))\|_2^2 \\
&= \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(g(\mathbf{x}_i)\mathbb{E}[r(y_i)|\mathbf{x}_i] - g_0(\mathbf{x}_i)\mathbb{E}[r_0(y_i)|\mathbf{x}_i])^2] \\
&\leq \frac{1}{n}\sum_{i=1}^n \mathbb{E}[(g(\mathbf{x}_i)r(y_i) - g_0(\mathbf{x}_i)r_0(y_i))^2] = \|G_n(g, r) - G_n(g_0, r_0)\|_2^2.
\end{aligned}$$

Thus,

$$\|\|X_n(g\,\theta(\cdot, r)) - X_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}(g\,\theta(\cdot, r))\|_2\|_{\mathcal{G}\times\mathcal{R}} \leq \|\|G_n - G_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_2\|_{\mathcal{G}\times\mathcal{R}}.$$

Lemma SA.25 implies that if we define $\mathcal{G} \times \overline{\mathcal{R}} = \{g(r - \theta(\cdot, r)) : g \in \mathcal{G}, r \in \mathcal{R}\}$, then

$$\mathtt{N}_{\mathcal{G}\times\overline{\mathcal{R}}, \mathcal{X}\times\mathcal{Y}}(\delta, \mathtt{M}_\mathcal{G} M_\mathcal{R}) \leq 2\mathtt{N}(\delta).$$

The conclusion then follows by applying the same empirical process argument to $\|X_n(g\,\theta(\cdot, r)) - X_n \circ \pi_{(\mathcal{G}\times\mathcal{R})_\delta}\|_{\mathcal{G}\times\mathcal{R}}$ as in Lemma SA.16. □

SA-IV.1.4 Strong Approximation Errors

To simplify notation, the parameters of $\mathcal{G}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{X}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{X} \times \mathcal{Y}$, and the index $\mathcal{X} \times \mathcal{Y}$ is omitted where there is no ambiguity. Recall we also define

$$J(\delta) = \sqrt{2}J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}), \qquad \delta \in (0,1],$$
$$\mathtt{N}(\delta) = \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}})\mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}}), \qquad \delta \in (0,1].$$

Lemma SA.27. *Suppose Assumption SA.2 holds, a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, 1)$ is given, $(Z_n^R(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ and $(\Pi_2 Z_n^R(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ are the Gaussian processes constructed as in Equations (SA-13) and (SA-14) on a possibly enlarged probability space, and $(\mathcal{G} \times \mathcal{R})_\delta$ is chosen in Section SA-III.1.4. Then for all $t > 0$,*

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

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

Proof of Lemma SA.27. We have shown in the proof of Lemma SA.17 that for any $(g,r) \in \mathcal{G} \times \mathcal{R}$,

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

It then follows from the relation between $\gamma_{j,k}$ and $\eta_{j,k}$ in Equation (SA-17) that for any $(g,r) \in \mathcal{G} \times \mathcal{R}$,

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

and hence by Lemma SA.23, for any $x > 0$, with probability at least $1 - 2\exp(-x)$,

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

The conclusion then follows from a union bound on $(\mathcal{G} \times \mathcal{R})_\delta$. □

Lemma SA.28. *Suppose Assumption SA.2 holds, a cylindered quasi-dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)$ is given with $\rho > 1$, $(Z_n^R(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ and $(\Pi_2 Z_n^R(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ are the Gaussian processes constructed as in Equations (SA-13) and (SA-14) on a possibly enlarged probability space, and $(\mathcal{G} \times \mathcal{R})_\delta$ is chosen in Section SA-III.1.4. Then for all $t > 0$,*

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

where $C_\rho > 0$ is a constant that only depends on ρ and $c_{\mathtt{v},\alpha} = \mathtt{v}(1 + (2\alpha)^{\alpha/2})$.

Proof. Since $\mathcal{C}_{M,N}(\mathbb{P}_Z, \rho)$ is a cylindered quasi-dyadic expansion, $\rho^{-1}2^{-M-N+j} \leq \mathbb{P}_Z(\mathcal{C}_{j,k}) \leq \rho 2^{-M-N+j}$, 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.17, for any $g \in \mathcal{G}, r \in \mathcal{R}$,

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

The result then follows from Lemma SA.14. □

SA-IV.1.5 Projection Error

To simplify notation, the parameters of $\mathcal{G}$ and $\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}$ (Definitions 4 to 12, SA.1, SA.2) are taken with $\mathcal{C} = \mathcal{X}$, and the index $\mathcal{X}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{X} \times \mathcal{Y}$, and the index $\mathcal{X} \times \mathcal{Y}$ is omitted where there is no ambiguity. Recall we also define

$$J(\delta) = \sqrt{2} J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2} J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}), \qquad \delta \in (0,1],$$
$$\mathtt{N}(\delta) = \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}}) \mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}}), \qquad \delta \in (0,1].$$

The projection errors for the R_n and Z_n^R processes can be decomposed by the observation that, for any $g \in \mathcal{G}$ and $r \in \mathcal{R}$,

$$\Pi_2 R_n(g,r) - R_n(g,r) = \Big(\Pi_1 G_n(g,r) - G_n(g,r)\Big) - \Big(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] X_n(g\,\theta(\cdot,r)) - X_n(g\,\theta(\cdot,r))\Big),$$
$$\Pi_2 Z_n^R(g,r) - Z_n^R(g,r) = \Big(\Pi_1 Z_n^G(g,r) - Z_n^G(g,r)\Big) - \Big(\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] Z_n^X(g\,\theta(\cdot,r)) - Z_n^X(g\,\theta(\cdot,r))\Big), \quad \text{(SA-21)}$$

where, in each line, the first term in parentheses is the projection error for the G_n-process, as discussed in Section SA-III.1.6, and the second term is the projection error for the X_n-process, detailed in Section SA-II.1.6, with

$$X_n(g\,\theta(\cdot,r)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[g(\mathbf{x}_i)\theta(\mathbf{x}_i,r) - \mathbb{E}[g(\mathbf{x}_i)\theta(\mathbf{x}_i,r)]\Big],$$
$$\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] X_n(g\,\theta(\cdot,r)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})](g\,\theta(\cdot,r))(\mathbf{x}_i) - \mathbb{E}[\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})](g\,\theta(\cdot,r))(\mathbf{x}_i)]\Big].$$

This decomposition allows us to leverage previously established error bounds and convergence results for G_n and X_n processes, thus facilitating the analysis of the R_n and Z_n^R processes. By utilizing known results from Sections SA-III.1.6 and SA-II.1.6, this approach simplifies the treatment of the projection errors for these new processes.

Lemma SA.29. *Suppose Assumption SA.2 holds, a cylindered dyadic expansion $\mathcal{C}_{M,N}(\mathbb{P}_Z, 1)$ is given, $(Z_n^R(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ and $(\Pi_2 Z_n^R(g,r) : g \in \mathcal{G}, r \in \mathcal{R})$ are the Gaussian processes constructed as in Equations (SA-19) on a possibly enlarged probability space, and $(\mathcal{G} \times \mathcal{R})_\delta$ is chosen in Section SA-III.1.4. Suppose $\mathbb{P}_X$ admits a Lebesgue density f_X supported on $\mathcal{X} \subseteq \mathbb{R}^d$. Then for all $t > N$, with probability at* *least* $1 - 4\mathtt{N}(\delta) n e^{-t}$,

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

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

where $c_{\mathtt{v},\alpha} = \mathtt{v}(1 + (2\alpha)^{\frac{\alpha}{2}})$, $c_{\mathtt{v},2\alpha} = \mathtt{v}^2(1 + (4\alpha)^\alpha)$, *and*

$$\mathtt{V}_\mathcal{G} = \min\{2\mathtt{M}_\mathcal{G}, \mathtt{L}_\mathcal{G} \|\mathcal{V}_M\|_\infty\} \Big(\sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})\Big)^2 2^M \mathfrak{m}(\mathcal{V}_M) \|\mathcal{V}_M\|_\infty \mathtt{TV}_\mathcal{G}^*,$$

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

with $\mathcal{V}_M = \cup_{0\le l<2^M} (\mathcal{X}_{0,l} - \mathcal{X}_{0,l})$ *the upper level quasi-dyadic variation set as in Section SA-II.1.6.*

Proof. By Equations (SA-16) and (SA-18), we can show the decomposition in Equation (SA-21) holds. The terms $\Pi_1 G_n - G_n$ and $\Pi_1 Z_n^G - Z_n^G$ can be bounded from results in Lemma SA.21. Recall $\mathcal{G}\cdot\mathcal{V}_\mathcal{R} = \{g\,\theta(\cdot, r) : g \in \mathcal{G}, r \in \mathcal{R}\}$. We know from Lemma SA.9 for all $t > 0$,

$$\mathbb{P}\left(|\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] X_n(g\,\theta(\cdot,r)) - X_n(g\,\theta(\cdot,r))| \ge 2\sqrt{\mathtt{V}_{\mathcal{G}\cdot\mathcal{V}_\mathcal{R}} t} + \frac{4}{3}\cdot\frac{\mathtt{M}_{\mathcal{G}\cdot\mathcal{V}_\mathcal{R}}}{\sqrt{n}} t\right) \le 2\exp(-t),$$

$$\mathbb{P}\left(|\Pi_0[\mathtt{p}_X(\mathcal{C}_{M,N})] Z_n^X(g\,\theta(\cdot,r)) - Z_n^X(g\,\theta(\cdot,r))| \ge 2\sqrt{\mathtt{V}_{\mathcal{G}\cdot\mathcal{V}_\mathcal{R}} t}\right) \le 2\exp(-t).$$

Moreover, suppose $\alpha > 0$ in (iv) from Assumption SA.2, $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \le 2$, hence by moment properties of sub-Gaussian random variables,

$$\sup_{r\in\mathcal{R}} \sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|r(y_i)||\mathbf{x}_i = \mathbf{x}] \le \mathtt{v}(1 + \sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[|y_i|^\alpha|\mathbf{x}_i = \mathbf{x}]) \le c_{\mathtt{v},\alpha}.$$

Hence $\mathtt{M}_{\mathcal{G}\cdot\mathcal{V}_\mathcal{R}} \le c_{\mathtt{v},\alpha}\mathtt{M}_\mathcal{G}$. Suppose $\alpha = 0$ from (iv) from Assumption SA.2 holds, $\sup_{r\in\mathcal{R}} \sup_{\mathbf{x}\in\mathcal{X}} |r(\mathbf{x})| \le 2\mathtt{v}$, hence we also have $\mathtt{M}_{\mathcal{G}\cdot\mathcal{V}_\mathcal{R}} \le c_{\mathtt{v},\alpha}\mathtt{M}_\mathcal{G}$. The result then follows from a union bound over $(\mathcal{G}\times\mathcal{R})_\delta$. □

SA-IV.2 General Result

The following theorem presents a generalization of Theorem 2 in the paper. To simplify notation, the parameters of $\mathcal{G}$ and $\mathcal{G}\cdot\mathcal{V}_\mathcal{R}$ (Definitions 4 to 12, SA.1, SA.2) are taken with $\mathcal{C} = \mathcal{Q}_\mathcal{G}$, and the index $\mathcal{Q}_\mathcal{G}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G}\times\mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{Q}_\mathcal{G}\times\mathcal{Y}$, and the index $\mathcal{Q}_\mathcal{G}\times\mathcal{Y}$ is omitted where there is no ambiguity.

Theorem SA.2. *Suppose* $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \le i \le n)$ *are i.i.d. random vectors taking values in* $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$ *with common law* $\mathbb{P}_Z$*, where* $\mathbf{x}_i$ *has distribution* $\mathbb{P}_X$ *supported on* $\mathcal{X} \subseteq \mathbb{R}^d$*,* y_i *has distribution* $\mathbb{P}_Y$ *supported on* $\mathcal{Y} \subseteq \mathbb{R}$*, and the following conditions hold.*

(i) $\mathcal{G}$ *is a real-valued pointwise measurable class of functions on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.

(ii) *There exists a surrogate measure* $\mathbb{Q}_\mathcal{G}$ *for* $\mathbb{P}_X$ *with respect to* $\mathcal{G}$ *such that* $\mathbb{Q}_\mathcal{G} = \mathfrak{m}\circ\phi_\mathcal{G}$*, where the normalizing transformation* $\phi_\mathcal{G} : \mathcal{Q}_\mathcal{G} \mapsto [0,1]^d$ *is a diffeomorphism.*

(iii) $\mathtt{M}_{\mathcal{G}} < \infty$ *and* $J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, 1) < \infty$.

(iv) $\mathcal{R}$ *is a real-valued pointwise measurable class of functions on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$.

(v) $J(\mathcal{R}, M_{\mathcal{R}}, 1) < \infty$, *where* $M_{\mathcal{R}}(y) + \mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \leq \mathtt{v}(1 + |y|^{\alpha})$ *for all* $y \in \mathcal{Y}$, *for some* $\mathtt{v} > 0$, *and for some* $\alpha \geq 0$. *Furthermore, if* $\alpha > 0$, *then* $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

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 trajectories on* $(\mathcal{G} \times \mathcal{R}, \mathfrak{d}_{\mathbb{P}_X, \mathbb{P}_Y})$ *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* $(g_1, r_1), (g_2, r_2) \in \mathcal{G} \times \mathcal{R}$.
- $\mathbb{P}[\|R_n - Z_n^R\|_{\mathcal{G} \times \mathcal{R}} > C_1 C_{\mathtt{v},\alpha} \mathsf{T}_n^R(t)] \leq C_2 e^{-t}$ *for all* $t > 0$,

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

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

and

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

Proof of Theorem SA.2. To simplify notation, we will use $\mathbb{E}[\cdot|\mathcal{X}_{0,l}]$ in short for $\mathbb{E}[\cdot|\mathbf{x}_i \in \mathcal{X}_{0,l}]$, and $\mathbb{E}[\cdot|\mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in short for $\mathbb{E}[\cdot|(\mathbf{x}_i, y_i) \in \mathcal{X}_{0,l} \times \mathcal{Y}_{l,j,m}]$ in this proof.

First, we make a reduction via the surrogate measure and normalizing transformation. Since $\mathrm{Supp}(\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}) \subseteq \mathrm{Supp}(\mathcal{G})$, we know $\mathcal{Q}_{\mathcal{G}}$ is also a surrogate measure for $\mathbb{P}_X$ with respect to $\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}$, and $\phi_{\mathcal{G}}$ remains a valid normalizing transformation. By the same argument as in the proof for Theorem 1, assumption (ii) implies that on a possibly enriched probability space, there exists $(\mathbf{u}_i : 1 \leq i \leq n)$ i.i.d distributed with law $\mathbb{P}_U = \mathsf{Uniform}([0,1]^d)$, and

$$g(\mathbf{x}_i) = g(\phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)), \qquad \forall g \in \mathcal{G}, 1 \leq i \leq n,$$

and if $g(\mathbf{x}_i) \neq 0$ for any $g \in \mathcal{G}$, then $\mathbf{x}_i = \phi_{\mathcal{H}}^{-1}(\mathbf{u}_i)$, $1 \leq i \leq n$.

Define $\widetilde{R}_n$ to be the empirical process based on $((\mathbf{u}_i, y_i) : 1 \leq i \leq n)$, and

$$\widetilde{R}_n(f,s) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big[f(\mathbf{u}_i)s(y_i) - \mathbb{E}[f(\mathbf{u}_i)s(y_i)|\mathbf{u}_i] \Big],$$

and take $\widetilde{\mathcal{G}} = \{g \circ \phi_{\mathcal{H}}^{-1} : g \in \mathcal{G}\}$, then

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

The relation between constants for $\widetilde{\mathcal{R}}$ and constants for $\mathcal{R}$ can be deduced from Lemma SA.10. Hence, without loss of generality, we assume $(\mathbf{x}_i : 1 \leq i \leq n)$ are i.i.d under common law $\mathbb{P}_X = \mathsf{Uniform}([0,1]^d)$ distributed and $\mathcal{X} = [0,1]^d$.

Take $\mathcal{A}_{M,N}(\mathbb{P}_Z, 1)$ to be the axis-aligned cylindered quasi-dyadic expansion of $\mathbb{R}^{d+1}$. By Lemma SA.27 and Lemma SA.29, for all $t > N$,

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

$$\mathbb{P}\Big[\|R_n - \Pi_2 R_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_{\mathtt{v},\alpha} \sqrt{2N^2 \mathtt{V} + 2^{-N} \mathtt{M}_{\mathcal{G}}^2}\, t^{\alpha+\frac{1}{2}} + C_{\mathtt{v},\alpha} \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1} \Big] \leq 4\mathtt{N}(\delta) n e^{-t},$$

$$\mathbb{P}\Big[\|Z_n^R - \Pi_2 Z_n^R\|_{(\mathcal{G}\times\mathcal{R})_\delta} > C_{\mathtt{v},\alpha} \sqrt{2N^2 \mathtt{V} + 2^{-N} \mathtt{M}_{\mathcal{G}}^2}\, t^{\frac{1}{2}} + C_{\mathtt{v},\alpha} \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t \Big] \leq 4\mathtt{N}(\delta) n e^{-t},$$

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

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

Let $f \in \Pi_2(\mathcal{G}\times\mathcal{R})$. Then there exists $g \in \mathcal{G}$ 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.1.1), $\mathcal{C}_{j,k} = \mathcal{X}_{j-N,k} \times \mathcal{Y}_{*,N,0}$. 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

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

It follows that

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

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

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

It follows that

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

Moreover, for all (j,k), we have $\widetilde{\beta}_{j,k}(g,r) \lesssim C_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G}} N^{\alpha}$. Hence $\mathtt{C}_{\Pi_2(\mathcal{G} \times \mathcal{R})} \lesssim (C_{\mathtt{v},\alpha} \mathtt{M}_{\mathcal{G}} N^{\alpha})^2$. The rest of the proofs follow from choosing optimal M, N and Lemma SA.16 in the same way as in the proof for Theorem SA.1. □

SA-IV.3 Proof of Theorem 2

The proof follows by Theorem SA.2 with $\delta = n^{-1/2}$, and

$$\mathtt{N}(n^{-1/2}) = \mathtt{N}_{\mathcal{G}}(1/\sqrt{2n}, \mathtt{M}_{\mathcal{G}}) \mathtt{N}_{\mathcal{R}}(1/\sqrt{2n}, M_{\mathcal{R}}) \leq \mathtt{c}_{\mathcal{G}} \mathtt{c}_{\mathcal{R}} (2\sqrt{n})^{\mathtt{d}_{\mathcal{G}} + \mathtt{d}_{\mathcal{R}}} = \mathtt{c}(2\sqrt{n})^{\mathtt{d}},$$

and

$$J(n^{-1/2}) = \sqrt{2} J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, 1/\sqrt{2n}) + \sqrt{2} J(\mathcal{R}, M_{\mathcal{R}}, 1/\sqrt{2n})$$
$$\leq 3n^{-1/2} \sqrt{\mathtt{d}_{\mathcal{G}} \log(\mathtt{c}_{\mathcal{G}} \sqrt{n})} + 3\delta \sqrt{\mathtt{d}_{\mathcal{R}} \log(\mathtt{c}_{\mathcal{R}} \sqrt{n})}$$
$$\leq 3\delta \sqrt{(\mathtt{d}_{\mathcal{G}} + \mathtt{d}_{\mathcal{R}}) \log(\mathtt{c}_{\mathcal{G}} \mathtt{c}_{\mathcal{R}} n)} \leq 3\delta \sqrt{\mathtt{d} \log(\mathtt{c} n)}.$$

This completes the proof. □

SA-IV.4 Proof of Corollary 4

Take $t = C \log n$ with $C > 1$ in Theorem 2. □

SA-IV.5 Example: Local Polynomial Estimators

The following lemma provides sufficient conditions for the rate of *non-linearity error* and *smoothing bias* claimed in Section 4.1.

Lemma SA.30. *Consider the setup of Section 4.1. Recall we assume that* $((\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ *are i.i.d random vectors taking values in* $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$, *with* $\mathbf{x}_i \sim \mathbb{P}_X$ *admitting a continuous Lebesgue density* f_X *on its support* $\mathcal{X} = [0,1]^d$. *Assume in addition that* $\mathbf{w} \mapsto \theta(\mathbf{w}; r)$ *is* $(\mathfrak{p}+1)$*-times continuously differentiable with* $(\mathfrak{p}+1)$*th partial derivatives bounded uniformly over* $\mathbf{w} \in \mathcal{W} \subseteq \mathcal{X}$ *and* $r \in \mathcal{R}_l$, $l = 1, 2$, *for some* $\mathfrak{p} \geq 0$.

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

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

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

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

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

Proof of Lemma SA.30. We concisely flash out the arguments that are standard from the empirical process literature.

Convergence rate for each entry of $\widehat{\mathbf{H}}_{\mathbf{w}} - \mathbf{H}_{\mathbf{w}}$: Consider $\mathbf{u}_1^\top(\widehat{\mathbf{H}}_{\mathbf{w}} - \mathbf{H}_{\mathbf{w}})\mathbf{u}_2$, 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

$$g_n(\xi,\mathbf{w}) = \left(\frac{\xi - \mathbf{w}}{h}\right)^{\mathbf{v}} \frac{1}{h^d} K\left(\frac{\xi - \mathbf{w}}{h}\right), \qquad \xi\in\mathcal{X}, \mathbf{w}\in\mathcal{W}.$$

Define $\mathcal{F} = \{g_n(\cdot,\mathbf{w}) : \mathbf{w}\in\mathcal{W}\}$. Then $\sup_{\mathbf{w}\in\mathcal{W}}|\mathbf{u}_1^\top(\widehat{\mathbf{H}}_{\mathbf{w}} - \mathbf{H}_{\mathbf{w}})\mathbf{u}_2| = \sup_{f\in\mathcal{F}}|\mathbb{E}_n[f(\mathbf{x}_i)] - \mathbb{E}[f(\mathbf{x}_i)]|$. By standard arguments from kernel regression literature, we can show $\mathcal{F}$ forms a VC-type class over $\mathcal{X}$ with exponent d and constant $\|\mathcal{X}\|_\infty/b$, with $\mathtt{M}_{\mathcal{F},\mathcal{X}} = O(b^{-d})$, $\sigma_n^2 = \sup_{f\in\mathcal{F}}\mathbb{V}[f(\mathbf{x}_i)] = O(b^{-d/2})$. By Corollary 5.1 in Chernozhukov *et al.* (2014), we can show $\mathbb{E}[\sup_{f\in\mathcal{F}}|\mathbb{E}_n[f(\mathbf{x}_i)] - \mathbb{E}[f(\mathbf{x}_i)]|] = O((nb^d)^{-1/2}\sqrt{\log n} + (nb^d)^{-1}\log n)$. 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 for any multi-indices $\mathbf{u}_1, \mathbf{u}_2$ with $|\mathbf{u}_1|, |\mathbf{u}_2| \leq p$,

$$\sup_{\mathbf{w}\in\mathcal{W}}|\mathbf{u}_1^\top(\widehat{\mathbf{H}}_{\mathbf{w}} - \mathbf{H}_{\mathbf{w}})\mathbf{u}_2| = O((nb^d)^{-1/2}\sqrt{\log n} + (nb^d)^{-1}\log n), \text{ a.s.}$$

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

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

Convergence rate for $\sup_{\mathbf{w}\in\mathcal{W}}\sup_{r\in\mathcal{R}_\ell}\|\mathbf{S}_{\mathbf{w},r}\|$, $\ell = 1,2$: Consider $\mathbf{v}^\top\mathbf{S}_{\mathbf{w},r}$ where $|\mathbf{v}| \leq p$. Define $\mathcal{H}_\ell = \{(\mathbf{z},y) \mapsto g_n(\mathbf{z},\mathbf{w})(r(y) - \theta(\mathbf{z},r)) : \mathbf{w}\in\mathcal{W}, r\in\mathcal{R}_\ell\}$, $\ell = 1,2$. It is not hard to check both $\mathcal{H}_1$ and $\mathcal{H}_2$ are VC-type classes over $\mathcal{X}$. By similar arguments as in $\widehat{\mathbf{H}}_{\mathbf{w}} - \mathbf{H}_{\mathbf{w}}$, for all $t > 0$,

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

where C_2 is a constant that does not depend on n. 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 $\mathbf{S}_{\mathbf{w},r}$,

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

Putting together for Non-Linearity Errors:

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

Smoothing Error: Take $\mathbf{R}_{\mathbf{w},r} = \mathbb{E}_n\left[\mathbf{r}_p\left(\frac{\mathbf{X}_i-\mathbf{w}}{h}\right)K_h(\mathbf{X}_i - \mathbf{w})\mathfrak{r}_{\mathbf{w}}(\mathbf{X}_i; r)\right]$ where

$$\mathfrak{r}_{\mathbf{w}}(\xi; r) = \theta(\xi; r) - \sum_{0\leq|\boldsymbol{\nu}|\leq\mathfrak{p}} \frac{\partial_{\boldsymbol{\nu}}\theta(\mathbf{w}; r)}{\boldsymbol{\nu}!}(\xi - \mathbf{w})^{\boldsymbol{\nu}}.$$

Since all $\theta(\cdot; r), r \in \mathcal{R}_\ell$ are $(\mathfrak{p}+1)$-times continuously differentiable with derivatives bounded uniformly over $\mathcal{X}$ and $\mathcal{R}_\ell$, we have almost surely $\sup_{r\in\mathcal{R}_\ell}\sup_{\mathbf{w}\in\mathcal{W}} |\mathbf{R}_{\mathbf{w},r}| = O(b^{\mathfrak{p}+1})$, $\ell = 1, 2$. We have proved that $\inf_{\mathbf{w}\in\mathcal{W}} \sigma_d(\widehat{\mathbf{H}}_{\mathbf{w}}) = \Omega(1)$ a.s.. Hence

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

This completes the proof. □

The following two examples provide 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 the law of* $\mathbf{b}_i = (\mathbf{x}_i, u_i)$, $\mathbb{P}_B$, *has continuous and positive Lebesgue density* f_B *on its support* $\mathcal{B} = [0, 1]^{d+1}$.

(b) $\mathtt{M}_{\{\varphi\},\mathcal{B}} = O(1)$, $\mathtt{K}_{\{\varphi\},\mathcal{B}} = O(1)$, *and* $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\varphi\},\mathrm{Supp}(g)\times[0,1]} = O(\sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)))$.

(c) $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\mathcal{V}_{\mathcal{R}_l},\mathrm{Supp}(g)} = O(\sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)))$ *and* $\mathtt{K}_{\mathcal{V}_{\mathcal{R}_l},\mathcal{B}} = O(1)$, *for* $l = 1, 2$.

Recall $\mathcal{G} = \{b^{-d/2}\mathfrak{K}_{\mathbf{w}}(\frac{\cdot-\mathbf{w}}{b}) : \mathbf{w} \in \mathcal{W}\}$ *with* $\mathfrak{K}_{\mathbf{w}}(\mathbf{u}) = \mathbf{e}_1^\top\mathbf{H}_{\mathbf{w}}^{-1}\mathbf{p}(\mathbf{u})K(\mathbf{u})$. *For* $\mathcal{R}_1$, *take* $\mathcal{H}_1 = \{h \circ T_{\mathbb{P}_B}^{-1} : h \in \mathcal{H}_1^o\}$, *where* $\mathcal{H}_1^o = \{(\mathbf{x}, u) \in \mathcal{B} \mapsto g(\mathbf{x})\varphi(\mathbf{x}, u) - g(\mathbf{x})\theta(\mathbf{x}, \mathrm{Id}) : g \in \mathcal{G}\}$, $T_{\mathbb{P}_B}$ *is the Rosenblatt transformation* *based on $\mathbb{P}_B$ given in Section 3.1. Recall we denote $\mathcal{X} = [0,1]^d$. Then,*

$$\begin{aligned}
\mathtt{M}_{\mathcal{H}_1,\mathcal{B}} &\leq \mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{M}_{\{\varphi\},\mathcal{B}} = O(b^{-d/2}),\\
\mathtt{TV}_{\mathcal{H}_1,\mathcal{B}} &\leq \frac{\overline{f}_B^2}{\underline{f}_B}(\mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g \in \mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g))) = O(b^{d/2-1}),\\
\mathtt{K}_{\mathcal{H}_1,\mathcal{B}} &\leq (2\sqrt{d})^{d-1} \frac{\overline{f}_B^{d+1}}{\underline{f}_B^d}(\mathtt{M}_{\{\varphi\},\mathcal{B}} \mathtt{K}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{K}_{\{\varphi\},\mathcal{B}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{K}_{\mathcal{V}_{\mathcal{R}_1},\mathcal{X}}) = O(b^{-d/2}),\\
\mathtt{N}_{\mathcal{H}_1,\mathcal{B}}(\delta, \mathtt{M}_{\mathcal{H}_1,\mathcal{B}}) &= O(\delta^{-d-1}), \qquad 0 < \delta < 1,
\end{aligned} \tag{SA-22}$$

where $\overline{f}_B = \sup_{\mathbf{x} \in \mathcal{B}} f_B(\mathbf{x})$ and $\underline{f}_B = \inf_{\mathbf{x} \in \mathcal{B}} f_B(\mathbf{x})$. Rio (1994, Theorem 1.1) implies that $(X_n(h) : h \in \mathcal{H}_1) = (\sqrt{nb^d}\mathbf{e}_1^\top \mathbf{H}_{\mathbf{w}}^{-1} \mathbf{S}_{\mathbf{w},r} : \mathbf{w} \in \mathcal{W}, r \in \mathcal{R}_1)$ admits a uniform Gaussian strong approximation with rate

$$\mathsf{S}_n(t) = C_{d,\varphi,\mathbb{P}_B}(nb^{d+1})^{-1/(2d+2)}\sqrt{t + d\log n} + C_{d,\varphi,\mathbb{P}_B}(nb^d)^{-1/2}(t + d\log n),$$

where $C_{d,\varphi,\mathbb{P}_B}$ is a quantity that only depends on d, φ and $\mathbb{P}_B$.

For $\mathcal{R}_2$, take $\mathcal{H}_2 = \{h \circ T_{\mathbb{P}_B}^{-1} : h \in \mathcal{H}_2^o\}$, where $\mathcal{H}_2^o = \{(\mathbf{x}, u) \in \mathcal{B} \mapsto g(\mathbf{x}) r \circ \varphi(\mathbf{x}, u) - g(\mathbf{x})\theta(\mathbf{x}, r) : g \in \mathcal{G}, r \in \mathcal{R}_2\}$. Then

$$\begin{aligned}
\mathtt{M}_{\mathcal{H}_2} &= \mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{M}_{\{\varphi\},\mathcal{B}} = O(b^{-d/2}),\\
\mathtt{TV}_{\mathcal{H}_2} &\leq \frac{\overline{f}_B^2}{\underline{f}_B}(\mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{E}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g \in \mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g))) \max\{\mathtt{L}_{\{\varphi\},\mathcal{B}}, 1\}^{d-1} = O(b^{d/2-1}),\\
\mathtt{N}_{\mathcal{H}_2,\mathcal{B}}(\delta, \mathtt{M}_{\mathcal{H}_2,\mathcal{B}}) &= O(\delta^{-d-1}), \qquad 0 < \delta < 1.
\end{aligned}$$

Rio (1994, Theorem 1.1) implies that $(X_n(h) : h \in \mathcal{H}_2) = (\sqrt{nb^d}\mathbf{e}_1^\top \mathbf{H}_{\mathbf{w}}^{-1} \mathbf{S}_{\mathbf{w},r} : \mathbf{w} \in \mathcal{W}, r \in \mathcal{R}_2)$ admits a Gaussian strong approximation with rate function

$$\mathsf{S}_n(t) = C_{d,\varphi,\mathbb{P}_B}(nb^{d+1})^{-1/(2d+2)}\sqrt{t + d\log n} + C_{d,\varphi,\mathbb{P}_B}\sqrt{\frac{\log n}{nb^d}}(t + d\log n),$$

where $C_{d,\varphi,\mathbb{P}_B}$ is a quantity that only depends on d, φ and $\mathbb{P}_B$.

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

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

(1) Properties of $\mathcal{G}$ Since $\sup_{\mathbf{w} \in \mathcal{W}} \|\mathbf{H}_{\mathbf{w}}^{-1}\| \lesssim 1$ and K is continuous with compact support, we know

$$\mathtt{M}_{\mathcal{G},\mathcal{X}} = O(b^{-d/2}).$$

By a change of variables, we can show

$$\mathtt{E}_{\mathcal{G},\mathcal{X}} = \sup_{\mathbf{w} \in \mathcal{W}} \mathbb{E}\left[\left|b^{-d/2}\mathfrak{K}_{\mathbf{w}}\left(\frac{\mathbf{x}_i - \mathbf{w}}{b}\right)\right|\right] = O(b^{d/2}).$$

And $\sup_{\mathbf{w}\in\mathcal{W}} \sup_{\mathbf{u},\mathbf{u}'} \frac{|\mathbf{r}_p(\frac{\mathbf{u}-\mathbf{w}}{b})-\mathbf{r}_p(\frac{\mathbf{u}'-\mathbf{w}}{b})|}{\|\mathbf{u}-\mathbf{u}'\|_\infty} = O(b^{-1})$, and $\sup_{\mathbf{w}\in\mathcal{W}} \sup_{\mathbf{u},\mathbf{u}'} \frac{|K(\frac{\mathbf{u}-\mathbf{w}}{b})-K(\frac{\mathbf{u}'-\mathbf{w}}{b})|}{\|\mathbf{u}-\mathbf{u}'\|_\infty} = O(b^{-1})$, hence

$$\mathtt{L}_{\mathcal{G},\mathcal{X}} = O(b^{-\frac{d}{2}-1}).$$

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

$$\mathtt{TV}_{\mathcal{G},\mathcal{X}} \leq \mathtt{L}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)) = O(b^{\frac{d}{2}-1}).$$

Now we will show that $\mathtt{M}_{\mathcal{G},\mathcal{X}}^{-1}\mathcal{G}$ is a VC-type class. We know $\sup_{\mathbf{w},\mathbf{w}'\in\mathcal{W}}\|\mathbf{H}_{\mathbf{w}} - \mathbf{H}_{\mathbf{w}'}\|/\|\mathbf{w}-\mathbf{w}'\|_\infty = O(b^{-1})$. Since $\inf_{\mathbf{w}\in\mathcal{W}}\|\mathbf{H}_{\mathbf{w}}\| = \Omega(1)$, we also have $\sup_{\mathbf{w},\mathbf{w}'\in\mathcal{W}}\|\mathbf{H}_{\mathbf{w}}^{-1} - \mathbf{H}_{\mathbf{w}'}^{-1}\|/\|\mathbf{w}-\mathbf{w}'\|_\infty = O(b^{-1})$. It follows that

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

To upper bound $\mathtt{K}_{\mathcal{G},\mathcal{X}}$, let $\mathcal{D} \subseteq \mathcal{X}$ be a cube with edges of length $\mathtt{a}$ parallel to the coordinate axises. Consider the following two cases: (i) if $\mathtt{a} < b$, then $\mathtt{TV}_{\mathcal{G},\mathcal{D}} \leq C_K b^{-d/2-1}\mathtt{a}^d \leq C_K b^{-d/2}\mathtt{a}^{d-1}$; (ii) if $\mathtt{a} > b$, then $\mathtt{TV}_{\mathcal{G},\mathcal{D}} \leq C_K \sup_{\mathbf{w}\in\mathcal{W}} \mathfrak{m}(\mathrm{Supp}(\mathfrak{K}_{\mathbf{w}}))\mathtt{L}_{\mathcal{G},\mathcal{X}} \leq C_K b^d b^{-d/2-1} \leq C_K b^{-d/2} b^{d-1} \leq C_K b^{-d/2}\mathtt{a}^{d-1}$. This shows

$$\mathtt{K}_{\mathcal{G},\mathcal{X}} \leq C_K b^{-d/2}.$$

Consider $h_{\mathbf{w}}(\cdot) = \sqrt{b^d}\mathbf{e}_1^T\mathbf{H}_{\mathbf{w}}^{-1}\mathbf{r}_p(\cdot)K(\cdot)$, $\mathbf{w}\in\mathcal{W}$. Then $b^{-d/2}\mathfrak{K}_{\mathbf{w}}(\frac{\cdot-\mathbf{w}}{b}) = h_{\mathbf{w}}\left(\frac{\cdot-\mathbf{w}}{b}\right)$, $\mathbf{w}\in\mathcal{W}$. Recall that $\mathbf{x}_i$ has common law $\mathbb{P}_X$ with Lebesgue density f_X. Then there exists a constant $\mathbf{c}$ only depending on $\sup_{\mathbf{x}\in\mathcal{X}} K(\mathbf{x})$, $\mathtt{L}_{\{K\},\mathcal{X}}$, $\sigma_K = (\int K(\mathbf{x})d\mathbf{x})^{1/2}$, $\overline{f}_X = \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})$, $\underline{f}_X = \inf_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})$ that

$$\begin{aligned}
&\sup_{\mathbf{w}\in\mathcal{W}} \|h_{\mathbf{w}}\|_\infty \leq \mathbf{c},\\
&\sup_{\mathbf{w}\in\mathcal{W}} \sup_{\mathbf{u},\mathbf{v}\in\mathcal{W}} \frac{|h_{\mathbf{w}}(\mathbf{u}) - h_{\mathbf{w}}(\mathbf{v})|}{\|\mathbf{u}-\mathbf{v}\|_\infty} \leq \mathbf{c},\\
&\sup_{\mathbf{w},\mathbf{w}'\in\mathcal{W}} \sup_{\mathbf{u}\in\mathcal{W}} \frac{|h_{\mathbf{w}}(\mathbf{u}) - h_{\mathbf{w}'}(\mathbf{u})|}{\|\mathbf{w}-\mathbf{w}'\|_\infty} \leq \mathbf{c}.
\end{aligned}$$

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

$$N_{\mathcal{X}}(\mathtt{M}_{\mathcal{G},\mathcal{X}}^{-1}\mathcal{G}, \|\cdot\|_{\mathbb{P}_X,2}, \delta) \leq \mathbf{c}\delta^{-2d-2} + 1.$$

(2) Properties of $\mathcal{H}_1^o$ Let $g \in \mathcal{G}$. Take $\mathcal{H}_1^a = \{g\cdot\varphi : g\in\mathcal{G}\}$ and $\mathcal{H}_1^b = \{g\cdot\theta(\cdot,\mathrm{Id}) : g\in\mathcal{G}\}$. Then

$$\mathtt{M}_{\mathcal{H}_1^a,\mathcal{B}} \leq \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{M}_{\{\varphi\},\mathcal{B}}.$$

We have shown that all functions in $\mathcal{G}$ are Lipschitz and $\mathtt{L}_{\mathcal{G},\mathcal{X}} = O(b^{-d/2-1})$, Ambrosio *et al.* (2000, Proposition 3.2 (b)) then implies

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

Let $\mathcal{C}$ be any cube of side-length a in $\mathbb{R}^{d+1}$. By Ambrosio *et al.* (2000, Proposition 3.2 (b)),

$$\mathtt{TV}_{\mathcal{H}_1^a,\mathcal{C}} \leq \mathtt{M}_{\{\varphi\},\mathcal{B}}\mathtt{TV}_{\mathcal{G},\mathcal{C}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathrm{TV}_{\{\varphi\},\mathrm{Supp}(g)\times[0,1]\cap\mathcal{C}} \leq \mathtt{M}_{\{\varphi\},\mathcal{B}}\mathtt{K}_{\mathcal{G},\mathcal{X}}a^d + \mathtt{K}_{\{\varphi\},\mathcal{B}}\mathtt{M}_{\mathcal{G},\mathcal{X}}a^d,$$

which implies

$$\mathtt{K}_{\mathcal{H}_1^a,\mathcal{B}} \leq \mathtt{M}_{\{\varphi\},\mathcal{B}}\mathtt{K}_{\mathcal{G},\mathcal{X}} + \mathtt{K}_{\{\varphi\},\mathcal{B}}\mathtt{M}_{\mathcal{G},\mathcal{X}}.$$

Similar argument shows

$$\mathtt{M}_{\mathcal{H}_1^b,\mathcal{X}} \leq \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{M}_{\{\varphi\},\mathcal{B}}, \qquad \mathtt{TV}_{\mathcal{H}_1^b,\mathcal{X}} \leq \mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\theta(\cdot,\mathrm{Id})\},\mathrm{Supp}(g)},$$
$$\mathtt{K}_{\mathcal{H}_1^b,\mathcal{X}} \leq \mathtt{M}_{\{\varphi\},\mathcal{B}}\mathtt{K}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{K}_{\{\theta(\cdot,\mathrm{Id})\},\mathcal{X}}.$$

Then by assumptions $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\varphi\},\mathrm{Supp}(g)\times[0,1]} = O(\sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)))$ and $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\theta(\cdot,\mathrm{Id})\},\mathrm{Supp}(g)} = O(\sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)))$, we have

$$\mathtt{M}_{\mathcal{H}_1^o} \leq \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{M}_{\{\varphi\},\mathcal{B}}, \qquad \mathtt{TV}_{\mathcal{H}_1^o} = O(\mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g))),$$
$$\mathtt{K}_{\mathcal{H}_1^o} \leq \mathtt{M}_{\{\varphi\},\mathcal{B}}\mathtt{K}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{K}_{\{\varphi\},\mathcal{B}} + \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{K}_{\{\theta(\cdot,\mathrm{Id})\},\mathcal{X}}.$$

By standard empirical process argument, $\mathcal{H}_1^o$ is a VC-type class with constant $\mathbf{c}2^{d+1}$ and exponent $2d+2$ with respect to envelope function $\mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{M}_{\{\varphi\},\mathcal{B}}$ over $\mathcal{B}$.

(3) Properties of $\mathcal{H}_2^o$ The main challenge is that $\mathcal{R}_2$ contains non-differentiable indicator. First, we study properties of $\mathcal{G}\cdot\mathcal{R}_2$. By Definition 4,

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

where $\mathcal{D}_{d+1}([0,1]^{d+1})$ denotes the space of infinitely differentiable functions from $[0,1]^{d+1}$ to $\mathbb{R}^{d+1}$, and $\mathcal{D}_d([0,1]^d)$ is analogously defined. Similar argument as in the proof for properties of $\mathcal{H}_1^o$ gives

$$\mathtt{TV}_{\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}},\mathcal{B}} \leq \mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathtt{TV}_{\mathcal{V}_{\mathcal{R}},\mathrm{Supp}(g)} = O(\mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g))).$$

It follows that

$$\mathtt{TV}_{\mathcal{G}\cdot\mathcal{R}_2+\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}},\mathcal{B}} = O(\mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{E}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g))).$$

Consider the change of variables function $T : [0,1]^{d+1} \to \mathbb{R}^{d+1}$ given by $T(\mathbf{x}, u) = (\mathbf{x}, \varphi(\mathbf{x}, u))$. Since $\mathtt{L}_{\{T\},\mathcal{B}} \leq \max\{\mathtt{L}_{\{\varphi\},\mathcal{B}}, 1\}$, Theorem 3.16 from Ambrosio *et al.* (2000) implies

$$\mathtt{TV}_{\mathcal{H}_2^o,\mathcal{B}} \leq \mathtt{L}_{\{T\},\mathcal{B}}^{d-1} \mathtt{TV}_{\mathcal{G}\cdot\mathcal{R}_2+\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}},\mathcal{B}} = O(\max\{\mathtt{L}_{\{\varphi\},\mathcal{B}}, 1\}^{d-1} \mathtt{TV}_{\mathcal{G}\cdot\mathcal{R}_2+\mathcal{G}\cdot\mathcal{V}_{\mathcal{R}},\mathcal{B}}).$$

By standard empirical process argument, $\mathcal{H}_2^o$ is a VC-type class with constant $C_1 2^{d+1}$ and exponent $2d+2$ with respect to envelope function $\mathtt{M}_{\mathcal{G},\mathcal{X}}$, where C_1 is a constant that does not depend on n.

(4) Effects of Rosenblatt Transformation By Lemma SA.10 with $\mathbb{Q}_{\mathcal{G}} = \mathbb{P}_X$ and $\phi_{\mathcal{G}} = \mathrm{Id}$, we have $\mathtt{TV}_{\mathcal{H}_1} \leq \mathtt{TV}_{\mathcal{H}_1^o} \overline{f}_B^2 \underline{f}_B^{-1}$, $\mathtt{TV}_{\mathcal{H}_2} \leq \mathtt{TV}_{\mathcal{H}_2^o} \overline{f}_B^2 \underline{f}_B^{-1}$, $\mathtt{M}_{\mathcal{H}_1} = \mathtt{M}_{\mathcal{H}_1^o}$, $\mathtt{M}_{\mathcal{H}_2} = \mathtt{M}_{\mathcal{H}_2^o}$. Moreover, $\mathcal{H}_1$ and $\mathcal{H}_2$ are VC-type classes with constant $C_2 2^{d+1}$ and exponent $2d+2$ with respect to envelope functions $\mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{M}_{\{\varphi\},\mathcal{B}}$ and $\mathtt{M}_{\mathcal{G},\mathcal{X}}$ respectively, with C_2 a constant that does not depend on n.

(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

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

where $C_{d,\varphi}$ 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

$$C_{d,\varphi} \sqrt{\frac{d\overline{f}_B^2}{\underline{f}_B}} \frac{\sqrt{\mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{TV}_{\mathcal{G},\mathcal{X},\{\varphi\},\mathcal{B}}}}{n^{\frac{1}{2d+2}}} \sqrt{t + d\log n} + C_{d,\varphi} \frac{\mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{M}_{\{\varphi\},\mathcal{B}}}{\sqrt{n}} (t + d\log n),$$

where $\mathtt{TV}_{\mathcal{G},\mathcal{X},\{\varphi\},\mathcal{B}} = \max\{\mathtt{L}_{\{\varphi\},\mathcal{B}}, 1\}^{d-1} (\mathtt{TV}_{\mathcal{G},\mathcal{X}} + \mathtt{E}_{\mathcal{G},\mathcal{X}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)))$. □

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 the law of* $\mathbf{b}_i = (\mathbf{x}_i, u_i)$, $\mathbb{P}_B$, *has a continuous and positive Lebesgue density* f_B *on its support* $\mathcal{B} = [0,1]^{d+1}$.

(b) $\mathtt{M}_{\{\varphi\},\mathcal{B}} = O(1)$, $\sup_{g\in\mathcal{G}} \mathtt{TV}_{\{\varphi\},\mathrm{Supp}(g)} = O(\sup_{g\in\mathcal{G}} \mathfrak{m}(\mathrm{Supp}(g)))$, $\mathtt{K}_{\{\varphi\},\mathcal{B}} = O(1)$, *and* $\mathtt{L}_{\{\varphi\},\mathcal{B}} = O(1)$.

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

Recall $\mathcal{G} = \{b^{-d/2} \mathfrak{K}_{\mathbf{w}}(\frac{\cdot - \mathbf{w}}{b}) : \mathbf{w} \in \mathcal{W}\}$ *with* $\mathfrak{K}_{\mathbf{w}}(\mathbf{u}) = \mathbf{e}_1^\top \mathbf{H}_{\mathbf{w}}^{-1} \mathbf{p}(\mathbf{u}) K(\mathbf{u})$. *For* $\mathcal{R}_1$, *take* $\mathcal{H}_1 = \{h \circ T_{\mathbb{P}_B}^{-1} : h \in \mathcal{H}_1^o\}$, *where* $\mathcal{H}_1^o = \{(\mathbf{x}, u) \in \mathcal{B} \mapsto g(\mathbf{x})\varphi(\mathbf{x}, u) - g(\mathbf{x})\theta(\mathbf{x}, \mathrm{Id}) : g \in \mathcal{G}\}$, $T_{\mathbb{P}_B}$ *is the Rosenblatt transformation based on* $\mathbb{P}_B$ *given in Section 3.1. Recall we denote* $\mathcal{X} = [0,1]^d$. *Then, Equation* (SA-22) *holds, and*

$$\mathtt{L}_{\mathcal{H}_1} \leq \mathtt{L}_{\mathcal{H}_1^o} \frac{\overline{f}_B}{\underline{f}_B} \leq (\mathtt{L}_{\mathcal{G},\mathcal{X}} \mathtt{M}_{\{\varphi\},\mathcal{B}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{L}_{\{\varphi\},\mathcal{B}} + \mathtt{M}_{\mathcal{G},\mathcal{X}} \mathtt{L}_{\mathcal{V}_{\mathcal{R}_1},\mathcal{X}}) \frac{\overline{f}_B}{\underline{f}_B} = O(b^{-d/2-1}).$$

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

$$\mathsf{S}_n(t) = C_{d,\varphi,\mathbb{P}_B}(nb^{d+1})^{-1/(d+1)}\sqrt{t + d\log n} + C_{d,\varphi,\mathbb{P}_B}(nb^d)^{-1/2}(t + d\log n),$$

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

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

Proof of Example SA.2. Besides the properties given in the proof of Example SA.1, using product rule we can show $\mathtt{L}_{\mathcal{H}_1^\circ} \leq \mathtt{L}_{\mathcal{G},\mathcal{X}}\mathtt{M}_{\{\varphi\},\mathcal{B}} + \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{L}_{\{\varphi\},\mathcal{B}} + \mathtt{M}_{\mathcal{G},\mathcal{X}}\mathtt{L}_{\mathcal{V}_{\mathcal{R}_1},\mathcal{X}} = O(b^{-d/2-1})$. By the discussion on Rosenblatt transformation in Section 3.1, $\mathtt{L}_{\mathcal{H}_1,\mathcal{X}} \leq \mathtt{L}_{\mathcal{H}_1^\circ,\mathcal{X}}\overline{f}_B/\underline{f}_B$. Take the surrogate measure to be $\mathcal{Q}_{\mathcal{H}_1} = \mathsf{Uniform}([0,1]^{d+1})$ with $\phi_{\mathcal{H}_1} = \mathrm{Id}$. The result then follows from application of Theorem SA.1 on

$$X_n(h) = \frac{1}{\sqrt{n}}\sum_{i=1}^n [h(\mathbf{x}_i, u_i) - \mathbb{E}[h(\mathbf{x}_i, u_i)]], \qquad h \in \mathcal{H}_1.$$

This completes the proof. □

Example SA.3 (Strong Approximation via Theorem 2)**.** *Consider the setup of Section 4.1 and assume the following regularity conditions hold:*

(a) $\mathbf{x}_i$ *has* $\mathbb{P}_X$ *with Lebesgue density* f_X *continuous on its support* $\mathcal{X}$*, which is a compact subset of* $\mathbb{R}^d$*, and* $\sup_{\mathbf{x}\in\mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

(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* $\ell = 1, 2$.

Recall that $\mathcal{G} = \{b^{-d/2}\mathfrak{K}_\mathbf{x}(\frac{\cdot - \mathbf{x}}{b}) : \mathbf{x} \in \mathcal{X}\}$. *Take the surrogate measure* $\mathbb{Q}_\mathcal{G} = \mathbb{P}_X$ *and the normalizing transformation* $\phi_\mathcal{G} = \mathrm{Id}$. *Then, using the notation introduced in the paper,*

$$\mathtt{c}_1 = d\frac{\overline{f}_X^2}{\underline{f}_X}, \qquad \mathtt{c}_2 = \frac{\overline{f}_X}{\underline{f}_X},$$

where $\overline{f}_X = \sup_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})$, $\underline{f}_X = \inf_{\mathbf{x}\in\mathcal{X}} f_X(\mathbf{x})$, *and*

$$\mathtt{M}_\mathcal{G} = O(b^{-d/2}), \quad \mathtt{E}_\mathcal{G} = O(b^{d/2}), \quad \mathtt{TV}_\mathcal{G} = O(b^{d/2-1}), \quad \mathtt{L}_\mathcal{G} = O(b^{-d/2-1}),$$
$$\mathtt{N}_\mathcal{G}(\delta) = O(\delta^{-d-1}), \quad 0 < \delta < 1.$$

Theorem 2 implies that $(R_n(g, r) : g \in \mathcal{G}, r \in \mathcal{R}_1) = (\sqrt{nb^d}\mathbf{e}_1^\top \mathbf{H}_\mathbf{x}^{-1}\mathbf{S}_{\mathbf{w},r} : \mathbf{x} \in [0,1]^d, r \in \mathcal{R}_1)$ *admits a uniform Gaussian strong approximation with rate function*

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

If, in addition, $\sup_{\mathbf{x}\in[0,1]^d} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$*, then Theorem 2 implies* $(R_n(g, r) : g \in \mathcal{G}, r \in \mathcal{R}_1) =$

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

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

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

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 the conditions in Theorem 2 with $\mathtt{c}_{\mathcal{R}_1} = 1$, $\mathtt{d}_{\mathcal{R}_1} = 1$ and $\alpha = 1$. Moreover, $\mathcal{R}_2$ satisfies the conditions in Theorem 2 with $\mathtt{c}_{\mathcal{R}_2}$ some universal 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 2. □

SA-V Quasi-Uniform Haar Basis

This section provides the proofs and additional results for Section 5. In Section SA-V.1, we present the proofs of Theorem 3 and Corollary 5, and verify the claims for Example 2. In Section SA-V.2, we present the proofs of Theorem 4, Corollary 6 and the Haar Partitioning-based Regression example in Section 5.3, with the additional results for M_n and R_n processes under generic entropy conditions.

SA-V.1 General Empirical Process

SA-V.1.1 Proof of Theorem 3

First, we make a reduction through the surrogate measure $\mathbb{Q}_\mathcal{H}$ (Definition 2). Denote $\mathcal{E}_\mathcal{H} = \mathcal{X} \cap \mathrm{Supp}(\mathcal{H})$. The definition of surrogate measure implies $\mathbb{P}_X|_{\mathcal{E}_\mathcal{H}} = \mathbb{Q}_\mathcal{H}|_{\mathcal{E}_\mathcal{H}}$. We use a coupling argument similar to the proof of Theorem 1. Define a probability measure $\mathbb{O}$ on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^{2d}))$ such that for all $A \in \mathcal{B}(\mathbb{R}^{2d})$, $\mathbb{O}|_{\mathcal{E}_\mathcal{H}\times\mathcal{E}_\mathcal{H}}(A) = \mathbb{P}_X(\Pi_{1:d}(A \cap \{(\mathbf{x},\mathbf{x}) : \mathbf{x} \in \mathcal{E}_\mathcal{H}\}))$, $\mathbb{O}|_{\mathcal{E}_\mathcal{H}\times\mathcal{E}_\mathcal{H}^c}(A) = \mathbb{O}|_{\mathcal{E}_\mathcal{H}^c\times\mathcal{E}_\mathcal{H}}(A) = 0$, $\mathbb{O}|_{\mathcal{E}_\mathcal{H}^c\times\mathcal{E}_\mathcal{H}^c}(A) = \int_{\mathcal{E}_\mathcal{H}^c} \mathbb{P}_X(A^\mathbf{y} \cap \mathcal{E}_\mathcal{H}^c) d\mathbb{Q}(\mathbf{y})$ where $A^\mathbf{y} = \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x},\mathbf{y}) \in A\}$, where we take $\Pi_{1:d}(E) = \{\mathbf{x} \in \mathbb{R}^d : (\mathbf{x},\mathbf{y}) \in E \text{ for some } \mathbf{y} \in \mathbb{R}^d\}$ for any $E \in \mathcal{B}(\mathbb{R}^{2d})$.

The definition of $\mathbb{O}$ implies the marginals are $\mathbb{P}_X$ and $\mathbb{Q}_\mathcal{H}$, respectively. By Skorohod embedding (Dudley, 2014, Lemma 3.35), on a possibly enlarged probability space, there exists $(\mathbf{z}_i : 1 \le i \le n)$ i.i.d. with law $\mathbb{Q}_\mathcal{H}$ such that $(\mathbf{x}_i, \mathbf{z}_i)$ has joint law $\mathbb{O}$ for each $1 \le i \le n$. In particular, when $\mathbf{x}_i \in \mathcal{E}_\mathcal{H}$, $\mathbf{z}_i = \mathbf{x}_i$; and $\mathbb{O}(\{\mathbf{x}_i \in \mathcal{E}_\mathcal{H}^c\} \triangle \{\mathbf{z}_i \in \mathcal{E}_\mathcal{H}^c\}) = 0$. Moreover, since $\mathbb{Q}_\mathcal{H}(\mathrm{Supp}(\mathcal{H}) \setminus \mathcal{X}) = 0$, and the definition of $\mathbb{O}$ on $\mathcal{E}_\mathcal{H} \times \mathcal{E}_\mathcal{H}$ as a product measure between $\mathbb{P}_X$ and $\mathbb{Q}_\mathcal{H}$, we know $\mathbb{O}(\{\mathbf{x}_i \in \mathcal{E}_\mathcal{H}^c\} \triangle \{\mathbf{z}_i \in \mathrm{Supp}(\mathcal{H})^c\}) = 0$. This allows for the reduction to $\mathbf{z}_i$-based processes, since for $1 \le i \le n$, almost surely

$$\begin{aligned}
h(\mathbf{x}_i) &= h(\mathbf{x}_i)\mathbb{1}(\mathbf{x}_i \in \mathcal{E}_\mathcal{H}) + 0 \cdot \mathbb{1}(\mathbf{x}_i \in \mathcal{E}_\mathcal{H}^c) \\
&= h(\mathbf{z}_i)\mathbb{1}(\mathbf{z}_i \in \mathcal{E}_\mathcal{H}) + 0 \cdot \mathbb{1}(\mathbf{z}_i \in \mathrm{Supp}(\mathcal{H})^c) \\
&= h(\mathbf{z}_i)\mathbb{1}(\mathbf{z}_i \in \mathcal{E}_\mathcal{H}) + h(\mathbf{z}_i) \cdot \mathbb{1}(\mathbf{z}_i \in \mathrm{Supp}(\mathcal{H})^c) \\
&= h(\mathbf{z}_i)\mathbb{1}(\mathbf{z}_i \in \mathcal{E}_\mathcal{H}) + h(\mathbf{z}_i) \cdot \mathbb{1}(\mathbf{z}_i \in \mathcal{E}_\mathcal{H}^c) \\
&= h(\mathbf{z}_i), \qquad \forall h \in \mathcal{H},
\end{aligned}$$

where the first line is due to $h = 0$ on $\text{Supp}(\mathcal{H})^c$, the second line is by $\mathbb{O}(\{\mathbf{x}_i \in \mathcal{E}_{\mathcal{H}}^c\} \triangle \{\mathbf{z}_i \in \text{Supp}(\mathcal{H})^c\}) = 0$, the third line is due to $h = 0$ on $\text{Supp}(\mathcal{H})^c$, and the fourth line by $\mathbb{Q}_{\mathcal{H}}(\text{Supp}(\mathcal{H}) \setminus \mathcal{X}) = 0$. Almost surely,

$$X_n(h) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \big[h(\mathbf{x}_i) - \mathbb{E}[h(\mathbf{x}_i)]\big] = \frac{1}{\sqrt{n}} \sum_{i=1}^n \big[h(\mathbf{z}_i) - \mathbb{E}[h(\mathbf{z}_i)]\big], \qquad \forall h \in \mathcal{H}.$$

Hence we reduce the problem to coupling for $(\widetilde{X}_n(h) : h \in \mathcal{H})$, with the process defined by

$$\widetilde{X}_n(h) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \big[h(\mathbf{z}_i) - \mathbb{E}[h(\mathbf{z}_i)]\big], \qquad h \in \mathcal{H},$$

with $(\mathbf{z}_i : 1 \le i \le n)$ i.i.d $\sim \mathbb{Q}_{\mathcal{H}}$. Suppose $2^K \le L < 2^{K+1}$. For each $l \in \{1, 2, \ldots, d\}$, we can divide at most 2^K cells into two intervals of equal measure under $\mathbb{Q}_{\mathcal{H}}$ such that we get a new partition of $\mathcal{Q}_{\mathcal{H}} = \sqcup_{0 \le j < 2^{K+1}} \Delta'_l$ and satisfies

$$\frac{\max_{0 \le l < 2^{K+1}} \mathbb{Q}_{\mathcal{H}}(\Delta'_l)}{\min_{0 \le l < 2^{K+1}} \mathbb{Q}_{\mathcal{H}}(\Delta'_l)} \le 2\rho.$$

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

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

and $\mathcal{H} \subseteq \text{Span}\{\mathbb{1}_{\Delta_j} : 0 \le j < L\} \subseteq \text{Span}\{\mathbb{1}_{\mathcal{C}_{0,k}} : 0 \le k < 2^{K+1}\}$. Now we consider the term $\mathtt{C}_{\mathcal{H}}$ from Lemma SA.5. 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 $|c_q| \le \mathtt{M}_{\{h\}}$. 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

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

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.5, we get there exists a mean-zero Gaussian process $\widetilde{Z}_n^X$ with the same covariance structure as $\widetilde{X}_n$ such that with probability at least $1 - 2\exp(-t) - 2^{K+1}\exp(-C_\rho n 2^{-K-1})$,

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

where $K \le \log_2(L)$, and C_ρ is a constant that only depends on ρ. The conclusion then follows from taking $(Z_n(h) : h \in \mathcal{H}) = (\widetilde{Z}_n(h) : h \in \mathcal{H})$ and the fact that $(X_n(h) : h \in \mathcal{H}) = (\widetilde{X}_n(h) : h \in \mathcal{H})$ almost surely. □

SA-V.1.2 Proof of Corollary 5

Take $t = C \log n$ with $C > 1$ and $\delta = n^{-\frac{1}{2}}$ in Theorem 3. $\square$

SA-V.1.3 Example 2: Histogram Density Estimation

Recall for $\mathbf{w} \in \mathcal{W}$, we define

$$h_{\mathbf{w}}(\mathbf{u}) = \sqrt{L} \sum_{0 \leq l < P} \mathbb{1}(\mathbf{u} \in \Delta_l) \mathbb{1}(\mathbf{w} \in \Delta_l), \qquad \mathbf{u} \in \mathcal{X}, \mathbf{w} \in \mathcal{W},$$

and $\mathcal{H} = \{h_{\mathbf{w}}(\cdot) : \mathbf{w} \in \mathcal{W}\}$. Then $\mathcal{H} \subseteq \operatorname{Span}(\mathbb{1}_{\Delta_l} : 0 \leq l < P)$. In particular, for every $\mathbf{u} \in \mathcal{X}$ and $\mathbf{w} \in \mathcal{W}$, at most one of $\mathbb{1}(\mathbf{u} \in \Delta_l) \mathbb{1}(\mathbf{x} \in \Delta_l)$ will be non-zero. Hence $\mathtt{M}_{\mathcal{H},\mathbb{R}^d} = L^{1/2}$. Each function in $\mathcal{H}$ can be written as $c\mathbb{1}(\Delta_l)$ for some $l \leq L$, which implies we can take $\mathtt{S}_{\mathcal{H}} = 1$.

If $\mathcal{W} = \mathcal{X}$, since we assume the partition is quasi-uniform of $\mathcal{Q}_{\mathcal{H}} = \mathcal{X}$ with $\mathbb{Q}_{\mathcal{H}} = \mathbb{P}_X$, we know $\max_{0 \leq l < P} \mathbb{P}_X(\Delta_l) \leq c_\rho L^{-1}$ for some constant $c_\rho > 0$ that only depends on ρ, which implies

$$\mathtt{E}_{\mathcal{H}} \leq \max_{0 \leq l < P} \mathbb{P}_X(\Delta_l) \cdot \mathtt{M}_{\mathcal{H}} \leq c_\rho L^{-1} \sqrt{L} \leq c_\rho L^{-1/2},$$

where in this case $P = L$.

If $\mathcal{W} \subsetneq \mathcal{X}$, take $\mathring{P}$ to be the unique number in $\mathbb{Z}$ such that $\mathring{P} \leq \frac{\mathbb{P}_X((\cup_{0 \leq l < P} \Delta_l)^c)}{\min_{0 \leq l < P} \mathbb{P}_X(\Delta_l)} < \mathring{P} + 1$. Then we consider the following two cases. Set $L = P + \mathring{P}$.

Case 1: $\mathring{P} \geq 1$. The construction in Example 2 implies for every $P \leq l < L$,

$$1 \leq \frac{\mathbb{Q}_{\mathcal{H}}(\Delta_l)}{\min_{0 \leq k < P} \mathbb{P}_X(\Delta_k)} \leq 1 + \mathring{P}^{-1} \leq 2.$$

In particular, $\min_{0 \leq k < L} \mathbb{P}_X(\Delta_l) = \min_{0 \leq k < P} \mathbb{P}_X(\Delta_l)$. Combined with quasi-uniformity of $\{\Delta_l : 0 \leq l < P\}$, we have

$$\frac{\max_{0 \leq k < L} \mathbb{P}_X(\Delta_k)}{\min_{0 \leq k < L} \mathbb{P}_X(\Delta_k)} \leq \max\{\rho, 2\}.$$

Since $\mathbb{P}_X$ agrees with $\mathbb{Q}_{\mathcal{H}}$ on $\sqcup_{0 \leq l < P} \Delta_l$, and $\sqcup_{P \leq l < L} \Delta_l \subseteq \mathcal{X} \cup \operatorname{Supp}(\mathcal{H})^c$, $\mathbb{Q}_{\mathcal{H}}$ is a surrogate measure of $\mathbb{P}_X$ with respect to $\mathcal{H}$. And we verified that $\{\Delta_l : 0 \leq l < L\}$ is a quasi-uniform partition of $\mathcal{Q}_{\mathcal{H}}$ with respect to $\mathbb{Q}_{\mathcal{H}}$.

Case 2: $\mathring{P} = 0$. Then for any $0 \leq l < P$, there exists $\mathring{P}_l \in \mathbb{N}$ such that

$$\mathring{P}_l \leq \frac{\mathbb{P}_X(\Delta_l)}{\mathbb{P}_X((\sqcup_{0 \leq l < P} \Delta_l)^c)} < \mathring{P}_l + 1.$$

Taking arbitrary Δ_P with $\mathbb{P}_X(\Delta_P) = \mathbb{P}_X((\sqcup_{0 \leq l < P} \Delta_l)^c)$, and for $0 \leq l < P$ break Δ_l into $\mathring{P}_l$ pieces of equal measure by $\mathbb{P}_X$, we can show by similar arguments as above that the refined cells with the additional Δ_P together forms a quasi-uniform partition of $\mathcal{X}$ with respect to $\mathbb{P}_X$. Suppose also in this case, the number of cells in the quasi-uniform partition is L after refinement.

In both cases, we know $\max_{0 \leq l < L} \mathbb{P}_X(\Delta_l) \leq c_\rho L^{-1}$ for some constant c_ρ that only depends on ρ, which implies

$$\mathtt{E}_{\mathcal{H}} \leq \max_{0 \leq l < P} \mathbb{P}_X(\Delta_l) \cdot \mathtt{M}_{\mathcal{H}} \leq c_\rho L^{-1}\sqrt{L} \leq c_\rho L^{-1/2}.$$

We can then apply Theorem 3 to get the stated rates. □

SA-V.2 Residual-Based (and Multiplicative Separable) Empirical Process

For $\delta \in (0,1]$, define

$$\mathtt{N}(\delta) = \mathtt{N}_{\mathcal{G}}(\delta/\sqrt{2}, \mathtt{M}_{\mathcal{G}})\mathtt{N}_{\mathcal{R}}(\delta/\sqrt{2}, M_{\mathcal{R}})$$

and

$$J(\delta) = \sqrt{2}J(\mathcal{G}, \mathtt{M}_{\mathcal{G}}, \delta/\sqrt{2}) + \sqrt{2}J(\mathcal{R}, M_{\mathcal{R}}, \delta/\sqrt{2}).$$

To simplify notation, the parameters of $\mathcal{G}$ and $\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}$ (Definitions 4 to 12, SA.1, SA.2) are taken with $\mathcal{C} = \mathcal{Q}_{\mathcal{G}}$, and the index $\mathcal{Q}_{\mathcal{G}}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{Q}_{\mathcal{G}} \times \mathcal{Y}$, and the index $\mathcal{Q}_{\mathcal{G}} \times \mathcal{Y}$ is omitted where there is no ambiguity.

Theorem SA.3. *Suppose $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ are i.i.d. random vectors taking values in $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$, where $\mathbf{x}_i$ has distribution $\mathbb{P}_X$ supported on $\mathcal{X} \subseteq \mathbb{R}^d$, y_i has distribution $\mathbb{P}_Y$ supported on $\mathcal{Y} \subseteq \mathbb{R}$, and the following conditions hold.*

(i) $\mathcal{G} \subseteq \operatorname{Span}\{\mathbb{1}_{\Delta_l} : 0 \leq l < L\}$ *is a class of Haar functions on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.

(ii) There exists a surrogate measure $\mathbb{Q}_{\mathcal{G}}$ for $\mathbb{P}_X$ with respect to $\mathcal{G}$ such that $\{\Delta_l : 0 \leq l < L\}$ forms a quasi-uniform partition of $\mathcal{Q}_{\mathcal{G}}$ with respect to $\mathbb{Q}_{\mathcal{G}}$:

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

(iii) $\mathcal{G}$ is a VC-type class with envelope function $\mathtt{M}_{\mathcal{G}}$ over $\mathcal{Q}_{\mathcal{G}}$ with $\mathtt{c}_{\mathcal{G}} \geq e$ and $\mathtt{d}_{\mathcal{G}} \geq 1$.

(iv) $\mathcal{R}$ is a real-valued pointwise measurable class of functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$.

(v) $\mathcal{R}$ is a VC-type class with envelope $M_{\mathcal{R},\mathcal{Y}}$ over $\mathcal{Y}$ with $\mathtt{c}_{\mathcal{R},\mathcal{Y}} \geq e$ and $\mathtt{d}_{\mathcal{R},\mathcal{Y}} \geq 1$, where $M_{\mathcal{R},\mathcal{Y}}(y) + \mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \leq \mathtt{v}(1 + |y|^\alpha)$ for all $y \in \mathcal{Y}$, for some $\mathtt{v} > 0$, and for some $\alpha \geq 0$. Furthermore, if $\alpha > 0$, then $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

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

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

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

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

and recall

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

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

Proof of Theorem SA.3. First, we make a reduction through the surrogate measure and normalizing transformation. Let $\mathcal{Z}_{\mathcal{G}} = \mathcal{X}\cap\mathrm{Supp}(\mathcal{G})$. Definition 2 implies $\mathbb{P}_X|_{\mathcal{Z}_{\mathcal{G}}} = \mathbb{Q}_{\mathcal{G}}|_{\mathcal{Z}_{\mathcal{G}}}$. Define a joint probability measure $\mathbb{O}$ on $(\mathbb{R}^d\times\mathbb{R}^d, \mathcal{B}(\mathbb{R}^{2d}))$ such that for all $A\in\mathcal{B}(\mathbb{R}^{2d})$

$$\mathbb{O}(A\cap(\mathcal{Z}_{\mathcal{G}}\times\mathcal{Z}_{\mathcal{G}})) = \mathbb{P}_X(\Pi_{1:d}(A\cap\{(\mathbf{x},\mathbf{x}):\mathbf{x}\in\mathcal{Z}_{\mathcal{G}}\})),$$
$$\mathbb{O}(A\cap(\mathcal{Z}_{\mathcal{G}}\times\mathcal{Z}_{\mathcal{G}}^c)) = \mathbb{O}(A\cap(\mathcal{Z}_{\mathcal{G}}^c\times\mathcal{Z}_{\mathcal{G}})) = 0,$$
$$\mathbb{O}(A\cap(\mathcal{Z}_{\mathcal{G}}^c\times\mathcal{Z}_{\mathcal{G}}^c)) = \int_{\mathcal{Z}_{\mathcal{G}}^c}\mathbb{P}_X(A^{\mathbf{y}}\cap\mathcal{Z}_{\mathcal{G}}^c)d\mathbb{Q}_{\mathcal{G}}(\mathbf{y}),$$

where for $A\in\mathcal{B}(\mathbb{R}^{2d})$, $\Pi_{1:d}(A) = \{\mathbf{x}\in\mathbb{R}^d : (\mathbf{x},\mathbf{y})\in A \text{ for some } \mathbf{y}\in\mathbb{R}^d\}$, $A^{\mathbf{y}} = \{\mathbf{x}\in\mathbb{R}^d : (\mathbf{x},\mathbf{y})\in A\}$.

Then we can check that (i) the marginals of $\mathbb{O}$ are $\mathbb{P}_X$ and $\mathbb{Q}_{\mathcal{G}}$, respectively; (ii) $\mathbb{O}|_{\mathcal{Z}_{\mathcal{G}}\times\mathbb{R}^d\cup\mathbb{R}^d\times\mathcal{Z}_{\mathcal{G}}}$ is supported on $\{(\mathbf{x},\mathbf{x}):\mathbf{x}\in\mathcal{Z}_{\mathcal{G}}\}$. By Skorohod embedding (Dudley, 2014, Lemma 3.35), on a possibly enlarged probability space, there exists a $\mathbf{u}_i, 1\le i\le n$ i.i.d. $\sim\mathbb{Q}_{\mathcal{G}}$ such that $(\mathbf{x}_i,\mathbf{u}_i)$ has joint law $\mathbb{O}$. In particular, if $\mathbf{x}_i\in\mathcal{Z}_{\mathcal{G}}$, then $\mathbf{x}_i = \mathbf{u}_i$; if $\mathbf{x}_i\in\mathcal{Z}_{\mathcal{G}}^c$, then $\mathbf{u}_i\in\mathcal{Z}_{\mathcal{G}}^c$, and since $\mathcal{Q}_{\mathcal{G}}\subseteq\mathcal{X}\cup(\cap_{g\in\mathcal{G}}\mathrm{Supp}(g)^c)$, $\mathbf{u}_i\in\cap_{g\in\mathcal{G}}\mathrm{Supp}(g)^c$. Thus for any $g\in\mathcal{G}$, $r\in\mathcal{R}$,

$$G_n(g,r) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g(\mathbf{x}_i)r(y_i) - \mathbb{E}[g(\mathbf{x}_i)r(y_i)]\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g(\mathbf{u}_i)r(y_i) - \mathbb{E}[g(\mathbf{u}_i)r(y_i)]\right],$$

where the second equality follows because $\mathbf{x}_i = \mathbf{u}_i$ on the event $\{\mathbf{x}_i\in\mathcal{Z}_{\mathcal{G}}\}$, and $g(\mathbf{x}_i) = g(\mathbf{u}_i) = 0$ (a.s.) on the event $\{\mathbf{x}_i\in\mathcal{Z}_{\mathcal{G}}^c\}$. Hence, we work with an equivalent empirical process

$$\widetilde{G}_n(g,r) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g(\mathbf{u}_i)r(y_i) - \mathbb{E}[g(\mathbf{u}_i)r(y_i)]\right], \qquad g\in\mathcal{G}, r\in\mathcal{R}.$$

In particular $(\widetilde{G}_n(g,r): g\in\mathcal{G}, r\in\mathcal{R}) = (G_n(g,r): g\in\mathcal{G}, r\in\mathcal{R})$. Hence w.l.o.g. assume $\mathbb{Q}_{\mathcal{G}} = \mathbb{P}_X$ and we work with the $(G_n(g,r): g\in\mathcal{G}, r\in\mathcal{R})$ process.

Suppose $2^M\le L<2^{M+1}$. For each $l\in\{1,2,\ldots,d\}$, 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 $\mathcal{X} = \sqcup_{0\le j<2^{M+1}}\Delta_l'$ and satisfies

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

By construction, for each $N \in \mathbb{N}$, there exists an axis-aligned quasi-dyadic expansion $\mathcal{A}_{M+1,N}(\mathbb{P}_Z, 2\rho) = \{\mathcal{C}_{j,k} : 0 \leq j \leq M+1+N, 0 \leq k < 2^{M+1+N-j}\}$ 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 $\mathcal{G} \subseteq \mathrm{Span}\{\mathbb{1}_{\Delta_j} : 0 \leq j < J\} \subseteq \mathrm{Span}\{\mathbb{1}_{\mathcal{X}_{0,k}} : 0 \leq k < 2^{M+1}\}$. Hence

$$\Pi_0(g,r) = \Pi_1(g,r) = \sum_{0 \leq l < 2^{K+1}} \sum_{0 \leq m < 2^N} \mathbb{1}(\mathcal{X}_{0,l} \times \mathcal{Y}_{j,l,m}) g|_{\mathcal{X}_{0,l}} \mathbb{E}[r(y_i)|\mathbf{x}_i \in \mathcal{X}_{0,l}, y_i \in \mathcal{Y}_{j,l,m}]. \qquad \text{(SA-23)}$$

Again, consider $(\mathcal{G} \times \mathcal{R})_\delta$ which is a $\delta\|\mathtt{M}_\mathcal{G} M_\mathcal{R}\|_{\mathbb{P}_Z}$ of $\mathcal{G} \times \mathcal{R}$ of cardinality no greater than $\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta, \mathtt{M}_\mathcal{G} M_\mathcal{R})$, $0 < \delta \leq 1$. The SA error for projected process on the δ-net is given by Lemma SA.18: For all $t > 0$,

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

Now we find an upper bound for $\mathtt{C}_{\Pi_1(\mathcal{G}\times\mathcal{R}),M+N}$. 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 $|\widetilde{\gamma}_{j',m'}(\mathbb{1}(\Delta'_{l_q}), r)| \leq C_{\mathtt{v},\alpha}2^{-l+1}$ if $\Delta'_{l_q} \subseteq \mathcal{X}_{j'-N,m'}$ where $C_{\mathtt{v},\alpha} = \mathtt{v}(1+(2\sqrt{\alpha})^\alpha)$. For j' such that $N \leq j' \leq j$,

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

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

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

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

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

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

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

It follows that

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

It follows that

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

By the characterization of projections in Equation SA-23, we know the mis-specification error is zero, that is, $\Pi_1 G_n(g,r) = \Pi_0 G_n(g,r)$ and $\Pi_1 Z_n^G(g,r) = \Pi_0 G_n(g,r)$. 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.20 and the argument for truncation error in the proof for Lemma SA.21, for all $t > N$,

$$\mathbb{P}\left(\|G_n - \Pi_1 G_n\|_{(\mathcal{G}\times\mathcal{R})_\delta} + \|Z_n^G - \Pi_1 Z_n^G\|_{(\mathcal{G}\times\mathcal{R})_\delta} \geq N \sqrt{2^{-N} \mathtt{M}_{\mathcal{G}}^2} t^{\alpha+\frac{1}{2}} + \frac{\mathtt{M}_{\mathcal{G}}}{\sqrt{n}} t^{\alpha+1} \right) \leq 4 \mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta, \mathtt{M}_{\mathcal{G}} M_{\mathcal{R}}) n e^{-t}. \tag{SA-24}$$

Then apply Lemma SA.18, we get there exists a mean-zero Gaussian process Z_n^G with the same covariance structure as G_n such that with probability at least $1 - 2\exp(-t) - 2^{M+1}\exp(-C_\rho n 2^{-M-1})$,

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

where $C_\rho > 0$ is a constant that only depends on ρ. □

The following theorem presents a generalization of Theorem 4 in the paper. To simplify notation, the parameters of $\mathcal{G}$ and $\mathcal{G} \cdot \mathcal{V}_{\mathcal{R}}$ (Definitions 4 to 12, SA.1, SA.2) are taken with $\mathcal{C} = \mathcal{Q}_{\mathcal{G}}$, and the index $\mathcal{Q}_{\mathcal{G}}$ is omitted where there is no ambiguity; the parameters of $\mathcal{R}$ (Definitions 4 to 12) are taken with $\mathcal{C} = \mathcal{Y}$, and the index $\mathcal{Y}$ is omitted where there is no ambiguity; and the parameters of $\mathcal{G} \times \mathcal{R}$ (Definitions 4 to 12, SA.3, SA.4) are taken with $\mathcal{C} = \mathcal{Q}_{\mathcal{G}} \times \mathcal{Y}$, and the index $\mathcal{Q}_{\mathcal{G}} \times \mathcal{Y}$ is omitted where there is no ambiguity.

Theorem SA.4. *Suppose $(\mathbf{z}_i = (\mathbf{x}_i, y_i) : 1 \leq i \leq n)$ are i.i.d. random vectors taking values in $(\mathbb{R}^{d+1}, \mathcal{B}(\mathbb{R}^{d+1}))$, where $\mathbf{x}_i$ has distribution $\mathbb{P}_X$ supported on $\mathcal{X} \subseteq \mathbb{R}^d$, y_i has distribution $\mathbb{P}_Y$ supported on $\mathcal{Y} \subseteq \mathbb{R}$, and the following conditions hold.*

(i) $\mathcal{G} \subseteq \mathrm{Span}\{\mathbb{1}_{\Delta_l} : 0 \leq l < L\}$ *is a class of Haar functions on* $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mathbb{P}_X)$.

(ii) There exists a surrogate measure $\mathbb{Q}_{\mathcal{G}}$ for $\mathbb{P}_X$ with respect to $\mathcal{G}$ such that $\{\Delta_l : 0 \leq l < L\}$ forms a quasi-uniform partition of $\mathcal{Q}_{\mathcal{G}}$ with respect to $\mathbb{Q}_{\mathcal{G}}$:

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

(iii) $\mathcal{G}$ is a VC-type class with envelope function $\mathtt{M}_{\mathcal{G}}$ over $\mathcal{Q}_{\mathcal{G}}$ with $\mathtt{c}_{\mathcal{G}} \geq e$ and $\mathtt{d}_{\mathcal{G}} \geq 1$.

(iv) $\mathcal{R}$ is a real-valued pointwise measurable class of functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_Y)$.

(v) $\mathcal{R}$ is a VC-type class with envelope $M_{\mathcal{R},\mathcal{Y}}$ over $\mathcal{Y}$ with $\mathtt{c}_{\mathcal{R},\mathcal{Y}} \geq e$ and $\mathtt{d}_{\mathcal{R},\mathcal{Y}} \geq 1$, where $M_{\mathcal{R},\mathcal{Y}}(y) + \mathtt{pTV}_{\mathcal{R},(-|y|,|y|)} \leq \mathtt{v}(1 + |y|^\alpha)$ for all $y \in \mathcal{Y}$, for some $\mathtt{v} > 0$, and for some $\alpha \geq 0$. Furthermore, if $\alpha > 0$, then $\sup_{\mathbf{x} \in \mathcal{X}} \mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \leq 2$.

Then, on a possibly enlarged probability space, there exists 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*
- $\mathbb{P}[\|R_n - Z_n^R\|_{\mathcal{G} \times \mathcal{R}} > C_1 C_{\mathtt{v},\alpha} C_\rho \min_{\delta \in (0,1)} (\mathsf{H}_n^R(t, \delta) + \mathsf{F}_n^R(t, \delta)) + \mathsf{W}_n(t))] \leq C_2 e^{-t} + L e^{-C_\rho n / L}$ *for all* $t > 0$,

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

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

with $\mathcal{V}_{\mathcal{R}} = \{\theta(\cdot, r) : \mathbf{x} \mapsto \mathbb{E}[r(y_i)|\mathbf{x}_i = \mathbf{x}], \mathbf{x} \in \mathcal{X}, r \in \mathcal{R}\}$ and $N^* = \lceil \log_2(\frac{n \mathtt{M}_{\mathcal{G}}}{2^L \mathtt{E}_{\mathcal{G}}}) \rceil$.

Proof of Theorem SA.4. By the same reduction through surrogate measure, we can w.l.o.g. assume $\mathbb{Q}_{\mathcal{G}} = \mathbb{P}_X$. Suppose $2^M \leq J < 2^{M+1}$. By the same cell divisions in the proof for Theorem SA.3, there exists a quasi-dyadic expansion $\mathcal{C}_{M+1,N}$ such that

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

By definition, the projection error can be decomposed as

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

where Π_0 denotes the L_2-projection from $L_2(\mathbb{R}^d)$ to $\mathrm{Span}(\{\mathbb{1}(\mathcal{X}_{0,l}) : 0 \le l < 2^{M+1}\})$. For any $g \in \mathcal{G}$, since $g \in \mathrm{Span}(\{\mathbb{1}(\mathcal{X}_{0,l}) : 0 \le l < 2^{M+1}\})$,

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

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 $t > 0$,

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

Combining Lemma SA.18 and Equation (SA-24), and the same calculation as in the proof for Theorem SA.2 to get $\mathtt{C}_{\Pi_2(\mathcal{G},\mathcal{R})} \lesssim (C_{\mathtt{v},\alpha}\mathtt{M}_\mathcal{G} N^\alpha)^2$, for all $t > N_*$, with probability at least $1 - 2\mathtt{N}_{\mathcal{G}\times\mathcal{R}}(\delta, \mathtt{M}_\mathcal{G} M_\mathcal{R})e^{-t} - 2^M\exp(-C_\rho n 2^{-M})$,

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

The rest follows from the error for fluctuation off the δ-net given in Lemma SA.16. The "bias" term $\sqrt{\mathtt{M}_\mathcal{G}\mathtt{E}_\mathcal{G}}\max_{0\le j<J}\|\Delta_j\|_\infty \mathtt{L}_{\mathcal{V}_\mathcal{R}}\sqrt{t}$ 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 $\varepsilon_i = \mathrm{sign}(r(y_i) - \mathbb{E}[r(y_i)|\mathbf{x}_i])|r(y_i) - \mathbb{E}[r(y_i)|\mathbf{x}_i]|^{1/\alpha}$. Take $\widetilde{r}(u) = \mathrm{sign}(u)|u|^\alpha, u \in \mathbb{R}$. 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 $\sup_{\mathbf{x}\in\mathcal{X}}\mathbb{E}[\exp(|y_i|)|\mathbf{x}_i = \mathbf{x}] \le 2$ 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 $g \in \mathcal{G}$,

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

where G_n' denotes the empirical process based on random sample $((\mathbf{x}_i,\varepsilon_i) : 1 \le i \le n)$. The result then follows from Theorem SA.3. By similar arguments as in the proof of Theorem SA.4,

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

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

SA-V.2.1 Proof of Theorem 4

By standard empirical process arguments, $\mathtt{N}_\mathcal{G}(\delta) \le \mathtt{c}_\mathcal{G}\delta^{-\mathtt{d}_\mathcal{G}}$ and $\mathtt{N}_\mathcal{R}(\delta) \le \mathtt{c}_\mathcal{R}\delta^{-\mathtt{d}_\mathcal{R}}$ for $\delta \in (0,1]$, and the result follows by Lemma SA.4. □

SA-V.2.2 Proof of Corollary 6

Take $t = C \log n$ with $C > 1$ in Theorem 4. $\square$

SA-V.3 Example: Haar Partitioning-based Regression

The following lemma gives precise regularity conditions for the example in Section 5.3 of the paper.

Lemma SA.31 (Haar Basis Regression Estimators)**.** *Consider the setup in Example 3, and assume in addition that* $\sup_{r\in\mathcal{R}_\ell}\sup_{\mathbf{x},\mathbf{y}\in\mathcal{X}} |\theta(\mathbf{x},r)-\theta(\mathbf{y},r)|/\|\mathbf{x}-\mathbf{y}\|_\infty < \infty$ *for* $\ell = 1, 2$.

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

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

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

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

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

Proof of Lemma SA.31. We use the notation $\mathbb{P}_X(\Delta_l) = \mathbb{P}(\mathbf{x}_i \in \Delta_l)$, and $\widehat{\mathbb{P}}_X(\Delta_l) = n^{-1}\sum_{i=1}^n \mathbb{1}(\mathbf{x}_i\in\Delta_l)$, $0 \le l < L$.

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

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

where $\epsilon_i(r) = r(y_i) - \mathbb{E}[r(y_i)|\mathbf{x}_i]$. By maximal inequality for sub-Gaussian random variables (van der Vaart and Wellner, 2013, Lemma 2.2.2), $\max_{0\le l<L}|L\widehat{\mathbb{P}}_X(\Delta_l) - L\mathbb{P}_X(\Delta_l)| = O(\sqrt{\frac{\log(nL)}{n/L}})$ a.s.. Since $\{\Delta_l : 0 \le l < L\}$ is a quasi-uniform partition of $\mathcal{X}$ with respect to $\mathbb{P}_X$, $\min_{0\le l<L} L\mathbb{P}_X(\Delta_l) = \Omega(1)$. Hence

$$\max_{0\le l<L}|L^{-1}\widehat{\mathbb{P}}_X(\Delta_l)^{-1}-L^{-1}\mathbb{P}_X(\Delta_l)^{-1}| = O(\sqrt{(n/L)^{-1}\log(nL)}), \quad a.s.. \tag{SA-25}$$

Take $\mathcal{H}_\ell = \{(\mathbf{w},y)\mapsto L\mathbb{1}(\mathbf{w}\in\Delta_l)(r(y)-\theta(\mathbf{w},r)) : 0\le l<L, r\in\mathcal{R}_\ell\}$, for $\ell = 1,2$. In particular, if we take $\mathcal{G} = \{L\mathbb{1}(\cdot\in\Delta_l) : 0 \le l < L\}$, 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 universal constant, $\mathtt{d}_{\mathcal{R}_2} = 2$. By standard empirical process arguments, both $\mathcal{H}_\ell$'s are VC-type class with $\mathtt{c}_{\mathcal{H}_1} = L$, $\mathtt{d}_{\mathcal{H}_1} = 1$, $\mathtt{c}_{\mathcal{H}_2} = O(L)$, $\mathtt{d}_{\mathcal{H}_2} = 2$. Since $\sup_{r\in\mathcal{R}_\ell}\max_{0\le l<L}|\frac{1}{n}\sum_{i=1}^n L\mathbb{1}(\mathbf{x}_i\in\Delta_l)\epsilon_i(r)| = \sup_{h\in\mathcal{H}_\ell}|\mathbb{E}_n[h(\mathbf{x}_i,y_i)]-\mathbb{E}[h(\mathbf{x}_i,y_i)]|$ is the suprema of empirical process, by Corollary 5.1 in Chernozhukov *et al.* (2014),

$$\begin{aligned}\sup_{r\in\mathcal{R}_1}\max_{0\le l<L}\left|\frac{1}{n}\sum_{i=1}^n L\mathbb{1}(\mathbf{x}_i\in\Delta_l)\epsilon_i(r)\right| &= O\left(\sqrt{\frac{\log(nL)}{n/L}} + \log(n)\frac{\log(nL)}{n/L}\right) \quad a.s.,\\ \sup_{r\in\mathcal{R}_2}\max_{0\le l<L}\left|\frac{1}{n}\sum_{i=1}^n L\mathbb{1}(\mathbf{x}_i\in\Delta_l)\epsilon_i(r)\right| &= O\left(\sqrt{\frac{\log(nL)}{n/L}}\right) \quad a.s..\end{aligned} \tag{SA-26}$$

Putting together Equations (SA-25), (SA-26), we have

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

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

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

□

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.

Ambrosio, L., Fusco, N., and Pallara, D. (2000). *Functions of bounded variation and free discontinuity problems*: Oxford university press.

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*, *30*(4), 3193–3223.

Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). "Gaussian approximation of suprema of empirical processes," *Annals of Statistics*, *42*(4), 1564–1597.

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.