SUPPLEMENT TO "YURINSKII'S COUPLING FOR MARTINGALES"

BY MATIAS D. CATTANEO[1,a], RICARDO P. MASINI[2,b] AND WILLIAM G. UNDERWOOD[3,c]

[1]*Department of Operations Research and Financial Engineering, Princeton University,* [a]*cattaneo@princeton.edu*

[2]*Department of Statistics, University of California, Davis,* [b]*rmasini@ucdavis.edu*

[3]*Statistical Laboratory, University of Cambridge,* [c]*wgu21@cam.ac.uk*

Contents

APPENDIX SA: PROOFS OF MAIN RESULTS

SA.1. Preliminary lemmas We give a sequence of preliminary lemmas which are useful for establishing our main results. Firstly, we present a conditional version of Strassen's theorem for the ℓ^p-norm [6, Theorem B.2; 16, Theorem 4], stated for completeness as Lemma SA.1.

LEMMA SA.1 (A conditional Strassen theorem for the ℓ^p-norm). *Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space supporting the $\mathbb{R}^d$-valued random variable X for some $d \geq 1$. Let $\mathcal{H}'$ be a countably generated sub-σ-algebra of $\mathcal{H}$ and suppose there exists a* $\mathrm{Unif}[0,1]$ *random variable on $(\Omega, \mathcal{H}, \mathbb{P})$ which is independent of the σ-algebra generated by X and $\mathcal{H}'$. Consider a regular conditional distribution $F(\cdot \mid \mathcal{H}')$ satisfying the following. Firstly, $F(A \mid \mathcal{H}')$ is an $\mathcal{H}'$-measurable random variable for all Borel sets $A \in \mathcal{B}(\mathbb{R}^d)$. Secondly, $F(\cdot \mid \mathcal{H}')(\omega)$ is a Borel probability measure on $\mathbb{R}^d$ for all $\omega \in \Omega$. Taking $\eta, \rho > 0$ and $p \in [1, \infty]$, with $\mathbb{E}^*$ the outer expectation, if*

$$\mathbb{E}^* \left[\sup_{A \in \mathcal{B}(\mathbb{R}^d)} \left\{ \mathbb{P}\big(X \in A \mid \mathcal{H}'\big) - F\big(A_p^\eta \mid \mathcal{H}'\big) \right\} \right] \leq \rho,$$

where $A_p^\eta = \{x \in \mathbb{R}^d : \|x - A\|_p \leq \eta\}$ and $\|x - A\|_p = \inf_{x' \in A} \|x - x'\|_p$, then there exists an $\mathbb{R}^d$-valued random variable Y with $Y \mid \mathcal{H}' \sim F(\cdot \mid \mathcal{H}')$ and $\mathbb{P}(\|X - Y\|_p > \eta) \leq \rho$.

PROOF (Lemma SA.1). By Theorem B.2 in Chen and Kato [6], noting that the σ-algebra generated by Z is countably generated and using the metric induced by the ℓ^p-norm. □

Next, we present in Lemma SA.2 an analytic result concerning the smooth approximation of Borel set indicator functions, similar to that given in Belloni et al. [3, Lemma 39].

LEMMA SA.2 (Smooth approximation of Borel indicator functions). *Let $A \subseteq \mathbb{R}^d$ be a Borel set and $Z \sim \mathcal{N}(0, I_d)$. For $\sigma, \eta > 0$ and $p \in [1, \infty]$, define*

$$g_{A\eta}(x) = \left(1 - \frac{\|x - A^\eta\|_p}{\eta}\right) \vee 0 \qquad \text{and} \qquad f_{A\eta\sigma}(x) = \mathbb{E}\big[g_{A\eta}(x + \sigma Z)\big].$$

Then f is infinitely differentiable and with $\varepsilon = \mathbb{P}(\|Z\|_p > \eta/\sigma)$, for all $k \geq 0$, any multi-index $\kappa = (\kappa_1, \ldots, \kappa_d) \in \mathbb{N}^d$, and all $x, y \in \mathbb{R}^d$, we have $|\partial^\kappa f_{A\eta\sigma}(x)| \leq \frac{\sqrt{\kappa!}}{\sigma^{|\kappa|}}$ and

$$\left|f_{A\eta\sigma}(x+y) - \sum_{|\kappa|=0}^{k} \frac{1}{\kappa!} \partial^\kappa f_{A\eta\sigma}(x) y^\kappa\right| \leq \frac{\|y\|_p \|y\|_2^k}{\sigma^k \eta \sqrt{k!}},$$

$$(1-\varepsilon)\mathbb{I}\big\{x \in A\big\} \leq f_{A\eta\sigma}(x) \leq \varepsilon + (1-\varepsilon)\mathbb{I}\big\{x \in A^{3\eta}\big\}.$$

PROOF (Lemma SA.2). Drop the subscripts on $g_{A\eta}$ and $f_{A\eta\sigma}$. By Taylor's theorem with Lagrange remainder, for a $t \in [0,1]$,

$$\left|f(x+y) - \sum_{|\kappa|=0}^{k} \frac{1}{\kappa!} \partial^\kappa f(x) y^\kappa\right| \leq \left|\sum_{|\kappa|=k} \frac{y^\kappa}{\kappa!} \big(\partial^\kappa f(x+ty) - \partial^\kappa f(x)\big)\right|.$$

Now with $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$,

$$f(x) = \mathbb{E}\big[g(x + \sigma W)\big] = \int_{\mathbb{R}^d} g(x + \sigma u) \prod_{j=1}^d \phi(u_j) \, \mathrm{d}u = \frac{1}{\sigma^d} \int_{\mathbb{R}^d} g(u) \prod_{j=1}^d \phi\left(\frac{u_j - x_j}{\sigma}\right) \mathrm{d}u$$

and since the integrand is bounded, we exchange differentiation and integration to compute

$$\begin{aligned}
\partial^\kappa f(x) &= \left(\frac{-1}{\sigma}\right)^{|\kappa|} \frac{1}{\sigma^d} \int_{\mathbb{R}^d} g(u) \prod_{j=1}^d \partial^{\kappa_j} \phi\left(\frac{u_j - x_j}{\sigma}\right) \mathrm{d}u \\
&= \left(\frac{-1}{\sigma}\right)^{|\kappa|} \int_{\mathbb{R}^d} g(x + \sigma u) \prod_{j=1}^d \partial^{\kappa_j} \phi(u_j) \, \mathrm{d}u \\
&= \left(\frac{-1}{\sigma}\right)^{|\kappa|} \mathbb{E}\left[g(x + \sigma Z) \prod_{j=1}^d \frac{\partial^{\kappa_j} \phi(Z_j)}{\phi(Z_j)}\right], \qquad (1)
\end{aligned}$$

where $Z \sim \mathcal{N}(0, I_d)$. Recalling that $|g(x)| \leq 1$ and applying the Cauchy–Schwarz inequality,

$$|\partial^\kappa f(x)| \leq \frac{1}{\sigma^{|\kappa|}} \prod_{j=1}^d \mathbb{E}\left[\left(\frac{\partial^{\kappa_j} \phi(Z_j)}{\phi(Z_j)}\right)^2\right]^{1/2} \leq \frac{1}{\sigma^{|\kappa|}} \prod_{j=1}^d \sqrt{\kappa_j!} = \frac{\sqrt{\kappa!}}{\sigma^{|\kappa|}},$$

as the expected square of the Hermite polynomial of degree κ_j against the standard Gaussian measure is $\kappa_j!$. By the reverse triangle inequality, $|g(x + ty) - g(x)| \leq t\|y\|_p/\eta$, so by (1),

$$\begin{aligned}
&\left|\sum_{|\kappa|=k} \frac{y^\kappa}{\kappa!} \big(\partial^\kappa f(x+ty) - \partial^\kappa f(x)\big)\right| \\
&\quad = \left|\sum_{|\kappa|=k} \frac{y^\kappa}{\kappa!} \frac{1}{\sigma^{|\kappa|}} \mathbb{E}\left[\big(g(x + ty + \sigma Z) - g(x + \sigma Z)\big) \prod_{j=1}^d \frac{\partial^{\kappa_j} \phi(Z_j)}{\phi(Z_j)}\right]\right|
\end{aligned}$$

$$\leq \frac{t\|y\|_p}{\sigma^k \eta} \mathbb{E}\left[\left| \sum_{|\kappa|=k} \frac{y^\kappa}{\kappa!} \prod_{j=1}^d \frac{\partial^{\kappa_j} \phi(Z_j)}{\phi(Z_j)} \right|\right].$$

Therefore by the Cauchy–Schwarz inequality,

$$\left(\sum_{|\kappa|=k} \frac{y^\kappa}{\kappa!} \big(\partial^\kappa f(x+ty) - \partial^\kappa f(x)\big) \right)^2 \leq \frac{t^2\|y\|_p^2}{\sigma^{2k}\eta^2} \mathbb{E}\left[\left(\sum_{|\kappa|=k} \frac{y^\kappa}{\kappa!} \prod_{j=1}^d \frac{\partial^{\kappa_j}\phi(Z_j)}{\phi(Z_j)} \right)^2\right]$$

$$= \frac{t^2\|y\|_p^2}{\sigma^{2k}\eta^2} \sum_{|\kappa|=k} \sum_{|\kappa'|=k} \frac{y^{\kappa+\kappa'}}{\kappa!\kappa'!} \prod_{j=1}^d \mathbb{E}\left[\frac{\partial^{\kappa_j}\phi(Z_j)}{\phi(Z_j)} \frac{\partial^{\kappa'_j}\phi(Z_j)}{\phi(Z_j)} \right].$$

Orthogonality of Hermite polynomials gives zero if $\kappa_j \neq \kappa'_j$. By the multinomial theorem,

$$\left| f(x+y) - \sum_{|\kappa|=0}^k \frac{1}{\kappa!} \partial^\kappa f(x) y^\kappa \right| \leq \frac{\|y\|_p}{\sigma^k\eta} \left(\sum_{|\kappa|=k} \frac{y^{2\kappa}}{\kappa!} \right)^{1/2}$$

$$\leq \frac{\|y\|_p}{\sigma^k\eta\sqrt{k!}} \left(\sum_{|\kappa|=k} \frac{k!}{\kappa!} y^{2\kappa} \right)^{1/2} \leq \frac{\|y\|_p \|y\|_2^k}{\sigma^k\eta\sqrt{k!}}.$$

For the final result, since $f(x) = \mathbb{E}[g(x+\sigma Z)]$ and $\mathbb{I}\{x \in A^\eta\} \leq g(x) \leq \mathbb{I}\{x \in A^{2\eta}\}$,

$$\begin{aligned} f(x) &\leq \mathbb{P}\left(x + \sigma Z \in A^{2\eta}\right) \\ &\leq \mathbb{P}\left(\|Z\|_p > \frac{\eta}{\sigma}\right) + \mathbb{I}\left\{x \in A^{3\eta}\right\} \mathbb{P}\left(\|Z\|_p \leq \frac{\eta}{\sigma}\right) = \varepsilon + (1-\varepsilon)\mathbb{I}\left\{x \in A^{3\eta}\right\}, \\ f(x) &\geq \mathbb{P}\left(x + \sigma Z \in A^{\eta}\right) \leq \mathbb{I}\left\{x \in A\right\} \mathbb{P}\left(\|Z\|_p \leq \frac{\eta}{\sigma}\right) = (1-\varepsilon)\mathbb{I}\left\{x \in A\right\}. \end{aligned}$$

□

We provide a useful Gaussian inequality in Lemma SA.3 which helps bound the $\beta_{\infty,k}$ moment terms appearing in several places throughout the paper.

LEMMA SA.3 (A useful Gaussian inequality). *Let $X \sim \mathcal{N}(0,\Sigma)$ where $\sigma_j^2 = \Sigma_{jj} \leq \sigma^2$ for all $1 \leq j \leq d$. Then*

$$\mathbb{E}\left[\|X\|_2^2 \|X\|_\infty\right] \leq 4\sigma\sqrt{\log 2d} \sum_{j=1}^d \sigma_j^2 \quad \textit{and} \quad \mathbb{E}\left[\|X\|_2^3 \|X\|_\infty\right] \leq 8\sigma\sqrt{\log 2d} \left(\sum_{j=1}^d \sigma_j^2 \right)^{3/2}.$$

PROOF (Lemma SA.3). By Cauchy–Schwarz, with $k \in \{2,3\}$, we have $\mathbb{E}\left[\|X\|_2^k \|X\|_\infty\right] \leq \mathbb{E}\left[\|X\|_2^{2k}\right]^{1/2} \mathbb{E}\left[\|X\|_\infty^2\right]^{1/2}$. For the first term, by Hölder's inequality and the fourth and sixth moments of the normal distribution,

$$\mathbb{E}\left[\|X\|_2^4\right] = \mathbb{E}\left[\left(\sum_{j=1}^d X_j^2\right)^2\right] = \sum_{j=1}^d \sum_{k=1}^d \mathbb{E}\left[X_j^2 X_k^2\right] \leq \left(\sum_{j=1}^d \mathbb{E}\left[X_j^4\right]^{\frac{1}{2}}\right)^2 = 3\left(\sum_{j=1}^d \sigma_j^2\right)^2,$$

$$\mathbb{E}\left[\|X\|_2^6\right] = \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d \mathbb{E}\left[X_j^2 X_k^2 X_l^2\right] \leq \left(\sum_{j=1}^d \mathbb{E}\left[X_j^6\right]^{\frac{1}{3}}\right)^3 = 15\left(\sum_{j=1}^d \sigma_j^2\right)^3.$$

For the second term, by Jensen's inequality and the χ^2 moment generating function,

$$\mathbb{E}\left[\|X\|_\infty^2\right] = \mathbb{E}\left[\max_{1\le j\le d} X_j^2\right] \le 4\sigma^2 \log \sum_{j=1}^d \mathbb{E}\left[e^{X_j^2/(4\sigma^2)}\right] \le 4\sigma^2 \log \sum_{j=1}^d \sqrt{2} \le 4\sigma^2 \log 2d.$$

□

We provide an ℓ^p-norm tail probability bound for Gaussian variables in Lemma SA.4, motivating the definition of the term $\phi_p(d)$.

LEMMA SA.4 (Gaussian ℓ^p-norm bound). *Let $X \sim \mathcal{N}(0,\Sigma)$ where $\Sigma \in \mathbb{R}^{d\times d}$ is positive semi-definite. Then $\mathbb{E}\left[\|X\|_p\right] \le \phi_p(d) \max_{1\le j\le d} \sqrt{\Sigma_{jj}}$ where $\phi_p(d) = \sqrt{pd^{2/p}}$ for $p \in [1,\infty)$ and $\phi_\infty(d) = \sqrt{2\log 2d}$.*

PROOF (Lemma SA.4). For $p \in [1,\infty)$, as each X_j is Gaussian, we have $\left(\mathbb{E}\left[|X_j|^p\right]\right)^{1/p} \le \sqrt{p\,\mathbb{E}[X_j^2]} = \sqrt{p\Sigma_{jj}}$. Therefore

$$\mathbb{E}\left[\|X\|_p\right] \le \left(\sum_{j=1}^d \mathbb{E}\left[|X_j|^p\right]\right)^{1/p} \le \left(\sum_{j=1}^d p^{p/2}\Sigma_{jj}^{p/2}\right)^{1/p} \le \sqrt{pd^{2/p}} \max_{1\le j\le d} \sqrt{\Sigma_{jj}}$$

by Jensen's inequality. For $p = \infty$, with $\sigma^2 = \max_j \Sigma_{jj}$, for $t > 0$,

$$\begin{aligned}\mathbb{E}\left[\|X\|_\infty\right] &\le t \log \sum_{j=1}^d \mathbb{E}\left[e^{|X_j|/t}\right] \le t \log \sum_{j=1}^d \mathbb{E}\left[2e^{X_j/t}\right] \\ &\le t \log\left(2de^{\sigma^2/(2t^2)}\right) \le t \log 2d + \frac{\sigma^2}{2t},\end{aligned}$$

again by Jensen's inequality. Setting $t = \frac{\sigma}{\sqrt{2\log 2d}}$ gives $\mathbb{E}\left[\|X\|_\infty\right] \le \sigma\sqrt{2\log 2d}$. □

We give a Gaussian–Gaussian ℓ^p-norm approximation as Lemma SA.5, useful for ensuring approximations remain valid upon substituting an estimator for the true variance matrix.

LEMMA SA.5 (Gaussian–Gaussian approximation in ℓ^p-norm). *Let $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d\times d}$ be positive semi-definite and take $Z \sim \mathcal{N}(0, I_d)$. For $p \in [1,\infty]$ we have*

$$\mathbb{P}\left(\left\|\left(\Sigma_1^{1/2} - \Sigma_2^{1/2}\right) Z\right\|_p > t\right) \le 2d \exp\left(\frac{-t^2}{2d^{2/p}\left\|\Sigma_1^{1/2} - \Sigma_2^{1/2}\right\|_2^2}\right).$$

PROOF (Lemma SA.5). Let $\Sigma \in \mathbb{R}^{d\times d}$ be positive semi-definite and write $\sigma_j^2 = \Sigma_{jj}$. For $p \in [1,\infty)$ by a union bound and Gaussian tail probabilities,

$$\begin{aligned}\mathbb{P}\left(\left\|\Sigma^{1/2} Z\right\|_p > t\right) &= \mathbb{P}\left(\sum_{j=1}^d \left|\left(\Sigma^{1/2}Z\right)_j\right|^p > t^p\right) \le \sum_{j=1}^d \mathbb{P}\left(\left|\left(\Sigma^{1/2}Z\right)_j\right|^p > \frac{t^p\sigma_j^p}{\|\sigma\|_p^p}\right) \\ &= \sum_{j=1}^d \mathbb{P}\left(|\sigma_j Z_j|^p > \frac{t^p\sigma_j^p}{\|\sigma\|_p^p}\right) = \sum_{j=1}^d \mathbb{P}\left(|Z_j| > \frac{t}{\|\sigma\|_p}\right) \le 2d\,\exp\left(\frac{-t^2}{2\|\sigma\|_p^2}\right).\end{aligned}$$

The same result holds for $p=\infty$ since

$$\mathbb{P}\left(\left\|\Sigma^{1/2}Z\right\|_\infty > t\right) = \mathbb{P}\left(\max_{1\le j\le d}\left|\left(\Sigma^{1/2}Z\right)_j\right| > t\right) \le \sum_{j=1}^d \mathbb{P}\left(\left|\left(\Sigma^{1/2}Z\right)_j\right| > t\right)$$
$$= \sum_{j=1}^d \mathbb{P}\left(|\sigma_j Z_j| > t\right) \le 2\sum_{j=1}^d \exp\left(\frac{-t^2}{2\sigma_j^2}\right) \le 2d\exp\left(\frac{-t^2}{2\|\sigma\|_\infty^2}\right).$$

Now we apply this to the matrix $\Sigma = \left(\Sigma_1^{1/2}-\Sigma_2^{1/2}\right)^2$. For $p\in[1,\infty)$,

$$\|\sigma\|_p^p = \sum_{j=1}^d (\Sigma_{jj})^{p/2} = \sum_{j=1}^d \left(\left(\Sigma_1^{1/2}-\Sigma_2^{1/2}\right)^2\right)_{jj}^{p/2} \le d\max_{1\le j\le d}\left(\left(\Sigma_1^{1/2}-\Sigma_2^{1/2}\right)^2\right)_{jj}^{p/2}$$
$$\le d\left\|\left(\Sigma_1^{1/2}-\Sigma_2^{1/2}\right)^2\right\|_2^{p/2} = d\left\|\Sigma_1^{1/2}-\Sigma_2^{1/2}\right\|_2^p$$

Similarly for $p=\infty$ we have

$$\|\sigma\|_\infty = \max_{1\le j\le d}(\Sigma_{jj})^{1/2} = \max_{1\le j\le d}\left(\left(\Sigma_1^{1/2}-\Sigma_2^{1/2}\right)^2\right)_{jj}^{1/2} \le \left\|\Sigma_1^{1/2}-\Sigma_2^{1/2}\right\|_2.$$

Thus for all $p\in[1,\infty]$ we have $\|\sigma\|_p \le d^{1/p}\left\|\Sigma_1^{1/2}-\Sigma_2^{1/2}\right\|_2$, with $d^{1/\infty}=1$. Hence

$$\mathbb{P}\left(\left\|\left(\Sigma_1^{1/2}-\Sigma_2^{1/2}\right)Z\right\|_p > t\right) \le 2d\exp\left(\frac{-t^2}{2\|\sigma\|_p^2}\right) \le 2d\exp\left(\frac{-t^2}{2d^{2/p}\left\|\Sigma_1^{1/2}-\Sigma_2^{1/2}\right\|_2^2}\right).$$

□

We also include, for completeness, a variance bound (Lemma SA.6) and an exponential concentration inequality (Lemma SA.7) for α-mixing random variables.

LEMMA SA.6 (Variance bounds for α-mixing random variables). *Let $X_1,\dots,X_n$ be real-valued α-mixing random variables with mixing coefficients $\alpha(j)$. Then*

(i) *If for constants M_i we have $|X_i|\le M_i$ a.s. then*

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] \le 4\sum_{j=1}^\infty \alpha(j)\sum_{i=1}^n M_i^2.$$

(ii) *If $\alpha(j)\le e^{-2j/C_\alpha}$ then for any $r>2$ there is a constant C_r depending only on r such that*

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] \le C_r C_\alpha \sum_{i=1}^n \mathbb{E}\left[|X_i|^r\right]^{2/r}.$$

PROOF (Lemma SA.6). Define $\alpha^{-1}(t)=\inf\{j\in\mathbb{N}:\alpha(j)\le t\}$ and $Q_i(t)=\inf\{s\in\mathbb{R}:\mathbb{P}(|X_i|>s)\le t\}$. By Corollary 1.1 in Rio [18] and Hölder’s inequality for $r>2$,

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] \le 4\sum_{i=1}^n\int_0^1 \alpha^{-1}(t)Q_i(t)^2\,\mathrm{d}t$$
$$\le 4\sum_{i=1}^n\left(\int_0^1 \alpha^{-1}(t)^{\frac{r}{r-2}}\,\mathrm{d}t\right)^{\frac{r-2}{r}}\left(\int_0^1 |Q_i(t)|^r\,\mathrm{d}t\right)^{\frac{2}{r}}\,\mathrm{d}t.$$

Now note that if $U \sim \mathrm{Unif}[0,1]$ then $Q_i(U)$ has the same distribution as X_i. Therefore

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] \leq 4\left(\int_0^1 \alpha^{-1}(t)^{\frac{r}{r-2}}\,\mathrm{d}t\right)^{\frac{r-2}{r}} \sum_{i=1}^n \mathbb{E}[|X_i|^r]^{\frac{2}{r}}.$$

If $\alpha(j) \leq e^{-2j/C_\alpha}$ then $\alpha^{-1}(t) \leq \frac{-C_\alpha \log t}{2}$ so, for some constant C_r depending only on r,

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] \leq 2C_\alpha\left(\int_0^1 (-\log t)^{\frac{r}{r-2}}\,\mathrm{d}t\right)^{\frac{r-2}{r}} \sum_{i=1}^n \mathbb{E}[|X_i|^r]^{\frac{2}{r}} \leq C_r C_\alpha \sum_{i=1}^n \mathbb{E}[|X_i|^r]^{\frac{2}{r}}.$$

Alternatively, if for constants M_i we have $|X_i| \leq M_i$ a.s. then

$$\mathrm{Var}\left[\sum_{i=1}^n X_i\right] \leq 4\int_0^1 \alpha^{-1}(t)\,\mathrm{d}t \sum_{i=1}^n M_i^2 \leq 4\sum_{j=1}^\infty \alpha(j) \sum_{i=1}^n M_i^2.$$

□

LEMMA SA.7 (Exponential concentration inequalities for α-mixing random variables). *Let $X_1, \ldots, X_n$ be zero-mean real-valued variables with α-mixing coefficients $\alpha(j) \leq e^{-2j/C_\alpha}$.*

(i) *Suppose $|X_i| \leq M$ a.s. for each $1 \leq i \leq n$. Then for all $t > 0$ there is a constant C_1 with*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > C_1 M\big(\sqrt{nt} + (\log n)(\log\log n)t\big)\right) \leq C_1 e^{-t}.$$

(ii) *Suppose further $\sum_{j=1}^n |\mathrm{Cov}[X_i, X_j]| \leq \sigma^2$. Then for all $t > 0$ there is a constant C_2 with*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \geq C_2\big((\sigma\sqrt{n} + M)\sqrt{t} + M(\log n)^2 t\big)\right) \leq C_2 e^{-t}.$$

PROOF (Lemma SA.7). We apply results from Merlevède, Peligrad and Rio [15], adjusting constants where necessary.

(i) By Theorem 1 in Merlevède, Peligrad and Rio [15],

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq \exp\left(-\frac{C_1 t^2}{nM^2 + Mt(\log n)(\log\log n)}\right).$$

Replace t by $M\sqrt{nt} + M(\log n)(\log\log n)t$.

(ii) By Theorem 2 in Merlevède, Peligrad and Rio [15],

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| > t\right) \leq \exp\left(-\frac{C_2 t^2}{n\sigma^2 + M^2 + Mt(\log n)^2}\right).$$

Replace t by $\sigma\sqrt{n}\sqrt{t} + M\sqrt{t} + M(\log n)^2 t$.

□

SA.2. Main results To establish Theorem 2.1, we first give the analogous result for martingales as Lemma SA.8. Our approach is similar to that used in modern versions of Yurinskii's coupling for independent data, as in Theorem 1 in Le Cam [13] and Theorem 10 in Chapter 10 of Pollard [17]. The proof of Lemma SA.8 relies on constructing a "modified" martingale, which is close to the original martingale, but which has an $\mathcal{H}_0$-measurable terminal quadratic variation.

LEMMA SA.8 (Strong approximation for vector-valued martingales). *Let $X_1, \ldots, X_n$ be $\mathbb{R}^d$-valued square-integrable random vectors adapted to a countably generated filtration $\mathcal{H}_0, \ldots, \mathcal{H}_n$. Suppose that $\mathbb{E}[X_i \mid \mathcal{H}_{i-1}] = 0$ for all $1 \leq i \leq n$ and define the martingale $S = \sum_{i=1}^n X_i$. Let $V_i = \operatorname{Var}[X_i \mid \mathcal{H}_{i-1}]$ and $\Omega = \sum_{i=1}^n V_i - \Sigma$ where Σ is a positive semi-definite $\mathcal{H}_0$-measurable $d \times d$ random matrix. For each $\eta > 0$ and $p \in [1, \infty]$ there is $T \mid \mathcal{H}_0 \sim \mathcal{N}(0, \Sigma)$ with*

$$\mathbb{P}\big(\|S - T\|_p > 5\eta\big) \leq \inf_{t>0} \left\{ 2\mathbb{P}\big(\|Z\|_p > t\big) + \min\left\{ \frac{\beta_{p,2} t^2}{\eta^3}, \frac{\beta_{p,3} t^3}{\eta^4} + \frac{\pi_3 t^3}{\eta^3} \right\} \right\}$$
$$+ \inf_{M \succeq 0} \big\{ 2\gamma(M) + \delta_p(M, \eta) + \varepsilon_p(M, \eta) \big\},$$

where the second infimum is over all positive semi-definite $d \times d$ non-random matrices, and

$$\beta_{p,k} = \sum_{i=1}^n \mathbb{E}\left[\|X_i\|_2^k \|X_i\|_p + \|V_i^{1/2} Z_i\|_2^k \|V_i^{1/2} Z_i\|_p \right], \qquad \gamma(M) = \mathbb{P}\big(\Omega \npreceq M\big),$$

$$\delta_p(M, \eta) = \mathbb{P}\left(\left\| \big((\Sigma + M)^{1/2} - \Sigma^{1/2}\big) Z \right\|_p \geq \eta \right), \qquad \pi_3 = \sum_{i=1}^{n+m} \sum_{|\kappa|=3} \mathbb{E}\Big[\big| \mathbb{E}\left[X_i^\kappa \mid \mathcal{H}_{i-1}\right] \big| \Big],$$

$$\varepsilon_p(M, \eta) = \mathbb{P}\left(\left\| (M - \Omega)^{1/2} Z \right\|_p \geq \eta, \ \Omega \preceq M \right),$$

for $k \in \{2, 3\}$, with $Z, Z_1, \ldots, Z_n$ i.i.d. standard Gaussian on $\mathbb{R}^d$ independent of $\mathcal{H}_n$.

PROOF (Lemma SA.8). **Part 1: constructing a modified martingale**
Take $M \succeq 0$ a fixed positive semi-definite $d \times d$ matrix. We start by constructing a new martingale based on S whose quadratic variation is $\Sigma + M$. Take $m \geq 1$ and define

$$H_k = \Sigma + M - \sum_{i=1}^k V_i, \qquad \tau = \sup\big\{ k \in \{0, 1, \ldots, n\} : H_k \succeq 0 \big\},$$

$$\tilde{X}_i = X_i \mathbb{I}\{i \leq \tau\} + \frac{1}{\sqrt{m}} H_\tau^{1/2} Z_i \mathbb{I}\{n+1 \leq i \leq n+m\}, \qquad \tilde{S} = \sum_{i=1}^{n+m} \tilde{X}_i,$$

where $Z_{n+1}, \ldots, Z_{n+m}$ is an i.i.d. sequence of standard Gaussian vectors in $\mathbb{R}^d$ independent of $\mathcal{H}_n$, noting that $H_0 = \Sigma + M \succeq 0$ a.s. Define the filtration $\tilde{\mathcal{H}}_0, \ldots, \tilde{\mathcal{H}}_{n+m}$, where $\tilde{\mathcal{H}}_i = \mathcal{H}_i$ for $0 \leq i \leq n$ and is the σ-algebra generated by $\mathcal{H}_n$ and $Z_{n+1}, \ldots, Z_i$ for $n+1 \leq i \leq n+m$. Observe that τ is a stopping time with respect to $\tilde{\mathcal{H}}_i$ because $H_{i+1} - H_i = -V_{i+1} \preceq 0$ almost surely, so $\{\tau \leq i\} = \{H_{i+1} \nsucceq 0\}$ for $0 \leq i < n$. This depends only on $V_1, \ldots, V_{i+1}$ and Σ which are $\tilde{\mathcal{H}}_i$-measurable. Similarly, $\{\tau = n\} = \{H_n \succeq 0\} \in \tilde{\mathcal{H}}_{n-1}$. Let $\tilde{V}_i = V_i \mathbb{I}\{i \leq \tau\}$ for $1 \leq i \leq n$ and $\tilde{V}_i = H_\tau / m$ for $n+1 \leq i \leq n+m$. Note that $\tilde{X}_i$ is $\tilde{\mathcal{H}}_i$-measurable and $\tilde{V}_i$ is $\tilde{\mathcal{H}}_{i-1}$-measurable. Further, $\mathbb{E}\big[\tilde{X}_i \mid \tilde{\mathcal{H}}_{i-1}\big] = 0$ and $\mathbb{E}\big[\tilde{X}_i \tilde{X}_i^\mathsf{T} \mid \tilde{\mathcal{H}}_{i-1}\big] = \tilde{V}_i$.

Part 2: bounding the difference between the original and modified martingales
By the triangle inequality,

$$\|S - \tilde{S}\|_p \leq \left\| \sum_{i=\tau+1}^{n} X_i \right\|_p + \left\| \frac{1}{\sqrt{m}} \sum_{i=n+1}^{m} H_\tau^{1/2} Z_i \right\|_p .$$

The first term on the right vanishes on $\{\tau = n\} = \{H_n \succeq 0\} = \{\Omega \preceq M\}$. For the second term, note that $\frac{1}{\sqrt{m}} \sum_{i=n+1}^{m} H_\tau^{1/2} Z_i$ is distributed as $H_\tau^{1/2} Z$, where Z is an independent standard Gaussian. Also $\mathbb{P}\big(\|H_\tau^{1/2} Z\|_p > \eta\big) \leq \mathbb{P}\big(\|H_n^{1/2} Z\|_p > \eta,\, \Omega \preceq M\big) + \mathbb{P}\big(\Omega \npreceq M\big)$. Therefore

$$\begin{aligned} \mathbb{P}\big(\|S - \tilde{S}\|_p > \eta\big) &\leq 2\mathbb{P}\big(\Omega \npreceq M\big) + \mathbb{P}\big(\|(M - \Omega)^{1/2} Z\|_p > \eta,\, \Omega \preceq M\big) \\ &= 2\gamma(M) + \varepsilon_p(M, \eta). \end{aligned} \tag{2}$$

Part 3: strong approximation of the modified martingale
Let $\tilde{Z}_1, \ldots, \tilde{Z}_{n+m}$ be i.i.d. $\mathcal{N}(0, I_d)$ and independent of $\tilde{\mathcal{H}}_{n+m}$. Define $\check{X}_i = \tilde{V}_i^{1/2} \tilde{Z}_i$ and $\check{S} = \sum_{i=1}^{n+m} \check{X}_i$. Fix a Borel set $A \subseteq \mathbb{R}^d$ and $\sigma, \eta > 0$ and let $f = f_{A\eta\sigma}$ be the function defined in Lemma SA.2. By the Lindeberg method, write the telescoping sum

$$\mathbb{E}\Big[f\big(\tilde{S}\big) - f\big(\check{S}\big) \mid \mathcal{H}_0\Big] = \sum_{i=1}^{n+m} \mathbb{E}\Big[f\big(Y_i + \tilde{X}_i\big) - f\big(Y_i + \check{X}_i\big) \mid \mathcal{H}_0\Big]$$

where $Y_i = \sum_{j=1}^{i-1} \tilde{X}_j + \sum_{j=i+1}^{n+m} \check{X}_j$. By Lemma SA.2 we have for $k \geq 0$

$$\begin{aligned} &\left| \mathbb{E}\big[f(Y_i + \tilde{X}_i) - f(Y_i + \check{X}_i) \mid \mathcal{H}_0\big] - \sum_{|\kappa|=0}^{k} \frac{1}{\kappa!} \mathbb{E}\left[\partial^\kappa f(Y_i) \left(\tilde{X}_i^\kappa - \check{X}_i^\kappa\right) \mid \mathcal{H}_0\right] \right| \\ &\quad \leq \frac{1}{\sigma^k \eta \sqrt{k!}} \mathbb{E}\left[\|\tilde{X}_i\|_p \|\tilde{X}_i\|_2^k + \|\check{X}_i\|_p \|\check{X}_i\|_2^k \mid \mathcal{H}_0\right]. \end{aligned}$$

With $k \in \{2, 3\}$, we bound each summand. With $|\kappa| = 0$ we have $\tilde{X}_i^\kappa = \check{X}_i^\kappa$, so consider $|\kappa| = 1$. Noting that $\sum_{i=1}^{n+m} \tilde{V}_i = \Sigma + M$, define

$$\tilde{Y}_i = \sum_{j=1}^{i-1} \tilde{X}_j + \tilde{Z}_i \left(\sum_{j=i+1}^{n+m} \tilde{V}_j \right)^{1/2} = \sum_{j=1}^{i-1} \tilde{X}_j + \tilde{Z}_i \left(\Sigma + M - \sum_{j=1}^{i} \tilde{V}_j \right)^{1/2}$$

and let $\check{\mathcal{H}}_i$ be the σ-algebra generated by $\tilde{\mathcal{H}}_{i-1}$ and $\tilde{Z}_i$. Note that $\tilde{Y}_i$ is $\check{\mathcal{H}}_i$-measurable and that Y_i and $\tilde{Y}_i$ have the same distribution conditional on $\tilde{\mathcal{H}}_{n+m}$. So

$$\begin{aligned} \sum_{|\kappa|=1} \frac{1}{\kappa!} \mathbb{E}\Big[\partial^\kappa f(Y_i)\big(\tilde{X}_i^\kappa - \check{X}_i^\kappa\big) \mid \mathcal{H}_0\Big] &= \mathbb{E}\Big[\nabla f(Y_i)^\mathsf{T} \big(\tilde{X}_i - \tilde{V}_i^{1/2} \tilde{Z}_i\big) \mid \mathcal{H}_0\Big] \\ &= \mathbb{E}\Big[\nabla f(\tilde{Y}_i)^\mathsf{T} \tilde{X}_i \mid \mathcal{H}_0\Big] - \mathbb{E}\Big[\nabla f(Y_i)^\mathsf{T} \tilde{V}_i^{1/2} \tilde{Z}_i \mid \mathcal{H}_0\Big] \\ &= \mathbb{E}\Big[\nabla f(\tilde{Y}_i)^\mathsf{T} \mathbb{E}\Big[\tilde{X}_i \mid \check{\mathcal{H}}_i\Big] \mid \mathcal{H}_0\Big] - \mathbb{E}\Big[\tilde{Z}_i\Big] \mathbb{E}\Big[\nabla f(Y_i)^\mathsf{T} \tilde{V}_i^{1/2} \mid \mathcal{H}_0\Big] \\ &= \mathbb{E}\Big[\nabla f(\tilde{Y}_i)^\mathsf{T} \mathbb{E}\Big[\tilde{X}_i \mid \tilde{\mathcal{H}}_{i-1}\Big] \mid \mathcal{H}_0\Big] - 0 = 0. \end{aligned}$$

Next, if $|\kappa| = 2$ then

$$\sum_{|\kappa|=2} \frac{1}{\kappa!} \mathbb{E}\left[\partial^\kappa f(Y_i) \left(\tilde{X}_i^\kappa - \check{X}_i^\kappa\right) \mid \mathcal{H}_0\right]$$

$$
\begin{aligned}
&= \frac{1}{2}\mathbb{E}\left[\tilde{X}_i^\mathsf{T}\nabla^2 f(Y_i)\tilde{X}_i - \tilde{Z}_i^\mathsf{T}\tilde{V}_i^{1/2}\nabla^2 f(Y_i)\tilde{V}_i^{1/2}\tilde{Z}_i \mid \mathcal{H}_0\right] \\
&= \frac{1}{2}\mathbb{E}\left[\mathbb{E}\left[\operatorname{Tr}\nabla^2 f(\check{Y}_i)\tilde{X}_i\tilde{X}_i^\mathsf{T} \mid \check{\mathcal{H}}_i\right] \mid \mathcal{H}_0\right] - \frac{1}{2}\mathbb{E}\left[\operatorname{Tr}\tilde{V}_i^{1/2}\nabla^2 f(Y_i)\tilde{V}_i^{1/2} \mid \mathcal{H}_0\right]\mathbb{E}\left[\tilde{Z}_i\tilde{Z}_i^\mathsf{T}\right] \\
&= \frac{1}{2}\mathbb{E}\left[\operatorname{Tr}\nabla^2 f(Y_i)\mathbb{E}\left[\tilde{X}_i\tilde{X}_i^\mathsf{T} \mid \tilde{\mathcal{H}}_{i-1}\right] \mid \mathcal{H}_0\right] - \frac{1}{2}\mathbb{E}\left[\operatorname{Tr}\nabla^2 f(Y_i)\tilde{V}_i \mid \mathcal{H}_0\right] = 0.
\end{aligned}
$$

Finally if $|\kappa| = 3$, then since $\check{X}_i \sim \mathcal{N}(0, \tilde{V}_i)$ conditional on $\tilde{\mathcal{H}}_{n+m}$, we have by symmetry of the Gaussian distribution and Lemma SA.2,

$$
\begin{aligned}
&\left|\sum_{|\kappa|=3}\frac{1}{\kappa!}\mathbb{E}\left[\partial^\kappa f(Y_i)\left(\tilde{X}_i^\kappa - \check{X}_i^\kappa\right) \mid \mathcal{H}_0\right]\right| \\
&\quad= \left|\sum_{|\kappa|=3}\frac{1}{\kappa!}\left(\mathbb{E}\left[\partial^\kappa f(\check{Y}_i)\mathbb{E}\left[\tilde{X}_i^\kappa \mid \check{\mathcal{H}}_i\right] \mid \mathcal{H}_0\right] - \mathbb{E}\left[\partial^\kappa f(Y_i)\,\mathbb{E}\left[\check{X}_i^\kappa \mid \tilde{\mathcal{H}}_{n+m}\right] \mid \mathcal{H}_0\right]\right)\right| \\
&\quad= \left|\sum_{|\kappa|=3}\frac{1}{\kappa!}\mathbb{E}\left[\partial^\kappa f(Y_i)\,\mathbb{E}\left[\tilde{X}_i^\kappa \mid \tilde{\mathcal{H}}_{i-1}\right] \mid \mathcal{H}_0\right]\right| \le \frac{1}{\sigma^3}\sum_{|\kappa|=3}\mathbb{E}\left[\left|\mathbb{E}\left[\tilde{X}_i^\kappa \mid \tilde{\mathcal{H}}_{i-1}\right]\right| \mid \mathcal{H}_0\right].
\end{aligned}
$$

Combining these and summing over i with $k = 2$ shows

$$
\mathbb{E}\left[f(\tilde{S}) - f(\check{S}) \mid \mathcal{H}_0\right] \le \frac{1}{\sigma^2\eta\sqrt{2}}\sum_{i=1}^{n+m}\mathbb{E}\left[\|\tilde{X}_i\|_p\|\tilde{X}_i\|_2^2 + \|\check{X}_i\|_p\|\check{X}_i\|_2^2 \mid \mathcal{H}_0\right]
$$

On the other hand, taking $k = 3$ gives

$$
\begin{aligned}
\mathbb{E}\left[f(\tilde{S}) - f(\check{S}) \mid \mathcal{H}_0\right] &\le \frac{1}{\sigma^3\eta\sqrt{6}}\sum_{i=1}^{n+m}\mathbb{E}\left[\|\tilde{X}_i\|_p\|\tilde{X}_i\|_2^3 + \|\check{X}_i\|_p\|\check{X}_i\|_2^3 \mid \mathcal{H}_0\right] \\
&\quad+ \frac{1}{\sigma^3}\sum_{i=1}^{n+m}\sum_{|\kappa|=3}\mathbb{E}\left[\left|\mathbb{E}\left[\tilde{X}_i^\kappa \mid \tilde{\mathcal{H}}_{i-1}\right]\right| \mid \mathcal{H}_0\right].
\end{aligned}
$$

For $1 \le i \le n$ we have $\|\tilde{X}_i\| \le \|X_i\|$ and $\|\check{X}_i\| \le \|V_i^{1/2}\tilde{Z}_i\|$. For $n+1 \le i \le n+m$ we have $\tilde{X}_i = H_\tau^{1/2}Z_i/\sqrt{m}$ and $\check{X}_i = H_\tau^{1/2}\tilde{Z}_i/\sqrt{m}$ which are equal in distribution given $\mathcal{H}_0$. Therefore with

$$
\tilde{\beta}_{p,k} = \sum_{i=1}^{n}\mathbb{E}\left[\|X_i\|_p\|X_i\|_2^k + \|V_i^{1/2}Z_i\|_p\|V_i^{1/2}Z_i\|_2^k \mid \mathcal{H}_0\right],
$$

we have, since $k \in \{2, 3\}$,

$$
\sum_{i=1}^{n+m}\mathbb{E}\left[\|\tilde{X}_i\|_p\|\tilde{X}_i\|_2^k + \|\check{X}_i\|_p\|\check{X}_i\|_2^k \mid \mathcal{H}_0\right] \le \tilde{\beta}_{p,k} + \frac{2}{\sqrt{m}}\mathbb{E}\left[\|H_\tau^{1/2}Z\|_p\|H_\tau^{1/2}Z\|_2^k \mid \mathcal{H}_0\right].
$$

Since H_i is weakly decreasing under the semi-definite partial order, we have $H_\tau \preceq H_0 = \Sigma + M$ implying that $|(H_\tau)_{jj}| \le \|\Sigma + M\|_{\max}$ and $\mathbb{E}\left[|(H_\tau^{1/2}Z)_j|^3 \mid \mathcal{H}_0\right] \le \sqrt{8/\pi}\,\|\Sigma + M\|_{\max}^{3/2}$. Hence as $p \ge 1$ and $k \in \{2, 3\}$,

$$
\mathbb{E}\left[\|H_\tau^{1/2}Z\|_p\|H_\tau^{1/2}Z\|_2^k \mid \mathcal{H}_0\right] \le \mathbb{E}\left[\|H_\tau^{1/2}Z\|_1^{k+1} \mid \mathcal{H}_0\right]
$$

$$\le d^{k+1} \max_{1\le j\le d} \mathbb{E}\left[|(H_\tau^{1/2}Z)_j|^{k+1} \mid \mathcal{H}_0\right]$$
$$\le 3d^4 \,\|\Sigma + M\|_{\max}^{(k+1)/2} \le 6d^4 \,\|\Sigma\|_{\max}^{(k+1)/2} + 6d^4\|M\|.$$

Assuming some X_i is not identically zero so the result is non-trivial, and supposing that Σ is bounded a.s. (replacing Σ by $\Sigma \cdot \mathbb{I}\{\|\Sigma\|_{\max} \le C\}$ for an appropriately large C if necessary), take m large enough that

$$\frac{2}{\sqrt{m}}\mathbb{E}\left[\|H_\tau^{1/2}Z\|_p\|H_\tau^{1/2}Z\|_2^k \mid \mathcal{H}_0\right] \le \frac{1}{4}\beta_{p,k}. \tag{3}$$

Further, if $|\kappa| = 3$ then $\left|\mathbb{E}\big[\tilde{X}_i^\kappa \mid \tilde{\mathcal{H}}_{i-1}\big]\right| \le \left|\mathbb{E}\left[X_i^\kappa \mid \mathcal{H}_{i-1}\right]\right|$ for $1 \le i \le n$ while by symmetry of the Gaussian distribution $\mathbb{E}\left[\tilde{X}_i^\kappa \mid \tilde{\mathcal{H}}_{i-1}\right] = 0$ for $n+1 \le i \le n+m$. Hence with

$$\tilde{\pi}_3 = \sum_{i=1}^{n+m}\sum_{|\kappa|=3}\mathbb{E}\Big[\big|\mathbb{E}\left[X_i^\kappa \mid \mathcal{H}_{i-1}\right]\big| \mid \mathcal{H}_0\Big],$$

we have

$$\mathbb{E}\left[f\big(\tilde{S}\big) - f\big(\check{S}\big) \mid \mathcal{H}_0\right] \le \min\left\{\frac{3\tilde{\beta}_{p,2}}{4\sigma^2\eta} + \frac{\beta_{p,2}}{4\sigma^2\eta}, \frac{3\tilde{\beta}_{p,3}}{4\sigma^3\eta} + \frac{\beta_{p,3}}{4\sigma^3\eta} + \frac{\tilde{\pi}_3}{\sigma^3}\right\}.$$

Along with Lemma SA.2, and with $\sigma = \eta/t$ and $\varepsilon = \mathbb{P}(\|Z\|_p > t)$, we conclude that

$$\mathbb{P}(\tilde{S} \in A \mid \mathcal{H}_0) = \mathbb{E}\big[\mathbb{I}\{\tilde{S} \in A\} - f(\tilde{S}) \mid \mathcal{H}_0\big] + \mathbb{E}\big[f(\tilde{S}) - f\big(\check{S}\big) \mid \mathcal{H}_0\big] + \mathbb{E}\big[f\big(\check{S}\big) \mid \mathcal{H}_0\big]$$
$$\le \varepsilon\,\mathbb{P}(\tilde{S} \in A \mid \mathcal{H}_0) + \min\left\{\frac{3\tilde{\beta}_{p,2}}{4\sigma^2\eta} + \frac{\beta_{p,2}}{4\sigma^2\eta}, \frac{3\tilde{\beta}_{p,3}}{4\sigma^3\eta} + \frac{\beta_{p,3}}{4\sigma^3\eta} + \frac{\tilde{\pi}_3}{\sigma^3}\right\}$$
$$+ \varepsilon + (1-\varepsilon)\mathbb{P}\big(\check{S} \in A_p^{3\eta} \mid \mathcal{H}_0\big)$$
$$\le \mathbb{P}\big(\check{S} \in A_p^{3\eta} \mid \mathcal{H}_0\big) + 2\mathbb{P}(\|Z\|_p > t) + \min\left\{\frac{3\tilde{\beta}_{p,2}t^2}{4\eta^3} + \frac{\beta_{p,2}t^2}{4\eta^3}, \frac{3\tilde{\beta}_{p,3}t^3}{4\eta^4} + \frac{\beta_{p,3}t^3}{4\eta^4} + \frac{\tilde{\pi}_3 t^3}{\eta^3}\right\}.$$

Taking a supremum and an outer expectation yields with $\beta_{p,k} = \mathbb{E}\big[\tilde{\beta}_{p,k}\big]$ and $\pi_3 = \mathbb{E}[\tilde{\pi}_3]$,

$$\mathbb{E}^*\left[\sup_{A\in\mathcal{B}(\mathbb{R}^d)}\Big\{\mathbb{P}(\tilde{S} \in A \mid \mathcal{H}_0) - \mathbb{P}\big(\check{S} \in A_p^{3\eta} \mid \mathcal{H}_0\big)\Big\}\right]$$
$$\le 2\mathbb{P}(\|Z\|_p > t) + \min\left\{\frac{\beta_{p,2}t^2}{\eta^3}, \frac{\beta_{p,3}t^3}{\eta^4} + \frac{\pi_3 t^3}{\eta^3}\right\}.$$

Finally, since $\check{S} = \sum_{i=1}^n \tilde{V}_i^{1/2}\tilde{Z}_i \sim \mathcal{N}(0, \Sigma + M)$ conditional on $\mathcal{H}_0$, the conditional Strassen theorem in Lemma SA.1 ensures the existence of $\tilde{S}$ and $\tilde{T} \mid \mathcal{H}_0 \sim \mathcal{N}(0, \Sigma + M)$ such that

$$\mathbb{P}\left(\|\tilde{S} - \tilde{T}\|_p > 3\eta\right) \le \inf_{t>0}\left\{2\mathbb{P}(\|Z\|_p > t) + \min\left\{\frac{\beta_{p,2}t^2}{\eta^3}, \frac{\beta_{p,3}t^3}{\eta^4} + \frac{\pi_3 t^3}{\eta^3}\right\}\right\}, \tag{4}$$

since the infimum is attained by continuity of $\|Z\|_p$.

Part 4: conclusion

We show how to write $\tilde{T} = (\Sigma + M)^{1/2} W$ where $W \sim \mathcal{N}(0, I_d)$ and use this representation to construct $T \mid \mathcal{H}_0 \sim \mathcal{N}(0, \Sigma)$. By the spectral theorem, let $\Sigma + M = U \Lambda U^\mathsf{T}$ where U is a $d \times d$ orthogonal random matrix and Λ is a diagonal $d \times d$ random matrix with diagonal entries satisfying $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and $\lambda_{r+1} = \cdots = \lambda_d = 0$ where $r = \operatorname{rank}(\Sigma + M)$. Let Λ^+ be the Moore–Penrose pseudo-inverse of Λ (obtained by inverting its non-zero elements) and define $W = U(\Lambda^+)^{1/2} U^\mathsf{T} \tilde{T} + U \tilde{W}$, where the first r elements of $\tilde{W}$ are zero and the last $d - r$ elements are i.i.d. $\mathcal{N}(0,1)$ independent from $\tilde{T}$. Then, it is easy to check that $W \sim \mathcal{N}(0, I_d)$ and that $\tilde{T} = (\Sigma + M)^{1/2} W$. Now define $T = \Sigma^{1/2} W$ so

$$\mathbb{P}\big(\|T - \tilde{T}\|_p > \eta\big) = \mathbb{P}\Big(\big\|\big((\Sigma + M)^{1/2} - \Sigma^{1/2}\big) W\big\|_p > \eta\Big) = \delta_p(M, \eta). \tag{5}$$

Finally (2), (4), (5), the triangle inequality and a union bound conclude the proof since by taking an infimum over $M \succeq 0$, and by possibly reducing the constant of $1/4$ in (3) to account for this infimum being potentially unattainable,

$$\begin{aligned}
\mathbb{P}\big(\|S - T\|_p > 5\eta\big) &\leq \mathbb{P}\big(\|\tilde{S} - \tilde{T}\|_p > 3\eta\big) + \mathbb{P}\big(\|S - \tilde{S}\|_p > \eta\big) + \mathbb{P}\big(\|T - \tilde{T}\|_p > \eta\big) \\
&\leq \inf_{t > 0} \left\{ 2\mathbb{P}\big(\|Z\|_p > t\big) + \min\left\{ \frac{\beta_{p,2} t^2}{\eta^3}, \frac{\beta_{p,3} t^3}{\eta^4} + \frac{\pi_3 t^3}{\eta^3} \right\} \right\} \\
&\quad + \inf_{M \succeq 0} \big\{ 2\gamma(M) + \delta_p(M, \eta) + \varepsilon_p(M, \eta) \big\}.
\end{aligned}$$

□

Applying Lemma SA.8 and the martingale approximation immediately yields Theorem 2.1.

PROOF (Theorem 2.1). Apply Lemma SA.8 to the martingale $\sum_{i=1}^n \tilde{X}_i$, noting that $S - \sum_{i=1}^n \tilde{X}_i = U$. □

Bounding the quantities in Theorem 2.1 gives a user-friendly version as Proposition 2.1.

PROOF (Proposition 2.1). We set $M = \nu^2 I_d$ and bound each term appearing on the right-hand side of the main inequality in Proposition 2.1

Part 1: bounding $\mathbb{P}(\|Z\|_p > t)$

By Markov's inequality and Lemma SA.4, we have $\mathbb{P}(\|Z\|_p > t) \leq \mathbb{E}[\|Z\|_p]/t \leq \phi_p(d)/t$.

Part 2: bounding $\gamma(M)$

With $M = \nu^2 I_d$ and by Markov, $\gamma(M) = \mathbb{P}\big(\Omega \not\preceq M\big) = \mathbb{P}\big(\|\Omega\|_2 > \nu^2\big) \leq \nu^{-2} \mathbb{E}[\|\Omega\|_2]$.

Part 3: bounding $\delta(M, \eta)$

By Markov's inequality and Lemma SA.4, using $\max_j |M_{jj}| \leq \|M\|_2$ for $M \succeq 0$,

$$\delta_p(M, \eta) = \mathbb{P}\left(\big\|\big((\Sigma + M)^{1/2} - \Sigma^{1/2}\big) Z\big\|_p \geq \eta \right) \leq \frac{\phi_p(d)}{\eta} \mathbb{E}\left[\big\|(\Sigma + M)^{1/2} - \Sigma^{1/2}\big\|_2 \right].$$

For semi-definite matrices the eigenvalue operator commutes with smooth matrix functions so

$$\|(\Sigma + M)^{1/2} - \Sigma^{1/2}\|_2 = \max_{1 \leq j \leq d} \left| \sqrt{\lambda_j(\Sigma) + \nu^2} - \sqrt{\lambda_j(\Sigma)} \right| \leq \nu$$

and hence $\delta_p(M, \eta) \leq \phi_p(d) \nu / \eta$.

Part 4: bounding $\varepsilon(M,\eta)$
Note that $(M-\Omega)^{1/2}Z$ is a centered Gaussian conditional on $\mathcal{H}_n$, on the event $\{\Omega \preceq M\}$. We thus have by Markov's inequality, Lemma SA.4 and Jensen's inequality that

$$\begin{aligned}
\varepsilon_p(M,\eta) &= \mathbb{P}\left(\left\|(M-\Omega)^{1/2}Z\right\|_p \geq \eta,\ \Omega \preceq M\right) \\
&\leq \frac{1}{\eta}\mathbb{E}\left[\mathbb{I}\{\Omega \preceq M\}\mathbb{E}\left[\left\|(M-\Omega)^{1/2}Z\right\|_p \mid \mathcal{H}_n\right]\right] \\
&\leq \frac{\phi_p(d)}{\eta}\mathbb{E}\left[\mathbb{I}\{\Omega \preceq M\}\max_{1\leq j\leq d}\sqrt{(M-\Omega)_{jj}}\right] \leq \frac{\phi_p(d)}{\eta}\mathbb{E}\left[\sqrt{\|M-\Omega\|_2}\right] \\
&\leq \frac{\phi_p(d)}{\eta}\mathbb{E}\left[\sqrt{\|\Omega\|_2}+\nu\right] \leq \frac{\phi_p(d)}{\eta}\left(\sqrt{\mathbb{E}[\|\Omega\|_2]}+\nu\right).
\end{aligned}$$

Thus by Theorem 2.1 and the previous parts,

$$\begin{aligned}
\mathbb{P}\big(\|S-T\|_p > 6\eta\big) &\leq \inf_{t>0}\left\{2\mathbb{P}\big(\|Z\|_p > t\big) + \min\left\{\frac{\beta_{p,2}t^2}{\eta^3}, \frac{\beta_{p,3}t^3}{\eta^4} + \frac{\pi_3 t^3}{\eta^3}\right\}\right\} \\
&\quad + \inf_{M\succeq 0}\big\{2\gamma(M) + \delta_p(M,\eta) + \varepsilon_p(M,\eta)\big\} + \mathbb{P}\big(\|U\|_p > \eta\big) \\
&\leq \inf_{t>0}\left\{\frac{2\phi_p(d)}{t} + \min\left\{\frac{\beta_{p,2}t^2}{\eta^3}, \frac{\beta_{p,3}t^3}{\eta^4} + \frac{\pi_3 t^3}{\eta^3}\right\}\right\} \\
&\quad + \inf_{\nu>0}\left\{\frac{2\mathbb{E}\left[\|\Omega\|_2\right]}{\nu^2} + \frac{2\phi_p(d)\nu}{\eta}\right\} + \frac{\phi_p(d)\sqrt{\mathbb{E}\left[\|\Omega\|_2\right]}}{\eta} + \mathbb{P}\big(\|U\|_p > \eta\big).
\end{aligned}$$

In general, set $t = 2^{1/3}\phi_p(d)^{1/3}\beta_{p,2}^{-1/3}\eta$ and $\nu = \mathbb{E}[\|\Omega\|_2]^{1/3}\phi_p(d)^{-1/3}\eta^{1/3}$, replacing η with $\eta/6$ to see

$$\mathbb{P}\big(\|S-T\|_p > 6\eta\big) \leq 24\left(\frac{\beta_{p,2}\phi_p(d)^2}{\eta^3}\right)^{1/3} + 17\left(\frac{\mathbb{E}\left[\|\Omega\|_2\right]\phi_p(d)^2}{\eta^2}\right)^{1/3} + \mathbb{P}\left(\|U\|_p > \frac{\eta}{6}\right).$$

Whenever $\pi_3 = 0$ we can set $t = 2^{1/4}\phi_p(d)^{1/4}\beta_{p,3}^{-1/4}\eta$, and with ν as above we obtain

$$\mathbb{P}\big(\|S-T\|_p > \eta\big) \leq 24\left(\frac{\beta_{p,3}\phi_p(d)^3}{\eta^4}\right)^{1/4} + 17\left(\frac{\mathbb{E}\left[\|\Omega\|_2\right]\phi_p(d)^2}{\eta^2}\right)^{1/3} + \mathbb{P}\left(\|U\|_p > \frac{\eta}{6}\right).$$

□

After establishing Proposition 2.1, Corollaries 2.1, 2.2 and 2.3 follow as in the main text.

PROOF (Corollary 2.1). Proposition 2.1 with $\mathbb{P}(\|U\|_p > \frac{\eta}{6}) \leq \frac{6}{\eta}\sum_{i=1}^n c_i(\zeta_i + \zeta_{n-i+1})$. □

PROOF (Corollary 2.2). By Proposition 2.1 with $U = 0$ a.s. □

PROOF (Corollary 2.3). By Corollary 2.2 with $\Omega = 0$ a.s. □

We conclude this section with a discussion expanding on the comments made in Remark 1 on deriving bounds in probability from Yurinskii's coupling. Consider for illustration the independent data second-order result given in Corollary 2.3: for each $\eta > 0$, there exists $T_n \mid \mathcal{H}_0 \sim \mathcal{N}(0,\Sigma)$ satisfying

$$\mathbb{P}\big(\|S_n - T_n\|_p > \eta\big) \leq 24\left(\frac{\beta_{p,2}\phi_p(d)^2}{\eta^3}\right)^{1/3},$$

where here we make explicit the dependence on the sample size n for clarity. The naive approach to converting this into a probability bound for $\|S_n - T_n\|_p$ is to select η to ensure the right-hand side is of order 1, arguing that the probability can then be made arbitrarily small by taking, in this case, η to be a large enough multiple of $\beta_{p,2}^{1/3}\phi_p(d)^{2/3}$. However, the somewhat subtle mistake is in neglecting the fact that the realization of the coupling variable T_n will in general depend on η, rendering the resulting bound invalid. As an explicit example of this phenomenon, take $\eta > 1$ and suppose $\|S_n - T_n(\eta)\| = \eta$ with probability $1 - 1/\eta$ and $\|S_n - T_n(\eta)\| = n$ with probability $1/\eta$. Then $\mathbb{P}\big(\|S_n - T_n(\eta)\| > \eta\big) = 1/\eta$ but it is not true for any η that $\|S_n - T_n(\eta)\| \lesssim_{\mathbb{P}} 1$.

We propose in Remark 1 the following fix. Instead of selecting η to ensure the right-hand side is of order 1, we instead choose it so the bound converges (slowly) to zero. This is easily achieved by taking the naive and incorrect bound and multiplying by some divergent sequence R_n. The resulting inequality reads, in the case of Corollary 2.3 with $\eta = \beta_{p,2}^{1/3}\phi_p(d)^{2/3}R_n$,

$$\mathbb{P}\Big(\|S_n - T_n\|_p > \beta_{p,2}^{1/3}\phi_p(d)^{2/3}R_n\Big) \le \frac{24}{R_n} \to 0.$$

We thus recover, for the price of a rate which is slower by an arbitrarily small amount, a valid upper bound in probability, as we can immediately conclude that

$$\|S_n - T_n\|_p \lesssim_{\mathbb{P}} \beta_{p,2}^{1/3}\phi_p(d)^{2/3}R_n.$$

SA.3. Strong approximation for martingale empirical processes We begin by presenting some calculations omitted from the main text relating to the motivating example of kernel density estimation with i.i.d. data. First, the bias of this estimator is bounded as

$$\big|\mathbb{E}\big[\hat{g}(x)\big] - g(x)\big| = \left|\int_{\frac{-x}{h}}^{\frac{1-x}{h}} K(\xi)\,\mathrm{d}\xi - 1\right| \le 2\int_{\frac{a}{h}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}\,\mathrm{d}\xi \le \frac{h}{a}\sqrt{\frac{2}{\pi}}\, e^{-\frac{a^2}{2h^2}}.$$

Next, we do the calculations necessary to apply Corollary 2.3. Define $k_{ij} = \frac{1}{nh}K\left(\frac{X_i - x_j}{h}\right)$ and $k_i = (k_{ij} : 1 \le j \le N)$. Then $\|k_i\|_\infty \le \frac{1}{nh\sqrt{2\pi}}$ a.s. and $\mathbb{E}[\|k_i\|_2^2] \le \frac{N}{n^2 h}\int_{-\infty}^{\infty} K(\xi)^2\,\mathrm{d}\xi \le \frac{N}{2n^2 h\sqrt{\pi}}$. Let $V = \mathrm{Var}[k_i] \in \mathbb{R}^{N\times N}$, so assuming that $1/h \ge \log 2N$, by Lemma SA.3,

$$\begin{aligned}
\beta_{\infty,2} &= n\mathbb{E}\left[\|k_i\|_2^2\|k_i\|_\infty\right] + n\mathbb{E}\left[\|V^{1/2}Z\|_2^2\|V^{1/2}Z\|_\infty\right] \\
&\le \frac{N}{\sqrt{8}n^2h^2\pi} + \frac{4N\sqrt{\log 2N}}{\sqrt{8}n^2h^{3/2}\pi^{3/4}} \le \frac{N}{n^2h^2}.
\end{aligned}$$

Finally, we verify the stochastic continuity bounds. By the Lipschitz property of K, it is easy to show that for $x, x' \in \mathcal{X}$ we have $\left|\frac{1}{h}K\left(\frac{X_i - x}{h}\right) - \frac{1}{h}K\left(\frac{X_i - x'}{h}\right)\right| \lesssim \frac{|x-x'|}{h^2}$ almost surely, and also that $\mathbb{E}\Big[\left|\frac{1}{h}K\left(\frac{X_i - x}{h}\right) - \frac{1}{h}K\left(\frac{X_i - x'}{h}\right)\right|^2\Big] \lesssim \frac{|x-x'|^2}{h^3}$. By chaining with the Bernstein–Orlicz norm and polynomial covering numbers,

$$\sup_{|x-x'|\le\delta} \big\|S(x) - S(x')\big\|_\infty \lesssim_{\mathbb{P}} \delta\sqrt{\frac{\log n}{nh^3}}$$

whenever $\log(N/h) \lesssim \log n$ and $nh \gtrsim \log n$. By a Gaussian process maximal inequality [20, Corollary 2.2.8] the same bound holds for $T(x)$ with

$$\sup_{|x-x'|\le\delta} \big\|T(x) - T(x')\big\|_\infty \lesssim_{\mathbb{P}} \delta\sqrt{\frac{\log n}{nh^3}}.$$

PROOF (Lemma 3.1). For $x, x' \in [a, 1-a]$, the scaled covariance function of this nonparametric estimator is

$$\begin{aligned} nh \operatorname{Cov}\left[\hat{g}(x), \hat{g}(x')\right] &= \frac{1}{h}\mathbb{E}\left[K\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x'}{h}\right)\right] \\ &\quad - \frac{1}{h}\mathbb{E}\left[K\left(\frac{X_i - x}{h}\right)\right]\mathbb{E}\left[K\left(\frac{X_i - x'}{h}\right)\right] \\ &= \frac{1}{2\pi}\int_{\frac{-x}{h}}^{\frac{1-x}{h}} \exp\left(-\frac{t^2}{2}\right)\exp\left(-\frac{1}{2}\left(t + \frac{x - x'}{h}\right)^2\right) \mathrm{d}t - hI(x)I(x') \end{aligned}$$

where $I(x) = \frac{1}{\sqrt{2\pi}}\int_{-x/h}^{(1-x)/h} e^{-t^2/2}\,\mathrm{d}t$. Completing the square and a substitution gives

$$nh \operatorname{Cov}\left[\hat{g}(x), \hat{g}(x')\right] = \frac{1}{2\pi}\exp\left(-\frac{1}{4}\left(\frac{x - x'}{h}\right)^2\right)\int_{\frac{-x-x'}{2h}}^{\frac{2-x-x'}{2h}} \exp\left(-t^2\right)\,\mathrm{d}t - hI(x)I(x').$$

Now we show that since x, x' are not too close to the boundary of $[0,1]$, the limits in the above integral can be replaced by $\pm\infty$. Note that $\frac{-x-x'}{2h} \le \frac{-a}{h}$ and $\frac{2-x-x'}{2h} \ge \frac{a}{h}$ so

$$\int_{-\infty}^{\infty} \exp\left(-t^2\right)\,\mathrm{d}t - \int_{\frac{-x-x'}{2h}}^{\frac{2-x-x'}{2h}} \exp\left(-t^2\right)\,\mathrm{d}t \le 2\int_{a/h}^{\infty} \exp\left(-t^2\right)\,\mathrm{d}t \le \frac{h}{a}\exp\left(-\frac{a^2}{h^2}\right).$$

Therefore since $\int_{-\infty}^{\infty} e^{-t^2}\,\mathrm{d}t = \sqrt{\pi}$,

$$\left| nh \operatorname{Cov}\left[\hat{g}(x), \hat{g}(x')\right] - \frac{1}{2\sqrt{\pi}}\exp\left(-\frac{1}{4}\left(\frac{x - x'}{h}\right)^2\right) + hI(x)I(x')\right| \le \frac{h}{2\pi a}\exp\left(-\frac{a^2}{h^2}\right).$$

Define the $N \times N$ matrix $\tilde{\Sigma}_{ij} = \frac{1}{2\sqrt{\pi}}\exp\left(-\frac{1}{4}\left(\frac{x_i - x_j}{h}\right)^2\right)$. By Baxter [1, Proposition 2.4, Proposition 2.5 and Equation 2.10], with $\mathcal{B}_k = \left\{b \in \mathbb{R}^{\mathbb{Z}} : \sum_{i\in\mathbb{Z}} \mathbb{I}\{b_i \ne 0\} \le k\right\}$,

$$\inf_{k\in\mathbb{N}} \inf_{b\in\mathbb{R}^k} \frac{\sum_{i=1}^k \sum_{j=1}^k b_i b_j\, e^{-\lambda(i-j)^2}}{\sum_{i=1}^k b_i^2} = \sqrt{\frac{\pi}{\lambda}}\sum_{i=-\infty}^{\infty}\exp\left(-\frac{(\pi e + 2\pi i)^2}{4\lambda}\right).$$

We use Riemann sums, noting that $\pi e + 2\pi x = 0$ at $x = -e/2 \approx -1.359$. Consider the substitutions $\mathbb{Z}\cap(-\infty,-3] \mapsto (-\infty,-2]$, $\{-2,-1\} \mapsto \{-2,-1\}$ and $\mathbb{Z}\cap[0,\infty) \mapsto [-1,\infty)$.

$$\begin{aligned} \sum_{i\in\mathbb{Z}} e^{-(\pi e+2\pi i)^2/4\lambda} &\le \int_{-\infty}^{-2} e^{-(\pi e+2\pi x)^2/4\lambda}\,\mathrm{d}x + e^{-(\pi e-4\pi)^2/4\lambda} \\ &\quad + e^{-(\pi e-2\pi)^2/4\lambda} + \int_{-1}^{\infty} e^{-(\pi e+2\pi x)^2/4\lambda}\,\mathrm{d}x. \end{aligned}$$

Now use the substitution $t = \frac{\pi e+2\pi x}{2\sqrt{\lambda}}$ and suppose $\lambda < 1$, yielding

$$\begin{aligned} \sum_{i\in\mathbb{Z}} e^{-(\pi e+2\pi i)^2/4\lambda} &\le \frac{\sqrt{\lambda}}{\pi}\int_{-\infty}^{\frac{\pi e-4\pi}{2\sqrt{\lambda}}} e^{-t^2}\,\mathrm{d}t + e^{-(\pi e-4\pi)^2/4\lambda} \\ &\quad + e^{-(\pi e-2\pi)^2/4\lambda} + \frac{\sqrt{\lambda}}{\pi}\int_{\frac{\pi e-2\pi}{2\sqrt{\lambda}}}^{\infty} e^{-t^2}\,\mathrm{d}t \end{aligned}$$

$$\leq \left(1+\frac{1}{\pi}\frac{\lambda}{4\pi-\pi e}\right)e^{-(\pi e-4\pi)^2/4\lambda}+\left(1+\frac{1}{\pi}\frac{\lambda}{\pi e-2\pi}\right)e^{-(\pi e-2\pi)^2/4\lambda}$$

$$\leq \frac{13}{12}e^{-(\pi e-4\pi)^2/4\lambda}+\frac{8}{7}e^{-(\pi e-2\pi)^2/4\lambda}\leq \frac{9}{4}\exp\left(-\frac{5}{4\lambda}\right).$$

Therefore

$$\inf_{k\in\mathbb{N}}\inf_{b\in\mathcal{B}_k}\frac{\sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}}b_i b_j\, e^{-\lambda(i-j)^2}}{\sum_{i\in\mathbb{Z}}b_i^2}<\frac{4}{\sqrt{\lambda}}\exp\left(-\frac{5}{4\lambda}\right)<4e^{-1/\lambda}.$$

From this and since $\tilde{\Sigma}_{ij}=\frac{1}{2\sqrt{\pi}}e^{-\lambda(i-j)^2}$ with $\lambda=\frac{1}{4(N-1)^2h^2}\leq\frac{\delta^2}{h^2}$, for each h and some $\delta\leq h$,

$$\lambda_{\min}(\tilde{\Sigma})\leq 2e^{-h^2/\delta^2}.$$

Recall that

$$\left|\Sigma_{ij}-\tilde{\Sigma}_{ij}+hI(x_i)I(x_j)\right|\leq\frac{h}{2\pi a}\exp\left(-\frac{a^2}{h^2}\right).$$

Now for any positive semi-definite $N\times N$ matrices A and B and vector v we have $\lambda_{\min}(A-vv^{\mathsf{T}})\leq\lambda_{\min}(A)$ and $\lambda_{\min}(B)\leq\lambda_{\min}(A)+\|B-A\|_2\leq\lambda_{\min}(A)+N\|B-A\|_{\max}$. Hence with $I_i=I(x_i)$,

$$\lambda_{\min}(\Sigma)\leq\lambda_{\min}(\tilde{\Sigma}-hII^{\mathsf{T}})+\frac{Nh}{2\pi a}\exp\left(-\frac{a^2}{h^2}\right)\leq 2e^{-h^2/\delta^2}+\frac{h}{\pi a\delta}e^{-a^2/h^2}.$$

□

PROOF (Proposition 3.1). Let $\mathcal{F}_\delta$ be a δ-cover of $(\mathcal{F},d)$. Using a union bound, we can write

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}\left|S(f)-T(f)\right|\geq 2t+\eta\right)\leq\mathbb{P}\left(\sup_{f\in\mathcal{F}_\delta}\left|S(f)-T(f)\right|\geq\eta\right)$$

$$+\mathbb{P}\left(\sup_{d(f,f')\leq\delta}\left|S(f)-S(f')\right|\geq t\right)+\mathbb{P}\left(\sup_{d(f,f')\leq\delta}\left|T(f)-T(f')\right|\geq t\right).$$

Part 1: bounding the difference on $\mathcal{F}_\delta$

We apply Corollary 2.2 with $p=\infty$ to the martingale difference sequence $\mathcal{F}_\delta(X_i)=\big(f(X_i):f\in\mathcal{F}_\delta\big)$ which takes values in $\mathbb{R}^{|\mathcal{F}_\delta|}$. Square integrability can be assumed otherwise $\beta_\delta=\infty$. Note $\sum_{i=1}^n\mathcal{F}_\delta(X_i)=S(\mathcal{F}_\delta)$ and $\phi_\infty(\mathcal{F}_\delta)\leq\sqrt{2\log 2|\mathcal{F}_\delta|}$. Therefore there exists a conditionally Gaussian vector $T(\mathcal{F}_\delta)$ with the same covariance structure as $S(\mathcal{F}_\delta)$ conditional on $\mathcal{H}_0$ satisfying

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}_\delta}\left|S(f)-T(f)\right|\geq\eta\right)\leq\frac{24\beta_\delta^{\frac{1}{3}}(2\log 2|\mathcal{F}_\delta|)^{\frac{1}{3}}}{\eta}+17\left(\frac{\sqrt{2\log 2|\mathcal{F}_\delta|}\sqrt{\mathbb{E}\left[\|\Omega_\delta\|_2\right]}}{\eta}\right)^{\frac{2}{3}}.$$

Part 2: bounding the fluctuations in $S(f)$

Since $\left\|S(f)-S(f')\right\|_\psi\leq Ld(f,f')$, by Theorem 2.2.4 in van der Vaart and Wellner [20]

$$\left\|\sup_{d(f,f')\leq\delta}\left|S(f)-S(f')\right|\right\|_\psi\leq C_\psi L\left(\int_0^\delta\psi^{-1}(N_\varepsilon)\,\mathrm{d}\varepsilon+\delta\psi^{-1}(N_\delta^2)\right)=C_\psi LJ_\psi(\delta).$$

Then, by Markov's inequality and the definition of the Orlicz norm,

$$\mathbb{P}\left(\sup_{d(f,f')\leq\delta}\left|S(f)-S(f')\right|\geq t\right)\leq\psi\left(\frac{t}{C_\psi LJ_\psi(\delta)}\right)^{-1}.$$

Part 3: bounding the fluctuations in $T(f)$
By the Vorob'ev–Berkes–Philipp theorem [10], $T(\mathcal{F}_\delta)$ extends to a conditionally Gaussian process $T(f)$. Firstly since $\big\|\big|T(f)-T(f')\big|\big\|_2 \le L d(f,f')$ conditionally on $\mathcal{H}_0$, and $T(f)$ is a conditional Gaussian process, we have $\big\|T(f)-T(f')\big\|_{\psi_2} \le 2Ld(f,f')$ conditional on $\mathcal{H}_0$ by van der Vaart and Wellner [20, Chapter 2.2, Complement 1], where $\psi_2(x)=\exp(x^2)-1$. Thus again by Theorem 2.2.4 in van der Vaart and Wellner [20], again conditioning on $\mathcal{H}_0$,

$$\left\|\sup_{d(f,f')\le\delta}\big|T(f)-T(f')\big|\right\|_{\psi_2} \le C_1 L\int_0^\delta \sqrt{\log N_\varepsilon}\,\mathrm{d}\varepsilon = C_1 L J_2(\delta)$$

for some universal constant $C_1>0$, where we used $\psi_2^{-1}(x)=\sqrt{\log(1+x)}$ and monotonicity of covering numbers. Then by Markov's inequality and the definition of the Orlicz norm,

$$\begin{aligned}\mathbb{P}\left(\sup_{d(f,f')\le\delta}\big|T(f)-T(f')\big|\ge t\right) &\le \left(\exp\left(\frac{t^2}{C_1^2L^2J_2(\delta)^2}\right)-1\right)^{-1}\vee 1\\ &\le 2\exp\left(\frac{-t^2}{C_1^2L^2J_2(\delta)^2}\right).\end{aligned}$$

Part 4: conclusion
The result follows by scaling t and η and enlarging constants if necessary. □

SA.4. Applications to nonparametric regression

PROOF (Proposition 4.1). We proceed according to the decomposition given in Section 4.1. By stationarity and Lemma SA-2.1 in Cattaneo, Farrell and Feng [5], we have $\sup_w \|p(w)\|_1 \lesssim 1$ and also $\|H\|_1 \lesssim n/k$ and $\|H^{-1}\|_1 \lesssim k/n$.

Part 1: bounding $\beta_{\infty,2}$ and $\beta_{\infty,3}$
Set $X_i = p(W_i)\varepsilon_i$ so $S=\sum_{i=1}^n X_i$ and set $\sigma_i^2=\sigma^2(W_i)$ and $V_i=\operatorname{Var}[X_i \mid \mathcal{H}_{i-1}] = \sigma_i^2 p(W_i)p(W_i)^\mathsf{T}$. Recall from Corollary 2.2 that for $r\in\{2,3\}$,

$$\beta_{\infty,r}=\sum_{i=1}^n \mathbb{E}\left[\|X_i\|_2^r\|X_i\|_\infty + \|V_i^{1/2}Z_i\|_2^r\|V_i^{1/2}Z_i\|_\infty\right]$$

with $Z_i\sim\mathcal{N}(0,1)$ i.i.d. and independent of V_i. For the first term, we use $\sup_w\|p(w)\|_2\lesssim 1$ and bounded third moments of ε_i:

$$\mathbb{E}\left[\|X_i\|_2^r\|X_i\|_\infty\right] \le \mathbb{E}\left[|\varepsilon_i|^3\|p(W_i)\|_2^{r+1}\right]\lesssim 1.$$

For the second term, apply Lemma SA.3 conditionally on $\mathcal{H}_n$ with $\sup_w\|p(w)\|_2\lesssim 1$ to see

$$\begin{aligned}\mathbb{E}\left[\|V_i^{1/2}Z_i\|_2^r\|V_i^{1/2}Z_i\|_\infty\right] &\lesssim \sqrt{\log 2k}\,\mathbb{E}\left[\max_{1\le j\le k}(V_i)_{jj}^{1/2}\left(\sum_{j=1}^k (V_i)_{jj}\right)^{r/2}\right]\\ &\lesssim \sqrt{\log 2k}\,\mathbb{E}\left[\sigma_i^{r+1}\max_{1\le j\le k}p(W_i)_j\left(\sum_{j=1}^k p(W_i)_j^2\right)^{r/2}\right]\\ &\lesssim \sqrt{\log 2k}\,\mathbb{E}\left[\sigma_i^{r+1}\right]\lesssim\sqrt{\log 2k}.\end{aligned}$$

Putting these together yields $\beta_{\infty,2}\lesssim n\sqrt{\log 2k}$ and $\beta_{\infty,3}\lesssim n\sqrt{\log 2k}$.

Part 2: bounding Ω
Set $\Omega = \sum_{i=1}^n \big(V_i - \mathbb{E}[V_i]\big)$ as in Lemma SA.8 so

$$\Omega = \sum_{i=1}^n \left(\sigma_i^2 p(W_i)p(W_i)^\mathsf{T} - \mathbb{E}\left[\sigma_i^2 p(W_i)p(W_i)^\mathsf{T}\right]\right).$$

Observe that Ω_{jl} is the sum of a zero-mean strictly stationary α-mixing sequence and so $\mathbb{E}[\Omega_{jl}^2] \lesssim n$ by Lemma SA.6(i). Since the basis functions satisfy Assumption 3 in Cattaneo, Farrell and Feng [5], Ω has a bounded number of non-zero entries in each row, and so by Jensen’s inequality

$$\mathbb{E}\left[\|\Omega\|_2\right] \leq \mathbb{E}\left[\|\Omega\|_\mathrm{F}\right] \leq \left(\sum_{j=1}^k \sum_{l=1}^k \mathbb{E}\left[\Omega_{jl}^2\right]\right)^{1/2} \lesssim \sqrt{nk}.$$

Part 3: strong approximation
By Corollary 2.2 and the previous parts, with any sequence $R_n \to \infty$,

$$\begin{aligned}\|S-T\|_\infty &\lesssim_\mathbb{P} \beta_{\infty,2}^{1/3}(\log 2k)^{1/3} R_n + \sqrt{\log 2k}\sqrt{\mathbb{E}[\|\Omega\|_2]}R_n \\ &\lesssim_\mathbb{P} n^{1/3}\sqrt{\log 2k} R_n + (nk)^{1/4}\sqrt{\log 2k} R_n.\end{aligned}$$

If further $\mathbb{E}\left[\varepsilon_i^3 \mid \mathcal{H}_{i-1}\right] = 0$ then the third-order version of Corollary 2.2 applies since

$$\pi_3 = \sum_{i=1}^n \sum_{|\kappa|=3} \mathbb{E}\Big[\big|\mathbb{E}[X_i^\kappa \mid \mathcal{H}_{i-1}]\big|\Big] = \sum_{i=1}^n \sum_{|\kappa|=3} \mathbb{E}\Big[\big|p(W_i)^\kappa\, \mathbb{E}[\varepsilon_i^3 \mid \mathcal{H}_{i-1}]\big|\Big] = 0,$$

giving

$$\|S-T\|_\infty \lesssim_\mathbb{P} \beta_{\infty,3}^{1/4}(\log 2k)^{3/8} R_n + \sqrt{\log 2k}\sqrt{\mathbb{E}[\|\Omega\|_2]}R_n \lesssim_\mathbb{P} (nk)^{1/4}\sqrt{\log 2k} R_n.$$

By Hölder’s inequality and with $\|H^{-1}\|_1 \lesssim k/n$ we have

$$\sup_{w\in\mathcal{W}} \left|p(w)^\mathsf{T} H^{-1} S - p(w)^\mathsf{T} H^{-1} T\right| \leq \sup_{w\in\mathcal{W}} \|p(w)\|_1 \|H^{-1}\|_1 \|S-T\|_\infty \lesssim n^{-1}k\|S-T\|_\infty.$$

Part 4: convergence of $\hat{H}$
We have $\hat{H} - H = \sum_{i=1}^n \left(p(W_i)p(W_i)^\mathsf{T} - \mathbb{E}\left[p(W_i)p(W_i)^\mathsf{T}\right]\right)$. Observe that $(\hat{H}-H)_{jl}$ is the sum of a zero-mean strictly stationary α-mixing sequence and so $\mathbb{E}[(\hat{H}-H)_{jl}^2] \lesssim n$ by Lemma SA.6(i). Since the basis functions satisfy Assumption 3 in Cattaneo, Farrell and Feng [5], $\hat{H}-H$ has a bounded number of non-zero entries in each row and so by Jensen’s inequality

$$\mathbb{E}\left[\|\hat{H}-H\|_1\right] = \mathbb{E}\left[\max_{1\leq i\leq k}\sum_{j=1}^k \big|(\hat{H}-H)_{ij}\big|\right] \leq \mathbb{E}\left[\sum_{1\leq i\leq k}\left(\sum_{j=1}^k \big|(\hat{H}-H)_{ij}\big|\right)^2\right]^{\frac{1}{2}} \lesssim \sqrt{nk}.$$

Part 5: bounding the matrix term
Note $\|\hat{H}^{-1}\|_1 \leq \|H^{-1}\|_1 + \|\hat{H}^{-1}\|_1\|\hat{H}-H\|_1\|H^{-1}\|_1$ so by the previous part, we deduce

$$\|\hat{H}^{-1}\|_1 \leq \frac{\|H^{-1}\|_1}{1 - \|\hat{H}-H\|_1\|H^{-1}\|_1} \lesssim_\mathbb{P} \frac{k/n}{1-\sqrt{nk}\,k/n} \lesssim_\mathbb{P} \frac{k}{n}$$

as $k^3/n \to 0$. Also, note that by the martingale structure, since $p(W_i)$ is bounded and supported on a region with volume at most of the order $1/k$, and as W_i has a Lebesgue density,

$$\mathrm{Var}[T_j] = \mathrm{Var}[S_j] = \mathrm{Var}\left[\sum_{i=1}^n \varepsilon_i p(W_i)_j\right] = \sum_{i=1}^n \mathbb{E}\left[\sigma_i^2 p(W_i)_j^2\right] \lesssim \frac{n}{k}.$$

So by the Gaussian maximal inequality in Lemma SA.4, $\|T\|_\infty \lesssim_{\mathbb{P}} \sqrt{\frac{n \log 2k}{k}}$. Since $k^3/n \to 0$,

$$\begin{aligned}
\sup_{w \in \mathcal{W}} \left| p(w)^\mathsf{T} (\hat{H}^{-1} - H^{-1}) S \right| &\leq \sup_{w \in \mathcal{W}} \|p(w)^\mathsf{T}\|_1 \|\hat{H}^{-1}\|_1 \|\hat{H} - H\|_1 \|H^{-1}\|_1 \|S - T\|_\infty \\
&\quad + \sup_{w \in \mathcal{W}} \|p(w)^\mathsf{T}\|_1 \|\hat{H}^{-1}\|_1 \|\hat{H} - H\|_1 \|H^{-1}\|_1 \|T\|_\infty \\
&\lesssim_{\mathbb{P}} \frac{k}{n} \sqrt{nk} \frac{k}{n} \left(n^{1/3} \sqrt{\log 2k} + (nk)^{1/4} \sqrt{\log 2k} \right) \\
&\quad + \frac{k}{n} \sqrt{nk} \frac{k}{n} \sqrt{\frac{n \log 2k}{k}} \lesssim_{\mathbb{P}} \frac{k^2}{n} \sqrt{\log 2k}.
\end{aligned}$$

Part 6: conclusion of the main result

By the previous parts, with $G(w) = p(w)^\mathsf{T} H^{-1} T$,

$$\begin{aligned}
&\sup_{w \in \mathcal{W}} \left| \hat{\mu}(w) - \mu(w) - p(w)^\mathsf{T} H^{-1} T \right| \\
&\quad = \sup_{w \in \mathcal{W}} \left| p(w)^\mathsf{T} H^{-1} (S - T) + p(w)^\mathsf{T} (\hat{H}^{-1} - H^{-1}) S + \mathrm{Bias}(w) \right| \\
&\quad \lesssim_{\mathbb{P}} \frac{k}{n} \|S - T\|_\infty + \frac{k^2}{n} \sqrt{\log 2k} + \sup_{w \in \mathcal{W}} |\mathrm{Bias}(w)| \\
&\quad \lesssim_{\mathbb{P}} \frac{k}{n} \left(n^{1/3} \sqrt{\log 2k} + (nk)^{1/4} \sqrt{\log 2k} \right) R_n + \frac{k^2}{n} \sqrt{\log 2k} + \sup_{w \in \mathcal{W}} |\mathrm{Bias}(w)| \\
&\quad \lesssim_{\mathbb{P}} n^{-2/3} k \sqrt{\log 2k} R_n + n^{-3/4} k^{5/4} \sqrt{\log 2k} R_n + \frac{k^2}{n} \sqrt{\log 2k} + \sup_{w \in \mathcal{W}} |\mathrm{Bias}(w)| \\
&\quad \lesssim_{\mathbb{P}} n^{-2/3} k \sqrt{\log 2k} R_n + \sup_{w \in \mathcal{W}} |\mathrm{Bias}(w)|
\end{aligned}$$

since $k^3/n \to 0$. If further $\mathbb{E}\left[\varepsilon_i^3 \mid \mathcal{H}_{i-1}\right] = 0$ then

$$\begin{aligned}
\sup_{w \in \mathcal{W}} \left| \hat{\mu}(w) - \mu(w) - p(w)^\mathsf{T} H^{-1} T \right| &\lesssim_{\mathbb{P}} \frac{k}{n} \|S - T\|_\infty + \frac{k^2}{n} \sqrt{\log 2k} + \sup_{w \in \mathcal{W}} |\mathrm{Bias}(w)| \\
&\lesssim_{\mathbb{P}} n^{-3/4} k^{5/4} \sqrt{\log 2k} R_n + \sup_{w \in \mathcal{W}} |\mathrm{Bias}(w)|.
\end{aligned}$$

Finally, we verify the variance bounds for the Gaussian process. Since $\sigma^2(w)$ is bounded,

$$\begin{aligned}
\mathrm{Var}[G(w)] &= p(w)^\mathsf{T} H^{-1} \,\mathrm{Var}\left[\sum_{i=1}^n p(W_i) \varepsilon_i\right] H^{-1} p(w) \\
&= p(w)^\mathsf{T} H^{-1} \mathbb{E}\left[\sum_{i=1}^n p(W_i) p(W_i)^\mathsf{T} \sigma^2(W_i)\right] H^{-1} p(w) \\
&\lesssim \|p(w)\|_2^2 \|H^{-1}\|_2^2 \|H\|_2 \lesssim k/n.
\end{aligned}$$

Similarly, since $\sigma^2(w)$ is bounded away from zero,

$$\mathrm{Var}[G(w)] \gtrsim \|p(w)\|_2^2 \|H^{-1}\|_2^2 \|H^{-1}\|_2^{-1} \gtrsim k/n.$$

Part 7: bounding the bias
We delegate the task of deriving bounds on the bias to Cattaneo, Farrell and Feng [5], who provide a high-level assumption on the approximation error in Assumption 4 and then use it to derive bias bounds in Section 3 of the form $\sup_{w \in \mathcal{W}} |\mathrm{Bias}(w)| \lesssim_{\mathbb{P}} k^{-\gamma}$. This assumption is verified for B-splines, wavelets and piecewise polynomials in their supplemental appendix. □

PROOF (Proposition 4.2). **Part 1: infeasible supremum approximation**
Provided that the bias is negligible, for all $s > 0$ we have

$$\begin{aligned}
&\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\sup_{w \in \mathcal{W}} \left| \frac{\hat{\mu}(w) - \mu(w)}{\sqrt{\rho(w,w)}} \right| \leq t \right) - \mathbb{P}\left(\sup_{w \in \mathcal{W}} \left| \frac{G(w)}{\sqrt{\rho(w,w)}} \right| \leq t \right) \right| \\
&\quad \leq \sup_{t \in \mathbb{R}} \mathbb{P}\left(t \leq \sup_{w \in \mathcal{W}} \left| \frac{G(w)}{\sqrt{\rho(w,w)}} \right| \leq t + s \right) + \mathbb{P}\left(\sup_{w \in \mathcal{W}} \left| \frac{\hat{\mu}(w) - \mu(w) - G(w)}{\sqrt{\rho(w,w)}} \right| > s \right).
\end{aligned}$$

By the Gaussian anti-concentration result given as Corollary 2.1 in Chernozhukov, Chetverikov and Kato [8] applied to a discretization of $\mathcal{W}$, the first term is at most $s\sqrt{\log n}$ up to a constant factor, and the second term converges to zero whenever $\frac{1}{s}\left(\frac{k^3 (\log k)^3}{n}\right)^{1/6} \to 0$. Thus a suitable value of s exists whenever $\frac{k^3 (\log n)^6}{n} \to 0$.

Part 2: feasible supremum approximation
By Chernozhukov, Chetverikov and Kato [7, Lemma 3.1], with $\rho(w,w') = \mathbb{E}[\hat{\rho}(w,w')]$,

$$\begin{aligned}
&\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\sup_{w \in \mathcal{W}} \left| \frac{\hat{G}(w)}{\sqrt{\hat{\rho}(w,w)}} \right| \leq t \,\middle|\, \mathbf{W}, \mathbf{Y} \right) - \mathbb{P}\left(\left| \frac{G(w)}{\sqrt{\rho(w,w)}} \right| \leq t \right) \right| \\
&\quad \lesssim_{\mathbb{P}} \sup_{w,w' \in \mathcal{W}} \left| \frac{\hat{\rho}(w,w')}{\sqrt{\hat{\rho}(w,w)\hat{\rho}(w',w')}} - \frac{\rho(w,w')}{\sqrt{\rho(w,w)\rho(w',w')}} \right|^{1/3} (\log n)^{2/3} \\
&\quad \lesssim_{\mathbb{P}} \left(\frac{n}{k}\right)^{1/3} \sup_{w,w' \in \mathcal{W}} |\hat{\rho}(w,w') - \rho(w,w')|^{1/3} (\log n)^{2/3} \\
&\quad \lesssim_{\mathbb{P}} \left(\frac{n (\log n)^2}{k}\right)^{1/3} \sup_{w,w' \in \mathcal{W}} \left| p(w)^\mathsf{T} \hat{H}^{-1} \left(\hat{\mathrm{Var}}[S] - \mathrm{Var}[S] \right) \hat{H}^{-1} p(w') \right|^{1/3} \\
&\quad \lesssim_{\mathbb{P}} \left(\frac{k (\log n)^2}{n}\right)^{1/3} \left\| \hat{\mathrm{Var}}[S] - \mathrm{Var}[S] \right\|_2^{1/3},
\end{aligned}$$

and vanishes in probability when $\frac{k(\log n)^2}{n} \left\| \hat{\mathrm{Var}}[S] - \mathrm{Var}[S] \right\|_2 \to_{\mathbb{P}} 0$. For the plug-in estimator,

$$\begin{aligned}
&\left\| \hat{\mathrm{Var}}[S] - \mathrm{Var}[S] \right\|_2 = \left\| \sum_{i=1}^n p(W_i) p(W_i^\mathsf{T}) \hat{\sigma}^2(W_i) - n \mathbb{E}\left[p(W_i) p(W_i^\mathsf{T}) \sigma^2(W_i) \right] \right\|_2 \\
&\quad \lesssim_{\mathbb{P}} \sup_{w \in \mathcal{W}} |\hat{\sigma}^2(w) - \sigma^2(w)| \left\| \hat{H} \right\|_2 \\
&\qquad + \left\| \sum_{i=1}^n p(W_i) p(W_i^\mathsf{T}) \sigma^2(W_i) - n \mathbb{E}\left[p(W_i) p(W_i^\mathsf{T}) \sigma^2(W_i) \right] \right\|_2
\end{aligned}$$

$$\lesssim_{\mathbb{P}} \frac{n}{k} \sup_{w \in \mathcal{W}} |\hat{\sigma}^2(w) - \sigma^2(w)| + \sqrt{nk},$$

where the second term is bounded by the same argument used to bound $\|\hat{H} - H\|_1$. Thus, the feasible approximation is valid whenever $(\log n)^2 \sup_{w \in \mathcal{W}} |\hat{\sigma}^2(w) - \sigma^2(w)| \to_{\mathbb{P}} 0$ and $\frac{k^3 (\log n)^4}{n} \to 0$. The validity of the uniform confidence band follows immediately. □

PROOF (Proposition 4.3). We apply Proposition 3.1 with the metric $d(f_w, f_{w'}) = \|w - w'\|_2$ and the function class

$$\mathcal{F} = \left\{ (W_i, \varepsilon_i) \mapsto e_1^\mathsf{T} H(w)^{-1} K_h(W_i - w) p_h(W_i - w) \varepsilon_i : \ w \in \mathcal{W} \right\},$$

with ψ chosen as a suitable Bernstein–Orlicz function.

Part 1: bounding $H(w)^{-1}$
Recall that $H(w) = \sum_{i=1}^n \mathbb{E}[K_h(W_i - w) p_h(W_i - w) p_h(W_i - w)^\mathsf{T}]$ and let $a(w) \in \mathbb{R}^k$ with $\|a(w)\|_2 = 1$. Since the density of W_i is bounded away from zero on $\mathcal{W}$,

$$\begin{aligned} a(w)^\mathsf{T} H(w) a(w) &= n \mathbb{E}\left[\left(a(w)^\mathsf{T} p_h(W_i - w) \right)^2 K_h(W_i - w) \right] \\ &\gtrsim n \int_{\mathcal{W}} \left(a(w)^\mathsf{T} p_h(u - w) \right)^2 K_h(u - w) \, \mathrm{d}u \\ &\gtrsim n \int_{\frac{\mathcal{W} - w}{h}} \left(a(w)^\mathsf{T} p(u) \right)^2 K(u) \, \mathrm{d}u. \end{aligned}$$

This is continuous in $a(w)$ on the compact set $\|a(w)\|_2 = 1$ and $p(u)$ forms a polynomial basis so $a(w)^\mathsf{T} p(u)$ has finitely many zeroes. Since $K(u)$ is compactly supported and $h \to 0$, the above integral is eventually strictly positive for all $x \in \mathcal{W}$, and hence is bounded below uniformly in $w \in \mathcal{W}$ by a positive constant. Therefore $\sup_{w \in \mathcal{W}} \|H(w)^{-1}\|_2 \lesssim 1/n$.

Part 2: bounding β_δ
Let $\mathcal{F}_\delta$ be a δ-cover of $(\mathcal{F}, d)$ with cardinality $|\mathcal{F}_\delta| \asymp \delta^{-m}$ and let $\mathcal{F}_\delta(W_i, \varepsilon_i) = \big(f(W_i, \varepsilon_i) : f \in \mathcal{F}_\delta \big)$. Define the truncated errors $\tilde{\varepsilon}_i = \varepsilon_i \mathbb{I}\{-a \log n \leq \varepsilon_i \leq b \log n\}$ and note that $\mathbb{E}\big[e^{|\varepsilon_i|/C_\varepsilon}\big] < \infty$ implies that $\mathbb{P}(\exists i : \tilde{\varepsilon}_i \neq \varepsilon_i) \lesssim n^{1 - (a \vee b)/C_\varepsilon}$. Hence, by choosing a and b large enough, with high probability, we can replace all ε_i by $\tilde{\varepsilon}_i$. Further, it is always possible to increase either a or b along with some randomization to ensure that $\mathbb{E}[\tilde{\varepsilon}_i] = 0$. Since K is bounded and compactly supported, W_i has a bounded density and $|\tilde{\varepsilon}_i| \lesssim \log n$,

$$\begin{aligned} \big\| f(W_i, \tilde{\varepsilon}_i) \big\|_2 &= \mathbb{E}\left[\left| e_1^\mathsf{T} H(w)^{-1} K_h(W_i - w) p_h(W_i - w) \tilde{\varepsilon}_i \right|^2 \right]^{1/2} \\ &\leq \mathbb{E}\left[\|H(w)^{-1}\|_2^2 K_h(W_i - w)^2 \|p_h(W_i - w)\|_2^2 \sigma^2(W_i) \right]^{1/2} \\ &\lesssim n^{-1} \mathbb{E}\left[K_h(W_i - w)^2 \right]^{1/2} \lesssim n^{-1} h^{-m/2}, \\ \big\| f(W_i, \tilde{\varepsilon}_i) \big\|_\infty &\leq \big\| \|H(w)^{-1}\|_2 K_h(W_i - w) \|p_h(W_i - w)\|_2 |\tilde{\varepsilon}_i| \big\|_\infty \\ &\lesssim n^{-1} \big\| K_h(W_i - w) \big\|_\infty \log n \lesssim n^{-1} h^{-m} \log n. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E}\left[\|\mathcal{F}_\delta(W_i, \tilde{\varepsilon}_i)\|_2^2 \|\mathcal{F}_\delta(W_i, \tilde{\varepsilon}_i)\|_\infty \right] &\leq \sum_{f \in \mathcal{F}_\delta} \big\| f(W_i, \tilde{\varepsilon}_i) \big\|_2^2 \max_{f \in \mathcal{F}_\delta} \big\| f(W_i, \tilde{\varepsilon}_i) \big\|_\infty \\ &\lesssim n^{-3} \delta^{-m} h^{-2m} \log n. \end{aligned}$$

Let $V_i(\mathcal{F}_\delta) = \mathbb{E}\big[\mathcal{F}_\delta(W_i, \tilde{\varepsilon}_i)\mathcal{F}_\delta(W_i, \tilde{\varepsilon}_i)^\mathsf{T} \mid \mathcal{H}_{i-1}\big]$ and $Z_i \sim \mathcal{N}(0, I_d)$ be i.i.d. and independent of $\mathcal{H}_n$. Note that $V_i(f, f) = \mathbb{E}[f(W_i, \tilde{\varepsilon}_i)^2 \mid W_i] \lesssim n^{-2}h^{-2m}$ and $\mathbb{E}[V_i(f, f)] = \mathbb{E}[f(W_i, \tilde{\varepsilon}_i)^2] \lesssim n^{-2}h^{-m}$. Thus by Lemma SA.3,

$$
\begin{aligned}
\mathbb{E}\left[\left\|V_i(\mathcal{F}_\delta)^{1/2}Z_i\right\|_2^2\left\|V_i(\mathcal{F}_\delta)^{1/2}Z_i\right\|_\infty\right] &= \mathbb{E}\left[\mathbb{E}\left[\left\|V_i(\mathcal{F}_\delta)^{1/2}Z_i\right\|_2^2\left\|V_i(\mathcal{F}_\delta)^{1/2}Z_i\right\|_\infty \mid \mathcal{H}_n\right]\right] \\
&\le 4\sqrt{\log 2|\mathcal{F}_\delta|}\,\mathbb{E}\left[\max_{f\in\mathcal{F}_\delta}\sqrt{V_i(f,f)}\sum_{f\in\mathcal{F}_\delta}V_i(f,f)\right] \\
&\lesssim n^{-3}h^{-2m}\delta^{-m}\sqrt{\log(1/\delta)}.
\end{aligned}
$$

Thus since $\log(1/\delta) \asymp \log(1/h) \asymp \log n$,

$$
\begin{aligned}
\beta_\delta &= \sum_{i=1}^n \mathbb{E}\left[\|\mathcal{F}_\delta(W_i, \tilde{\varepsilon}_i)\|_2^2\|\mathcal{F}_\delta(W_i, \tilde{\varepsilon}_i)\|_\infty + \left\|V_i(\mathcal{F}_\delta)^{1/2}Z_i\right\|_2^2\left\|V_i(\mathcal{F}_\delta)^{1/2}Z_i\right\|_\infty\right] \\
&\lesssim \frac{\log n}{n^2h^{2m}\delta^m}.
\end{aligned}
$$

Part 3: bounding Ω_δ

Let $C_K > 0$ be the radius of a ℓ^2-ball containing the support of K and note that

$$
\begin{aligned}
|V_i(f, f')| &= \Big|\mathbb{E}\Big[e_1^\mathsf{T}H(w)^{-1}p_h(W_i - w)e_1^\mathsf{T}H(w')^{-1}p_h(W_i - w') \\
&\qquad \times K_h(W_i - w)K_h(W_i - w')\tilde{\varepsilon}_i^2 \,\Big|\, \mathcal{H}_{i-1}\Big]\Big| \\
&\lesssim n^{-2}K_h(W_i - w)K_h(W_i - w') \\
&\lesssim n^{-2}h^{-m}K_h(W_i - w)\mathbb{I}\{\|w - w'\|_2 \le 2C_Kh\}.
\end{aligned}
$$

Since W_i are α-mixing with $\alpha(j) < e^{-2j/C_\alpha}$, Lemma SA.6(ii) with $r = 3$ gives

$$
\begin{aligned}
&\mathrm{Var}\left[\sum_{i=1}^n V_i(f, f')\right] \\
&\lesssim \sum_{i=1}^n \mathbb{E}\left[|V_i(f, f')|^3\right]^{2/3} \lesssim n^{-3}h^{-2m}\mathbb{E}\left[K_h(W_i - w)^3\right]^{2/3}\mathbb{I}\{\|w - w'\|_2 \le 2C_Kh\} \\
&\lesssim n^{-3}h^{-2m}(h^{-2m})^{2/3}\mathbb{I}\{\|w - w'\|_2 \le 2C_Kh\} \\
&\lesssim n^{-3}h^{-10m/3}\mathbb{I}\{\|w - w'\|_2 \le 2C_Kh\}.
\end{aligned}
$$

Therefore, by Jensen's inequality,

$$
\begin{aligned}
\mathbb{E}\big[\|\Omega_\delta\|_2\big] \le \mathbb{E}\big[\|\Omega_\delta\|_\mathrm{F}\big] &\le \mathbb{E}\left[\sum_{f,f'\in\mathcal{F}_\delta}(\Omega_\delta)_{f,f'}^2\right]^{1/2} \le \left(\sum_{f,f'\in\mathcal{F}_\delta}\mathrm{Var}\left[\sum_{i=1}^n V_i(f, f')\right]\right)^{1/2} \\
&\lesssim n^{-3/2}h^{-5m/3}\left(\sum_{f,f'\in\mathcal{F}_\delta}\mathbb{I}\{\|w - w'\|_2 \le 2C_Kh\}\right)^{1/2} \\
&\lesssim n^{-3/2}h^{-5m/3}\big(h^m\delta^{-2m}\big)^{1/2} \lesssim n^{-3/2}h^{-7m/6}\delta^{-m}.
\end{aligned}
$$

Note that we could have used $\|\cdot\|_1$ rather than $\|\cdot\|_\mathrm{F}$, but this term is negligible either way.

Part 4: regularity of the stochastic processes
For each $f, f' \in \mathcal{F}$, define the zero-mean and α-mixing random variables

$$u_i(f,f') = e_1^\mathsf{T}\big(H(w)^{-1}K_h(W_i-w)p_h(W_i-w) - H(w')^{-1}K_h(W_i-w')p_h(W_i-w')\big)\tilde{\varepsilon}_i.$$

To bound this we use that for all $1 \le j \le k$, by the Lipschitz property of the kernel and monomials,

$$\begin{aligned}
&|K_h(W_i-w) - K_h(W_i-w')| \\
&\quad \lesssim h^{-m-1}\|w-w'\|_2\big(\mathbb{I}\{\|W_i-w\| \le C_K h\} + \mathbb{I}\{\|W_i-w'\| \le C_K h\}\big), \\
&|p_h(W_i-w)_j - p_h(W_i-w')_j| \lesssim h^{-1}\|w-w'\|_2,
\end{aligned}$$

to deduce that for any $1 \le j, l \le k$,

$$\begin{aligned}
\big|H(w)_{jl} - H(w')_{jl}\big| &= \big|n\mathbb{E}\big[K_h(W_i-w)p_h(W_i-w)_j p_h(W_i-w)_l \\
&\qquad - K_h(W_i-w')p_h(W_i-w')_j p_h(W_i-w')_l\big]\big| \\
&\le n\mathbb{E}\left[|K_h(W_i-w) - K_h(W_i-w')|\,|p_h(W_i-w)_j p_h(W_i-w)_l|\right] \\
&\quad + n\mathbb{E}\left[|p_h(W_i-w)_j - p_h(W_i-w')_j|\,|K_h(W_i-w')p_h(W_i-w)_l|\right] \\
&\quad + n\mathbb{E}\left[|p_h(W_i-w)_l - p_h(W_i-w')_l|\,|K_h(W_i-w')p_h(W_i-w')_j|\right] \\
&\lesssim nh^{-1}\|w-w'\|_2.
\end{aligned}$$

Therefore as the dimension of the matrix $H(w)$ is fixed,

$$\big\|H(w)^{-1} - H(w')^{-1}\big\|_2 \le \big\|H(w)^{-1}\big\|_2\big\|H(w')^{-1}\big\|_2\big\|H(w) - H(w')\big\|_2 \lesssim \frac{\|w-w'\|_2}{nh}.$$

Hence

$$\begin{aligned}
\big|u_i(f,f')\big| &\le \big\|H(w)^{-1}K_h(W_i-w)p_h(W_i-w) - H(w')^{-1}K_h(W_i-w')p_h(W_i-w')\tilde{\varepsilon}_i\big\|_2 \\
&\le \big\|H(w)^{-1} - H(w')^{-1}\big\|_2\big\|K_h(W_i-w)p_h(W_i-w)\tilde{\varepsilon}_i\big\|_2 \\
&\quad + \big|K_h(W_i-w) - K_h(W_i-w')\big|\big\|H(w')^{-1}p_h(W_i-w)\tilde{\varepsilon}_i\big\|_2 \\
&\quad + \big\|p_h(W_i-w) - p_h(W_i-w')\big\|_2\big\|H(w')^{-1}K_h(W_i-w')\tilde{\varepsilon}_i\big\|_2 \\
&\lesssim \frac{\|w-w'\|_2}{nh}\big|K_h(W_i-w)\tilde{\varepsilon}_i\big| + \frac{1}{n}\big|K_h(W_i-w) - K_h(W_i-w')\big|\,|\tilde{\varepsilon}_i| \\
&\lesssim \frac{\|w-w'\|_2 \log n}{nh^{m+1}},
\end{aligned}$$

and from the penultimate line, we also deduce that

$$\begin{aligned}
\mathrm{Var}[u_i(f,f')] &\lesssim \frac{\|w-w'\|_2^2}{n^2h^2}\mathbb{E}\left[K_h(W_i-w)^2\sigma^2(X_i)\right] \\
&\qquad + \frac{1}{n^2}\mathbb{E}\left[\big(K_h(W_i-w) - K_h(W_i-w')\big)^2\sigma^2(X_i)\right] \lesssim \frac{\|w-w'\|_2^2}{n^2h^{m+2}}.
\end{aligned}$$

Further, $\mathbb{E}[u_i(f,f')u_j(f,f')] = 0$ for $i \ne j$ so by Lemma SA.7(ii), for a constant $C_1 > 0$,

$$\mathbb{P}\left(\Big|\sum_{i=1}^n u_i(f,f')\Big| \ge \frac{C_1\|w-w'\|_2}{\sqrt{n}h^{m/2+1}}\left(\sqrt{t} + \sqrt{\frac{(\log n)^2}{nh^m}}\sqrt{t} + \sqrt{\frac{(\log n)^6}{nh^m}}t\right)\right) \le C_1 e^{-t}.$$

Therefore, adjusting the constant if necessary and since $nh^m \gtrsim (\log n)^7$,

$$\mathbb{P}\left(\Big| \sum_{i=1}^n u_i(f, f') \Big| \geq \frac{C_1 \|w - w'\|_2}{\sqrt{n} h^{m/2+1}} \left(\sqrt{t} + \frac{t}{\sqrt{\log n}} \right) \right) \leq C_1 e^{-t}.$$

By Lemma 2 in van de Geer and Lederer [19] with $\psi(x) = \exp\Big(\big(\sqrt{1 + 2x/\sqrt{\log n}} - 1\big)^2 \log n \Big) - 1$,

$$\Big\| \sum_{i=1}^n u_i(f, f') \Big\|_\psi \lesssim \frac{\|w - w'\|_2}{\sqrt{n} h^{m/2+1}}$$

so we take $L = \frac{1}{\sqrt{n} h^{m/2+1}}$. Noting $\psi^{-1}(t) = \sqrt{\log(1+t)} + \frac{\log(1+t)}{2\sqrt{\log n}}$ and $N_\delta \lesssim \delta^{-m}$,

$$J_\psi(\delta) = \int_0^\delta \psi^{-1}\big(N_\varepsilon\big) \, \mathrm{d}\varepsilon + \delta \psi^{-1}\big(N_\delta\big) \lesssim \frac{\delta \log(1/\delta)}{\sqrt{\log n}} + \delta \sqrt{\log(1/\delta)} \lesssim \delta \sqrt{\log n},$$

$$J_2(\delta) = \int_0^\delta \sqrt{\log N_\varepsilon} \, \mathrm{d}\varepsilon \lesssim \delta \sqrt{\log(1/\delta)} \lesssim \delta \sqrt{\log n}.$$

Part 5: strong approximation

Recalling that $\tilde{\varepsilon}_i = \varepsilon_i$ for all i with high probability, by Proposition 3.1, for all $t, \eta > 0$ there exists a zero-mean Gaussian process $T(w)$ satisfying

$$\mathbb{E}\left[\left(\sum_{i=1}^n f_w(W_i, \varepsilon_i) \right) \left(\sum_{i=1}^n f_{w'}(W_i, \varepsilon_i) \right) \right] = \mathbb{E}\big[T(w) T(w') \big]$$

for all $w, w' \in \mathcal{W}$ and

$$\begin{aligned}
&\mathbb{P}\left(\sup_{w \in \mathcal{W}} \left| \sum_{i=1}^n f_w(W_i, \varepsilon_i) - T(w) \right| \geq C_\psi (t + \eta) \right) \\
&\leq C_\psi \inf_{\delta > 0} \inf_{\mathcal{F}_\delta} \left\{ \frac{\beta_\delta^{1/3} (\log 2|\mathcal{F}_\delta|)^{1/3}}{\eta} + \left(\frac{\sqrt{\log 2|\mathcal{F}_\delta|} \sqrt{\mathbb{E}\left[\|\Omega_\delta\|_2\right]}}{\eta} \right)^{2/3} \right. \\
&\qquad \left. + \psi\left(\frac{t}{L J_\psi(\delta)} \right)^{-1} + \exp\left(\frac{-t^2}{L^2 J_2(\delta)^2} \right) \right\} \\
&\leq C_\psi \left\{ \frac{\left(\frac{\log n}{n^2 h^{2m} \delta^m} \right)^{1/3} (\log n)^{1/3}}{\eta} + \left(\frac{\sqrt{\log n} \sqrt{n^{-3/2} h^{-7m/6} \delta^{-m}}}{\eta} \right)^{2/3} \right. \\
&\qquad \left. + \psi\left(\frac{t}{\frac{1}{\sqrt{n} h^{m/2+1}} J_\psi(\delta)} \right)^{-1} + \exp\left(\frac{-t^2}{\left(\frac{1}{\sqrt{n} h^{m/2+1}} \right)^2 J_2(\delta)^2} \right) \right\} \\
&\leq C_\psi \left\{ \frac{(\log n)^{2/3}}{n^{2/3} h^{2m/3} \delta^{m/3} \eta} + \left(\frac{n^{-3/4} h^{-7m/12} \delta^{-m/2} \sqrt{\log n}}{\eta} \right)^{2/3} \right. \\
&\qquad \left. + \psi\left(\frac{t \sqrt{n} h^{m/2+1}}{\delta \sqrt{\log n}} \right)^{-1} + \exp\left(\frac{-t^2 n h^{m+2}}{\delta^2 \log n} \right) \right\}.
\end{aligned}$$

Noting $\psi(x) \geq e^{x^2/4}$ for $x \leq 4\sqrt{\log n}$, any $R_n \to \infty$ gives the probability bound

$$\sup_{w \in \mathcal{W}} \left| \sum_{i=1}^n f_w(W_i, \varepsilon_i) - T(w) \right| \lesssim_{\mathbb{P}} \frac{(\log n)^{2/3}}{n^{2/3} h^{2m/3} \delta^{m/3}} R_n + \frac{\sqrt{\log n}}{n^{3/4} h^{7m/12} \delta^{m/2}} R_n + \frac{\delta \sqrt{\log n}}{\sqrt{n} h^{m/2+1}}.$$

Optimizing over δ gives $\delta \asymp \left(\frac{\log n}{n h^{m-6}} \right)^{\frac{1}{2m+6}} = h \left(\frac{\log n}{n h^{3m}} \right)^{\frac{1}{2m+6}}$ and so

$$\sup_{w \in \mathcal{W}} \left| \sum_{i=1}^n f_w(W_i, \varepsilon_i) - T(w) \right| \lesssim_{\mathbb{P}} \left(\frac{(\log n)^{m+4}}{n^{m+4} h^{m(m+6)}} \right)^{\frac{1}{2m+6}} R_n.$$

Part 6: convergence of $\hat{H}(w)$

For $1 \leq j, l \leq k$ define the zero-mean random variables

$$\begin{aligned} u_{ijl}(w) &= K_h(W_i - w) p_h(W_i - w)_j p_h(W_i - w)_l \\ &\quad - \mathbb{E}\big[K_h(W_i - w) p_h(W_i - w)_j p_h(W_i - w)_l \big] \end{aligned}$$

and note that $|u_{ijl}(w)| \lesssim h^{-m}$. By Lemma SA.7(i) for a constant $C_2 > 0$ and all $t > 0$,

$$\mathbb{P}\left(\left| \sum_{i=1}^n u_{ijl}(w) \right| > C_2 h^{-m} \big(\sqrt{nt} + (\log n)(\log \log n) t \big) \right) \leq C_2 e^{-t}.$$

Further, note that by Lipschitz properties,

$$\left| \sum_{i=1}^n u_{ijl}(w) - \sum_{i=1}^n u_{ijl}(w') \right| \lesssim h^{-m-1} \|w - w'\|_2$$

so there is a δ-cover of $(\mathcal{W}, \|\cdot\|_2)$ with size at most $n^a \delta^{-a}$ for some $a > 0$. Adjusting C_2,

$$\mathbb{P}\left(\sup_{w \in \mathcal{W}} \left| \sum_{i=1}^n u_{ijl}(w) \right| > C_2 h^{-m} \big(\sqrt{nt} + (\log n)(\log \log n) t \big) + C_2 h^{-m-1} \delta \right) \leq C_2 n^a \delta^{-a} e^{-t}$$

and hence

$$\sup_{w \in \mathcal{W}} \left| \sum_{i=1}^n u_{ijl}(w) \right| \lesssim_{\mathbb{P}} h^{-m} \sqrt{n \log n} + h^{-m} (\log n)^3 \lesssim_{\mathbb{P}} \sqrt{\frac{n \log n}{h^{2m}}}.$$

Therefore

$$\sup_{w \in \mathcal{W}} \|\hat{H}(w) - H(w)\|_2 \lesssim_{\mathbb{P}} \sqrt{\frac{n \log n}{h^{2m}}}.$$

Part 7: bounding the matrix term

Firstly note that, since $\sqrt{\frac{\log n}{n h^{2m}}} \to 0$, we have that uniformly in $w \in \mathcal{W}$

$$\|\hat{H}(w)^{-1}\|_2 \leq \frac{\|H(w)^{-1}\|_2}{1 - \|\hat{H}(w) - H(w)\|_2 \|H(w)^{-1}\|_2} \lesssim_{\mathbb{P}} \frac{1/n}{1 - \sqrt{\frac{n \log n}{h^{2m}}} \frac{1}{n}} \lesssim_{\mathbb{P}} \frac{1}{n}.$$

Therefore

$$\begin{aligned} \sup_{w \in \mathcal{W}} \left| e_1^\mathsf{T} \big(\hat{H}(w)^{-1} - H(w)^{-1} \big) S(w) \right| &\leq \sup_{w \in \mathcal{W}} \left\| \hat{H}(w)^{-1} - H(w)^{-1} \right\|_2 \|S(w)\|_2 \\ &\leq \sup_{w \in \mathcal{W}} \left\| \hat{H}(w)^{-1} \right\|_2 \left\| H(w)^{-1} \right\|_2 \left\| \hat{H}(w) - H(w) \right\|_2 \|S(w)\|_2 \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n^3 h^{2m}}} \sup_{w \in \mathcal{W}} \|S(w)\|_2. \end{aligned}$$

Now for $1 \leq j \leq k$ write $u_{ij}(w) = K_h(W_i - w) p_h(W_i - w)_j \tilde{\varepsilon}_i$ so that $S(w)_j = \sum_{i=1}^n u_{ij}(w)$ with high probability. Note that $u_{ij}(w)$ are zero-mean with $\mathrm{Cov}[u_{ij}(w), u_{i'j}(w)] = 0$ for $i \neq i'$. Also $|u_{ij}(w)| \lesssim h^{-m} \log n$ and $\mathrm{Var}[u_{ij}(w)] \lesssim h^{-m}$. Thus by Lemma SA.7(ii) for a constant $C_3 > 0$,

$$\mathbb{P}\left(\Big| \sum_{i=1}^n u_{ij}(w) \Big| \geq C_3 \big((h^{-m/2} \sqrt{n} + h^{-m} \log n) \sqrt{t} + h^{-m} (\log n)^3 t \big) \right) \leq C_3 e^{-t},$$

$$\mathbb{P}\left(\Big| \sum_{i=1}^n u_{ij}(w) \Big| > C_3 \left(\sqrt{\frac{tn}{h^m}} + \frac{t (\log n)^3}{h^m} \right) \right) \leq C_3 e^{-t},$$

where we used $nh^m \gtrsim (\log n)^2$ and adjusted the constant if necessary. As before, $u_{ij}(w)$ is Lipschitz in w with a constant which is at most polynomial in n, so for some $a > 0$

$$\mathbb{P}\left(\sup_{w \in \mathcal{W}} \Big| \sum_{i=1}^n u_{ij}(w) \Big| > C_3 \left(\sqrt{\frac{tn}{h^m}} + \frac{t (\log n)^3}{h^m} \right) \right) \leq C_3 n^a e^{-t},$$

$$\sup_{w \in \mathcal{W}} \|S(w)\|_2 \lesssim_{\mathbb{P}} \sqrt{\frac{n \log n}{h^m}} + \frac{(\log n)^4}{h^m} \lesssim_{\mathbb{P}} \sqrt{\frac{n \log n}{h^m}}$$

as $nh^m \gtrsim (\log n)^7$. Finally

$$\sup_{w \in \mathcal{W}} \big| e_1^{\mathsf{T}} \big(\hat{H}(w)^{-1} - H(w)^{-1} \big) S(w) \big| \lesssim_{\mathbb{P}} \sqrt{\frac{\log n}{n^3 h^{2m}}} \sqrt{\frac{n \log n}{h^m}} \lesssim_{\mathbb{P}} \frac{\log n}{\sqrt{n^2 h^{3m}}}.$$

Part 8: bounding the bias

Since $\mu \in \mathcal{C}^\gamma$, we have, by the multivariate version of Taylor’s theorem,

$$\mu(W_i) = \sum_{|\kappa|=0}^{\gamma-1} \frac{1}{\kappa!} \partial^\kappa \mu(w) (W_i - w)^\kappa + \sum_{|\kappa|=\gamma} \frac{1}{\kappa!} \partial^\kappa \mu(w') (W_i - w)^\kappa$$

for some w' on the line segment connecting w and W_i. Now since $p_h(W_i - w)_1 = 1$,

$$\begin{aligned}
&e_1^{\mathsf{T}} \hat{H}(w)^{-1} \sum_{i=1}^n K_h(W_i - w) p_h(W_i - w) \mu(w) \\
&\quad = e_1^{\mathsf{T}} \hat{H}(w)^{-1} \sum_{i=1}^n K_h(W_i - w) p_h(W_i - w) p_h(W_i - w)^{\mathsf{T}} e_1 \mu(w) = e_1^{\mathsf{T}} e_1 \mu(w) = \mu(w).
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathrm{Bias}(w) &= e_1^{\mathsf{T}} \hat{H}(w)^{-1} \sum_{i=1}^n K_h(W_i - w) p_h(W_i - w) \mu(W_i) - \mu(w) \\
&= e_1^{\mathsf{T}} \hat{H}(w)^{-1} \sum_{i=1}^n K_h(W_i - w) p_h(W_i - w) \\
&\quad \times \left(\sum_{|\kappa|=0}^{\gamma-1} \frac{1}{\kappa!} \partial^\kappa \mu(w) (W_i - w)^\kappa + \sum_{|\kappa|=\gamma} \frac{1}{\kappa!} \partial^\kappa \mu(w') (W_i - w)^\kappa - \mu(w) \right) \\
&= \sum_{|\kappa|=1}^{\gamma-1} \frac{1}{\kappa!} \partial^\kappa \mu(w) e_1^{\mathsf{T}} \hat{H}(w)^{-1} \sum_{i=1}^n K_h(W_i - w) p_h(W_i - w) (W_i - w)^\kappa
\end{aligned}$$

$$+\sum_{|\kappa|=\gamma}\frac{1}{\kappa!}\partial^\kappa \mu(w')e_1^\mathsf{T}\hat{H}(w)^{-1}\sum_{i=1}^n K_h(W_i-w)p_h(W_i-w)(W_i-w)^\kappa$$

$$=\sum_{|\kappa|=\gamma}\frac{1}{\kappa!}\partial^\kappa \mu(w')e_1^\mathsf{T}\hat{H}(w)^{-1}\sum_{i=1}^n K_h(W_i-w)p_h(W_i-w)(W_i-w)^\kappa,$$

where we used that $p_h(W_i-w)$ is a vector containing monomials in W_i-w of order up to γ, so $e_1^\mathsf{T}\hat{H}(w)^{-1}\sum_{i=1}^n K_h(W_i-w)p_h(W_i-w)(W_i-w)^\kappa=0$ whenever $1\le|\kappa|\le\gamma$. Finally

$$\begin{aligned}
&\sup_{w\in\mathcal{W}}|\mathrm{Bias}(w)| \\
&\quad=\sup_{w\in\mathcal{W}}\left|\sum_{|\kappa|=\gamma}\frac{1}{\kappa!}\partial^\kappa \mu(w')e_1^\mathsf{T}\hat{H}(w)^{-1}\sum_{i=1}^n K_h(W_i-w)p_h(W_i-w)(W_i-w)^\kappa\right| \\
&\quad\lesssim_\mathbb{P}\sup_{w\in\mathcal{W}}\max_{|\kappa|=\gamma}|\partial^\kappa\mu(w')|\,\|\hat{H}(w)^{-1}\|_2\left\|\sum_{i=1}^n K_h(W_i-w)p_h(W_i-w)\right\|_2 h^\gamma \\
&\quad\lesssim_\mathbb{P}\frac{h^\gamma}{n}\sup_{w\in\mathcal{W}}\left\|\sum_{i=1}^n K_h(W_i-w)p_h(W_i-w)\right\|_2 .
\end{aligned}$$

Now write $\tilde{u}_{ij}(w)=K_h(W_i-w)p_h(W_i-w)_j$ and note that $|\tilde{u}_{ij}(w)|\lesssim h^{-m}$ and $\mathbb{E}[\tilde{u}_{ij}(w)]\lesssim 1$. By Lemma SA.7(i), for a constant C_4,

$$\mathbb{P}\left(\left|\sum_{i=1}^n\tilde{u}_{ij}(w)-\mathbb{E}\left[\sum_{i=1}^n\tilde{u}_{ij}(w)\right]\right|>C_4h^{-m}\big(\sqrt{nt}+(\log n)(\log\log n)t\big)\right)\le C_4e^{-t}.$$

As in previous parts, by Lipschitz properties, this implies

$$\sup_{w\in\mathcal{W}}\left|\sum_{i=1}^n\tilde{u}_{ij}(w)\right|\lesssim_\mathbb{P} n\left(1+\sqrt{\frac{\log n}{nh^{2m}}}\right)\lesssim_\mathbb{P} n.$$

Therefore $\sup_{w\in\mathcal{W}}|\mathrm{Bias}(w)|\lesssim_\mathbb{P} nh^\gamma/n\lesssim_\mathbb{P} h^\gamma$.

Part 9: conclusion

By the previous parts,

$$\begin{aligned}
\sup_{w\in\mathcal{W}}|\hat{\mu}(w)-\mu(w)-T(w)|&\le\sup_{w\in\mathcal{W}}\left|e_1^\mathsf{T}H(w)^{-1}S(w)-T(w)\right| \\
&\quad+\sup_{w\in\mathcal{W}}\left|e_1^\mathsf{T}\big(\hat{H}(w)^{-1}-H(w)^{-1}\big)S(w)\right|+\sup_{w\in\mathcal{W}}|\mathrm{Bias}(w)| \\
&\lesssim_\mathbb{P}\left(\frac{(\log n)^{m+4}}{n^{m+4}h^{m(m+6)}}\right)^{\frac{1}{2m+6}}R_n+\frac{\log n}{\sqrt{n^2h^{3m}}}+h^\gamma \\
&\lesssim_\mathbb{P}\frac{R_n}{\sqrt{nh^m}}\left(\frac{(\log n)^{m+4}}{nh^{3m}}\right)^{\frac{1}{2m+6}}+h^\gamma,
\end{aligned}$$

where the last inequality follows because $nh^{3m}\to\infty$ and $\frac{1}{2m+6}\le\frac{1}{2}$. Finally, we verify the upper and lower bounds on the variance of the Gaussian process. Since the spectrum of $H(w)^{-1}$ is bounded above and below by $1/n$,

$$\mathrm{Var}[T(w)]=\mathrm{Var}\left[e_1^\mathsf{T}H(w)^{-1}\sum_{i=1}^n K_h(W_i-w)p_h(W_i-w)\varepsilon_i\right]$$

$$
\begin{aligned}
&= e_1^{\mathsf{T}} H(w)^{-1} \operatorname{Var}\left[\sum_{i=1}^n K_h(W_i - w) p_h(W_i - w)\varepsilon_i\right] H(w)^{-1} e_1^{\mathsf{T}} \\
&\lesssim \|H(w)^{-1}\|_2^2 \max_{1 \le j \le k} \sum_{i=1}^n \operatorname{Var}\left[K_h(W_i - w) p_h(W_i - w)_j \sigma(W_i)\right] \\
&\lesssim \frac{1}{n^2} n \frac{1}{h^m} \lesssim \frac{1}{nh^m}.
\end{aligned}
$$

$\operatorname{Var}[T(w)] \gtrsim \frac{1}{nh^m}$ by the same argument given to bound the eigenvalues of $H(w)^{-1}$. □

APPENDIX SB: DISTRIBUTIONAL APPROXIMATION OF MARTINGALE ℓ^p-NORMS

We present some applications of the results derived in Appendix A. In certain empirical settings, including nonparametric significance tests [14] and nearest neighbor search procedures [4], an estimator or test statistic can be expressed under the null hypothesis as the ℓ^p-norm of a zero-mean (possibly high-dimensional) martingale for some $p \in [1, \infty]$. In the notation of Corollary 2.2, it is therefore of interest to bound Kolmogorov–Smirnov quantities of the form

$$
\sup_{t \ge 0} \left|\mathbb{P}(\|S\|_p \le t) - \mathbb{P}(\|T\|_p \le t)\right|.
$$

Let $\mathcal{B}_p$ be the class of closed ℓ^p-balls in $\mathbb{R}^d$ centered at the origin and set

$$
\Delta_p(\eta) := \Delta_p(\mathcal{B}_p, \eta) = \sup_{t \ge 0} \mathbb{P}(t < \|T\|_p \le t + \eta).
$$

PROPOSITION SB.1 (Distributional approximation of martingale ℓ^p-norms). *Assume the setup of Corollary 2.2, with Σ non-random. Then for $T \sim \mathcal{N}(0, \Sigma)$,*

$$
\sup_{t \ge 0} \left|\mathbb{P}(\|S\|_p \le t) - \mathbb{P}\left(\|T\|_p \le t\right)\right| \le \inf_{\eta > 0} \left\{\Gamma_p(\eta) + \Delta_p(\eta)\right\}. \tag{6}
$$

PROOF (Proposition SB.1). Applying Proposition A.1 with $\mathcal{A} = \mathcal{B}_p$ gives

$$
\begin{aligned}
\sup_{t \ge 0} \left|\mathbb{P}(\|S\|_p \le t) - \mathbb{P}\left(\|T\|_p \le t\right)\right| &= \sup_{A \in \mathcal{B}_p} \left|\mathbb{P}(S \in A) - \mathbb{P}(T \in A)\right| \\
&\le \inf_{\eta > 0} \left\{\Gamma_p(\eta) + \Delta_p(\mathcal{B}_p, \eta)\right\} \le \inf_{\eta > 0} \left\{\Gamma_p(\eta) + \Delta_p(\eta)\right\}.
\end{aligned}
$$

□

The right-hand side of (6) can be controlled in various ways. In the case of $p = \infty$, note that ℓ^∞-balls are rectangles so $\mathcal{B}_\infty \subseteq \mathcal{R}$, giving $\Delta_\infty(\eta) \le \eta(\sqrt{2 \log d} + 2)/\sigma_{\min}$ whenever $\min_j \Sigma_{jj} \ge \sigma_{\min}^2$. Alternatively, Giessing [11, Theorem 1] provides $\Delta_\infty(\eta) \lesssim \eta / \sqrt{\operatorname{Var}[\|T\|_\infty] + \eta^2}$. In fact, by Hölder duality of ℓ^p-norms, we can write $\|T\|_p = \sup_{\|u\|_q \le 1} u^{\mathsf{T}} T$ where $1/p + 1/q = 1$. Then, applying the Gaussian process anti-concentration result of Giessing [11, Theorem 2] yields the more general $\Delta_p(\eta) \lesssim \eta / \sqrt{\operatorname{Var}[\|T\|_p] + \eta^2}$. Thus, the problem can be reduced to that of obtaining lower bounds for $\operatorname{Var}\left[\|T\|_p\right]$, with techniques for doing so discussed, for example, in Giessing [11, Section 4]. Note that alongside the ℓ^p-norms, other functionals can be analyzed in this manner, including the maximum statistic and other order statistics [12, 11].

To conduct inference in this situation, we need to feasibly approximate the quantiles of $\|T\|_p$. To that end, take a significance level $\tau \in (0, 1)$ and define

$$
\hat{q}_p(\tau) = \inf\left\{t \in \mathbb{R} : \mathbb{P}(\|\hat{T}\|_p \le t \mid \mathbf{X}) \ge \tau\right\} \quad \text{where} \quad \hat{T} \mid \mathbf{X} \sim \mathcal{N}(0, \hat{\Sigma}),
$$

with $\hat{\Sigma}$ any $\mathbf{X}$-measurable positive semi-definite estimator of Σ. Note that for the canonical estimator $\hat{\Sigma} = \sum_{i=1}^n X_i X_i^\mathsf{T}$ we can write $\hat{T} = \sum_{i=1}^n X_i Z_i$ with $Z_1, \ldots, Z_n$ i.i.d. standard Gaussian independent of $\mathbf{X}$, yielding the Gaussian multiplier bootstrap. Now assuming the law of $\|\hat{T}\|_p \mid \mathbf{X}$ has no atoms, we can apply Proposition A.2 to see

$$\sup_{\tau \in (0,1)} \left| \mathbb{P}\left(\|S\|_p \leq \hat{q}_p(\tau) \right) - \tau \right| \leq \mathbb{E}\left[\sup_{t \geq 0} \left| \mathbb{P}(\|S\|_p \leq t) - \mathbb{P}(\|\hat{T}\|_p \leq t \mid \mathbf{X}) \right| \right]$$

$$\leq \inf_{\eta > 0} \left\{ \Gamma_p(\eta) + 2\Delta_p(\eta) + 2d\, \mathbb{E}\left[\exp\left(\frac{-\eta^2}{2d^{2/p} \big\| \hat{\Sigma}^{1/2} - \Sigma^{1/2} \big\|_2^2} \right) \right] \right\}$$

and hence the bootstrap is valid whenever $\|\hat{\Sigma}^{1/2} - \Sigma^{1/2}\|_2^2$ is sufficiently small. See the discussion in Appendix A regarding methods for bounding this object.

REMARK 1 (One-dimensional distributional approximations). In our application to distributional approximation of ℓ^p-norms, the object of interest $\|S\|_p$ is a one-dimensional functional of the high-dimensional martingale; contrast this with the more general Proposition A.1 which directly considers the d-dimensional random vector S. As such, our coupling-based approach may be improved in certain settings by applying a more carefully tailored smoothing argument. For example, Belloni and Oliveira [2] employ a "log sum exponential" bound [see also 7] for the maximum statistic $\max_{1 \leq j \leq d} S_j$, along with a coupling due to Chernozhukov, Chetverikov and Kato [9], to attain an improved dependence on the dimension. Naturally their approach does not permit the formulation of high-dimensional central limit theorems over arbitrary classes of Borel sets as in our Proposition A.1.

REFERENCES

[1] BAXTER, B. J. C. (1994). Norm estimates for inverses of Toeplitz distance matrices. *Journal of Approximation Theory* **79** 222–242.

[2] BELLONI, A. and OLIVEIRA, R. I. (2018). A high dimensional central limit theorem for martingales, with applications to context tree models. arXiv:1809.02741.

[3] BELLONI, A., CHERNOZHUKOV, V., CHETVERIKOV, D. and FERNÁNDEZ-VAL, I. (2019). Conditional quantile processes based on series or many regressors. *Journal of Econometrics* **213** 4–29.

[4] BIAU, G. and MASON, D. M. (2015). High-dimensional p-norms. In *Mathematical Statistics and Limit Theorems* 21–40. Springer.

[5] CATTANEO, M. D., FARRELL, M. H. and FENG, Y. (2020). Large sample properties of partitioning-based series estimators. *Annals of Statistics* **48** 1718–1741.

[6] CHEN, X. and KATO, K. (2020). Jackknife multiplier bootstrap: finite sample approximations to the U-process supremum with applications. *Probability Theory and Related Fields* **176** 1097–1163.

[7] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2013). Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Annals of Statistics* **41** 2786–2819.

[8] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014a). Anti-concentration and honest, adaptive confidence bands. *Annals of Statistics* **42** 1787–1818.

[9] CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014b). Gaussian approximation of suprema of empirical processes. *Annals of Statistics* **42** 1564–1597.

[10] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. *Cambridge Studies in Advanced Mathematics*. Cambridge University Press.

[11] GIESSING, A. (2023). Anti-concentration of suprema of Gaussian processes and Gaussian order statistics. arXiv:2310.12119.

[12] KOZBUR, D. (2021). Dimension-free anticoncentration bounds for Gaussian order statistics with discussion of applications to multiple testing. arXiv:2107.10766.

[13] LE CAM, L. (1988). On the Prokhorov distance between the empirical process and the associated Gaussian bridge. *Technical report*.

[14] LOPES, M. E., LIN, Z. and MÜLLER, H.-G. (2020). Bootstrapping max statistics in high dimensions: Near-parametric rates under weak variance decay and application to functional and multinomial data. *Annals of Statistics* **48** 1214–1229.

[15] MERLEVÈDE, F., PELIGRAD, M. and RIO, E. (2009). Bernstein inequality and moderate deviations under strong mixing conditions. In *High Dimensional Probability V: The Luminy Volume* 273–292. Institute of Mathematical Statistics.

[16] MONRAD, D. and PHILIPP, W. (1991). Nearby variables with nearby conditional laws and a strong approximation theorem for Hilbert space valued martingales. *Probability Theory and Related Fields* **88** 381–404.

[17] POLLARD, D. (2002). *A User's Guide to Measure Theoretic Probability*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.

[18] RIO, E. (2017). *Asymptotic Theory of Weakly Dependent Random processes* **80**. Springer.

[19] VAN DE GEER, S. and LEDERER, J. (2013). The Bernstein–Orlicz norm and deviation inequalities. *Probability Theory and Related Fields* **157** 225–250.

[20] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York.