Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL TO 'ISOTONIC REGRESSION IN GENERAL DIMENSIONS'

BY QIYANG HAN[*], TENGYAO WANG[†], SABYASACHI CHATTERJEE[‡,§] AND RICHARD J. SAMWORTH[†]

University of Washington[*], *University of Cambridge*[†], *University of Chicago*[‡] *and University of Illinois at Urbana-Champaign*[§]

APPENDIX A: PROOFS OF PREPARATORY PROPOSITIONS

PROOF OF PROPOSITION 7. For any $f : [0,1]^d \to \mathbb{R}$, define $\mathbb{M}_n f := 2\sum_{i=1}^n \epsilon_i\{f(X_i) - f_0(X_i)\} - \sum_{i=1}^n \{f(X_i) - f_0(X_i)\}^2$ and $Mf := \mathbb{E}\mathbb{M}_n f = -n\|f - f_0\|_{L_2(P)}^2$. By the definition of $\hat{f}_n$, we have that $\sum_{i=1}^n (\hat{f}_n(X_i) - f_0(X_i) - \epsilon_i)^2 \leq \sum_{i=1}^n \epsilon_i^2$, which implies that $\mathbb{M}_n \hat{f}_n \geq 0$. We therefore have that for any $r > 0$,

$$\begin{aligned}
&\mathbb{P}\big(\{\|\hat{f}_n - f_0\|_{L_2(P)} \geq r\} \cap \{\|\hat{f}_n - f_0\|_\infty \leq 6\log^{1/2} n\}\big) \\
&\quad\leq \sum_{\ell=1}^\infty \mathbb{P}\biggl(\sup_{f \in \mathcal{G}(f_0, 2^\ell r, 6\log^{1/2} n)\setminus\mathcal{G}(f_0, 2^{\ell-1} r, 6\log^{1/2} n)} (\mathbb{M}_n - M)f \geq n2^{2\ell-2}r^2\biggr) \\
&\quad\leq \sum_{\ell=1}^\infty \mathbb{P}\biggl(\sup_{f \in \mathcal{G}(f_0, 2^\ell r, 6\log^{1/2} n)} \biggl|\frac{1}{n^{1/2}}\sum_{i=1}^n \epsilon_i (f - f_0)(X_i)\biggr| \geq 2^{2\ell-4} n^{1/2} r^2\biggr) \\
&\qquad\qquad + \sum_{\ell=1}^\infty \mathbb{P}\biggl(\sup_{f \in \mathcal{G}(f_0, 2^\ell r, 6\log^{1/2} n)} \Bigl|\mathbb{G}_n (f - f_0)^2\Bigr| \geq 2^{2\ell-3} n^{1/2} r^2\biggr). \qquad (1)
\end{aligned}$$

By a moment inequality for empirical processes (Giné, Latała and Zinn, 2000, Proposition 3.1) and (18) in the main text, we have for all $p \geq 1$ that

$$\begin{aligned}
&\mathbb{E}\biggl[\sup_{f \in \mathcal{G}(f_0, 2^\ell r, 6\log^{1/2} n)} \biggl|\frac{1}{n^{1/2}}\sum_{i=1}^n \epsilon_i\{f(X_i) - f_0(X_i)\}\biggr|^p\biggr]^{1/p} \\
&\qquad\qquad \lesssim K\phi_n(2^\ell r) + 2^\ell r p^{1/2} + n^{-1/2} p \log n. \qquad (2)
\end{aligned}$$

For any $C' > 0$ and $r \geq C' K r_n$, we have $\phi_n(2^\ell r) \leq 2^\ell (r/r_n)\phi_n(r_n) \leq 2^\ell n^{1/2} r_n r \leq (C'K)^{-1} 2^\ell n^{1/2} r^2$. It therefore follows from (2) and Lemma 11 that there exist universal constants $C, C' > 0$ such that for all $\ell \in \mathbb{N}$ and

$r \geq C'Kr_n$,

$$\mathbb{P}\biggl(\sup_{f \in \mathcal{G}(f_0, 2^\ell r, 6\log^{1/2} n)} \biggl|\frac{1}{n^{1/2}} \sum_{i=1}^n \epsilon_i\{f(X_i) - f_0(X_i)\}\biggr| \geq 2^{2\ell-4} n^{1/2} r^2\biggr) \leq C \exp\biggl(-\frac{2^{2\ell} n r^2}{C \log n}\biggr). \tag{3}$$

Similarly, by a symmetrisation inequality (cf. van der Vaart and Wellner (1996, Lemma 2.3.1)), (19) in the main text and the same argument as above, and by increasing C, C' if necessary, we have that for all $\ell \in \mathbb{N}$ and $r \geq C'Kr_n$,

$$\mathbb{P}\biggl(\sup_{f \in \mathcal{G}(f_0, 2^\ell r, 6\log^{1/2} n)} \bigl|\mathbb{G}_n(f - f_0)^2\bigr| \geq 2^{2\ell-3} n^{1/2} r^2\biggr) \leq C \exp\biggl(-\frac{2^{2\ell} n r^2}{C \log n}\biggr). \tag{4}$$

Substituting (3) and (4) into (1), we obtain that for all $r \geq C'Kr_n$,

$$\mathbb{P}\bigl(\{\|\hat{f}_n - f_0\|_{L_2(P)} \geq r\} \cap \{\|\hat{f}_n - f_0\|_\infty \leq 6\log^{1/2} n\}\bigr) \lesssim \sum_{\ell=1}^\infty \exp\biggl(-\frac{2^{2\ell} n r^2}{C \log n}\biggr) \lesssim \exp\biggl(-\frac{n r^2}{C \log n}\biggr).$$

It follows that

$$\begin{aligned} &\mathbb{E}\bigl(\|\hat{f}_n - f_0\|_{L_2(P)}^2 \mathbb{1}_{\{\|\hat{f}_n - f_0\|_\infty \leq 6\log^{1/2} n\}}\bigr) \\ &= \int_0^\infty 2t\mathbb{P}\bigl(\bigl\{\bigl\|\hat{f}_n - f_0\bigr\|_{L_2(P)} \geq t\bigr\} \cap \{\|\hat{f}_n - f_0\|_\infty \leq 6\log^{1/2} n\}\bigr)\,\mathrm{d}t \\ &\lesssim K^2 r_n^2 + \int_{C'Kr_n}^\infty 2t \exp\biggl(-\frac{t^2}{Cr_n^2}\biggr)\,\mathrm{d}t \lesssim K^2 r_n^2, \end{aligned}$$

as desired, where we have used $r_n^2 \geq n^{-1}\log n$ in the penultimate inequality. □

PROOF OF PROPOSITION 8. [Upper bound] It is convenient here to work with the class of block decreasing functions $\mathcal{F}_{d,\downarrow} := \{f : [0,1]^d \to \mathbb{R} : -f \in \mathcal{F}_d\}$ instead. We write $\mathcal{F}_d^+ := \{f \in \mathcal{F}_d : f \geq 0\}$ and $\mathcal{F}_{d,\downarrow}^+ := \{f \in \mathcal{F}_{d,\downarrow} : f \geq 0\}$. By replacing f with $-f$ and decomposing any function f into its positive and negative parts, it suffices to prove the result with $\mathcal{G}_\downarrow^+(0,r,1) := \mathcal{F}_{d,\downarrow}^+ \cap B_2(r,P) \cap B_\infty(1)$ in place of $\mathcal{G}(0,r,1)$. Since $\mathcal{G}_\downarrow^+(0,r,1) = \mathcal{G}_\downarrow^+(0,1,1)$ for $r \geq 1$, we may also assume without loss of generality that $r \leq 1$. We handle the cases $d = 2$ and $d \geq 3$ separately.

Case $d = 2$. We apply Lemma 7 with $\eta = r/(2n)$ and Lemma 8 to obtain

$$\mathbb{E} \sup_{f \in \mathcal{F}^+_{2,\downarrow} \cap B_2(r,P) \cap B_\infty(1)} \biggl| \frac{1}{n^{1/2}} \sum_{i=1}^n \xi_i f(X_i) \biggr| \lesssim_{d,m_0,M_0} n^{1/2}\eta + \log^3 n \int_\eta^r \frac{r}{\varepsilon}\,\mathrm{d}\varepsilon + \frac{(\log^4 n)(\log\log n)^2}{n^{1/2}} \lesssim r \log^4 n,$$

as desired.

Case $d \geq 3$. We assume without loss of generality that $n = n_1^d$ for some $n_1 \in \mathbb{N}$. We define strips $I_\ell := [0,1]^{d-1} \times [\frac{\ell-1}{n_1}, \frac{\ell}{n_1}]$ for $\ell = 1, \ldots, n_1$, so that $[0,1]^d = \cup_{\ell=1}^{n_1} I_\ell$. Our strategy is to analyse the expected supremum of the symmetrised empirical process when restricted to each strip. To this end, define $S_\ell := \{X_1, \ldots, X_n\} \cap I_\ell$ and $N_\ell := |S_\ell|$, and let $\Omega_0 := \{m_0 n^{1-1/d}/2 \leq \min_\ell N_\ell \leq \max_\ell N_\ell \leq 2M_0 n^{1-1/d}\}$. Then by Hoeffding's inequality,

$$\mathbb{P}(\Omega_0^c) \leq \sum_{\ell=1}^{n_1} \mathbb{P}\biggl(\Bigl| N_\ell - \mathbb{E}N_\ell \Bigr| > \frac{m_0 n}{2n_1} \biggr) \leq 2n_1 \exp\bigl(-m_0^2 n^{1-2/d}/8\bigr).$$

Hence we have

$$\mathbb{E} \sup_{f \in \mathcal{F}^+_{d,\downarrow} \cap B_2(r,P) \cap B_\infty(1)} \biggl| \frac{1}{n^{1/2}} \sum_{i=1}^n \xi_i f(X_i) \biggr| \leq \mathbb{E}\biggl(\sum_{\ell : N_\ell \geq 1} \frac{N_\ell^{1/2}}{n^{1/2}} E_\ell \, \mathbb{1}_{\Omega_0} \biggr) + C \exp\bigl(-m_0^2 n^{1-2/d}/16\bigr), \tag{5}$$

where

$$E_\ell := \mathbb{E}\biggl\{ \sup_{f \in \mathcal{F}^+_{d,\downarrow} \cap B_2(r,P) \cap B_\infty(1)} \biggl| \frac{1}{N_\ell^{1/2}} \sum_{i : X_i \in S_\ell} \xi_i f(X_i) \biggr| \biggm| N_1, \ldots, N_{n_1} \biggr\}.$$

By Lemma 9, for any $f \in \mathcal{F}^+_{d,\downarrow} \cap B_2(r,P) \cap B_\infty(1)$ and $\ell \in \{1, \ldots, n_1\}$, we have $\int_{I_\ell} f^2 \,\mathrm{d}P \leq 7(M_0/m_0)\ell^{-1} r^2 \log^d n =: r_{n,\ell}^2$. Consequently, we have by Lemma 7 that for any $\eta \in [0, r_{n,\ell}/3)$,

$$E_\ell \lesssim N_\ell^{1/2}\eta + \int_\eta^{r_{n,\ell}} H_{[\,],\ell}^{1/2}(\varepsilon)\,\mathrm{d}\varepsilon + \frac{H_{[\,],\ell}(r_{n,\ell})}{N_\ell^{1/2}}, \tag{6}$$

where $H_{[\,],\ell}(\varepsilon) := \log N_{[\,]}\bigl(\varepsilon, \mathcal{F}^+_{d,\downarrow}(I_\ell) \cap B_2(r_{n,\ell}, P; I_\ell) \cap B_\infty(1; I_\ell), \|\cdot\|_{L_2(P;I_\ell)}\bigr)$. Here, the set $\mathcal{F}^+_{d,\downarrow}(I_\ell)$ is the class of non-negative functions on I_ℓ that are block decreasing, $B_\infty(1; I_\ell)$ is the class of functions on I_ℓ that are bounded by 1 and $B_2(r_{n,\ell}, P; I_\ell)$ is the class of measurable functions f on I_ℓ with $\|f\|_{L_2(P;I_\ell)} \leq r_{n,\ell}$. Note that any $g \in \mathcal{F}^+_{d,\downarrow}(I_\ell) \cap B_2(r_{n,\ell}, P; I_\ell) \cap B_\infty(1; I_\ell)$ can be rescaled into a function $f_g \in \mathcal{F}^+_{d,\downarrow} \cap B_2\bigl(n_1^{1/2}(M_0/m_0)^{1/2} r_{n,\ell}, P\bigr) \cap B_\infty(1)$ via the invertible map $f_g(x_1, \ldots, x_{d-1}, x_d) := g(x_1, \ldots, x_{d-1}, (x_d + \ell - 1)/n_1)$. Moreover, we have $\int_{[0,1]^d} (f_g - f_{g'})^2 \, dP \geq n_1 (m_0/M_0) \int_{I_\ell} (g - g')^2 \, dP$. Thus, by Lemma 8, for $\varepsilon \in [\eta, r_{n,\ell}]$,

$$\begin{aligned} H_{[\,],\ell}(\varepsilon) &\leq \log N_{[\,]}\bigl(n^{1/(2d)}(m_0/M_0)^{1/2}\varepsilon, \\ &\qquad \mathcal{F}^+_{d,\downarrow} \cap B_2\bigl(n^{1/(2d)}(M_0/m_0)^{1/2} r_{n,\ell}, P\bigr) \cap B_\infty(1), \|\cdot\|_{L_2(P)}\bigr) \\ &\lesssim_{d,m_0,M_0} \left(\frac{r_{n,\ell}}{\varepsilon}\right)^{2(d-1)} \log_+^{d^2}(1/\varepsilon). \end{aligned}$$

Substituting the above bound into (6), and choosing $\eta = n^{-1/(2d)} r_{n,\ell}$, we obtain

$$\begin{aligned} E_\ell &\lesssim_{d,m_0,M_0} N_\ell^{1/2}\eta + \log^{d^2/2} n \int_\eta^{r_{n,\ell}} \left(\frac{r_{n,\ell}}{\varepsilon}\right)^{d-1} d\varepsilon + \frac{\log^{d^2} n}{N_\ell^{1/2}} \\ &\lesssim N_\ell^{1/2}\eta + \frac{r_{n,\ell}^{d-1} \log^{d^2/2} n}{\eta^{d-2}} + \frac{\log^{d^2} n}{N_\ell^{1/2}}. \end{aligned}$$

Hence

$$\begin{aligned} E_\ell \mathbb{1}_{\Omega_0} &\lesssim_{d,m_0,M_0} r_{n,\ell}\, n^{1/2-1/d} \log^{d^2/2} n + n^{-1/2+1/(2d)} \log^{d^2} n \\ &\lesssim_{m_0,M_0} r_{n,\ell}\, n^{1/2-1/d} \log^{d^2/2} n, \end{aligned} \tag{7}$$

where in the final inequality we used the conditions that $d \geq 3$ and $r \geq n^{-(1-2/d)} \log^{(d^2-d)/2} n$. Combining (5) and (7), we have that

$$\begin{aligned} &\mathbb{E} \sup_{f \in \mathcal{F}^+_{d,\downarrow} \cap B_2(r,P) \cap B_\infty(1)} \left| \frac{1}{n^{1/2}} \sum_{i=1}^n \xi_i f(X_i) \right| \\ &\lesssim_{d,m_0,M_0} r n^{1/2-3/(2d)} \log^{(d^2+d)/2} n \sum_{\ell=1}^{n_1} \ell^{-1/2} \lesssim r n^{1/2-1/d} \log^{(d^2+d)/2} n, \end{aligned}$$

which completes the proof.

[Lower bound] Assume without of loss of generality that $n = n_1^d$ for some $n_1 \in \mathbb{N}$. For a multi-index $w = (w_1, \ldots, w_d) \in \mathbb{L}_{d,n}$, let $L_w := \prod_{j=1}^d (w_j - 1/n_1, w_j]$ and $N_w := |\{X_1, \ldots, X_n\} \cap L_w|$. We define $W := \{(w_1, \ldots, w_d) :$

$\sum_{j=1}^d w_j = 1\}$ to be indices of a mutually incomparable collection of cubelets and define $\tilde{W} := \{w \in W : N_w \geq 1\}$ to be the (random) set of indices of cubelets in this collection that contain at least one design point. For each $w \in \tilde{W}$, associate $i_w := \min\{i : X_i \in L_w\}$. For each realisation of the Rademacher random variables $\xi = (\xi_i)_{i=1}^n$ and design points $X = \{X_i\}_{i=1}^n$, define $f_{\xi,X} : [0,1]^d \to [-1,1]$ to be the function such that

$$f_{\xi,X}(x) := \begin{cases} r\,\xi_{i_w} & \text{if } x \in L_w,\ w \in \tilde{W} \\ r & \text{if } x \in L_w \text{ with } \sum_{j=1}^d w_j > n_1 \\ -r & \text{otherwise.} \end{cases}$$

For $r \leq 1$, we have $f_{\xi,X} \in \mathcal{F}_d \cap B_2(r,P) \cap B_\infty(1)$. Therefore,

$$\begin{aligned} \mathbb{E} \sup_{f \in \mathcal{F}_d \cap B_2(r,P) \cap B_\infty(1)} \sum_{i=1}^n \xi_i f(X_i) &\geq \mathbb{E} \sum_{i=1}^n \xi_i f_{\xi,X}(X_i) \\ &\geq \mathbb{E}\biggl[\mathbb{E}\biggl\{\sum_{i=1}^n \xi_i f_{\xi,X}(X_i) \biggm| X_1,\ldots,X_n,\{\xi_{i_w} : w \in \tilde{W}\}\biggr\}\biggr] \\ &= \mathbb{E} \sum_{w \in \tilde{W}} \xi_{i_w} f_{\xi,X}(X_{i_w}) = r\mathbb{E}|\tilde{W}|. \end{aligned}$$

The desired lower bound follows since $\mathbb{E}|\tilde{W}| \geq \{1-(1-m_0/n)^n\}|W| \geq (1-e^{-m_0})|W| \gtrsim_{d,m_0} n^{1-1/d}$, where the final bound follows as in the proof of Proposition 5. □

Proof of Proposition 9. Let $r_n := n^{-1/d}\log^{\gamma_d} n$. We write

$$\mathbb{E}\|\tilde{f}_n\|^2_{L_2(\mathbb{P}_n)} = \mathbb{E}\bigl\{\|\tilde{f}_n\|^2_{L_2(\mathbb{P}_n)}\mathbb{1}_{\{\|\hat{f}_n\|_{L_2(P)} \leq r_n\}}\bigr\} + \mathbb{E}\bigl\{\|\tilde{f}_n\|^2_{L_2(\mathbb{P}_n)}\mathbb{1}_{\{\|\hat{f}_n\|_{L_2(P)} > r_n\}}\bigr\} \tag{8}$$

and control the two terms on the right hand side of (8) separately. For the first term, we have

$$\begin{aligned} \mathbb{E}\bigl\{\|\tilde{f}_n\|^2_{L_2(\mathbb{P}_n)}\mathbb{1}_{\{\|\hat{f}_n\|_{L_2(P)} \leq r_n\}}\bigr\} &\leq \mathbb{E} \sup_{f \in \mathcal{F}_d \cap B_2(r_n,P) \cap B_\infty(6\log^{1/2} n)} \frac{1}{n}\sum_{i=1}^n f^2(X_i) \\ &\lesssim r_n^2 + \frac{1}{n}\mathbb{E} \sup_{f \in \mathcal{F}_d \cap B_2(r_n,P) \cap B_\infty(6\log^{1/2} n)} \biggl|\sum_{i=1}^n \xi_i f^2(X_i)\biggr| \\ &\lesssim r_n^2 + \frac{\log^{1/2} n}{n}\mathbb{E} \sup_{f \in \mathcal{F}_d \cap B_2(r_n,P) \cap B_\infty(6\log^{1/2} n)} \biggl|\sum_{i=1}^n \xi_i f(X_i)\biggr| \\ &\lesssim_{d,m_0,M_0} r_n^2 + r_n n^{-1/d}\log^{\gamma_d} n \lesssim r_n^2, \end{aligned} \tag{9}$$

where the second line uses the symmetrisation inequality (cf. van der Vaart and Wellner, 1996, Lemma 2.3.1), the third inequality follows from Lemma 6 and the penultimate inequality follows from Proposition 8. For the second term on the right-hand side of (8), we first claim that there exists $C'_{d,m_0,M_0} > 0$, depending only on d, m_0 and M_0, such that

$$\mathbb{P}(\mathcal{E}^c) \leq \frac{2}{n^2}, \tag{10}$$

where

$$\mathcal{E} := \biggl\{ \sup_{f\in\mathcal{F}_d\cap B_2(r_n,P)^c\cap B_\infty(6\log^{1/2} n)} \biggl| \frac{\mathbb{P}_n f^2}{Pf^2} - 1 \biggr| \leq C'_{d,m_0,M_0} \biggr\}.$$

To see this, we adopt a peeling argument as follows. Let $\mathcal{F}_{d,\ell} := \{f \in \mathcal{F}_d \cap B_\infty(6\log^{1/2} n) : 2^{\ell-1}r_n^2 < Pf^2 \leq 2^\ell r_n^2\}$ and let m be the largest integer such that $2^m r_n^2 < 32\log n$ (so that $m \asymp \log n$). We have that

$$\sup_{\substack{f\in\mathcal{F}_d\cap B_\infty(6\log^{1/2} n)\\ \|f\|_{L_2(P)}>r_n}} \biggl| \frac{\mathbb{P}_n f^2}{Pf^2} - 1 \biggr| \leq \frac{2}{n^{1/2}} \max_{\ell=1,\ldots,m} \biggl\{ (2^\ell r_n^2)^{-1} \sup_{f\in\mathcal{F}_{d,\ell}} |\mathbb{G}_n f^2| \biggr\}.$$

By Talagrand's concentration inequality for empirical processes (Talagrand, 1996), in the form given by Massart (2000, Theorem 3), applied to the class $\{f^2 : f \in \mathcal{F}_{d,\ell}\}$, we have that for any $s_\ell > 0$,

$$\mathbb{P}\biggl\{ \sup_{f\in\mathcal{F}_{d,\ell}} |\mathbb{G}_n f^2| > 2\mathbb{E} \sup_{f\in\mathcal{F}_{d,\ell}} |\mathbb{G}_n f^2| + 12\sqrt{2}\bigl(2^\ell s_\ell \log n\bigr)^{1/2} r_n + \frac{1242 s_\ell \log n}{n^{1/2}} \biggr\} \leq e^{-s_\ell}.$$

Here we have used the fact that $\sup_{f\in\mathcal{F}_{d,\ell}} \mathrm{Var}_P f^2 \leq \sup_{f\in\mathcal{F}_{d,\ell}} Pf^2 \|f\|_\infty^2 \leq 36 \cdot 2^\ell r_n^2 \log n$. Further, by the symmetrisation inequality again, Lemma 6 and Proposition 8, we have that

$$\begin{aligned}\mathbb{E} \sup_{f\in\mathcal{F}_{d,\ell}} |\mathbb{G}_n f^2| &\leq \frac{2}{n^{1/2}} \mathbb{E} \sup_{f\in\mathcal{F}_{d,\ell}} \biggl| \sum_{i=1}^n \xi_i f^2(X_i) \biggr| \leq \frac{48\log^{1/2} n}{n^{1/2}} \mathbb{E} \sup_{f\in\mathcal{F}_{d,\ell}} \biggl| \sum_{i=1}^n \xi_i f(X_i) \biggr| \\ &\lesssim_{d,m_0,M_0} 2^{\ell/2} r_n n^{1/2-1/d} \log^{\gamma_d} n.\end{aligned}$$

By a union bound, we have that with probability at least $1 - \sum_{\ell=1}^m e^{-s_\ell}$,

$$\begin{aligned}&\sup_{f\in\mathcal{F}_d\cap B_2(r_n,P)^c\cap B_\infty(6\log^{1/2} n)} \biggl| \frac{\mathbb{P}_n f^2}{Pf^2} - 1 \biggr| \\ &\qquad \lesssim_{d,m_0,M_0} \max_{\ell=1,\ldots,m} \biggl\{ \frac{n^{1/2-1/d}\log^{\gamma_d} n + s_\ell^{1/2}\log^{1/2} n}{2^{\ell/2} n^{1/2} r_n} + \frac{s_\ell \log n}{2^\ell n r_n^2} \biggr\}.\end{aligned}$$

By choosing $s_\ell := 2^\ell \log n$, we see that on an event of probability at least $1 - \sum_{\ell=1}^m e^{-s_\ell} \geq 1 - \sum_{\ell=1}^\infty n^{-\ell-1} \geq 1 - 2n^{-2}$, we have

$$\sup_{f \in \mathcal{F}_d \cap B_2(r_n,P)^c \cap B_\infty(6\log^{1/2} n)} \left| \frac{\mathbb{P}_n f^2}{P f^2} - 1 \right| \lesssim_{d,m_0,M_0} 1,$$

which verifies (10). Thus

$$\begin{aligned}\mathbb{E}\bigl\{\bigl\|\tilde{f}_n\bigr\|^2_{L_2(\mathbb{P}_n)} \mathbb{1}_{\{\|\hat{f}_n\|_{L_2(P)} > r_n\}}\bigr\} &\leq \mathbb{E}\bigl\{\bigl\|\tilde{f}_n\bigr\|^2_{L_2(\mathbb{P}_n)} \mathbb{1}_{\{\|\hat{f}_n\|_{L_2(P)} > r_n\}} \mathbb{1}_{\mathcal{E}}\bigr\} + \frac{72\log n}{n^2} \\ &\leq (C'_{d,m_0,M_0} + 1)\mathbb{E}\bigl\|\tilde{f}_n\bigr\|^2_{L_2(P)} + \frac{72 \log n}{n^2}. \qquad (11)\end{aligned}$$

Combining (8), (9) and (11), we obtain

$$\mathbb{E}\bigl\|\tilde{f}_n\bigr\|^2_{L_2(\mathbb{P}_n)} \lesssim_{d,m_0,M_0} r_n^2 + \mathbb{E}\bigl\|\tilde{f}_n\bigr\|^2_{L_2(P)},$$

as desired. □

Proof of Proposition 10. Let $r_n := n^{-1/d} \log^{\gamma_d} n$ and observe that by Lemma 5 and Proposition 8, we have that for $r \geq r_n$,

$$\mathbb{E} \sup_{f \in \mathcal{F}_d \cap B_2(r,P) \cap B_\infty(6\log^{1/2} n)} \left| \frac{1}{n^{1/2}} \sum_{i=1}^n \epsilon_i f(X_i) \right| \lesssim_{d,m_0,M_0} r n^{1/2-1/d} \log^{\gamma_d} n.$$

On the other hand, by Lemma 6 and Proposition 8, for $r \geq r_n$,

$$\mathbb{E} \sup_{f \in \mathcal{F}_d \cap B_2(r,P) \cap B_\infty(6\log^{1/2} n)} \left| \frac{1}{n^{1/2}} \sum_{i=1}^n \xi_i f^2(X_i) \right| \lesssim_{d,m_0,M_0} r n^{1/2-1/d} \log^{\gamma_d} n.$$

It follows that the conditions of Proposition 7 are satisfied for this choice of r_n with $\phi_n(r) := r n^{1/2-1/d} \log^{\gamma_d} n$ and $K \lesssim_{d,m_0,M_0} 1$. By Lemma 10, Propositions 9 and 7, we have that

$$\begin{aligned}R_n(\hat{f}_n, 0) &\leq \mathbb{E}\|\tilde{f}_n\|^2_{L_2(\mathbb{P}_n)} + n^{-1} \\ &\lesssim_{d,m_0,M_0} n^{-2/d} \log^{2\gamma_d} n + \mathbb{E}\|\tilde{f}_n\|^2_{L_2(P)} \lesssim_{d,m_0,M_0} n^{-2/d} \log^{2\gamma_d} n,\end{aligned}$$

as desired. □

APPENDIX B: AUXILIARY LEMMAS

We collect here various auxiliary results used in the proofs in the main document (Han et al., 2018).

The proof of Corollary 1 in the main document requires the following lemma on Riemann approximation of block increasing functions.

LEMMA 1. *Suppose $n_1 = n^{1/d}$ is a positive integer. For any $f \in \mathcal{F}_d$, define $f_L(x_1,\ldots,x_d) := f\bigl(n_1^{-1}\lfloor n_1x_1\rfloor,\ldots,n_1^{-1}\lfloor n_1x_d\rfloor\bigr)$ and $f_U(x_1,\ldots,x_d) := f\bigl(n_1^{-1}\lceil n_1x_1\rceil,\ldots,n_1^{-1}\lceil n_1x_d\rceil\bigr)$. Then*

$$\int_{[0,1]^d}(f_U - f_L)^2 \leq 4dn^{-1/d}\|f\|_\infty^2.$$

PROOF. For $x = (x_1,\ldots,x_d)^\top$ and $x' = (x_1',\ldots,x_d')^\top$ in $\mathbb{L}_{d,n}$, we say x and x' are equivalent if and only if $x_j - x_1 = x_j' - x_1'$ for $j = 1,\ldots,d$. Let $\mathbb{L}_{d,n} = \bigsqcup_{r=1}^N P_r$ be the partition of $\mathbb{L}_{d,n}$ into equivalence classes. Since each P_r has non-empty intersection with a different element of the set $\{(x_1,\ldots,x_d) \in \mathbb{L}_{d,n} : \max_j x_j = 1\}$, we must have $N \leq dn^{1-1/d}$. Therefore, we have

$$\begin{aligned}
&\int_{[0,1]^d}(f_U - f_L)^2 = \sum_{r=1}^N \int_{P_r + n_1^{-1}(-1,0]^d}(f_U - f_L)^2 \\
&\leq \frac{2}{n}\|f\|_\infty \sum_{r=1}^N \sum_{x=(x_1,\ldots,x_d)^\top \in P_r}\bigl\{f(x_1,\ldots,x_d) - f\bigl(x_1 - n_1^{-1},\ldots,x_d - n_1^{-1}\bigr)\bigr\} \\
&\leq \frac{2N}{n}\|f\|_\infty\bigl(f(1,\ldots,1) - f(0,\ldots,0)\bigr) \leq 4dn^{-1/d}\|f\|_\infty^2,
\end{aligned}$$

as desired. □

The following is a simple generalisation of Jensen's inequality.

LEMMA 2. *Suppose $h : [0,\infty) \to (0,\infty)$ is a non-decreasing function satisfying the following:*

(i) *There exists $x_0 \geq 0$ such that h is concave on $[x_0,\infty)$.*
(ii) *There exists some $x_1 > x_0$ such that $h(x_1) - x_1h_+'(x_1) \geq h(x_0)$, where h_+' is the right derivative of h.*

Then there exists $C_h > 0$, depending only on h, such that for any nonnegative random variable X with $\mathbb{E}X < \infty$, we have

$$\mathbb{E}h(X) \leq C_h h(\mathbb{E}X).$$

PROOF. Define $H : [0, \infty) \to [h(0), \infty)$ by

$$H(x) := \begin{cases} h(x_1) - x_1 h'_+(x_1) + x h'_+(x_1) & \text{if } x \in [0, x_1) \\ h(x) & \text{if } x \in [x_1, \infty). \end{cases}$$

Then H is a concave majorant of h. Moreover, we have $H \leq (h(x_1)/h(0))h$. Hence, by Jensen's inequality, we have

$$\mathbb{E}h(X) \leq \mathbb{E}H(X) \leq H(\mathbb{E}X) \leq \frac{h(x_1)}{h(0)} h(\mathbb{E}X),$$

as desired. □

We need the following lower bound on the metric entropy of $\mathcal{M}(\mathbb{L}_{2,n}) \cap B_2(1)$ for the proof of Proposition 2.

LEMMA 3. *There exist universal constants $c > 0$ and $\varepsilon_0 > 0$ such that*

$$\log N\bigl(\varepsilon_0, \mathcal{M}(\mathbb{L}_{2,n}) \cap B_2(1), \|\cdot\|_2\bigr) \geq c \log^2 n.$$

PROOF. It suffices to prove the equivalent result that there exist universal constants $c, \varepsilon_0 > 0$ such that the packing number $D\bigl(\varepsilon_0, \mathcal{M}(\mathbb{L}_{2,n}) \cap B_2(1), \|\cdot\|_2\bigr)$ (i.e. the maximum number of disjoint open Euclidean balls of radius ε_0 that can be fitted into $\mathcal{M}(\mathbb{L}_{2,n}) \cap B_2(1)$) is at least $\exp(c \log^2 n)$. Without loss of generality, we may also assume that $n_1 := n^{1/2} = 2^\ell - 1$ for some $\ell \in \mathbb{N}$, so that $\ell \asymp \log n$. Now, for $r = 1, \ldots, \ell$, let $I_r := n_1^{-1}\{2^{r-1}, \ldots, 2^r - 1\}$ and consider the set

$$\bar{\mathcal{M}} := \biggl\{\theta \in \mathbb{R}^{\mathbb{L}_{2,n}} : \theta_{I_r \times I_s} \in \Bigl\{\frac{-\mathbf{1}_{I_r \times I_s}}{\sqrt{2^{r+s+1}} \log n}, \frac{-\mathbf{1}_{I_r \times I_s}}{\sqrt{2^{r+s}} \log n}\Bigr\}\biggr\} \subseteq \mathcal{M}(\mathbb{L}_{2,n}) \cap B_2(1),$$

where $\mathbf{1}_{I_r \times I_s}$ denotes the all-one vector on $I_r \times I_s$. Define a bijection $\psi : \bar{\mathcal{M}} \to \{0, 1\}^{\ell^2}$ by

$$\psi(\theta) := \Bigl(\mathbb{1}_{\{\theta_{I_r \times I_s} = -\mathbf{1}_{I_r \times I_s}/\sqrt{2^{r+s+1}} \log n\}}\Bigr)_{r,s=1}^{\ell}.$$

Then, for $\theta, \theta' \in \bar{\mathcal{M}}$,

$$\|\theta - \theta'\|_2^2 = \frac{d_{\mathrm{H}}(\psi(\theta), \psi(\theta'))}{\log^2 n} \frac{1}{4}\Bigl(1 - \frac{1}{2^{1/2}}\Bigr)^2,$$

where $d_{\mathrm{H}}(\cdot,\cdot)$ denotes the Hamming distance. On the other hand, by the Gilbert–Varshamov inequality (e.g. Massart, 2007, Lemma 4.7), there exists a subset $\mathcal{I} \subseteq \{0,1\}^{\ell^2}$ such that $|\mathcal{I}| \geq \exp(\ell^2/8)$ and $d_{\mathrm{H}}(v,v') \geq \ell^2/4$ for any distinct $v, v' \in \mathcal{I}$. Then the set $\psi^{-1}(\mathcal{I}) \subseteq \bar{\mathcal{M}}$ has cardinality at least $\exp(\ell^2/8) \geq \exp(\log^2 n/32)$, and each pair of distinct elements have squared ℓ_2 distance at least $\varepsilon_0 := \frac{\ell^2/4}{\log^2 n}\frac{1}{4}(1-\frac{1}{2^{1/2}})^2 \gtrsim 1$, as desired. □

Lemma 4 below gives a lower bound on the size of the maximal antichain (with respect to the natural partial ordering on $\mathbb{R}^d$) among independent and identically distributed $X_1,\ldots,X_n$.

LEMMA 4. *Let $d \geq 2$. Let $X_1,\ldots,X_n \stackrel{\mathrm{iid}}{\sim} P$, where P is a distribution on $[0,1]^d$ with Lebesgue density bounded above by $M_0 \in [1,\infty)$. Then with probability at least $1 - e^{-ed^{-1}(M_0 n)^{1/d}\log(M_0 n)}$, there is an antichain in G_X with cardinality at least $n^{1-1/d}/(2eM_0^{1/d})$.*

PROOF. By Dilworth's Theorem (Dilworth, 1950), for each realisation of the directed acyclic graph G_X, there exists a covering of $V(G_X)$ by chains $\mathcal{C}_1,\ldots,\mathcal{C}_M$, where M denotes the cardinality of a maximum antichain of G_X. Thus, it suffices to show that with the given probability, the maximum chain length of G_X is at most $k := \lceil e(M_0 n)^{1/d}\rceil \leq 2e(M_0 n)^{1/d}$. By a union bound, we have that

$$\begin{aligned}\mathbb{P}(\exists \text{ a chain of length } k \text{ in } G_X) &\leq \frac{n!}{(n-k)!}\mathbb{P}(X_1 \preceq \cdots \preceq X_k)\\ &\leq \binom{n}{k}(k!)^{-(d-1)}M_0^k \leq \left(\frac{en}{k}\right)^k\left(\frac{k}{e}\right)^{-k(d-1)}M_0^k\\ &\leq (M_0 n)^{-k/d} \leq e^{-ed^{-1}(M_0 n)^{1/d}\log(M_0 n)},\end{aligned}$$

as desired. □

The following two lemmas control the empirical processes in (18) and (19) in the main text by the symmetrised empirical process in (20) in the main text.

LEMMA 5. *Let $n \geq 2$, and suppose that $X_1,\ldots,X_n,\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_n$ are independent, with $X_1,\ldots,X_n$ identically distributed on $\mathcal{X}$ and $\tilde{\epsilon}_1,\ldots,\tilde{\epsilon}_n$ identically distributed, with $|\tilde{\epsilon}_1|$ stochastically dominated by $|\epsilon_1|$. Then for any countable class $\mathcal{F}$ of measurable, real-valued functions defined on $\mathcal{X}$, we have*

$$\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \tilde{\epsilon}_i f(X_i)\right| \leq 2\log^{1/2} n\,\mathbb{E}\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \xi_i f(X_i)\right|.$$

PROOF. Let $\alpha_0 := 0$, and for $k = 1, \ldots, n$, let $\alpha_k := \mathbb{E}|\tilde{\epsilon}_{(k)}|$, where $|\tilde{\epsilon}_{(1)}| \leq \cdots \leq |\tilde{\epsilon}_{(n)}|$ are the order statistics of $\{|\tilde{\epsilon}_1|, \ldots, |\tilde{\epsilon}_n|\}$, so that $\alpha_n \leq (2\log n)^{1/2}$. Observe that for any $k = 1, \ldots, n$,

$$\begin{aligned}
\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^k \xi_i f(X_i)\biggr| &= \mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^k \xi_i f(X_i) + \mathbb{E}\sum_{i=k+1}^n \xi_i f(X_i)\biggr| \\
&\leq \mathbb{E}\sup_{f\in\mathcal{F}}\mathbb{E}\biggl\{\biggl|\sum_{i=1}^n \xi_i f(X_i)\biggr| \biggm| X_1, \ldots, X_k, \xi_1, \ldots, \xi_k\biggr\} \\
&\leq \mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f(X_i)\biggr|. \qquad (12)
\end{aligned}$$

We deduce from Han and Wellner (2017, Proposition 5) and (12) that

$$\begin{aligned}
\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \tilde{\epsilon}_i f(X_i)\biggr| &\leq 2^{1/2}\sum_{k=1}^n (\alpha_{n+1-k} - \alpha_{n-k})\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^k \xi_i f(X_i)\biggr| \\
&\leq 2^{1/2}\alpha_n \mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f(X_i)\biggr|,
\end{aligned}$$

as required. □

LEMMA 6. *Let $X_1, \ldots, X_n$ be random variables taking values in $\mathcal{X}$ and $\mathcal{F}$ be a countable class of measurable functions $f : \mathcal{X} \to [-1, 1]$. Then*

$$\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f^2(X_i)\biggr| \leq 4\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f(X_i)\biggr|.$$

PROOF. By Ledoux and Talagrand (2013, Theorem 4.12), applied to $\phi_i(y) = y^2/2$ for $i = 1, \ldots, n$ (note that $y \mapsto y^2/2$ is a contraction on $[0, 1]$), we have

$$\begin{aligned}
\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f^2(X_i)\biggr| &= \mathbb{E}\biggl\{\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f^2(X_i)\biggr| \biggm| X_1, \ldots, X_n\biggr\} \\
&\leq 4\mathbb{E}\biggl\{\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f(X_i)\biggr| \biggm| X_1, \ldots, X_n\biggr\} = 4\mathbb{E}\sup_{f\in\mathcal{F}}\biggl|\sum_{i=1}^n \xi_i f(X_i)\biggr|,
\end{aligned}$$

as required. □

The following is a local maximal inequality for empirical processes under bracketing entropy conditions. This result is well known for $\eta = 0$ in the literature, but we provide a proof for the general case $\eta \geq 0$ for the convenience of the reader.

LEMMA 7. *Let $X_1, \ldots, X_n \overset{\mathrm{iid}}{\sim} P$ on $\mathcal{X}$ with empirical distribution $\mathbb{P}_n$, and, for some $r > 0$, let $\mathcal{G} \subseteq B_2(r, P) \cap B_\infty(1)$ be a countable class of measurable functions. Then for any $\eta \in [0, r/3)$, we have*

$$\mathbb{E} \sup_{f \in \mathcal{G}} |\mathbb{G}_n f| \lesssim n^{1/2}\eta + \int_\eta^r \log_+^{1/2} N_{[\,]}(\varepsilon, \mathcal{G}, \|\cdot\|_{L_2(P)}) \, \mathrm{d}\varepsilon + \frac{1}{n^{1/2}} \log_+ N_{[\,]}(r, \mathcal{G}, \|\cdot\|_{L_2(P)}).$$

The above inequality also holds if we replace $\mathbb{G}_n f$ with the symmetrised empirical process $n^{-1/2}\sum_{i=1}^n \xi_i f(X_i)$.

PROOF. Writing $N_r := N_{[\,]}(r, \mathcal{G}, \|\cdot\|_{L_2(P)})$, there exists $\{(f_\ell^L, f_\ell^U) : \ell = 1, \ldots, N_r\}$ that form an r-bracketing set for $\mathcal{G}$ in the $L_2(P)$ norm. Letting $\mathcal{G}_1 := \{f \in \mathcal{G} : f_1^L \leq f \leq f_1^U\}$ and $\mathcal{G}_\ell := \{f \in \mathcal{G} : f_\ell^L \leq f \leq f_\ell^U\} \setminus \cup_{j=1}^{\ell-1} \mathcal{G}_j$ for $\ell = 2, \ldots, N_r$, we see that $\{\mathcal{G}_\ell\}_{\ell=1}^{N_r}$ is a partition of $\mathcal{G}$ such that the $L_2(P)$-diameter of each $\mathcal{G}_\ell$ is at most r. It follows by van der Vaart and Wellner (1996, Lemma 2.14.3) that for any choice of $f_\ell \in \mathcal{G}_\ell$, we have that

$$\begin{aligned}\mathbb{E} \sup_{f \in \mathcal{G}} |\mathbb{G}_n f| &\lesssim n^{1/2}\eta + \int_\eta^r \log_+^{1/2} N_{[\,]}(\varepsilon, \mathcal{G}, \|\cdot\|_{L_2(P)}) \, \mathrm{d}\varepsilon \\ &\quad + \mathbb{E} \max_{\ell=1,\ldots,N_r} |\mathbb{G}_n f_\ell| + \mathbb{E} \max_{\ell=1,\ldots,N_r} \Big| \mathbb{G}_n \Big(\sup_{f \in \mathcal{G}_\ell} |f - f_\ell| \Big) \Big|. \end{aligned} \tag{13}$$

The third and fourth terms of (13) can be controlled by Bernstein's inequality (in the form of (2.5.5) in van der Vaart and Wellner (1996)):

$$\mathbb{E} \max_{\ell=1,\ldots,N_r} |\mathbb{G}_n f_\ell| \vee \mathbb{E} \max_{\ell=1,\ldots,N_r} \Big| \mathbb{G}_n \Big(\sup_{f \in \mathcal{G}_\ell} |f - f_\ell| \Big) \Big| \lesssim \frac{\log_+ N_r}{n^{1/2}} + r \log_+^{1/2} N_r.$$

Since $\eta < r/3$, the last term $r \log_+^{1/2} N_r$ in the above display can be assimilated into the entropy integral in (13), which establishes the claim for $\mathbb{E} \sup_{f \in \mathcal{G}} |\mathbb{G}_n f|$.

We now study the symmetrised empirical process. For $f \in \mathcal{G}$, we define $e \otimes f : \{-1, 1\} \times \mathcal{X} \to \mathbb{R}$ by $(e \otimes f)(t, x) := t f(x)$, and apply the previous result to the function class $e \otimes \mathcal{G} := \{e \otimes f : f \in \mathcal{G}\} \subseteq B_2(r, P_\xi \otimes P) \cap B_\infty(1)$, where P_ξ denotes the Rademacher distribution on $\{-1, 1\}$. Here the randomness is induced by the independently and identically distributed pairs $(\xi_i, X_i)_{i=1}^n$. For any $f \in \mathcal{G}$ and any ε-bracket $[\underline{f}, \bar{f}]$ containing f, we have that $[e_+ \otimes \underline{f} - e_- \otimes \bar{f}, e_+ \otimes \bar{f} - e_- \otimes \underline{f}]$ is an ε-bracket for $e \otimes f$ in the $L_2(P_\xi \otimes P)$ metric, where $e_+(t) := \max\{e(t), 0\} = \max(t, 0)$ and $e_-(t) := \max(-t, 0)$. It follows that for every $\epsilon > 0$,

$$N_{[\,]}(\varepsilon, e \otimes \mathcal{G}, L_2(P_\xi \otimes P)) \leq N_{[\,]}(\varepsilon, \mathcal{G}, L_2(P)),$$

which proves the claim for the symmetrised empirical process. □

In the next two lemmas, we assume, as in the main text, that P is a distribution on $[0,1]^d$ with Lebesgue density bounded above and below by $M_0 \in [1, \infty)$ and $m_0 \in (0, 1]$ respectively. As in the proof of Proposition 8, let $\mathcal{F}_{d,\downarrow}^+ = \{f : -f \in \mathcal{F}_d, f \geq 0\}$. The following result is used to control the bracketing entropy terms that appear in Lemma 7 when we apply it in the proof of Proposition 8.

LEMMA 8. *There exists a constant $C_d > 0$, depending only on d, such that for any $r, \epsilon > 0$,*

$$\begin{aligned}\log N_{[\,]}\big(\varepsilon, &\mathcal{F}_{d,\downarrow}^+ \cap B_2(r, P) \cap B_\infty(1), \|\cdot\|_{L_2(P)}\big) \\ &\leq C_d \begin{cases} (r/\varepsilon)^2 \frac{M_0}{m_0} \log^2\big(\frac{M_0}{m_0}\big) \log_+^4(1/\varepsilon) \log_+^2\big(\frac{r \log_+(1/\varepsilon)}{\varepsilon}\big) & \text{if } d = 2, \\ (r/\varepsilon)^{2(d-1)} \big(\frac{M_0}{m_0}\big)^{d-1} \log_+^{d^2}(1/\varepsilon) & \text{if } d \geq 3. \end{cases}\end{aligned}$$

PROOF. We first claim that for any $\eta \in (0, 1/4]$,

$$\begin{aligned}\log N_{[\,]}\big(\varepsilon, &\mathcal{F}_{d,\downarrow}^+ \cap B_2(r, P), \|\cdot\|_{L_2(P;[\eta,1]^d)}\big) \\ &\lesssim_d \begin{cases} \big(\frac{r}{\varepsilon}\big)^2 \frac{M_0}{m_0} \log^2\big(\frac{M_0}{m_0}\big) \log^4(1/\eta) \log_+^2\big(\frac{r \log(1/\eta)}{\varepsilon}\big) & \text{if } d = 2, \\ \big(\frac{r}{\varepsilon}\big)^{2(d-1)} \big(\frac{M_0}{m_0}\big)^{d-1} \log^{d^2}(1/\eta) & \text{if } d \geq 3. \end{cases}\end{aligned} \tag{14}$$

By the cone property of $\mathcal{F}_{d,\downarrow}^+$, it suffices to establish the above claim when $r = 1$. We denote by $\mathrm{vol}(S)$ the d-dimensional Lebesgue measure of a measurable set $S \subseteq [0,1]^d$. By Gao and Wellner (2007, Theorem 1.1) and a scaling argument, we have for any $\delta, M > 0$ and any hyperrectangle $A \subseteq [0,1]^d$ that

$$\log N_{[\,]}\big(\delta, \mathcal{F}_{d,\downarrow}^+ \cap B_\infty(M), \|\cdot\|_{L_2(P;A)}\big) \lesssim_d \begin{cases} (\gamma/\delta)^2 \log_+^2(\gamma/\delta) & \text{if } d = 2, \\ (\gamma/\delta)^{2(d-1)} & \text{if } d \geq 3, \end{cases} \tag{15}$$

where $\gamma := M_0^{1/2} M \mathrm{vol}^{1/2}(A)$. We define $m := \lfloor \log_2(1/\eta) \rfloor$ and set $I_\ell := [2^\ell \eta, 2^{\ell+1}\eta] \cap [0,1]$ for each $\ell = 0, \ldots, m$. Then for $\ell_1, \ldots, \ell_d \in \{0, \ldots, m\}$, any $f \in \mathcal{F}_{d,\downarrow}^+ \cap B_2(1,P)$ is uniformly bounded by $\bigl\{m_0 \prod_{j=1}^d (2^{\ell_j}\eta)\bigr\}^{-1/2}$ on the hyperrectangle $\prod_{j=1}^d I_{\ell_j}$. Then by (15) we see that for any $\delta > 0$,

$$\log N_{[\,]}\bigl(\delta, \mathcal{F}_{d,\downarrow}^+ \cap B_2(1,P), \|\cdot\|_{L_2(P;\prod_{j=1}^d I_{\ell_j})}\bigr) \lesssim_d \begin{cases} \delta^{-2}(M_0/m_0)\log^2(\frac{M_0}{m_0})\log_+^2(1/\delta) & \text{if } d=2, \\ \delta^{-2(d-1)}(M_0/m_0)^{d-1} & \text{if } d \geq 3, \end{cases}$$

where we have used the fact that $\log_+(ax) \leq 2\log_+(a)\log_+(x)$ for any $a, x > 0$. Note that these bounds do not depend on η, since the dependence of M and $\mathrm{vol}(A)$ on η is such that it cancels in the expression for γ. Global brackets for $\mathcal{F}_{d,\downarrow}^+ \cap B_2(1)$ on $[\eta,1]^d$ can then be constructed by taking all possible combinations of local brackets on $I_{\ell_1} \times \cdots \times I_{\ell_d}$ for $\ell_1, \ldots, \ell_d \in \{0, \ldots, m\}$. Overall, for any $\varepsilon > 0$, setting $\delta = (m+1)^{-d/2}\varepsilon$ establishes the claim (14) in the case $r = 1$.

We conclude that if we fix any $\varepsilon > 0$, take $\eta = \varepsilon^2/(4d) \wedge 1/4$ and take a single bracket consisting of the constant functions 0 and 1 on $[0,1]^d \setminus [\eta,1]^d$, we have

$$\begin{aligned} \log N_{[\,]}\bigl(\varepsilon, &\mathcal{F}_{d,\downarrow}^+ \cap B_2(r,P) \cap B_\infty(1), \|\cdot\|_{L_2(P)}\bigr) \\ &\leq \log N_{[\,]}\bigl(\varepsilon/2, \mathcal{F}_{d,\downarrow}^+ \cap B_2(r,P), \|\cdot\|_{L_2(P;[\eta,1]^d)}\bigr) \\ &\lesssim_d \begin{cases} (r/\varepsilon)^2 \frac{M_0}{m_0}\log^2(\frac{M_0}{m_0})\log_+^4(1/\varepsilon)\log_+^2\bigl(\frac{r\log_+(1/\varepsilon)}{\varepsilon}\bigr) & \text{if } d=2, \\ (r/\varepsilon)^{2(d-1)}(\frac{M_0}{m_0})^{d-1}\log_+^{d^2}(1/\varepsilon) & \text{if } d \geq 3, \end{cases} \end{aligned}$$

completing the proof. □

For $0 < r < 1$, let F_r be the envelope function of $\mathcal{F}_{d,\downarrow}^+ \cap B_2(r,P) \cap B_\infty(1)$. The lemma below controls the $L_2(P)$ norm of F_r when restricted to strips of the form $I_\ell := [0,1]^{d-1} \times [\frac{\ell-1}{n_1}, \frac{\ell}{n_1}]$ for $\ell = 1, \ldots, n_1$.

LEMMA 9. *For any $r \in (0,1]$ and $\ell = 1, \ldots, n_1$, we have*

$$\int_{I_\ell} F_r^2 \,\mathrm{d}P \leq \frac{7M_0 r^2 \log_+^d(1/r^2)}{m_0 \ell}.$$

PROOF. By monotonicity and the $L_2(P)$ and L_∞ constraints, we have $F_r^2(x_1, \ldots, x_d) \leq \frac{r^2}{m_0 x_1 \cdots x_d} \wedge 1$. We first claim that for any $d \in \mathbb{N}$,

$$\int_{[0,1]^d} \biggl(\frac{t}{x_1 \cdots x_d} \wedge 1\biggr) \,\mathrm{d}x_1 \cdots \mathrm{d}x_d \leq 5t\log_+^d(1/t).$$

To see this, we define $S_d := \big\{(x_1,\ldots,x_d) : \prod_{j=1}^d x_j \geq t\big\}$ and set $a_d := \int_{S_d} \frac{t}{x_1\cdots x_d}\,\mathrm{d}x_1\cdots\mathrm{d}x_d$ and $b_d := \int_{S_d} \mathrm{d}x_1\cdots\mathrm{d}x_d$. By integrating out the last coordinate, we obtain the following relation

$$b_d = \int_{S_{d-1}} \Big(1 - \frac{t}{x_1\cdots x_{d-1}}\Big)\,\mathrm{d}x_1\cdots\mathrm{d}x_{d-1} = b_{d-1} - a_{d-1}. \tag{16}$$

On the other hand, we have by direct computation that

$$\begin{aligned} a_d &= \int_t^1 \cdots \int_{\frac{t}{x_1\cdots x_{d-1}}}^1 \frac{t}{x_1\cdots x_d}\,\mathrm{d}x_d\cdots\mathrm{d}x_1 \\ &\leq a_{d-1}\log(1/t) \leq \cdots \leq a_1 \log^{d-1}(1/t) = t\log^d(1/t). \end{aligned} \tag{17}$$

Combining (16) and (17), we have

$$\begin{aligned} \int_{[0,1]^d} \Big(\frac{t}{x_1\cdots x_d} \wedge 1\Big)\,\mathrm{d}x_1\cdots\mathrm{d}x_d &= a_d + 1 - b_d \\ &\leq \min\{a_d + 1, a_d + a_{d-1} + \cdots + a_1 + 1 - b_1\} \\ &\leq \min\Big\{t\log^d(1/t) + 1, \frac{t\log^{d+1}(1/t)}{\log(1/t) - 1}\Big\} \leq 5t\log_+^d(1/t), \end{aligned}$$

as claimed, where the final inequality follows by considering the cases $t \in [1/e, 1]$, $t \in [1/4, 1/e)$ and $t \in [0, 1/4)$ separately. Consequently, for $\ell = 2, \ldots, n_1$, we have that

$$\begin{aligned} \int_{I_\ell} F_r^2\,\mathrm{d}P &\leq \frac{M_0}{m_0} \int_{(\ell-1)/n_1}^{\ell/n_1} \int_{[0,1]^{d-1}} \Big(\frac{r^2/x_d}{x_1\cdots x_{d-1}} \wedge 1\Big)\,\mathrm{d}x_1\cdots\mathrm{d}x_{d-1}\mathrm{d}x_d \\ &\leq \frac{M_0}{m_0} \int_{(\ell-1)/n_1}^{\ell/n_1} 5(r^2/x_d)\log_+^{d-1}(x_d/r^2)\,\mathrm{d}x_d \\ &\leq \frac{M_0}{m_0} 5r^2 \log_+^{d-1}(1/r^2)\log\big(\ell/(\ell-1)\big) \leq \frac{7M_0 r^2 \log_+^{d-1}(1/r^2)}{m_0 \ell}, \end{aligned}$$

as desired. For the remaining case $\ell = 1$, we have

$$\int_{I_1} F_r^2\,\mathrm{d}P \leq M_0 \int_{[0,1]^d} F_r^2\,\mathrm{d}x_1\cdots\mathrm{d}x_d \leq \frac{5M_0}{m_0} r^2 \log_+^d(1/r^2),$$

which is also of the correct form. □

LEMMA 10. *For any Borel measurable* $f_0 : [0,1]^d \to [-1,1]$ *and any* $a > 2$, *we have* $\mathbb{P}(\|\hat{f}_n - f_0\|_\infty > a) \leq ne^{-(a-2)^2/2}$. *Consequently,*

$$\mathbb{E}\Big(\|\hat{f}_n - f_0\|_\infty^2 \mathbb{1}_{\{\|\hat{f}_n - f_0\|_\infty > a\}}\Big) \leq n\big(a^2 + 2 + 2\sqrt{2\pi}\big)e^{-(a-2)^2/2}.$$

PROOF. Recall that we say $U \subseteq \mathbb{R}^d$ is an *upper set* if whenever $x \in U$ and $x \preceq y$, we have $y \in U$; we say, $L \subseteq \mathbb{R}^d$ is a *lower set* if $-L$ is an upper set. We write $\mathcal{U}$ and $\mathcal{L}$ respectively for the collections of upper and lower sets in $[0,1]^d$. The least squares estimator $\hat{f}_n$ over $\mathcal{F}_d$ then has a well-known min-max representation (Robertson, Wright and Dykstra, 1988, Theorem 1.4.4):

$$\hat{f}_n(X_i) = \min_{L \in \mathcal{L}, L \ni X_i} \max_{U \in \mathcal{U}, U \ni X_i} \overline{Y_{L \cap U}},$$

where $\overline{Y_{L \cap U}}$ denotes the average value of the elements of $\{Y_i : X_i \in L \cap U\}$. Thus we have

$$\|\hat{f}_n\|_\infty = \max_{1 \le i \le n} |\hat{f}_n(X_i)| \le \max_{1 \le i \le n} |Y_i|.$$

Since $Y_i = f_0(X_i) + \epsilon_i$ and $\|f_0\|_\infty \le 1$, we have by a union bound that

$$\mathbb{P}\big(\|\hat{f}_n - f_0\|_\infty \ge t\big) \le n\mathbb{P}(|\epsilon_1| \ge t - 2).$$

The first claim follows using the fact that $\mathbb{P}(\epsilon_1 \ge t) \le \frac{1}{2}e^{-t^2/2}$ for any $t \ge 0$. Moreover, for any $a > 2$,

$$\begin{aligned}
\mathbb{E}\Big(\|\hat{f}_n - f_0\|_\infty^2 \mathbb{1}_{\{\|\hat{f}_n - f_0\|_\infty > a\}}\Big) &= \int_0^\infty 2t\, \mathbb{P}\big(\|\hat{f}_n - f_0\|_\infty \ge \max\{a, t\}\big)\, \mathrm{d}t \\
&\le na^2 \mathbb{P}(|\epsilon_1| \ge a - 2) + n \int_a^\infty 2t\, \mathbb{P}(|\epsilon_1| \ge t - 2)\, \mathrm{d}t \\
&\le n\big(a^2 + 2 + 2\sqrt{2\pi}\big)e^{-(a-2)^2/2},
\end{aligned}$$

as desired. □

LEMMA 11. *If Y is a non-negative random variable such that $(\mathbb{E}Y^p)^{1/p} \le A_1 p + A_2 p^{1/2} + A_3$ for all $p \in [1, \infty)$ and some $A_1, A_2 > 0$, $A_3 \ge 0$, then for every $t \ge 0$,*

$$\mathbb{P}(Y \ge t + eA_3) \le e \exp\bigg(-\min\bigg\{\frac{t}{2eA_1}, \frac{t^2}{4e^2 A_2^2}\bigg\}\bigg).$$

PROOF. Let $s := \min\{t/(2eA_1), t^2/(2eA_2)^2\}$. For values of t such that $s \ge 1$, we have by Markov's inequality that

$$\mathbb{P}(Y \ge t + eA_3) \le \bigg(\frac{A_1 s + A_2 s^{1/2} + A_3}{t + eA_3}\bigg)^s \le e^{-s} \le e^{1-s}.$$

For values of t such that $s < 1$, we trivially have $\mathbb{P}(Y \ge t + eA_3) \le \mathbb{P}(Y \ge t) \le e^{1-s}$, as desired. □

LEMMA 12. *Let X be a non-negative random variable satisfying $X \le b$ almost surely. Then*

$$\mathbb{E}e^X \le \exp\biggl\{\frac{e^b - 1}{b}\mathbb{E}X\biggr\}.$$

PROOF. We have

$$\mathbb{E}e^X = \sum_{r=0}^{\infty} \frac{\mathbb{E}(X^r)}{r!} \le 1 + \sum_{r=1}^{\infty} \frac{b^{r-1}\mathbb{E}X}{r!} = 1 + \frac{\mathbb{E}X}{b}(e^b - 1) \le \exp\biggl\{\frac{e^b - 1}{b}\mathbb{E}X\biggr\},$$

as required. □

REFERENCES

Dilworth, R. P. (1950) A decomposition theorem for partially ordered sets. *Ann. Math.*, **51**, 161–166.

Gao, F. and Wellner, J. A. (2007) Entropy estimate for high-dimensional monotonic functions. *J. Mult. Anal.*, **98**, 1751–1764.

Giné, E., Latała, R and Zinn, J. (2000) Exponential and moment inequalities for U-statistics. In *High dimensional probability, II (Seattle, WA, 1999)*, volume 47 of *Progr. Probab.*, pp. 13–38. Birkhäuser Boston, Boston, MA, 2000.

Han, Q., Wang, T., Chatterjee, S. and Samworth, R. J. (2018) Isotonic regression in general dimensions. *Ann. Statist.*, to appear.

Han, Q. and Wellner, J. A. (2017) Convergence rates of least squares regression estimators with heavy-tailed errors. *Ann. Statist.*, to appear. Available at arXiv:1706.02410.

Ledoux, M. and Talagrand, M. (2013) *Probability in Banach Spaces: Isoperimetry and Processes*. Springer-Verlag, Berlin.

Massart, P. (2000) About the constant in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, **28**, 863–884.

Massart, P. (2007) *Concentration Inequalities and Model Selection*, Springer, Berlin.

Robertson, T., Wright, F. T. and Dykstra, R. L. (1988) *Order Restricted Statistical Inference*. John Wiley & Sons, Ltd., Chichester.

Talagrand, M. (1996) New concentration inequalities in product spaces. *Invent. Math.*, **126**, 505–563.

van der Vaart, A. W. and Wellner, J. A. (1996) *Weak Convergence and Empirical Processes*. Springer, New York.

DEPARTMENT OF STATISTICS, BOX 354322
UNIVERSITY OF WASHINGTON
SEATTLE, WA 98195
UNITED STATES
E-MAIL: royhan@uw.edu

STATISTICAL LABORATORY
WILBERFORCE ROAD
CAMBRIDGE, CB3 0WB
UNITED KINGDOM
E-MAIL: t.wang@statslab.cam.ac.uk
E-MAIL: r.samworth@statslab.cam.ac.uk

ILLINI HALL
725 S WRIGHT ST #101
CHAMPAIGN, IL 61820
UNITED STATES
E-MAIL: sc1706@illinois.edu