Submitted to Bernoulli

Supplement to "The Goldenshluger–Lepski Method for Constrained Least-Squares Estimators over RKHSs"

STEPHEN PAGE [1,*] and STEFFEN GRÜNEWÄLDER [2]

[1] *STOR-i, Lancaster University, Lancaster, LA1 4YF, United Kingdom.*
E-mail: [*]s.page@lancaster.ac.uk

[2] *Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, United Kingdom. E-mail:* s.grunewalder@lancaster.ac.uk

In this supplement, we provide the proofs of the results for the two regression problems in [3], including the majorants, along with some technical results.

Appendix A: Proof of the Regression Results for a Fixed RKHS

We bound the distance between $\hat{h}_r$ and h_r in the $L^2(P_n)$ norm for $r \geq 0$ and $h_r \in rB_H$ to prove the following.

Lemma 22 *Assume (Y) and (H). Let $t \geq 1$ and $A_{1,t} \in \mathcal{F}$ be the set on which*

$$\|\hat{h}_r - h_r\|^2_{L^2(P_n)} \leq \frac{20\|k\|^{1/2}_{\mathsf{diag}}\sigma r t^{1/2}}{n^{1/2}} + 4\|h_r - g\|^2_\infty$$

simultaneously for all $r \geq 0$ and all $h_r \in rB_H$. We have $\mathbb{P}(A_{1,t}) \geq 1 - e^{-t}$.

Proof The result is trivial for $r = 0$. By Lemma 2 of [2], we have

$$\|\hat{h}_r - h_r\|^2_{L^2(P_n)} \leq \frac{4}{n}\sum_{i=1}^n (Y_i - g(X_i))(\hat{h}_r(X_i) - h_r(X_i)) + 4\|h_r - g\|^2_{L^2(P_n)}$$

for all $r > 0$ and all $h_r \in rB_H$. We now bound the right-hand side. We have

$$\|h_r - g\|^2_{L^2(P_n)} \leq \|h_r - g\|^2_\infty.$$

Furthermore,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))(\hat{h}_r(X_i) - h_r(X_i)) \\
&\leq \sup_{f \in 2rB_H} \left| \frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))f(X_i) \right| \\
&= \sup_{f \in 2rB_H} \left| \left\langle \frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))k_{X_i}, f \right\rangle_H \right| \\
&= 2r \left\| \frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))k_{X_i} \right\|_H \\
&= 2r \left(\frac{1}{n^2}\sum_{i,j=1}^{n}(Y_i - g(X_i))(Y_j - g(X_j))k(X_i, X_j) \right)^{1/2}
\end{aligned}$$

by the reproducing kernel property and the Cauchy–Schwarz inequality. Let K be the $n \times n$ matrix with $K_{i,j} = k(X_i, X_j)$ and let ε be the vector of the $Y_i - g(X_i)$. Then

$$\frac{1}{n^2}\sum_{i,j=1}^{n}(Y_i - g(X_i))(Y_j - g(X_j))k(X_i, X_j) = \varepsilon^{\mathsf{T}}(n^{-2}K)\varepsilon.$$

Furthermore, since k is a measurable function on $(S \times S, \mathcal{S} \otimes \mathcal{S})$, we have that $n^{-2}K$ is an $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrix on $(\Omega, \mathcal{F})$ and non-negative-definite. Let a_i for $1 \leq i \leq n$ be the eigenvalues of $n^{-2}K$. Then

$$\max_{1\leq i\leq n} a_i \leq \operatorname{tr}(n^{-2}K) \leq n^{-1}\|k\|_{\mathsf{diag}}$$

and

$$\operatorname{tr}((n^{-2}K)^2) = \|a\|_2^2 \leq \|a\|_1^2 \leq n^{-2}\|k\|_{\mathsf{diag}}^2.$$

Therefore, by Lemma 36 of [2], we have

$$\varepsilon^{\mathsf{T}}(n^{-2}K)\varepsilon \leq \|k\|_{\mathsf{diag}}\sigma^2 n^{-1}(1 + 2t + 2(t^2 + t)^{1/2})$$

and

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - g(X_i))(\hat{h}_r(X_i) - h_r(X_i)) \leq \frac{5\|k\|_{\mathsf{diag}}^{1/2}\sigma r t^{1/2}}{n^{1/2}}$$

with probability at least $1 - e^{-t}$. The result follows. ∎

The following lemma, which is Lemma 25 of [2], is useful for proving Lemma 24.

Lemma 23 *Let $D > 0$ and $A \subseteq L^\infty$ be separable with $\|f\|_\infty \leq D$ for all $f \in A$. Let*

$$Z = \sup_{f \in A} \left| \|f\|^2_{L^2(P_n)} - \|f\|^2_{L^2(P)} \right|.$$

Then, for $t > 0$, we have

$$Z \leq \mathbb{E}(Z) + \left(\frac{2D^4 t}{n} + \frac{4D^2\,\mathbb{E}(Z) t}{n} \right)^{1/2} + \frac{2D^2 t}{3n}$$

with probability at least $1 - e^{-t}$.

We bound the supremum of the difference in the $L^2(P_n)$ norm and the $L^2(P)$ norm over rB_H for $r \geq 0$ to prove the next result. We make use of the notion of a *contraction vanishing at* 0. A function $\varphi : \mathbb{R} \to \mathbb{R}$ is called a contraction vanishing at 0 if it is Lipschitz continuous (with Lipschitz constant 1) and if $\varphi(0) = 0$.

Lemma 24 *Assume (H). Let $t \geq 1$ and $A_{2,t} \in \mathcal{F}$ be the set on which*

$$\sup_{f_1, f_2 \in rB_H} \left| \|Vf_1 - Vf_2\|^2_{L^2(P_n)} - \|Vf_1 - Vf_2\|^2_{L^2(P)} \right| \leq \frac{97\|k\|^{1/2}_{\mathsf{diag}} Crt^{1/2}}{n^{1/2}} + \frac{8\|k\|^{1/2}_{\mathsf{diag}} Crt}{3n}$$

simultaneously for all $r \geq 0$. We have $\mathbb{P}(A_{2,t}) \geq 1 - e^{-t}$.

Proof The result is trivial for $r = 0$. Let

$$Z = \sup_{r > 0} \sup_{f_1, f_2 \in rB_H} \frac{1}{r} \left| \|Vf_1 - Vf_2\|^2_{L^2(P_n)} - \|Vf_1 - Vf_2\|^2_{L^2(P)} \right|.$$

Furthermore, let the ε_i for $1 \leq i \leq n$ be i.i.d. Rademacher random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, independent of the X_i. Lemma 2.3.1 of [8] shows

$$\mathbb{E}(Z) \leq 2\,\mathbb{E} \left(\sup_{r > 0} \sup_{f_1, f_2 \in rB_H} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (r^{-1/2} Vf_1(X_i) - r^{-1/2} Vf_2(X_i))^2 \right| \right)$$

by symmetrisation. Since

$$|Vf_1(X_i) - Vf_2(X_i)| \leq 2C$$

for all $r > 0$ and all $f_1, f_2 \in rB_H$, we find

$$\frac{(r^{-1/2} Vf_1(X_i) - r^{-1/2} Vf_2(X_i))^2}{4C}$$

is a contraction vanishing at 0 as a function of $r^{-1} Vf_1(X_i) - r^{-1} Vf_2(X_i)$ for all $1 \leq i \leq n$. By Theorem 3.2.1 of [1], we have

$$\mathbb{E} \left(\sup_{r > 0} \sup_{f_1, f_2 \in rB_H} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \frac{(r^{-1/2} Vf_1(X_i) - r^{-1/2} Vf_2(X_i))^2}{4C} \right| \middle| X \right)$$

is at most

$$2\,\mathbb{E}\left(\sup_{r>0}\sup_{f_1,f_2\in rB_H}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i(r^{-1}Vf_1(X_i)-r^{-1}Vf_2(X_i))\right|\,\middle|X\right)$$

almost surely. Therefore,

$$\begin{aligned}\mathbb{E}(Z)&\le 16C\,\mathbb{E}\left(\sup_{r>0}\sup_{f_1,f_2\in rB_H}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i(r^{-1}Vf_1(X_i)-r^{-1}Vf_2(X_i))\right|\right)\\&\le 32C\,\mathbb{E}\left(\sup_{r>0}\sup_{f\in rB_H}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i r^{-1}Vf(X_i)\right|\right)\end{aligned}$$

by the triangle inequality. Again, by Theorem 3.2.1 of [1], we have

$$\mathbb{E}(Z)\le 64C\,\mathbb{E}\left(\sup_{r>0}\sup_{f\in rB_H}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i r^{-1}f(X_i)\right|\right)$$

since V is a contraction vanishing at 0. We have

$$\begin{aligned}\sup_{r>0}\sup_{f\in rB_H}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i r^{-1}f(X_i)\right|&=\sup_{r>0}\sup_{f\in rB_H}\left|\left\langle\frac{1}{n}\sum_{i=1}^n\varepsilon_i k_{X_i},r^{-1}f\right\rangle_H\right|\\&=\left\|\frac{1}{n}\sum_{i=1}^n\varepsilon_i k_{X_i}\right\|_H\\&=\left(\frac{1}{n^2}\sum_{i,j=1}^n\varepsilon_i\varepsilon_j k(X_i,X_j)\right)^{1/2}.\end{aligned}$$

by the reproducing kernel property and the Cauchy–Schwarz inequality. By Jensen's inequality, we have

$$\begin{aligned}\mathbb{E}\left(\sup_{r>0}\sup_{f\in rB_H}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i r^{-1}f(X_i)\right|\,\middle|X\right)&\le\left(\frac{1}{n^2}\sum_{i,j=1}^n\operatorname{cov}(\varepsilon_i,\varepsilon_j|X)k(X_i,X_j)\right)^{1/2}\\&=\left(\frac{1}{n^2}\sum_{i=1}^n k(X_i,X_i)\right)^{1/2}\end{aligned}$$

almost surely and again, by Jensen's inequality, we have

$$\mathbb{E}\left(\sup_{r>0}\sup_{f\in rB_H}\left|\frac{1}{n}\sum_{i=1}^n\varepsilon_i r^{-1}f(X_i)\right|\right)\le\left(\frac{\|k\|_{\mathsf{diag}}}{n}\right)^{1/2}.$$

Hence, $\mathbb{E}(Z) \leq 64\|k\|_{\mathsf{diag}}^{1/2} C n^{-1/2}$.

Let

$$A = \left\{ r^{-1/2} V f_1 - r^{-1/2} V f_2 : r > 0 \text{ and } f_1, f_2 \in r B_H \right\}.$$

We have that $(0, \infty)$, the set indexing r, is separable. Furthermore, H is separable and so is separable in $C(S)$ as it can be continuously embedded in $C(S)$ due to its bounded kernel. Therefore, $rB_H \subseteq H$ is separable in $C(S)$ for $r > 0$. Hence, we have that $A \subseteq C(S)$ is separable. Furthermore,

$$\begin{aligned} \left\| r^{-1/2} V f_1 - r^{-1/2} V f_2 \right\|_\infty &\leq \min\left(2 C r^{-1/2}, 2\|k\|_{\mathsf{diag}}^{1/2} r^{1/2}\right) \\ &\leq 2\|k\|_{\mathsf{diag}}^{1/4} C^{1/2} \end{aligned}$$

for all $r > 0$ and all $f_1, f_2 \in rB_H$. The first term in the minimum comes from clipping using V, while the second term comes from the continuous embedding of H in $C(S)$ due to its bounded kernel. By Lemma 23, we have

$$Z \leq \mathbb{E}(Z) + \left(\frac{32\|k\|_{\mathsf{diag}} C^2 t}{n} + \frac{16\|k\|_{\mathsf{diag}}^{1/2} C\, \mathbb{E}(Z) t}{n} \right)^{1/2} + \frac{8\|k\|_{\mathsf{diag}}^{1/2} C t}{3n}$$

with probability at least $1 - e^{-t}$. We have $\mathbb{E}(Z) \leq 64\|k\|_{\mathsf{diag}}^{1/2} C n^{-1/2}$ from above. The result follows. ■

We move the bound on the distance between $V\hat{h}_r$ and Vh_r from the $L^2(P_n)$ norm to the $L^2(P)$ norm for $r \geq 0$ and $h_r \in rB_H$.

Corollary 25 *Assume (Y) and (H). Let $t \geq 1$ and recall the definitions of $A_{1,t}$ and $A_{2,t}$ from Lemmas 22 and 24. On the set $A_{1,t} \cap A_{2,t} \in \mathcal{F}$, for which $\mathbb{P}(A_{1,t} \cap A_{2,t}) \geq 1 - 2e^{-t}$, we have*

$$\|V\hat{h}_r - Vh_r\|_{L^2(P)}^2 \leq \frac{\|k\|_{\mathsf{diag}}^{1/2}(97C + 20\sigma) r t^{1/2}}{n^{1/2}} + \frac{8\|k\|_{\mathsf{diag}}^{1/2} C r t}{3n} + 4\|h_r - g\|_\infty^2$$

simultaneously for all $r \geq 0$ and all $h_r \in rB_H$.

Proof By Lemma 22, we have

$$\|\hat{h}_r - h_r\|_{L^2(P_n)}^2 \leq \frac{20\|k\|_{\mathsf{diag}}^{1/2} \sigma r t^{1/2}}{n^{1/2}} + 4\|h_r - g\|_\infty^2$$

for all $r \geq 0$ and all $h_r \in rB_H$, so

$$\|V\hat{h}_r - Vh_r\|_{L^2(P_n)}^2 \leq \frac{20\|k\|_{\mathsf{diag}}^{1/2} \sigma r t^{1/2}}{n^{1/2}} + 4\|h_r - g\|_\infty^2.$$

Since $\hat{h}_r, h_r \in rB_H$, by Lemma 24 we have

$$\begin{aligned}
&\|V\hat{h}_r - Vh_r\|^2_{L^2(P)} - \|V\hat{h}_r - Vh_r\|^2_{L^2(P_n)} \\
\le &\sup_{f_1, f_2 \in rB_H} \left| \|Vf_1 - Vf_2\|^2_{L^2(P_n)} - \|Vf_1 - Vf_2\|^2_{L^2(P)} \right| \\
\le &\frac{97\|k\|^{1/2}_{\mathsf{diag}} Crt^{1/2}}{n^{1/2}} + \frac{8\|k\|^{1/2}_{\mathsf{diag}} Crt}{3n}.
\end{aligned}$$

The result follows. ■

We assume $(g1)$ to bound the distance between $V\hat{h}_r$ and g in the $L^2(P)$ norm for $r \ge 0$ and prove Theorem 1.

Proof of Theorem 1 Note that $Vg = g$. We have

$$\begin{aligned}
\|V\hat{h}_r - g\|^2_{L^2(P)} &\le \left(\|V\hat{h}_r - Vh_r\|_{L^2(P)} + \|Vh_r - g\|_{L^2(P)} \right)^2 \\
&\le 2\|V\hat{h}_r - Vh_r\|^2_{L^2(P)} + 2\|Vh_r - g\|^2_{L^2(P)} \\
&\le 2\|V\hat{h}_r - Vh_r\|^2_{L^2(P)} + 2\|h_r - g\|^2_{L^2(P)}
\end{aligned}$$

for all $r \ge 0$ and all $h_r \in rB_H$. By Corollary 25, we have

$$\|V\hat{h}_r - Vh_r\|^2_{L^2(P)} \le \frac{\|k\|^{1/2}_{\mathsf{diag}}(97C + 20\sigma)rt^{1/2}}{n^{1/2}} + \frac{8\|k\|^{1/2}_{\mathsf{diag}} Crt}{3n} + 4\|h_r - g\|^2_\infty.$$

Hence,

$$\|V\hat{h}_r - g\|^2_{L^2(P)} \le \frac{2\|k\|^{1/2}_{\mathsf{diag}}(97C + 20\sigma)rt^{1/2}}{n^{1/2}} + \frac{16\|k\|^{1/2}_{\mathsf{diag}} Crt}{3n} + 10\|h_r - g\|^2_\infty.$$

Taking an infimum over $h_r \in rB_H$ proves the result. ■

Appendix B: Proof of the Regression Results for a Collection of RKHSs

Lemma 26 *Assume ($\mathcal{K}1$). We have that $\hat{h}_{k,r}$ is an $(C(S), \mathcal{B}(C(S)))$-valued measurable function on $(\Omega \times \mathcal{K} \times [0,\infty), \mathcal{F} \otimes \mathcal{B}(\mathcal{K}) \otimes \mathcal{B}([0,\infty)))$, where k varies in $\mathcal{K}$ and r varies in $[0,\infty)$.*

Proof Let K be the $n \times n$ symmetric matrix with $K_{i,j} = k(X_i, X_j)$ for $k \in \mathcal{K}$. Then K is a continuous function of k and X, hence it is an $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrix on $(\Omega \times \mathcal{K}, \mathcal{F} \otimes \mathcal{B}(\mathcal{K}))$, where k varies in $\mathcal{K}$. By Lemma 39, there exist an orthogonal matrix A and a diagonal matrix D which are both $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrices on $(\Omega\times\mathcal{K},\mathcal{F}\otimes\mathcal{B}(\mathcal{K}))$ such that $K=ADA^\mathsf{T}$. Since K is non-negative definite, the diagonal entries of D are non-negative, and we may assume that they are non-increasing. Let $m=\operatorname{rk}K$, which is measurable on $(\Omega\times\mathcal{K},\mathcal{F}\otimes\mathcal{B}(\mathcal{K}))$. For $r>0$, if

$$r^2<\sum_{i=1}^m D_{i,i}^{-1}(A^\mathsf{T}Y)_i^2,$$

then define $\mu(r)>0$ by

$$\sum_{i=1}^m \frac{D_{i,i}}{(D_{i,i}+n\mu(r))^2}(A^\mathsf{T}Y)_i^2=r^2.$$

Otherwise, let $\mu(r)=0$. Let $a\in\mathbb{R}^n$ be defined by

$$(A^\mathsf{T}a)_i=(D_{i,i}+n\mu(r))^{-1}(A^\mathsf{T}Y)_i$$

for $1\le i\le m$ and $(A^\mathsf{T}a)_i=0$ for $m+1\le i\le n$, noting that A^T has the inverse A since it is an orthogonal matrix. By Lemma 3 of [2],

$$\hat{h}_{k,r}=\sum_{i=1}^n a_i k_{X_i}$$

for $r>0$ and $\hat{h}_{k,0}=0$ for $k\in\mathcal{K}$.

Since $\mu(r)>0$ is strictly decreasing for

$$r^2<\sum_{i=1}^m D_{i,i}^{-1}(A^\mathsf{T}Y)_i^2$$

and $\mu(r)=0$ otherwise, we find

$$\{\mu(r)\le\mu\}=\left\{\sum_{i=1}^m \frac{D_{i,i}}{(D_{i,i}+n\mu)^2}(A^\mathsf{T}Y)_i^2\le r^2\right\}$$

for $\mu\in[0,\infty)$. Therefore, $\mu(r)$ is measurable on $(\Omega\times\mathcal{K}\times[0,\infty),\mathcal{F}\otimes\mathcal{B}(\mathcal{K})\otimes\mathcal{B}([0,\infty)))$, where k varies in $\mathcal{K}$ and r varies in $[0,\infty)$. Hence, the a above with $\mu=\mu(r)$ for $r>0$ is measurable on $(\Omega\times\mathcal{K}\times[0,\infty),\mathcal{F}\otimes\mathcal{B}(\mathcal{K})\otimes\mathcal{B}([0,\infty)))$. By Lemma 4.29 of [7], $\Phi_k:S\to H_k$ by $\Phi_k(x)=k_x$ is continuous for all $k\in\mathcal{K}$. Hence, $\Phi:\mathcal{K}\times S\to C(S)$ by $\Phi(k,x)=k_x$ is continuous and k_{X_i} for $1\le i\le n$ are $(C(S),\mathcal{B}(C(S)))$-valued measurable functions on $(\Omega\times\mathcal{K},\mathcal{F}\otimes\mathcal{B}(\mathcal{K}))$. Together, these show that $\hat{h}_{k,r}$ is an $(C(S),\mathcal{B}(C(S)))$-valued measurable function on $(\Omega\times\mathcal{K}\times[0,\infty),\mathcal{F}\otimes\mathcal{B}(\mathcal{K})\otimes\mathcal{B}([0,\infty)))$, where k varies in $\mathcal{K}$ and r varies in $[0,\infty)$, recalling that $\hat{h}_{k,0}=0$. ■

Let $\psi_1(x) = \exp(|x|) - 1$ for $x \in \mathbb{R}$ and

$$\|Z\|_{\psi_1} = \inf\{a \in (0, \infty) : \mathbb{E}(\psi_1(Z/a)) \leq 1\}$$

for any random variable Z on $(\Omega, \mathcal{F})$. Note that this infimum is attained by the monotone convergence theorem, and $\|Z\|_{\psi_1}$ increases as $|Z|$ increases pointwise. Let L^{ψ_1} be the set of random variables Z on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\|Z\|_{\psi_1} < \infty$. We have that $(L^{\psi_1}, \|\cdot\|_{\psi_1})$ is a Banach space known as an Orlicz space (see [5]).

Lemma 27 *Let $Z \in L^{\psi_1}$. We have*

$$\mathbb{E}(|Z|) \leq (\log 2)\|Z\|_{\psi_1}.$$

Let $t \geq 0$. We have

$$|Z| \leq \|Z\|_{\psi_1}(\log 2 + t)$$

with probability at least $1 - e^{-t}$.

Proof We have $\mathbb{E}(\exp(|Z|/\|Z\|_{\psi_1})) \leq 2$. The first result follows from Jensen's inequality. The second result follows from Chernoff bounding. ■

For $m \times n$ matrices U and V, define $U \circ V$ to be the $m \times n$ matrix with

$$(U \circ V)_{i,j} = U_{i,j} V_{i,j}.$$

Recall that

$$\begin{aligned} \mathcal{L} &= \{k/\|k\|_{\mathsf{diag}} : k \in \mathcal{K}\} \cup \{0\}, \\ D &= \sup_{f_1, f_2 \in \mathcal{L}} \|f_1 - f_2\|_\infty \leq 2, \\ J &= \left(162 \int_0^{D/2} \log(2N(a, \mathcal{L}, \|\cdot\|_\infty)) da + 1\right)^{1/2}. \end{aligned}$$

The following lemma is useful for proving Lemma 29.

Lemma 28 *Assume ($\mathcal{K}1$). Let the ε_i for $1 \leq i \leq n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ such that (X_i, ε_i) are i.i.d. and ε_i is σ^2-subgaussian given X_i. Let*

$$W(f) = \frac{1}{n^2} \sum_{i,j=1}^n \varepsilon_i \varepsilon_j f(X_i, X_j)$$

for $f \in \mathcal{L}$. We have

$$\left\| \sup_{f \in \mathcal{L}} W(f) \right\|_{\psi_1} \leq \frac{4J^2\sigma^2}{n}.$$

Proof Let F be the $n \times n$ matrix with $F_{i,j} = f(X_i, X_j)$, where F varies with $f \in \mathcal{L}$. Note that F is an $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrix on $(\Omega, \mathcal{F})$. Then $W(f) = n^{-2}\varepsilon^{\mathsf{T}}F\varepsilon$. Let $Z(f) = n^{-2}\varepsilon^{\mathsf{T}}(F - I \circ F)\varepsilon$ for $f \in \mathcal{L}$. Note that Z is continuous in f. We have

$$\|Z(f_1) - Z(f_2)\|_{\psi_1} \le 36\sigma^2 n^{-1}\|f_1 - f_2\|_\infty$$

for $f_1, f_2 \in \mathcal{L}$: let $F^{(1)}$ and $F^{(2)}$ be the matrices corresponding to f_1 and f_2. Observe that $\|Z(f_1) - Z(f_2)\|_{\psi_1} \le a$ by Lemma 41 when a is chosen such that

$$n^2 a = 2^{7/2}(\log 2)\sigma^2(\operatorname{tr}(F^{(1)} - F^{(2)})^2)^{1/2}/\log(5/4) \le 36\sigma^2 n\|f_1 - f_2\|_\infty.$$

Let $d(f_1, f_2) = 36\sigma^2 n^{-1}\|f_1 - f_2\|_\infty$ for $f_1, f_2 \in \mathcal{L}$ and

$$D_d = \sup_{f_1, f_2 \in \mathcal{L}} d(f_1, f_2).$$

By Lemma 38 with $M = \mathcal{L}$ and $s_0 = 0$, we find

$$\begin{aligned}\left\|\sup_{f\in\mathcal{L}}|Z(f)|\right\|_{\psi_1} &\le 18\int_0^{D_d/2}\log(2N(a, \mathcal{L}, d))da\\ &= \frac{648\sigma^2}{n}\int_0^{D/2}\log(2N(a, \mathcal{L}, \|\cdot\|_\infty))da.\end{aligned}$$

Hence,

$$\left\|\sup_{f\in\mathcal{L}}W(f)\right\|_{\psi_1} \le \left\|n^{-2}\sup_{f\in\mathcal{L}}\varepsilon^{\mathsf{T}}(I\circ F)\varepsilon\right\|_{\psi_1} + \frac{648\sigma^2}{n}\int_0^{D/2}\log(2N(a, \mathcal{L}, \|\cdot\|_\infty))da.$$

We have

$$n^{-2}\sup_{f\in\mathcal{L}}\varepsilon^{\mathsf{T}}(I\circ F)\varepsilon \le n^{-2}\varepsilon^{\mathsf{T}}\varepsilon,$$

noting that $F_{i,i} \in [0, 1]$ for $1 \le i \le n$ and $f \in \mathcal{L}$. Let δ_i for $1 \le i \le n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent of each other and the ε_i, with $\delta_i \sim \mathrm{N}(0, \sigma^2)$. Lemma 35 of [2] shows

$$\begin{aligned}\mathbb{E}\left(\exp\left(n^{-2}t\sup_{f\in\mathcal{L}}\varepsilon^{\mathsf{T}}(I\circ F)\varepsilon\right)\right) &\le \mathbb{E}\left(\exp\left(n^{-2}t\varepsilon^{\mathsf{T}}\varepsilon\right)\right)\\ &\le \mathbb{E}\left(\exp\left(n^{-2}t\delta^{\mathsf{T}}\delta\right)\right)\\ &= \prod_{i=1}^n\left(1 - 2\sigma^2 n^{-2}t\right)^{-1/2}\end{aligned}$$

for $0 \le 2\sigma^2 n^{-2} t < 1$ by computing the moment generating function of the δ_i^2. We have that $(1-x)^{-1/2} \le \exp(x)$ for $x \in [0, 1/2]$, so

$$\mathbb{E}\left(\exp\left(n^{-2} t \sup_{f \in \mathcal{L}} \varepsilon^{\mathsf{T}} (I \circ F) \varepsilon\right)\right) \le \prod_{i=1}^{n} \exp\left(2\sigma^2 n^{-2} t\right) = \exp\left(2\sigma^2 n^{-1} t\right)$$

for $0 \le 4\sigma^2 n^{-2} t \le 1$. This bound is at most 2 and valid for

$$t \le \min\left(\frac{n^2}{4\sigma^2}, \frac{(\log 2) n}{2\sigma^2}\right).$$

Hence,

$$\left\| n^{-2} \sup_{f \in \mathcal{L}} \varepsilon^{\mathsf{T}} (I \circ F) \varepsilon \right\|_{\psi_1} \le \max\left(\frac{4\sigma^2}{n^2}, \frac{2\sigma^2}{(\log 2) n}\right) \le \frac{4\sigma^2}{n}$$

and

$$\left\| \sup_{f \in \mathcal{L}} W(f) \right\|_{\psi_1} \le \frac{648\sigma^2}{n} \int_0^{D/2} \log(2N(a, \mathcal{L}, \|\cdot\|_\infty)) da + \frac{4\sigma^2}{n}.$$

The result follows. ∎

We bound the distance between $\hat{h}_{k,r}$ and $h_{k,r}$ in the $L^2(P_n)$ norm for $k \in \mathcal{K}$, $r \ge 0$ and $h_{k,r} \in r B_k$ to prove the following Lemma.

Lemma 29 *Assume (Y) and ($\mathcal{K}$1). Let $t \ge 1$. There exists a set $A_{3,t} \in \mathcal{F}$ with $\mathbb{P}(A_{3,t}) \ge 1 - e^{-t}$ on which*

$$\|\hat{h}_{k,r} - h_{k,r}\|_{L^2(P_n)}^2 \le \frac{21 J \|k\|_{\mathsf{diag}}^{1/2} \sigma r t^{1/2}}{n^{1/2}} + 4\|h_{k,r} - g\|_\infty^2$$

simultaneously for all $k \in \mathcal{K}$, all $r \ge 0$ and all $h_{k,r} \in r B_k$.

Proof The result is trivial for $r = 0$. By Lemma 2 of [2], we have

$$\|\hat{h}_{k,r} - h_{k,r}\|_{L^2(P_n)}^2 \le \frac{4}{n} \sum_{i=1}^{n} (Y_i - g(X_i))(\hat{h}_{k,r}(X_i) - h_{k,r}(X_i)) + 4\|h_{k,r} - g\|_{L^2(P_n)}^2$$

for all $k \in \mathcal{K}$, all $r > 0$ and all $h_{k,r} \in r B_k$. We now bound the right-hand side. We have

$$\|h_{k,r} - g\|_{L^2(P_n)}^2 \le \|h_{k,r} - g\|_\infty^2.$$

Furthermore,

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} (Y_i - g(X_i))(\hat{h}_{k,r}(X_i) - h_{k,r}(X_i)) \\ &\le \sup_{f \in 2r B_k} \left| \frac{1}{n} \sum_{i=1}^{n} (Y_i - g(X_i)) f(X_i) \right| \end{aligned}$$

$$= \sup_{f \in 2rB_k} \left| \left\langle \frac{1}{n} \sum_{i=1}^{n} (Y_i - g(X_i)) k_{X_i}, f \right\rangle_k \right|$$
$$= 2r \left\| \frac{1}{n} \sum_{i=1}^{n} (Y_i - g(X_i)) k_{X_i} \right\|_k$$
$$= 2r \left(\frac{1}{n^2} \sum_{i,j=1}^{n} (Y_i - g(X_i))(Y_j - g(X_j)) k(X_i, X_j) \right)^{1/2}$$

by the reproducing kernel property and the Cauchy–Schwarz inequality. Let

$$Z = \sup_{k \in \mathcal{K}} \left(\frac{1}{\|k\|_{\mathsf{diag}} n^2} \sum_{i,j=1}^{n} (Y_i - g(X_i))(Y_j - g(X_j)) k(X_i, X_j) \right).$$

By Lemma 28 with $\varepsilon_i = Y_i - g(X_i)$, we have $\|Z\|_{\psi_1} \leq 4J^2\sigma^2 n^{-1}$. By Lemma 27, we have $Z \leq 4J^2\sigma^2(\log 2 + t)n^{-1}$ with probability at least $1 - e^{-t}$. The result follows. ■

The following lemma is useful for proving Lemma 31.

Lemma 30 *Let*

$$A = \left\{ \|k\|_{\mathsf{diag}}^{-1/4} r^{-1/2} V f_1 - \|k\|_{\mathsf{diag}}^{-1/4} r^{-1/2} V f_2 : k \in \mathcal{K}, r > 0 \text{ and } f_1, f_2 \in rB_k \right\}.$$

Then A is separable as a subset of $C(S)$.

Proof By Theorem 4.21 of [7], we have that

$$\left\{ \sum_{i=1}^{m} a_i k_{s_i} : m \geq 1 \text{ and } a_i \in \mathbb{R}, s_i \in S \text{ for } 1 \leq i \leq m \right\}$$

is dense in H_k for $k \in \mathcal{K}$. Hence,

$$\left\{ \sum_{i=1}^{m} a_i k_{s_i} : m \geq 1 \text{ and } a_i \in \mathbb{R}, s_i \in S \text{ for } 1 \leq i \leq m \text{ with } \sum_{i,j=1}^{m} a_i a_j k(s_i, s_j) \leq r^2 \right\}$$

is dense in $rB_k \subseteq H_k$ for $k \in \mathcal{K}$ and $r > 0$. Since S is separable, it has a countable dense subset S_0. Let $D_{k,r}$ be

$$\left\{ \sum_{i=1}^{m} a_i k_{s_i} : m \geq 1 \text{ and } a_i \in \mathbb{Q}, s_i \in S_0 \text{ for } 1 \leq i \leq m \text{ with } \sum_{i,j=1}^{m} a_i a_j k(s_i, s_j) \leq r^2 \right\}$$

for $k \in \mathcal{K}$ and $r > 0$. Since the function $\Phi_k : S \to H_k$ by $\Phi_k(x) = k_x$ is continuous by Lemma 4.29 of [7], we have that $D_{k,r}$ is dense in $rB_k \subseteq H_k$ by suitable choices for $a_i \in \mathbb{Q}$

for $1 \leq i \leq m$. Since k is bounded for all $k \in \mathcal{K}$, as subsets of $C(S)$ we have that $D_{k,r}$ is dense in rB_k and

$$A = \mathrm{cl}\left(\left\{\|k\|_{\mathsf{diag}}^{-1/4} r^{-1/2}(Vf_1 - Vf_2) : k \in \mathcal{K}, r > 0 \text{ and } f_1, f_2 \in D_{k,r}\right\}\right).$$

Since $(\mathcal{K}, \|\cdot\|_\infty)$ is separable, it has a countable dense subset $\mathcal{K}_0$. Hence,

$$A = \mathrm{cl}\left(\left\{\|k\|_{\mathsf{diag}}^{-1/4} r^{-1/2}(Vf_1 - Vf_2) : k \in \mathcal{K}_0, r \in (0,\infty) \cap \mathbb{Q} \text{ and } f_1, f_2 \in D_{k,r}\right\}\right)$$

by suitable choices for $r \in (0,\infty) \cap \mathbb{Q}$. The result follows. ∎

We bound the supremum of the difference in the $L^2(P_n)$ norm and the $L^2(P)$ norm over rB_k for $k \in \mathcal{K}$ and $r \geq 0$ to prove the following Lemma.

Lemma 31 *Assume ($\mathcal{K}$1). Let $t \geq 1$ and $A_{4,t} \in \mathcal{F}$ be the set on which*

$$\sup_{f_1, f_2 \in rB_k} \left| \|Vf_1 - Vf_2\|_{L^2(P_n)}^2 - \|Vf_1 - Vf_2\|_{L^2(P)}^2 \right| \leq \frac{151J\|k\|_{\mathsf{diag}}^{1/2} Crt^{1/2}}{n^{1/2}} + \frac{8\|k\|_{\mathsf{diag}}^{1/2} Crt}{3n}$$

simultaneously for all $k \in \mathcal{K}$ and all $r \geq 0$. We have $\mathbb{P}(A_{4,t}) \geq 1 - e^{-t}$.

Proof The result is trivial for $r = 0$. Let

$$Z = \sup_{k \in \mathcal{K}} \sup_{r>0} \sup_{f_1, f_2 \in rB_k} \|k\|_{\mathsf{diag}}^{-1/2} r^{-1} \left| \|Vf_1 - Vf_2\|_{L^2(P_n)}^2 - \|Vf_1 - Vf_2\|_{L^2(P)}^2 \right|.$$

We have that Z is a random variable by Lemma 30. Furthermore, let the ε_i for $1 \leq i \leq n$ be i.i.d. Rademacher random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, independent of the X_i. Lemma 2.3.1 of [8] shows

$$\mathbb{E}(Z) \leq 2\,\mathbb{E}\left(\sup_{k \in \mathcal{K}} \sup_{r>0} \sup_{f_1, f_2 \in rB_k} \|k\|_{\mathsf{diag}}^{-1/2} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i (r^{-1/2} Vf_1(X_i) - r^{-1/2} Vf_2(X_i))^2 \right| \right)$$

by symmetrisation. Since

$$|Vf_1(X_i) - Vf_2(X_i)| \leq 2C$$

for all $k \in \mathcal{K}$, all $r > 0$ and all $f_1, f_2 \in rB_k$, we find

$$\frac{(r^{-1/2} Vf_1(X_i) - r^{-1/2} Vf_2(X_i))^2}{4C}$$

is a contraction vanishing at 0 as a function of $r^{-1}Vf_1(X_i) - r^{-1}Vf_2(X_i)$ for all $1 \leq i \leq n$. By Theorem 3.2.1 of [1], we have

$$\mathbb{E}\left(\sup_{k \in \mathcal{K}} \sup_{r>0} \sup_{f_1, f_2 \in rB_k} \|k\|_{\mathsf{diag}}^{-1/2} \left| \frac{1}{n} \sum_{i=1}^n \varepsilon_i \frac{(r^{-1/2} Vf_1(X_i) - r^{-1/2} Vf_2(X_i))^2}{4C} \right| \,\middle|\, X \right)$$

is at most

$$2\,\mathbb{E}\left(\sup_{k\in\mathcal{K}}\sup_{r>0}\sup_{f_1,f_2\in rB_k}\|k\|_{\mathsf{diag}}^{-1/2}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(r^{-1}Vf_1(X_i)-r^{-1}Vf_2(X_i))\right|\,\middle|\,X\right)$$

almost surely. Therefore,

$$\begin{aligned}\mathbb{E}(Z)&\le 16C\,\mathbb{E}\left(\sup_{k\in\mathcal{K}}\sup_{r>0}\sup_{f_1,f_2\in rB_k}\|k\|_{\mathsf{diag}}^{-1/2}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i(r^{-1}Vf_1(X_i)-r^{-1}Vf_2(X_i))\right|\right)\\&\le 32C\,\mathbb{E}\left(\sup_{k\in\mathcal{K}}\sup_{r>0}\sup_{f\in rB_k}\|k\|_{\mathsf{diag}}^{-1/2}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i r^{-1}Vf(X_i)\right|\right)\end{aligned}$$

by the triangle inequality. Again, by Theorem 3.2.1 of [1], we have

$$\mathbb{E}(Z)\le 64C\,\mathbb{E}\left(\sup_{k\in\mathcal{K}}\sup_{r>0}\sup_{f\in rB_k}\|k\|_{\mathsf{diag}}^{-1/2}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i r^{-1}f(X_i)\right|\right)$$

since V is a contraction vanishing at 0. We have

$$\begin{aligned}&\sup_{k\in\mathcal{K}}\sup_{r>0}\sup_{f\in rB_k}\|k\|_{\mathsf{diag}}^{-1/2}\left|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i r^{-1}f(X_i)\right|\\&=\sup_{k\in\mathcal{K}}\sup_{r>0}\sup_{f\in rB_k}\|k\|_{\mathsf{diag}}^{-1/2}\left|\left\langle\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i k_{X_i},r^{-1}f\right\rangle_k\right|\\&=\sup_{k\in\mathcal{K}}\|k\|_{\mathsf{diag}}^{-1/2}\left\|\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i k_{X_i}\right\|_k\\&=\sup_{k\in\mathcal{K}}\|k\|_{\mathsf{diag}}^{-1/2}\left(\frac{1}{n^2}\sum_{i,j=1}^{n}\varepsilon_i\varepsilon_j k(X_i,X_j)\right)^{1/2}\end{aligned}$$

by the reproducing kernel property and the Cauchy–Schwarz inequality. By Lemma 28 with $\sigma^2=1$, Lemma 27 and Jensen's inequality, we have $\mathbb{E}(Z)\le 107JCn^{-1/2}$.

Let

$$A=\left\{\|k\|_{\mathsf{diag}}^{-1/4}r^{-1/2}Vf_1-\|k\|_{\mathsf{diag}}^{-1/4}r^{-1/2}Vf_2 : k\in\mathcal{K}, r>0\text{ and }f_1,f_2\in rB_k\right\}.$$

We have that $A\subseteq C(S)$ is separable by Lemma 30. Furthermore,

$$\begin{aligned}\left\|\|k\|_{\mathsf{diag}}^{-1/4}r^{-1/2}Vf_1-\|k\|_{\mathsf{diag}}^{-1/4}r^{-1/2}Vf_2\right\|_\infty&\le\min\left(2C\|k\|_{\mathsf{diag}}^{-1/4}r^{-1/2},2\|k\|_{\mathsf{diag}}^{1/4}r^{1/2}\right)\\&\le 2C^{1/2}\end{aligned}$$

for all $k \in \mathcal{K}$, all $r > 0$ and all $f_1, f_2 \in rB_k$. By Lemma 23, we have

$$Z \leq \mathbb{E}(Z) + \left(\frac{32C^2 t}{n} + \frac{16C\,\mathbb{E}(Z)t}{n} \right)^{1/2} + \frac{8Ct}{3n}$$

with probability at least $1 - e^{-t}$. We have $\mathbb{E}(Z) \leq 107JCn^{-1/2}$ from above. The result follows. ■

We move the bound on the distance between $V\hat{h}_{k,r}$ and $Vh_{k,r}$ from the $L^2(P_n)$ norm to the $L^2(P)$ norm for $k \in \mathcal{K}$, $r \geq 0$ and $h_{k,r} \in rB_k$.

Corollary 32 *Assume (Y) and ($\mathcal{K}$1). Let $t \geq 1$ and recall the definitions of $A_{3,t}$ and $A_{4,t}$ from Lemmas 29 and 31. On the set $A_{3,t} \cap A_{4,t} \in \mathcal{F}$, for which $\mathbb{P}(A_{3,t} \cap A_{4,t}) \geq 1 - 2e^{-t}$, we have*

$$\|V\hat{h}_{k,r} - Vh_{k,r}\|^2_{L^2(P)} \leq \frac{J\|k\|^{1/2}_{\mathsf{diag}}(151C + 21\sigma)rt^{1/2}}{n^{1/2}} + \frac{8\|k\|^{1/2}_{\mathsf{diag}}Crt}{3n} + 4\|h_{k,r} - g\|^2_\infty$$

simultaneously for all $k \in \mathcal{K}$, all $r \geq 0$ and all $h_{k,r} \in rB_k$.

Proof By Lemma 29, we have

$$\|\hat{h}_{k,r} - h_{k,r}\|^2_{L^2(P_n)} \leq \frac{21J\|k\|^{1/2}_{\mathsf{diag}}\sigma rt^{1/2}}{n^{1/2}} + 4\|h_{k,r} - g\|^2_\infty$$

for all $k \in \mathcal{K}$, all $r \geq 0$ and all $h_{k,r} \in rB_k$, so

$$\|V\hat{h}_{k,r} - Vh_{k,r}\|^2_{L^2(P_n)} \leq \frac{21J\|k\|^{1/2}_{\mathsf{diag}}\sigma rt^{1/2}}{n^{1/2}} + 4\|h_{k,r} - g\|^2_\infty.$$

Since $\hat{h}_{k,r}, h_{k,r} \in rB_k$, by Lemma 31 we have

$$\begin{aligned}
&\|V\hat{h}_{k,r} - Vh_{k,r}\|^2_{L^2(P)} - \|V\hat{h}_{k,r} - Vh_{k,r}\|^2_{L^2(P_n)} \\
&\leq \sup_{f_1, f_2 \in rB_k} \left| \|Vf_1 - Vf_2\|^2_{L^2(P_n)} - \|Vf_1 - Vf_2\|^2_{L^2(P)} \right| \\
&\leq \frac{151J\|k\|^{1/2}_{\mathsf{diag}}Crt^{1/2}}{n^{1/2}} + \frac{8\|k\|^{1/2}_{\mathsf{diag}}Crt}{3n}.
\end{aligned}$$

The result follows. ■

We assume $(g1)$ to bound the distance between $V\hat{h}_{k,r}$ and g in the $L^2(P)$ norm for $k \in \mathcal{K}$ and $r \geq 0$ and prove Theorem 6.

Proof of Theorem 6 Note that $Vg = g$. We have

$$\begin{aligned}\|V\hat{h}_{k,r} - g\|^2_{L^2(P)} &\leq \left(\|V\hat{h}_{k,r} - Vh_{k,r}\|_{L^2(P)} + \|Vh_{k,r} - g\|_{L^2(P)}\right)^2 \\ &\leq 2\|V\hat{h}_{k,r} - Vh_{k,r}\|^2_{L^2(P)} + 2\|Vh_{k,r} - g\|^2_{L^2(P)} \\ &\leq 2\|V\hat{h}_{k,r} - Vh_{k,r}\|^2_{L^2(P)} + 2\|h_{k,r} - g\|^2_{L^2(P)}\end{aligned}$$

for all $k \in \mathcal{K}$, all $r \geq 0$ and all $h_{k,r} \in rB_k$. By Corollary 32, we have

$$\|V\hat{h}_{k,r} - Vh_{k,r}\|^2_{L^2(P)} \leq \frac{J\|k\|^{1/2}_{\mathsf{diag}}(151C + 21\sigma)rt^{1/2}}{n^{1/2}} + \frac{8\|k\|^{1/2}_{\mathsf{diag}}Crt}{3n} + 4\|h_{k,r} - g\|^2_\infty.$$

Hence,

$$\|V\hat{h}_{k,r} - g\|^2_{L^2(P)} \leq \frac{2J\|k\|^{1/2}_{\mathsf{diag}}(151C + 21\sigma)rt^{1/2}}{n^{1/2}} + \frac{16\|k\|^{1/2}_{\mathsf{diag}}Crt}{3n} + 10\|h_{k,r} - g\|^2_\infty.$$

Taking an infimum over $h_{k,r} \in rB_k$ proves the result. ■

Appendix C: Covering Numbers for Gaussian Kernels

Recall that

$$\mathcal{L} = \left\{f_\gamma(x_1, x_2) = \exp\left(-\|x_1 - x_2\|_2^2/\gamma^2\right) : \gamma \in \Gamma \text{ and } x_1, x_2 \in S\right\} \cup \{0\}.$$

for $\Gamma \subseteq [u, v]$ non-empty for $v \geq u > 0$. We prove a continuity result about the function $F : \Gamma \to \mathcal{L} \setminus \{0\}$ by $F(\gamma) = f_\gamma$. We also bound the covering numbers of $\mathcal{L}$.

Lemma 33 *Assume ($\mathcal{K}2$). Let $\gamma, \eta \in \Gamma$. We have*

$$\|f_\gamma - f_\eta\|_\infty \leq \frac{(\gamma^2 - \eta^2)^{1/2}}{\gamma \vee \eta}.$$

For $a \in (0, 1)$, we have $N(a, \mathcal{L}, \|\cdot\|_\infty) \leq \log(v/u)a^{-2} + 2$. For $a \geq 1$, we have $N(a, \mathcal{L}, \|\cdot\|_\infty) = 1$.

Proof Let $\gamma \geq \eta$ and $x_1, x_2 \in S$. We have

$$\begin{aligned}|f_\gamma(x_1, x_2) - f_\eta(x_1, x_2)| &= f_\gamma(x_1, x_2) - f_\eta(x_1, x_2) \\ &\leq \exp\left(-\|x_1 - x_2\|_2^2/\gamma^2\right).\end{aligned}$$

This is at most $a \in (0, 1)$ whenever $\|x_1 - x_2\|_2 > \gamma \log(1/a)^{1/2}$. Suppose $\|x_1 - x_2\|_2 \leq \gamma \log(1/a)^{1/2}$. We have

$$|f_\gamma(x_1, x_2) - f_\eta(x_1, x_2)| = f_\gamma(x_1, x_2) - f_\eta(x_1, x_2)$$

$$\begin{aligned}&\leq \exp\left(\|x_1 - x_2\|_2^2/\eta^2\right)\left(f_\gamma(x_1, x_2) - f_\eta(x_1, x_2)\right)\\&= \exp\left(\|x_1 - x_2\|_2^2\left(\eta^{-2} - \gamma^{-2}\right)\right) - 1\\&\leq \exp\left(\log(1/a)\left((\gamma/\eta)^2 - 1\right)\right) - 1.\end{aligned}$$

This is at most a whenever

$$\gamma \leq \left(1 + \frac{\log(1+a)}{\log(1/a)}\right)^{1/2}\eta. \tag{1}$$

Since $x/(1+x) \leq \log(1+x) \leq x$ for $x \geq 0$, we have

$$\begin{aligned}\left(1 + \frac{\log(1+a)}{\log(1/a)}\right)^{1/2} &= \left(1 + \frac{\log(1+a)}{\log(1+(1-a)/a)}\right)^{1/2}\\&\geq \left(1 + \frac{a/(1+a)}{(1-a)/a}\right)^{1/2}\\&= \left(1 + \frac{a^2}{1-a^2}\right)^{1/2}.\end{aligned}$$

Hence, (1) holds whenever

$$\gamma \leq \left(1 + \frac{a^2}{1-a^2}\right)^{1/2}\eta,$$

or

$$\log(\gamma) \leq \frac{1}{2}\log\left(1 + \frac{a^2}{1-a^2}\right) + \log(\eta).$$

The first result follows by rearranging for a.

Since

$$\log\left(1 + \frac{a^2}{1-a^2}\right) \geq \frac{a^2/(1-a^2)}{1 + a^2/(1-a^2)} = a^2,$$

(1) holds whenever $\log(\gamma) \leq a^2/2 + \log(\eta)$. Hence, for any $\gamma, \eta \in \Gamma$, we find $\|f_\gamma - f_\eta\|_\infty \leq a$ whenever $|\log(\gamma) - \log(\eta)| \leq a^2/2$. Let $b \geq 1$ and $\gamma_i \in \Gamma$ for $1 \leq i \leq b$. Recall that $\Gamma \subseteq [u, v]$. If we let

$$\log(\gamma_i) = \log(u) + a^2(2i-1)/2$$

and let b be such that

$$\log(v) - \left(\log(u) + a^2(2b-1)/2\right) \leq a^2/2,$$

then we find the f_{γ_i} for $1 \leq i \leq b$ form an a cover of $(\mathcal{L} \setminus \{0\}, \|\cdot\|_\infty)$. Rearranging the above shows that we can choose

$$b = \left\lceil \frac{\log(v/u)}{a^2} \right\rceil$$

and the second result follows by adding $\{0\}$ to the cover. The third result follows from the fact that $f_\gamma(x_1, x_2) \in (0, 1]$ for all $\gamma \in \Gamma$ and all $x_1, x_2 \in S$. ■

Lemma 34 *Assume (*$\mathcal{K}2$*). We have that* $\hat{h}_{\gamma,r}$ *is an* $(C(S), \mathcal{B}(C(S)))$*-valued measurable function on* $(\Omega \times \Gamma \times [0,\infty), \mathcal{F} \otimes \mathcal{B}(\Gamma) \otimes \mathcal{B}([0,\infty)))$*, where* γ *varies in* Γ *and* r *varies in* $[0,\infty)$*.*

We calculate an integral of these covering numbers.

Lemma 35 *Assume (*$\mathcal{K}2$*). We have*

$$\int_0^{1/2} \log N(a, \mathcal{L}, \|\cdot\|_\infty) da \leq \frac{\log(2 + 4\log(v/u))}{2} + 1.$$

Proof We have

$$\int_0^{1/2} \log N(a, \mathcal{L}, \|\cdot\|_\infty) da \leq \int_0^{1/2} \log\left(2 + \log(v/u) a^{-2}\right) da$$

by Lemma 33. Changing variables to $b = 2a$ gives

$$\begin{aligned}\frac{1}{2}\int_0^1 \log\left(2 + 4\log(v/u) b^{-2}\right) db &\leq \frac{1}{2}\int_0^1 \log\left((2 + 4\log(v/u)) b^{-2}\right) db \\ &= \frac{\log(2 + 4\log(v/u))}{2} + \int_0^1 \log(b^{-1}) db.\end{aligned}$$

Changing variables to $s = \log(b^{-1})$ shows

$$\int_0^1 \log\left(b^{-1}\right) db = \int_0^\infty s\exp(-s) ds = 1$$

since the last integral is the mean of an Exponential(1) random variable. ■

This lemma allows us directly to gain a bound on J.

Lemma 36 *Assume (*$\mathcal{K}2$*). We have*

$$J \leq \left(81(\log(8\log(v/u) + 4) + 2) + 1\right)^{1/2}.$$

Appendix D: The Orlicz Space L^{ψ_1}

Recall that $\psi_1(x) = \exp(|x|) - 1$ for $x \in \mathbb{R}$,

$$\|Z\|_{\psi_1} = \inf\{a \in (0,\infty) : \mathbb{E}(\psi_1(Z/a)) \leq 1\}$$

for any random variable Z on $(\Omega, \mathcal{F})$ and L^{ψ_1} is the set of random variables Z on $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\|Z\|_{\psi_1} < \infty$. We have that $(L^{\psi_1}, \|\cdot\|_{\psi_1})$ is a Banach space known as an Orlicz space (see [5]). For $t \geq 0$, also recall that

$$\mathbb{E}(|Z|) \leq (\log 2)\|Z\|_{\psi_1} \text{ and } |Z| \leq \|Z\|_{\psi_1}(\log 2 + t)$$

with probability at least $1 - e^{-t}$ by Lemma 27. We prove a maximal inequality in L^{ψ_1} using the same method as Lemma 2.3.3 of [1].

Lemma 37 *Let $Z_i \in L^{\psi_1}$ for $1 \leq i \leq I$. Then*

$$\left\| \max_{1 \leq i \leq I} |Z_i| \right\|_{\psi_1} \leq \frac{\log(2I)}{\log(5/4)} \max_{1 \leq i \leq I} \|Z_i\|_{\psi_1}.$$

Proof Let $M = \max_{1 \leq i \leq I} \|Z_i\|_{\psi_1}$. Also, let $C \geq 1$ and $a \in (0, \infty)$. By Lemma 27, we have

$$\begin{aligned}
\mathbb{E}\left(\exp\left(\max_{1 \leq i \leq I} |Z_i|/a\right)\right) &= \int_0^\infty \mathbb{P}\left(\max_{1 \leq i \leq I} |Z_i| > a \log t\right) dt \\
&\leq C + \int_C^\infty \mathbb{P}\left(\max_{1 \leq i \leq I} |Z_i| > a \log t\right) dt \\
&\leq C + \sum_{i=1}^I \int_C^\infty \mathbb{P}\left(|Z_i| > a \log t\right) dt \\
&\leq C + I \int_C^\infty 2t^{-a/M} dt.
\end{aligned}$$

Differentiating this bound with respect to C gives $1 - 2IC^{-a/M}$, so the bound is minimised by $C = (2I)^{M/a}$. For $a > M$, the bound becomes

$$\begin{aligned}
C + 2I \frac{M}{a - M} C^{-(a-M)/M} &= (2I)^{M/a} + \frac{M}{a - M}(2I)^{1-(a-M)/a} \\
&= \frac{a}{a - M}(2I)^{M/a}.
\end{aligned}$$

Let

$$a = \frac{M \log(2I)}{\log b}$$

for $b > 1$. We have

$$\mathbb{E}\left(\exp\left(\max_{1 \leq i \leq I} |Z_i|/a\right)\right) \leq 2$$

if $b^2 2^b \leq 4$, the hardest case being $I = 1$. This holds for $b = 5/4$ and the result follows. ■

We perform chaining in L^{ψ_1} using the same method as Theorem 2.3.6 of [1]. Recall that $N(a, M, d)$ is the minimum size of an $a > 0$ cover of a metric space (M, d).

Lemma 38 *Let Z be a stochastic process on $(\Omega, \mathcal{F})$ indexed by a separable metric space (M, d) on which Z is almost-surely continuous with $\|Z(s) - Z(t)\|_{\psi_1} \leq d(s,t)$ for all $s, t \in M$. Let $D = \sup_{s,t \in M} d(s,t)$. Fix $s_0 \in M$. Then*

$$\left\| \sup_{s \in M} |Z(s) - Z(s_0)| \right\|_{\psi_1} \leq \frac{4}{\log(5/4)} \int_0^{D/2} \log(2N(a, M, d)) da.$$

Proof Since (M, d) is separable, it has a countable dense subset M_0. We have

$$\left\| \sup_{s \in M} |Z(s) - Z(s_0)| \right\|_{\psi_1} = \left\| \sup_{s \in M_0} |Z(s) - Z(s_0)| \right\|_{\psi_1}$$

because Z is almost-surely continuous on M. Since M_0 is countable, there exists a sequence of increasing finite subsets $F_n \subseteq M$ for $n \geq 1$ whose union is M_0. We have

$$\left\| \sup_{s \in M} |Z(s) - Z(s_0)| \right\|_{\psi_1} = \lim_{n \to \infty} \left\| \max_{s \in F_n} |Z(s) - Z(s_0)| \right\|_{\psi_1}$$

by the monotone convergence theorem. Fix $n \geq 1$ and let $F = F_n$. Let $\delta_j = 2^{-j} D$ for $j \geq 0$. Since F is finite, there exists a minimum $J \geq 0$ such that

$$\{t \in F : d(s,t) \leq \delta_J\} = \{s\}$$

for all $s \in F$. Let A_j for $0 \leq j \leq J-1$ be a δ_j cover of (M, d) of size $N(\delta_j, M, d)$, where we let $A_0 = \{s_0\}$. We define the chain $C : F \times \{0, \ldots, J\} \to M$ as follows. Let $C(s, J) = s$ for all $s \in F$. For $1 \leq j \leq J$, given $C(s, j)$, let $C(s, j-1)$ be some closest point in A_{j-1} to $C(s, j)$, depending on s only through $C(s, j)$. We have

$$Z(s) - Z(s_0) = \sum_{j=1}^{J} Z(C(s,j)) - Z(C(s, j-1))$$

for $s \in F$. Hence,

$$\max_{s \in F} |Z(s) - Z(s_0)| \leq \sum_{j=1}^{J} \max_{s \in F} |Z(C(s,j)) - Z(C(s, j-1))|.$$

By Lemma 37, we have

$$\begin{aligned}
\left\| \max_{s \in F} |Z(s) - Z(s_0)| \right\|_{\psi_1} &\leq \sum_{j=1}^{J} \left\| \max_{s \in F} |Z(C(s,j)) - Z(C(s, j-1))| \right\|_{\psi_1} \\
&\leq \sum_{j=1}^{J} \frac{\log(2N(\delta_j, M, d)) \delta_{j-1}}{\log(5/4)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{4}{\log(5/4)} \sum_{j=1}^{J} (\delta_j - \delta_{j+1}) \log(2N(\delta_j, M, d)) \\
&\leq \frac{4}{\log(5/4)} \int_{\delta_{J+1}}^{\delta_1} \log(2N(a, M, d)) da \\
&\leq \frac{4}{\log(5/4)} \int_{0}^{D/2} \log(2N(a, M, d)) da.
\end{aligned}
$$

The result follows. ∎

Appendix E: Subgaussian Random Variables and Symmetric Matrices

The following result is Lemma 31 of [2], which is essentially Theorem 2.1 from [4].

Lemma 39 *Let M be a non-negative-definite matrix which is an $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrix on $(\Omega, \mathcal{F})$. There exist an orthogonal matrix A and a diagonal matrix D which are both $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrices on $(\Omega, \mathcal{F})$ such that $M = ADA^{\mathsf{T}}$.*

Recall that for $m \times n$ matrices U and V, we define $U \circ V$ to be the $m \times n$ matrix with

$$(U \circ V)_{i,j} = U_{i,j} V_{i,j}.$$

The following lemma is a conditional version of Theorem 1.1 of [6], but with explicit values for the constants derived here.

Lemma 40 *Let ε_i for $1 \leq i \leq n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent conditional on some sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and let*

$$\mathbb{E}(\exp(t\varepsilon_i) | \mathcal{G}) \leq \exp(\sigma^2 t^2/2)$$

almost surely for t a random variable on $(\Omega, \mathcal{G})$. Let M be an $n \times n$ symmetric matrix which is an $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrix on $(\Omega, \mathcal{G})$. We have

$$\mathbb{E}\left(\exp\left(t\varepsilon^{\mathsf{T}}(M - I \circ M)\varepsilon\right) \middle| \mathcal{G}\right) \leq \exp\left(16\sigma^4 \operatorname{tr}(M^2) t^2\right)$$

almost surely for t a random variable on $(\Omega, \mathcal{G})$ such that $32\sigma^4 \operatorname{tr}(M^2) t^2 \leq 1$.

Proof We follow the proof of Theorem 1.1 of [6]. Let

$$Z = \varepsilon^{\mathsf{T}}(M - I \circ M)\varepsilon = \sum_{i \neq j} M_{i,j} \varepsilon_i \varepsilon_j.$$

Also, let ϕ_i for $1 \leq i \leq n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent of each other, the ε_i and $\mathcal{G}$, with $\phi_i \sim \text{Bernoulli}(1/2)$. Furthermore, let

$$W = \sum_{i \neq j} \phi_i (1 - \phi_j) M_{i,j} \varepsilon_i \varepsilon_j.$$

We have $Z = 4\,\mathbb{E}(W | \mathcal{G}, \varepsilon)$ almost surely, which gives

$$\exp(tZ) \leq \mathbb{E}(\exp(4tW) | \mathcal{G}, \varepsilon)$$

almost surely for t a random variable on $(\Omega, \mathcal{G})$ by Jensen's inequality. Let

$$S = \{1 \leq i \leq n : \phi_i = 1\}.$$

We can write

$$W = \sum_{i \in S, j \in S^\mathsf{c}} M_{i,j} \varepsilon_i \varepsilon_j.$$

Since the ε_j are independent, we have

$$\begin{aligned}
\mathbb{E}(\exp(tZ)|\mathcal{G}) &\leq \mathbb{E}(\exp(4tW)|\mathcal{G}) \\
&= \mathbb{E}\left(\prod_{j \in S^\mathsf{c}} \mathbb{E}\left(\exp\left(4t \sum_{i \in S} M_{i,j} \varepsilon_i \varepsilon_j \right) \middle| \mathcal{G}, \phi \right) \middle| \mathcal{G} \right) \\
&\leq \mathbb{E}\left(\prod_{j \in S^\mathsf{c}} \exp\left(8t^2 \sigma^2 \left(\sum_{i \in S} M_{i,j} \varepsilon_i \right)^2 \right) \middle| \mathcal{G} \right) \\
&= \mathbb{E}\left(\exp\left(8t^2 \sigma^2 \sum_{j \in S^\mathsf{c}} \left(\sum_{i \in S} M_{i,j} \varepsilon_i \right)^2 \right) \middle| \mathcal{G} \right)
\end{aligned}$$

almost surely. Let δ_i for $1 \leq i \leq n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent of each other, the ε_i, the ϕ_i and $\mathcal{G}$, with $\delta_i \sim \mathrm{N}(0, \sigma^2)$. Since the ε_i are independent, we have

$$\begin{aligned}
\mathbb{E}(\exp(tZ)|\mathcal{G}) &\leq \mathbb{E}\left(\exp\left(4t \sum_{j \in S^\mathsf{c}} \sum_{i \in S} M_{i,j} \varepsilon_i \delta_j \right) \middle| \mathcal{G} \right) \\
&= \mathbb{E}\left(\prod_{i \in S} \mathbb{E}\left(\exp\left(4t \sum_{j \in S^\mathsf{c}} M_{i,j} \delta_j \varepsilon_i \right) \middle| \mathcal{G}, \phi \right) \middle| \mathcal{G} \right) \\
&\leq \mathbb{E}\left(\prod_{i \in S} \exp\left(8t^2 \sigma^2 \left(\sum_{j \in S^\mathsf{c}} M_{i,j} \delta_j \right)^2 \right) \middle| \mathcal{G} \right)
\end{aligned}$$

$$= \mathbb{E}\left(\exp\left(8t^2\sigma^2\sum_{i\in S}\left(\sum_{j\in S^{\mathsf{c}}} M_{i,j}\delta_j\right)^2\right)\middle|\mathcal{G}\right)$$

almost surely. Let F be the $n \times n$ matrix with $F_{i,j} = 1$ if $i = j \in S$ and 0 otherwise. Note that F is an $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrix on $(\Omega, \sigma(\phi))$. Then

$$\mathbb{E}(\exp(tZ)|\mathcal{G}) \leq \mathbb{E}\left(\exp\left(8t^2\sigma^2\delta^{\mathsf{T}}(I-F)MFM(I-F)\delta\right)\middle|\mathcal{G}\right)$$

almost surely. By Lemma 39, there exist an orthogonal matrix A and a diagonal matrix D which are both $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrices on $(\Omega, \sigma(\mathcal{G}, \phi))$ such that

$$(I-F)MFM(I-F) = ADA^{\mathsf{T}},$$

which is non-negative definite. Since $A^{\mathsf{T}}\delta$ and δ have the same distribution given $\mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}(\exp(tZ)|\mathcal{G}) &\leq \mathbb{E}\left(\exp\left(8t^2\sigma^2\delta^{\mathsf{T}}D\delta\right)\middle|\mathcal{G}\right)\\ &= \mathbb{E}\left(\prod_{i=1}^n \mathbb{E}\left(\exp\left(8t^2\sigma^2 D_{i,i}\delta_i^2\right)\middle|\mathcal{G},\phi\right)\middle|\mathcal{G}\right)\\ &= \mathbb{E}\left(\prod_{i=1}^n \left(1-16\sigma^4 D_{i,i}t^2\right)^{-1/2}\middle|\mathcal{G}\right)\end{aligned}$$

almost surely for $16\sigma^4(\max_{1\leq i\leq n} D_{i,i})t^2 < 1$ by computing the moment generating function of the δ_i^2. We have that $(1-x)^{-1/2} \leq \exp(x)$ for $x \in [0, 1/2]$, so

$$\mathbb{E}(\exp(tZ)|\mathcal{G}) \leq \mathbb{E}\left(\prod_{i=1}^n \exp\left(16\sigma^4 D_{i,i}t^2\right)\middle|\mathcal{G}\right) = \mathbb{E}\left(\exp\left(16\sigma^4\operatorname{tr}(D)t^2\right)\middle|\mathcal{G}\right)$$

almost surely for $32\sigma^4(\max_{1\leq i\leq n} D_{i,i})t^2 \leq 1$. We have

$$\operatorname{tr}(D) = \operatorname{tr}((I-F)MFM(I-F)) = \sum_{i\in S}\sum_{j\in S^{\mathsf{c}}} M_{i,j}^2 \leq \sum_{i=1}^n\sum_{j=1}^n M_{i,j}^2 = \operatorname{tr}(M^2)$$

and

$$\max_{1\leq i\leq n} D_{i,i} \leq \operatorname{tr}(D) \leq \operatorname{tr}(M^2).$$

The result follows. ■

We move the bound on the conditional moment generating function of $\varepsilon^{\mathsf{T}}(M - I\circ M)\varepsilon$ to that of $|\varepsilon^{\mathsf{T}}(M - I\circ M)\varepsilon|$.

Lemma 41 *Let ε_i for $1 \leq i \leq n$ be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ which are independent conditional on some sub-σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ and let*

$$\mathbb{E}(\exp(t\varepsilon_i)|\mathcal{G}) \leq \exp(\sigma^2t^2/2)$$

almost surely for t a random variable on $(\Omega, \mathcal{G})$. Let M be an $n \times n$ symmetric matrix which is an $(\mathbb{R}^{n\times n}, \mathcal{B}(\mathbb{R}^{n\times n}))$-valued measurable matrix on $(\Omega, \mathcal{G})$. We have

$$\mathbb{E}\left(\exp\left(t|\varepsilon^\mathsf{T}(M - I\circ M)\varepsilon|\right)\middle|\mathcal{G}\right) \leq \frac{1}{1 - 2^{7/2}\sigma^2\operatorname{tr}(M^2)^{1/2}t} 2^{2^{7/2}\sigma^2\operatorname{tr}(M^2)^{1/2}t}$$

almost surely for $t \geq 0$, a random variable on $(\Omega, \mathcal{G})$, such that $2^{7/2}\sigma^2\operatorname{tr}(M^2)^{1/2}t < 1$. Hence,

$$\mathbb{E}\left(\frac{\left|\varepsilon^\mathsf{T}(M - I\circ M)\varepsilon\right|}{2^{7/2}(\log 2)\sigma^2\operatorname{tr}(M^2)^{1/2}/\log(5/4)}\middle|\mathcal{G}\right) \leq 2.$$

Proof Let

$$Z = \varepsilon^\mathsf{T}(M - I\circ M)\varepsilon.$$

By Lemma 40, we have

$$\mathbb{E}(\exp(tZ)|\mathcal{G}) \leq \exp\left(16\sigma^4\operatorname{tr}(M^2)t^2\right)$$

almost surely for t a random variable on $(\Omega, \mathcal{G})$ such that $32\sigma^4\operatorname{tr}(A^2)t^2 \leq 1$. By Chernoff bounding, we have

$$\mathbb{P}(Z \geq z|\mathcal{G}) \leq \exp\left(-tz + 16\sigma^4\operatorname{tr}(M^2)t^2\right)$$

almost surely for $z \geq 0$, a random variable on $(\Omega, \mathcal{G})$, $t \geq 0$ and $32\sigma^4\operatorname{tr}(A^2)t^2 \leq 1$. Minimising over t gives

$$\mathbb{P}(Z \geq z|\mathcal{G}) \leq \exp\left(-\min\left(\frac{z^2}{2^6\sigma^4\operatorname{tr}(M^2)}, \frac{z}{2^{7/2}\sigma^2\operatorname{tr}(M^2)^{1/2}}\right)\right)$$

almost surely. The first term in the minimum is attained by $t = 2^{-5}\sigma^{-4}\operatorname{tr}(M^2)^{-1}z$ when $z < 2^{5/2}\sigma^2\operatorname{tr}(M^2)^{1/2}$, and the second term is attained by $t = 2^{-5/2}\sigma^{-2}\operatorname{tr}(M^2)^{-1/2}$ when $z \geq 2^{5/2}\sigma^2\operatorname{tr}(M^2)^{1/2}$. In the second case, note that

$$16\sigma^4\operatorname{tr}(M^2)t^2 = \frac{1}{2} \leq \frac{z}{2^{7/2}\sigma^2\operatorname{tr}(M^2)^{1/2}}.$$

The same result holds if we replace Z with $-Z$ by replacing M with $-M$. For $C \geq 1$ and $t \geq 0$, random variables on $(\Omega, \mathcal{G})$, we have

$$\begin{aligned}
\mathbb{E}(\exp(t|Z|)|\mathcal{G}) &= \int_0^\infty \mathbb{P}(|Z| \geq (\log s)/t|\mathcal{G})ds \\
&\leq C + \int_C^\infty \mathbb{P}(|Z| \geq (\log s)/t|\mathcal{G})ds \\
&\leq C + \int_C^\infty \mathbb{P}(Z \geq (\log s)/t|\mathcal{G})ds + \int_C^\infty \mathbb{P}(-Z \geq (\log s)/t|\mathcal{G})ds \\
&\leq C + 2\int_C^\infty \exp\left(-\min\left(\frac{(\log s)^2}{2^6\sigma^4\operatorname{tr}(M^2)t^2}, \frac{\log s}{2^{7/2}\sigma^2\operatorname{tr}(M^2)^{1/2}t}\right)\right)ds
\end{aligned}$$

almost surely. By letting $C \geq \exp(2^{5/2}\sigma^2 \operatorname{tr}(M^2)^{1/2}t)$, the bound becomes

$$C + 2\int_C^\infty s^{-(2^{7/2}\sigma^2 \operatorname{tr}(M^2)^{1/2}t)^{-1}} ds.$$

Let $u = 2^{7/2}\sigma^2 \operatorname{tr}(M^2)^{1/2}t$, a random variable on $(\Omega, \mathcal{G})$. Differentiating this bound with respect to C gives $1 - 2C^{-u^{-1}}$, so the bound is minimised by $C = 2^u$. This satisfies the condition on C above as

$$e^{2^{5/2}} \leq 3^6 \leq 2^{10} \leq 2^{2^{7/2}}.$$

For $u < 1$, the bound becomes

$$\begin{aligned} C + 2\frac{u}{1-u}C^{-(1-u)/u} &= 2^u + \frac{u}{1-u}2^{1-(1-u)} \\ &= \frac{1}{1-u}2^u. \end{aligned}$$

The first result follows. Let

$$u = \frac{\log b}{\log 2}$$

for $b > 1$. We have

$$\mathbb{E}(\exp(t|Z|)|\mathcal{G}) \leq 2$$

almost surely if $b^2 2^b \leq 4$. This holds for $b = 5/4$ and the second result follows. ■

References

[1] Evarist Giné and Richard Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, New York, 2016.

[2] Stephen Page and Steffen Grünewälder. Ivanov-regularised least-squares estimators over large RKHSs and their interpolation spaces. *Journal of Machine Learning Research*, 20(120):1–49, 2019.

[3] Stephen Page and Steffen Grünewälder. The Goldenshluger–Lepski method for constrained least-squares estimators over RKHSs. 2020.

[4] Yamilet Quintana and José M. Rodríguez. Measurable diagonalization of positive definite matrices. *J. Approx. Theory*, 185:91–97, 2014.

[5] Malempati Madhusudana Rao and Zhong Dao Ren. *Theory of Orlicz Spaces*. Marcel Dekker, New York, 1991.

[6] Mark Rudelson and Roman Vershynin. Hanson–Wright inequality and sub-Gaussian concentration. *Electron. Commun. Probab.*, 18, 2013.

[7] Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer–Verlag, New York, 2008.

[8] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer–Verlag, New York, 1996.