

A Appendix

To emphasize the underlying parameters of the NN, by some abuse of notation, we introduce

$$\mathcal{G}_k(\Theta) := \left\{ g : \mathbb{R}^d \rightarrow \mathbb{R} : g(x) = \sum_{i=1}^k \beta_i \phi(w_i \cdot x + b_i) + b_0, (\{\beta_i, w_i, b_i\}_{i=1}^k, b_0) \in \Theta \right\}, \quad (34a)$$

$$\Theta_k(\mathbf{a}) := \left\{ (\{\beta_i, w_i, b_i\}_{i=1}^k, b_0) : \begin{array}{l} w_i \in \mathbb{R}^d, b_0, b_i, \beta_i \in \mathbb{R}, \max_{\substack{i=1, \dots, k \\ j=1, \dots, d}} \{|w_{i,j}|, |b_i|\} \leq a_1 \\ |\beta_i| \leq a_2, \quad i = 1, \dots, k, \quad |b_0| \leq a_3 \end{array} \right\}. \quad (34b)$$

Also, throughout the Appendix, we denote $g(x) = \sum_{i=1}^k \beta_i \phi(w_i \cdot x + b_i) + b_0$ for $\theta = (\{\beta_i, w_i, b_i\}_{i=1}^k, b_0)$ by g_θ , whenever the underlying θ needs to be emphasized.

We first state an auxiliary result which will be useful in the proofs that follow. For $b \geq 0$, an integer $l \geq 0$, and an open set $\mathcal{U} \subseteq \mathbb{R}^d$ containing the origin, consider the class $\mathcal{S}_{l,b}(\mathcal{U})$ of square-integrable functions defined below:

$$\mathcal{S}_{l,b}(\mathcal{U}) := \left\{ f \in L^1(\mathcal{U}) \cup L^2(\mathcal{U}) : \begin{array}{l} |f(0)| \leq b, \quad D^\alpha f \text{ exists Lebesgue a.e. on } \mathcal{U} \quad \forall \alpha \text{ s.t. } |\alpha| = l, \\ \|D^\alpha f\|_{L^2(\mathcal{U})} \leq b \text{ for } |\alpha| \in \{1, l\} \end{array} \right\}. \quad (35)$$

The following lemma states that functions in $\mathcal{S}_{l,b}(\mathbb{R}^d)$ with sufficient smoothness order l belong to the Barron class. Its proof essentially follows using arguments from Barron (1993), where it was mentioned without explicit quantification. Below, we provide a proof for completeness.

Lemma 1 (Smoothness and Barron class). *If $f \in \mathcal{S}_{s,b}(\mathbb{R}^d)$ for $s := \lfloor \frac{d}{2} \rfloor + 2$, then we have*

$$B(f) \leq b \kappa_d \sqrt{d}, \quad (36a)$$

$$\kappa_d^2 := (d + d^s) \int_{\mathbb{R}^d} (1 + \|\omega\|^{2(s-1)})^{-1} d\omega < \infty. \quad (36b)$$

Consequently, $\mathcal{S}_{s,b}(\mathbb{R}^d) \subseteq \mathcal{B}_{b\kappa_d\sqrt{d}\vee b}$.

Proof. Since $f \in L^1(\mathbb{R}^d) \cup L^2(\mathbb{R}^d)$, its Fourier transform $\hat{f}(\omega)$ exists, and hence, $F(d\omega) = |\hat{f}(\omega)| d\omega$. Then, it follows that

$$B(f) = \int_{\mathbb{R}^d} \sup_{x \in \mathcal{X}} |\omega \cdot x| |\hat{f}(\omega)| d\omega \leq \sqrt{d} \int_{\mathbb{R}^d} \|\omega\| |\hat{f}(\omega)| d\omega, \quad (37)$$

where we used $\sup_{x \in \mathcal{X}} |\omega \cdot x| \leq \sqrt{d} \|\omega\|$ which holds by Cauchy-Schwarz inequality.

Next, recall that if the partial derivatives $D^\alpha f$, $|\alpha| = s$, exists on \mathbb{R}^d , then all partial derivatives $D^\alpha f$, $0 \leq |\alpha| \leq s$, also exists. Hence, if $\|D^\alpha f\|_{L^2(\mathbb{R}^d)} \leq b$ for all α with $|\alpha| \in \{1, s\}$, we have

$$\int_{\mathbb{R}^d} \|\omega\| |\hat{f}(\omega)| d\omega \stackrel{(a)}{\leq} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2(s-1)}} \right)^{\frac{1}{2}} \left(\int_{\mathbb{R}^d} (\|\omega\|^2 + \|\omega\|^{2s}) |\hat{f}(\omega)|^2 d\omega \right)^{\frac{1}{2}} \quad (38)$$

$$\begin{aligned} & \stackrel{(b)}{\leq} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2(s-1)}} \right)^{\frac{1}{2}} \left(\sum_{|\alpha|=1} \|D^\alpha f\|_{L^2(\mathbb{R}^d)}^2 + \sum_{|\alpha|=s} \|D^\alpha f\|_{L^2(\mathbb{R}^d)}^2 \right)^{\frac{1}{2}} \\ & \stackrel{(c)}{\leq} \kappa_d b, \end{aligned} \quad (39)$$

where

- (a) follows from Cauchy-Schwarz inequality;
- (b) is due to Plancherel's theorem;
- (c) follows since $|\{\alpha : |\alpha| = s\}| = d^s$ and $\|D^\alpha f\|_{L^2(\mathbb{R}^d)} \leq b$.

Combining (37) and (39) leads to (36a). The final claim follows from (5) and (36a) by noting that $|f(0)| \leq b$ by definition. \square

A.1 Proof of Theorem 2

The proof relies on arguments from Barron (1992) and Barron (1993), along with the uniform central limit theorem for uniformly bounded VC function classes. Fix an arbitrary (small) $\delta > 0$, and let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be such that $\tilde{f} = f|_{\mathcal{X}}$ and $B(f) \vee f(0) \leq c + \delta$. This is possible since $c_B^*(\tilde{f}) \leq c$. Then, it follows from the proof of Barron (1993, Theorem 2) that

$$f_0(x) := f(x) - f(0) = \int_{\omega \in \mathbb{R}^d \setminus \{0\}} \varrho(x, \omega) \mu(d\omega),$$

where

$$\begin{aligned} \varrho(x, \omega) &= \frac{B(f)}{\sup_{x \in \mathcal{X}} |\omega \cdot x|} (\cos(\omega \cdot x + \zeta(\omega)) - \cos(\zeta(\omega))), \\ B(f) &:= \int_{\mathbb{R}^d} \sup_{x \in \mathcal{X}} |\omega \cdot x| F(d\omega), \\ \mu(d\omega) &= \frac{\sup_{x \in \mathcal{X}} |\omega \cdot x| F(d\omega)}{B(f)}, \end{aligned}$$

and $\zeta : \mathbb{R}^d \rightarrow \mathbb{R}$. Note that $\mu \in \mathcal{P}(\mathbb{R}^d)$ is a probability measure.

Let $\tilde{\Theta}_1(k, B(f)) := \Theta_1(\sqrt{k} \log k, 2B(f), 0)$ (see (34b)). Then, it further follows from the proofs⁴ of Barron (1993, Lemma 2-Lemma 4, Theorem 3) that there exists a probability measure $\mu_k \in \mathcal{P}(\tilde{\Theta}_1(k, B(f)))$ (see Barron (1993, Eqns. (28)-(32))) such that

$$\left\| f_0 - \int_{\theta \in \tilde{\Theta}_1(k, B(f))} g_\theta(\cdot) \mu_k(d\theta) \right\|_{\infty, P, Q} \leq \frac{2(B(f) + 1)}{\sqrt{k}}, \quad (40)$$

where $g_\theta(x) = \beta \phi(w \cdot x + b)$ for $\theta = (\beta, w, b)$. Note that $\int_{\tilde{\Theta}_1(k, B(f))} \mu_k(d\theta) = 1 < \infty$.

Next, for each fixed x , let $v_x : \tilde{\Theta}_1(k, B(f)) \rightarrow \mathbb{R}$ be given by $v_x(\theta) := g_\theta(x)$, and consider the function class $\mathcal{V}_k(\tilde{\Theta}_1(k, B(f))) = \{v_x, x \in \mathbb{R}^d\}$. Note that every $v_x \in \mathcal{V}_k(\tilde{\Theta}_1(k, B(f)))$ is a composition of an affine function in θ with the bounded monotonic function $\beta \phi(\cdot)$. Hence, noting that $\mathcal{V}_k(\tilde{\Theta}_1(k, B(f)))$ is a VC function class (Van Der Vaart and Wellner (1996)), it follows from Van Der Vaart and Wellner (1996, Theorem 2.8.3) that it is a uniform Donsker class (in particular, μ_k -Donsker) for all probability measures $\mu \in \mathcal{P}(\tilde{\Theta}_1(k, B(f)))$. Furthermore, an application of Van Der Vaart and Wellner (1996, Corollary 2.2.8)) yields that there exists k parameter vectors, $\theta_i := (\beta_i, w_i, b_i) \in \tilde{\Theta}_1(k, B(f))$, $1 \leq i \leq k$, such that (see also Yukich et al. (1995, Theorem 2.1))

$$\sup_{x \in \mathbb{R}^d} \left| \int_{\theta \in \tilde{\Theta}_1(k, B(f))} g_\theta(x) \mu_k(d\theta) - \frac{1}{k} \sum_{i=1}^k g_{\theta_i}(x) \right| \leq \hat{c}_d B(f) k^{-\frac{1}{2}}, \quad (41)$$

where \hat{c}_d is a constant which depends only on d . Note that the R.H.S. of (41) is independent of μ_k and depends on f and \mathcal{X} only via $B(f)$.

From (40), (41) and triangle inequality, we obtain

$$\left\| f_0 - \frac{1}{k} \sum_{i=1}^k g_{\theta_i} \right\|_{\infty, P, Q} \leq (\hat{c}_d B(f) + 2B(f) + 2) k^{-\frac{1}{2}}.$$

⁴The claims in Barron (1993, Lemma 2- Lemma 4, Theorem 3) are stated for L_2 norm, but it is not hard to see from the proof therein that the same also holds for L^∞ norm, apart from the following subtlety. In the proof of Lemma 3, it is shown that $\varrho(x, \omega)$, $\omega \in \mathbb{R}^d$, lies in the convex closure of a certain class of step functions, whose discontinuity points are adjusted to coincide with the continuity points of the underlying measure μ . Similarly, here, the step discontinuities needs to be adjusted to coincide with the continuity points of both P and Q . Nevertheless, the same arguments hold since the common continuity points of P and Q form a dense set.

Setting $\theta = \left\{ \left\{ \left(\frac{\beta_i}{k}, w_i, b_i \right) \right\}_{i=1}^k, f(0) \right\}$ and $g_\theta(x) = f(0) + \frac{1}{k} \sum_{i=1}^k g_{\theta_i}(x)$, we have

$$\|f - g_\theta\|_{\infty, P, Q} \leq ((\hat{c}_d + 2)B(f) + 2) k^{-\frac{1}{2}} \leq ((\hat{c}_d + 2)(c + \delta) + 2) k^{-\frac{1}{2}}.$$

Next, note that $\|\tilde{f} - g_\theta\|_{\infty, P, Q} = \|f - g_\theta\|_{\infty, P, Q}$ and $g_\theta \in \mathcal{G}_k^*(B(f) \vee f(0)) \subseteq \mathcal{G}_k^*(c + \delta)$. Since $\delta > 0$ is arbitrary, we obtain that there exists $g_\theta \in \mathcal{G}_k^*(c)$

$$\|\tilde{f} - g_\theta\|_{\infty, P, Q} \leq ((\hat{c}_d + 2)c + 2) k^{-\frac{1}{2}} =: \tilde{C}_{d,c} k^{-\frac{1}{2}}, \quad (42)$$

thus proving the claim in (12).

On the other hand, it follows similar to (38) in Lemma 1 that for a fixed $\epsilon > 0$ and $l(\epsilon) = d/2 + 1 + \epsilon$, the set of functions $f \in \mathbb{R}^d \rightarrow \mathbb{R}$ such that $B(f) \leq c$ includes those whose Fourier transform $\hat{f}(\omega)$ satisfies

$$\int_{\mathbb{R}^d} \left(\|\omega\|^2 + \|\omega\|^{2l(\epsilon)} \right) |\hat{f}(\omega)|^2 d\omega \leq c^2 d^{-1} \left(\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2(l(\epsilon)-1)}} \right)^{-1}, \quad (43)$$

since $\int_{\mathbb{R}^d} \frac{d\omega}{1 + \|\omega\|^{2(l(\epsilon)-1)}} < \infty$. Then, (13) follows from the proof of Barron (1992)[Theorem 3]. Note from the proof therein that the constant in (13) may in general depend on d and ϵ .

A.2 Proof of Corollary 1

By Theorem 2, it suffices to show that there exists an extension f_e of f from \mathcal{U} to \mathbb{R}^d such that $B(f_e) \vee f_e(0) \leq \bar{c}_{b,c,d}$. Let α_j denote a multi-index of order j , and recall that $s := \lfloor \frac{d}{2} \rfloor + 2$. Consider an extension of $D^{\alpha_s} f$ from \mathcal{U} to \mathbb{R}^d for each α_s as follows:

$$D^{\alpha_s} f(x) := \inf_{x' \in \mathcal{U}} D^{\alpha_s} f(x') + c \|x - x'\|^\delta, \quad x \in \mathbb{R}^d \setminus \mathcal{U}. \quad (44)$$

Note that $D^{\alpha_s} f$ extended this way is Hölder continuous with the same constant c and exponent δ on \mathbb{R}^d . Fixing $D^{\alpha_s} f$ on \mathbb{R}^d induces an extension of all lower (and also higher) order derivatives $D^{\alpha_j} f$, $0 \leq j < s$ to \mathbb{R}^d , which can be defined recursively as $D^{\alpha_1} D^{\alpha_{s-j}} f(x) = D^{\alpha_1 + \alpha_{s-j}} f(x)$, $x \in \mathbb{R}^d$, for all α_1 , α_{s-j} and $j = 1, \dots, s$.

Let $\mathcal{U}' := \{x' \in \mathbb{R}^d : \|x' - x\| < 1 \text{ for some } x \in \mathcal{X}\}$. Suppose $\mathcal{U} \subset \mathcal{U}'$. By the mean value theorem, we have for any $x, x' \in \mathcal{U}'$ and $j = 1, \dots, s$,

$$\begin{aligned} |D^{\alpha_{s-j}} f(x')| &\leq |D^{\alpha_{s-j}} f(x)| + \max_{\substack{\tilde{x} \in \mathcal{U}', \\ \alpha_1}} |D^{\alpha_{s-j} + \alpha_1} f(\tilde{x})| \|x - x'\|_1 \\ &\leq |D^{\alpha_{s-j}} f(x)| + \max_{\substack{\tilde{x} \in \mathcal{U}', \\ \alpha_1}} |D^{\alpha_{s-j} + \alpha_1} f(\tilde{x})| \sqrt{d} \|x - x'\|, \end{aligned} \quad (45)$$

where the last step follows from $\|x - x'\|_1 \leq \sqrt{d} \|x - x'\|$. Also, note from (44) that $D^{\alpha_s} f(x) < b + c$ for all $x \in \mathcal{U}'$, and recall that since $f \in \mathcal{H}_{b,c}^{s,\delta}(\mathcal{U})$, we have $|D^{\alpha_{s-j}} f(x)| \leq b$ for all $x \in \mathcal{U}$. Then, for any $x' \in \mathcal{U}'$, taking $x \in \mathcal{X}$ satisfying $\|x - x'\| \leq 1$ (such an x exists by definition of \mathcal{U}') in (45) yields

$$|D^{\alpha_{s-1}} f(x')| \leq b + (b + c)\sqrt{d}. \quad (46)$$

Starting from (46) and recursively applying (45), we obtain for $j = 1, \dots, s$, and $x' \in \mathcal{U}'$,

$$|D^{\alpha_{s-j}} f(x')| \leq b \sum_{i=1}^j d^{\frac{i-1}{2}} + (b + c)d^{\frac{j}{2}} \leq b \frac{1 - d^{\frac{s}{2}}}{1 - \sqrt{d}} + (b + c)d^{\frac{s}{2}} =: \tilde{b}. \quad (47)$$

Thus, the extension f from \mathcal{U} to \mathbb{R}^d satisfies $f|_{\mathcal{U}'} \in \mathcal{H}_{b,c}^{s,\delta}(\mathcal{U}')$. If $\mathcal{U}' \subseteq \mathcal{U}$, then $f|_{\mathcal{U}'} \in \mathcal{H}_{b,c}^{s,\delta}(\mathcal{U}')$ by definition, and thus, in either case, $f|_{\mathcal{U}'} \in \mathcal{H}_{b,c}^{s,\delta}(\mathcal{U}')$.

The desired final extension is $f_e : \mathbb{R}^d \rightarrow \mathbb{R}$ given by $f_e(x) := f(x) \cdot f_C(x)$, where

$$f_C(x) := \mathbb{1}_{\mathcal{X}'} * \psi_{\frac{1}{2}}(x) := \int_{\mathbb{R}^d} \mathbb{1}_{\mathcal{X}'}(y) \psi_{\frac{1}{2}}(x-y) dy, \quad x \in \mathbb{R}^d, \quad (48)$$

$$\mathcal{X}' := \{x' \in \mathbb{R}^d : \|x' - x\| \leq 0.5 \text{ for some } x \in \mathcal{X}\},$$

$$\psi(x) := \begin{cases} u^{-1} e^{-\frac{1}{\frac{1}{2} - \|x\|^2}}, & \|x\| < \frac{1}{2}, \\ 0, & \text{otherwise,} \end{cases} \quad (49)$$

and u is the normalization constant such that $\int_{\mathbb{R}^d} \psi(x) dx = 1$. Note that $\psi \in C^\infty(\mathbb{R}^d)$, and consequently, $f_C \in C^\infty(\mathbb{R}^d)$ from (48) by dominated convergence theorem. Also, observe that $f_C(x) = 1$ for $x \in \mathcal{X}$, $f_C(x) = 0$ for $x \in \mathbb{R}^d \setminus \mathcal{U}'$ and $f_C(x) \in (0, 1)$ for $x \in \mathcal{U}' \setminus \mathcal{X}$. Hence, $f_e(x) = f(x)$ for $x \in \mathcal{X}$, $f_e(x) = 0$ for $x \in \mathbb{R}^d \setminus \mathcal{U}'$ and $|f_e(x)| \leq |f(x)|$ for $x \in \mathcal{U}' \setminus \mathcal{X}$, thus satisfying $f_e|_{\mathcal{X}} = f|_{\mathcal{X}} = f$ as required. Moreover, for all $j = 0, \dots, s$,

$$|D^{\alpha_j} f_e(x)| \stackrel{(a)}{\leq} 2^j \tilde{b} \max_{\substack{x \in \mathcal{U}', \\ \alpha: |\alpha| \leq j}} |D^{\alpha} f_C(x)| \stackrel{(b)}{\leq} 2^s \tilde{b} \max_{\substack{x: \|x\| \leq 0.5, \\ \alpha: |\alpha| \leq s}} |D^{\alpha} \psi(x)| =: \hat{b}, \quad x \in \mathcal{U}', \quad (50a)$$

$$D^{\alpha_j} f_e(x) = 0, \quad x \notin \mathcal{U}', \quad (50b)$$

where

(a) follows using chain rule for differentiation and (47);

(b) follows from the definition in (48).

Then, we have for $j = 0, \dots, s$,

$$\begin{aligned} \|D^{\alpha_j} f_e\|_{L^2(\mathbb{R}^d)}^2 &= \int_{\mathbb{R}^d} (D^{\alpha_j} f_e)^2(x) dx \\ &= \int_{\mathcal{U}'} (D^{\alpha_j} f_e)^2(x) dx \leq \hat{b}^2 \text{Vol}_d(0.5\sqrt{d} + 1) \\ &= \hat{b}^2 \frac{\pi^{\frac{d}{2}}}{\Gamma(\frac{d}{2} + 1)} (0.5\sqrt{d} + 1)^d =: \bar{b}, \end{aligned} \quad (51)$$

where $\text{Vol}_d(r)$ denotes the volume of a Euclidean ball in \mathbb{R}^d with radius r and Γ denotes the gamma function. Defining $b' := \sqrt{\bar{b}}$ and noting that $b' \geq \hat{b}$, we have from (50) and (51) that $f_e(x) \in \tilde{\mathcal{S}}_{s,b'}(\mathbb{R}^d)$, where

$$\tilde{\mathcal{S}}_{s,b'}(\mathbb{R}^d) := \left\{ f \in L^1(\mathbb{R}^d) \cup L^2(\mathbb{R}^d) : \begin{array}{l} |f(0)| \leq b', \quad D^{\alpha} f \text{ exists Lebesgue a.e. on } \mathbb{R}^d \forall \alpha \text{ s.t. } |\alpha| = s, \\ \|D^{\alpha} f\|_{L^2(\mathbb{R}^d)} \leq b' \text{ for } |\alpha| = 1, \dots, s \end{array} \right\}. \quad (52)$$

Observe that $\tilde{\mathcal{S}}_{s,b'}(\mathbb{R}^d) \subseteq \mathcal{S}_{s,b'}(\mathbb{R}^d)$ (see (35)). This implies via Lemma 1 that $B(f_e) \leq c' := \kappa_d \sqrt{d} b'$ and

$$f_e \in \mathcal{B}_{b' \vee c'} \cap \tilde{\mathcal{S}}_{s,b'}(\mathbb{R}^d) \subseteq \mathcal{B}_{b' \vee c'} \cap \mathcal{S}_{s,b'}(\mathbb{R}^d). \quad (53)$$

Then, by defining

$$\bar{c}_{b,c,d} := b' \vee c', \quad (54)$$

where

$$b' = \pi^{\frac{d}{4}} \Gamma^{-1/2}(0.5d + 1) (0.5\sqrt{d} + 1)^{\frac{d}{2}} 2^s \left(b \frac{1 - d^{\frac{s}{2}}}{1 - \sqrt{d}} + (b + c) d^{\frac{s}{2}} \right) \max_{\substack{x: \|x\| \leq 0.5, \\ \alpha: |\alpha| \leq s}} \psi^{(\alpha)}(x), \quad (55)$$

$$c' = \sqrt{d} \kappa_d b', \quad (56)$$

$$\kappa_d^2 = (d + d^s) \int_{\mathbb{R}^d} (1 + \|\omega\|^{2(s-1)})^{-1} d\omega,$$

it follows from Theorem 2 (see (42)) that there exists $g \in \mathcal{G}_k^*(\bar{c}_{b,c,d})$ such that

$$\|\tilde{f} - g\|_{\infty, P, Q} \leq \tilde{C}_{d, \bar{c}_{b,c,d}} k^{-\frac{1}{2}}. \quad (57)$$

This completes the proof.

A.3 Proof of Theorem 3

We will show that Theorem 3 holds with

$$V_{k,\mathbf{a},\gamma} := 4Ca_2^2 k R_{k,\mathbf{a},\gamma}^2, \quad (58)$$

$$E_{k,\mathbf{a},n,\gamma} := 2\sqrt{2}n^{-\frac{1}{2}}ka_2R_{k,\mathbf{a},\gamma} = 4\sqrt{2}n^{-\frac{1}{2}}k^{3/2}a_2\left(\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} + 1\right), \quad (59)$$

where

$$R_{k,\mathbf{a},\gamma} := 2\left(\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} + 1\right)\sqrt{k}, \quad (60)$$

and $\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})}$ is defined in (16). We have

$$\begin{aligned} & \hat{H}_{\gamma,\mathcal{G}_k(\mathbf{a})}(x^n, y^n) - H_{\gamma,\mathcal{G}_k(\mathbf{a})}(P, Q) \\ &= \sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} \frac{1}{n} \sum_{i=1}^n g_\theta(x_i) - \frac{1}{n} \sum_{i=1}^n \gamma(g_\theta(y_i)) - \left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} \mathbb{E}_P[g_\theta(X)] - \mathbb{E}_Q[\gamma(g_\theta(Y))] \right) \\ &\leq \sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} \frac{1}{n} \sum_{i=1}^n g_\theta(x_i) - \frac{1}{n} \sum_{i=1}^n \gamma(g_\theta(y_i)) - \mathbb{E}_P[g_\theta(X)] + \mathbb{E}_Q[\gamma(g_\theta(Y))]. \end{aligned} \quad (61)$$

Let

$$Z_\theta := \frac{1}{n} \sum_{i=1}^n g_\theta(X_i) - \frac{1}{n} \sum_{i=1}^n \gamma(g_\theta(Y_i)) - \mathbb{E}_P[g_\theta(X)] + \mathbb{E}_Q[\gamma(g_\theta(Y))]. \quad (62)$$

We have

$$\begin{aligned} |Z_\theta - Z_{\theta'}| &\leq \sum_{i=1}^n \frac{1}{n} |g_\theta(X_i) - g_{\theta'}(X_i) - \mathbb{E}_P[g_\theta(X) - g_{\theta'}(X)]| \\ &\quad + \frac{1}{n} |\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i)) - \mathbb{E}_Q[\gamma(g_\theta(Y)) - \gamma(g_{\theta'}(Y))]|. \end{aligned} \quad (63)$$

Since $0 \leq \phi(x) \leq 1$ for all $x \in \mathbb{R}^d$, for any $x, x' \in \mathcal{X}$ and $\theta = (\{\beta_i, w_i, b_i\}_{i=1}^k, b_0)$, $\theta' = (\{\beta'_i, w'_i, b'_i\}_{i=1}^k, b'_0) \in \Theta_k(\mathbf{a})$,

$$|g_\theta(x) - g_{\theta'}(x')| \leq \sum_{i=1}^k |\beta_i - \beta'_i| \leq \|\beta(\theta) - \beta(\theta')\|_1, \quad (64)$$

where $\beta(\theta) := (\beta_1, \dots, \beta_k)$. Moreover, an application of the mean value theorem yields that for all $\theta, \theta' \in \Theta_k(\mathbf{a})$,

$$|\gamma(g_\theta(x)) - \gamma(g_{\theta'}(x'))| \leq \bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} |g_\theta(x) - g_{\theta'}(x')| \leq \bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} \|\beta(\theta) - \beta(\theta')\|_1, \quad (65)$$

where $\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})}$ is defined in (16). Hence, with probability one

$$\begin{aligned} & \frac{1}{n} |g_\theta(X_i) - g_{\theta'}(X_i) - \mathbb{E}_P[g_\theta(X_i) - g_{\theta'}(X_i)]| + \frac{1}{n} |\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i)) - \mathbb{E}_Q[\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i))]| \\ &\leq \frac{1}{n} [|g_\theta(X_i) - g_{\theta'}(X_i)| + |\mathbb{E}_P[g_\theta(X_i) - g_{\theta'}(X_i)]| + |\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i))| + |\mathbb{E}_Q[\gamma(g_\theta(Y_i)) - \gamma(g_{\theta'}(Y_i))]|] \\ &\leq \frac{1}{n} s_{k,\mathbf{a},\gamma} \|\beta(\theta) - \beta(\theta')\|_1, \end{aligned} \quad (66)$$

where $s_{k,\mathbf{a},\gamma} := 2\left(\bar{\gamma}'_{\mathcal{G}_k(\mathbf{a})} + 1\right)$. Note that $\mathbb{E}[Z_\theta] = 0$ for all $\theta \in \Theta_k(\mathbf{a})$. Then, using the fact that $\|\beta(\theta) - \beta(\theta')\|_1 \leq \sqrt{k} \|\beta(\theta) - \beta(\theta')\|$, it follows from (63) and (66) via Hoeffding's lemma that

$$\mathbb{E} \left[e^{t(Z_\theta - Z_{\theta'})} \right] \leq e^{\frac{1}{2} t^2 d_{k,\mathbf{a},n,\gamma}(\theta, \theta')^2}, \quad (67)$$

where

$$\mathbf{d}_{k,\mathbf{a},n,\gamma}(\theta, \theta') := \frac{s_{k,\mathbf{a},\gamma} \sqrt{k} \|\beta(\theta) - \beta(\theta')\|}{\sqrt{n}} := \frac{R_{k,\mathbf{a},\gamma}}{\sqrt{n}} \|\beta(\theta) - \beta(\theta')\|. \quad (68)$$

It follows that $\{Z_\theta\}_{\theta \in \Theta_k(\mathbf{a})}$ is a separable subgaussian process on the metric space $(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}(\theta, \theta'))$. Next, note that $N(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}(\cdot, \cdot), \epsilon) = N([-a_2, a_2]^k, n^{-\frac{1}{2}} R_{k,\mathbf{a},\gamma} \|\cdot\|, \epsilon)$. Also, $[-a_2, a_2]^k \subseteq B^k(\sqrt{k} a_2)$. Hence, we have

$$\begin{aligned} N(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}(\cdot, \cdot), \epsilon) &\leq N\left(B^k(\sqrt{k} a_2), n^{-\frac{1}{2}} R_{k,\mathbf{a},\gamma} \|\cdot\|, \epsilon\right) \\ &= N\left(B^k(\sqrt{k} a_2), \|\cdot\|, \sqrt{n} R_{k,\mathbf{a},\gamma}^{-1} \epsilon\right) \\ &\leq \frac{(\sqrt{k} a_2 + \sqrt{n} R_{k,\mathbf{a},\gamma}^{-1} \epsilon)^k}{(\sqrt{n} R_{k,\mathbf{a},\gamma}^{-1} \epsilon)^k} \\ &= \left(1 + \frac{\sqrt{k} a_2 R_{k,\mathbf{a},\gamma}}{\sqrt{n} \epsilon}\right)^k, \end{aligned} \quad (69)$$

where, in (69), we used that the covering number of Euclidean ball $B^d(r)$ w.r.t. Euclidean norm satisfies

$$N(B^d(r), \|\cdot\|, \epsilon) \leq \left(\frac{r + \epsilon}{\epsilon}\right)^d. \quad (70)$$

Also, for $\epsilon \geq \text{diam}(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}) := \max_{\theta, \theta' \in \Theta_k(\mathbf{a})} \mathbf{d}_{k,\mathbf{a},n,\gamma}(\theta, \theta') = 2\sqrt{k} a_2 R_{k,\mathbf{a},\gamma} n^{-\frac{1}{2}}$, we have that $N(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}(\cdot, \cdot), \epsilon) = 1$. Then,

$$\begin{aligned} E_{k,\mathbf{a},n,\gamma} &:= \int_0^\infty \sqrt{\log N(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}(\cdot, \cdot), \epsilon)} d\epsilon \\ &= \int_0^{\text{diam}(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma})} \sqrt{\log N(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}(\cdot, \cdot), \epsilon)} d\epsilon \\ &\leq \sqrt{k} \int_0^{\text{diam}(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma})} \sqrt{\log \left(1 + \frac{a_2 \sqrt{k} R_{k,\mathbf{a},\gamma}}{\sqrt{n} \epsilon}\right)} d\epsilon \\ &\leq n^{-\frac{1}{4}} k^{\frac{3}{4}} \sqrt{a_2 R_{k,\mathbf{a},\gamma}} \int_0^{\text{diam}(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma})} \epsilon^{-\frac{1}{2}} d\epsilon \\ &= 2k^{\frac{3}{4}} n^{-\frac{1}{4}} \sqrt{a_2 R_{k,\mathbf{a},\gamma}} \text{diam}(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}), \end{aligned} \quad (71)$$

$$= 2k^{\frac{3}{4}} n^{-\frac{1}{4}} \sqrt{a_2 R_{k,\mathbf{a},\gamma}} \text{diam}(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma}), \quad (72)$$

where, we used the inequality $\log(1+x) \leq x$ (for $x \geq -1$) in (71). It follows from Theorem 1 that there exists a constant C such that for $\delta > 0$,

$$\begin{aligned} \mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} Z_\theta \geq C E_{k,\mathbf{a},n,\gamma} + \delta\right) &= \mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} Z_\theta - Z_0 \geq C E_{k,\mathbf{a},n,\gamma} + \delta\right) \\ &\leq C e^{-\frac{\delta^2}{C \text{diam}(\Theta_k(\mathbf{a}), \mathbf{d}_{k,\mathbf{a},n,\gamma})^2}} = C e^{-\frac{n \delta^2}{4 C a_2^2 R_{k,\mathbf{a},\gamma}^2 k}}, \end{aligned} \quad (73)$$

where $Z_0 = 0$. It follows similarly that for $\delta > 0$,

$$\mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} -Z_\theta \geq \delta + C E_{k,\mathbf{a},n,\gamma}\right) \leq C e^{-\frac{n \delta^2}{4 C a_2^2 R_{k,\mathbf{a},\gamma}^2 k}}. \quad (74)$$

Combining (73) and (74) yields

$$\mathbb{P}\left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} |Z_\theta| \geq \delta + C E_{k,\mathbf{a},n,\gamma}\right) \leq 2C e^{-\frac{n \delta^2}{4 C a_2^2 R_{k,\mathbf{a},\gamma}^2 k}}. \quad (75)$$

From (61), (62) and (75), we obtain that for $\delta > 0$,

$$\begin{aligned} & \mathbb{P} \left(\left| H_{\gamma, \mathcal{G}_k(\mathbf{a})}(P, Q) - \hat{H}_{\gamma, \mathcal{G}_k(\mathbf{a})}(X^n, Y^n) \right| \geq \delta + CE_{k, \mathbf{a}, n, \gamma} \right) \\ & \leq \mathbb{P} \left(\sup_{g_\theta \in \mathcal{G}_k(\mathbf{a})} |Z_\theta| \geq \delta + CE_{k, \mathbf{a}, n, \gamma} \right) \leq 2Ce^{-\frac{n\delta^2}{4Ca_2^2 R_{k, \mathbf{a}, \gamma}^2}}. \end{aligned} \quad (76)$$

B Appendix: KL divergence

B.1 Proof of Theorem 4

Let $D_{\mathcal{G}_k(\mathbf{a}_k)}(P, Q) := H_{\gamma_{\text{KL}}, \mathcal{G}_k(\mathbf{a}_k)}(P, Q)$. The proof of Theorem 4 relies on the following lemma, whose proof is given in Appendix B.1.1.

Lemma 2. *Let $P, Q \in \mathcal{P}(\mathcal{X})$ such that $P \ll Q$ and $D_{\text{KL}}(P\|Q) < \infty$. Then, for $X^n \sim P^{\otimes n}$ and $Y^n \sim Q^{\otimes n}$, the following holds for any $\alpha > 0$:*

$$\begin{aligned} \text{(i)} \quad & \text{For } n, k_n, \mathbf{a}_{k_n} = (a_{1, k_n}, a_{2, k_n}, a_{3, k_n}) \text{ such that } k_n^{\frac{3}{2}} a_{2, k_n} e^{k_n a_{2, k_n} + a_{3, k_n}} = O\left(n^{\frac{1-\alpha}{2}}\right), \\ & \hat{D}_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(X^n, Y^n) \xrightarrow[n \rightarrow \infty]{} D_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(P, Q), \quad \mathbb{P} - \text{a.s.} \end{aligned} \quad (77)$$

$$\begin{aligned} \text{(ii)} \quad & \text{For } n, k, \mathbf{a}_k = (a_{1, k}, a_{2, k}, a_{3, k}) \text{ such that } k^{\frac{3}{2}} a_{2, k} e^{ka_{2, k} + a_{3, k}} = O\left(n^{\frac{1-\alpha}{2}}\right) \\ & \mathbb{E} \left[\left| \hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n, Y^n) - D_{\mathcal{G}_k(\mathbf{a}_k)}(P, Q) \right| \right] = O\left(n^{-\frac{1}{2}} k^{\frac{3}{2}} a_{2, k} e^{ka_{2, k} + a_{3, k}}\right). \end{aligned} \quad (78)$$

We proceed to prove (20). Since $f_{\text{KL}} \in \mathcal{C}(\mathcal{X})$ for a compact set \mathcal{X} , it follows from Stinchcombe and White (1990, Theorem 2.8) that for any $\epsilon > 0$ and $k \geq k_0(\epsilon)$, there exists a $g_{\bar{\theta}} \in \mathcal{G}_k(\mathbf{1})$ such that

$$\sup_{x \in \mathcal{X}} |f_{\text{KL}}(x) - g_{\bar{\theta}}(x)| \leq \epsilon. \quad (79)$$

This implies that

$$\lim_{k \rightarrow \infty} D_{\mathcal{G}_k(\mathbf{1})}(P, Q) = D_{\text{KL}}(P\|Q). \quad (80)$$

To see this, note that

$$D_{\mathcal{G}_k(\mathbf{1})}(P, Q) \leq D_{\text{KL}}(P\|Q), \quad \forall k \in \mathbb{N}, \quad (81)$$

by (18) since g_θ is continuous and bounded ($|g_\theta| \leq k+1$). Moreover, the left hand side (L.H.S.) of (81) is monotonically increasing in k , and being bounded, has a limit point. Then, (80) will follow if we show that the limit point is $D_{\text{KL}}(P\|Q)$. Assume otherwise that $\lim_{k \rightarrow \infty} D_{\mathcal{G}_k(\mathbf{1})}(P, Q) < D_{\text{KL}}(P\|Q)$. Note that $\mathcal{G}_k(\mathbf{1})$ is a closed set and hence the supremum in the variational form of the L.H.S. of (81) is a maximum. Then, defining

$$D(g) := 1 + \mathbb{E}_P[g(X)] - \mathbb{E}_Q \left[e^{g(Y)} \right], \quad (82)$$

this implies that there exists $\delta > 0$ and

$$g_{\theta_k^*} := \arg \max_{g_\theta \in \mathcal{G}_k(\mathbf{1})} D(g_\theta), \quad (83)$$

such that for all k ,

$$D_{\text{KL}}(P\|Q) - D(g_{\theta_k^*}) \geq \delta. \quad (84)$$

However, it follows from (79) that for all $k \geq k_0(\epsilon)$,

$$D_{\text{KL}}(P\|Q) - D(g_{\theta_k^*}) \leq D_{\text{KL}}(P\|Q) - D(g_{\bar{\theta}})$$

$$\begin{aligned} &\leq \mathbb{E}_P [|f_{\text{KL}}(X) - g_{\hat{\theta}}(X)|] + \mathbb{E}_Q [|e^{f_{\text{KL}}(Y)} - e^{g_{\hat{\theta}}(Y)}|] \\ &\leq \mathbb{E}_P [|f_{\text{KL}}(X) - g_{\hat{\theta}}(X)|] + L_{P,Q} \mathbb{E}_Q [|1 - e^{g_{\hat{\theta}}(Y) - f_{\text{KL}}(Y)}|] \end{aligned} \quad (85)$$

$$\leq \epsilon + L_{P,Q}(e^\epsilon - 1), \quad (86)$$

where (86) follows from (79). Note that

$$0 \leq L_{P,Q} := \left\| \frac{dP}{dQ} \right\|_\infty < \infty, \quad (87)$$

since $e^{f_{\text{KL}}}$ is a continuous function and hence bounded over a compact support \mathcal{X} . Taking ϵ sufficiently small in (86) contradicts (84), thus proving (80). Next, for $a_{3,k} = a_{2,k} = a_{1,k} = 1$ and any $\eta > 0$, $k^{\frac{3}{2}} a_{2,k} e^{k a_{2,k} + a_{3,k}} < e^{k(1+\eta)}$ provided k is sufficiently large. Then, (20) follows from (77) and (80) by letting $k = k_n \rightarrow \infty$ (subject to constraint in Lemma 2(i)), and noting that $\eta > 0$ is arbitrary.

Next, we prove (21). Note that since $f_{\text{KL}} \in \mathcal{I}(M)$, we have from (42) that for k such that $m_k \geq M$, there exists $g_\theta \in \mathcal{G}_k^*(m_k)$ satisfying

$$\|f_{\text{KL}} - g_\theta\|_{\infty, P, Q} \leq \tilde{C}_{d,M} k^{-\frac{1}{2}} = ((\hat{c}_d + 2)M + 2) k^{-\frac{1}{2}}.$$

On the other hand, for k such that $m_k < M$, taking $g_0 = 0$ yields $\|f_{\text{KL}} - g_0\|_{\infty, P, Q} \leq M$. Hence, for all k , there exists $g_{\theta_k^*} \in \mathcal{G}_k^*(m_k)$ such that

$$\|f_{\text{KL}} - g_{\theta_k^*}\|_{\infty, P, Q} \leq D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}, \quad (88)$$

where $\mathbf{m} = \{m_k\}_{k \in \mathbb{N}}$,

$$D_{d,M,\mathbf{m}} := \tilde{C}_{d,M} \vee \sqrt{\bar{m}(M, \mathbf{m})} M, \quad (89)$$

$$\bar{m}(M, \mathbf{m}) := \min \{k \in \mathbb{N} : m_k \geq M\}. \quad (90)$$

Also, observe that $D_{\text{KL}}(P \| Q) \geq D_{\mathcal{G}_k^*(m_k)}(P, Q)$ since $g_{\theta_k^*} \in \mathcal{G}_k^*(m_k)$ is bounded. Then, the following chain of inequalities hold:

$$\begin{aligned} &|D_{\text{KL}}(P \| Q) - D_{\mathcal{G}_k^*(m_k)}(P, Q)| \\ &= D_{\text{KL}}(P \| Q) - D_{\mathcal{G}_k^*(m_k)}(P, Q) \\ &\stackrel{(a)}{\leq} \mathbb{E}_P [|f_{\text{KL}}(X) - g_{\theta_k^*}(X)|] + L_{P,Q} \mathbb{E}_Q [|1 - e^{g_{\theta_k^*}(Y) - f_{\text{KL}}(Y)}|] \\ &\stackrel{(b)}{\leq} D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + e^M \left(e^{D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}} - 1 \right), \end{aligned} \quad (91)$$

where

(a) follows similar to (85);

(b) is due to (88) and $L_{P,Q} \leq e^M$ since $f_{\text{KL}} \in \mathcal{I}(M)$.

On the other hand, taking $a_{1,k} = \sqrt{k} \log k$, $k a_{2,k} = a_{3,k} = m_k$, and k satisfying $\sqrt{k} e^{2m_k} = O\left(n^{\frac{1-\alpha}{2}}\right)$ for some $\alpha > 0$, we have

$$\begin{aligned} &\mathbb{E} \left[|\hat{D}_{\mathcal{G}_k^*(m_k)}(X^n, Y^n) - D_{\text{KL}}(P \| Q)| \right] \\ &\stackrel{(a)}{\leq} |D_{\mathcal{G}_k^*(m_k)}(P, Q) - D_{\text{KL}}(P \| Q)| + \mathbb{E} \left[|D_{\mathcal{G}_k^*(m_k)}(P, Q) - \hat{D}_{\mathcal{G}_k^*(M)}(X^n, Y^n)| \right] \\ &\stackrel{(b)}{\leq} D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + e^M \left(e^{D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}} - 1 \right) + O\left(e^{2m_k} \sqrt{k} n^{-\frac{1}{2}}\right) \end{aligned} \quad (92)$$

$$\stackrel{(c)}{=} O_M \left(e^{D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}} - 1 \right) + O\left(e^{2m_k} \sqrt{k} n^{-\frac{1}{2}}\right), \quad (93)$$

where

(a) is due to triangle inequality;

(b) follows from (78) and (91).

Choosing $m_k = 0.5 \log k$ in (93) yields

$$\mathbb{E} \left[\left| \hat{D}_{\mathcal{G}_k^*(0.5 \log k)}(X^n, Y^n) - D_{\text{KL}}(P \| Q) \right| \right] = O(k^{-\frac{1}{2}}) + O(k^{\frac{3}{2}} n^{-\frac{1}{2}}), \quad (94)$$

since for k sufficiently large,

$$e^{D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}} - 1 = \sum_{j=1}^{\infty} \frac{\left(D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}\right)^j}{j!} \leq \sum_{j=1}^{\infty} \left(D_{d,M,\mathbf{m}} k^{-\frac{1}{2}}\right)^j = O(k^{-\frac{1}{2}}).$$

This completes the proof.

Remark 10. Setting $m_k = M$ in (93) and via steps leading to (94), we obtain (22).

B.1.1 Proof of Lemma 2

Note that for $\gamma_{\text{KL}}(x) = e^x - 1$,

$$\begin{aligned} \bar{\gamma}'_{\mathcal{G}_k(\mathbf{a}_k)} &= \sup_{\substack{x \in \mathcal{X}, \\ g_{\theta} \in \mathcal{G}_k(\mathbf{a}_k)}} \gamma'_{\text{KL}}(g_{\theta}(x)) \leq e^{ka_{2,k} + a_{3,k}}, \\ R_{k,\mathbf{a}_k,\gamma} &\leq 2\sqrt{k} (e^{ka_{2,k} + a_{3,k}} + 1), \end{aligned}$$

where γ'_{KL} denotes the derivative of γ_{KL} . Since

$$E_{k,\mathbf{a}_k,n,\gamma} \leq 4\sqrt{2}n^{-\frac{1}{2}}k^{\frac{3}{2}}a_{2,k}(e^{ka_{2,k} + a_{3,k}} + 1) \xrightarrow{n \rightarrow \infty} 0, \quad (95)$$

for k, \mathbf{a}_k such that $k^{\frac{3}{2}}a_{2,k}e^{ka_{2,k} + a_{3,k}} = O(n^{\frac{1-\alpha}{2}})$ for $\alpha > 0$, it follows from (17) that for any $k \in \mathbb{N}$, $\delta > 0$, and n sufficiently large,

$$\mathbb{P} \left(\left| D_{\mathcal{G}_k(\mathbf{a}_k)}(P, Q) - \hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n, Y^n) \right| \geq \delta \right) \leq 2Ce^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k}^2k^2(e^{ka_{2,k} + a_{3,k}} + 1)^2}}. \quad (96)$$

Hence, for k_n, \mathbf{a}_{k_n} such that $k_n^{\frac{3}{2}}a_{2,k_n}e^{k_na_{2,k_n} + a_{1,k_n}} = O(n^{\frac{1-\alpha}{2}})$,

$$\sum_{n=1}^{\infty} \mathbb{P} \left(\left| D_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(P, Q) - \hat{D}_{\mathcal{G}_{k_n}(\mathbf{a}_{k_n})}(X^n, Y^n) \right| \geq \delta \right) \leq 2C \sum_{n=1}^{\infty} e^{-\frac{n(\delta - CE_{k_n,\mathbf{a}_{k_n},n,\gamma})^2}{16Ca_{2,k_n}^2k_n^2(e^{k_na_{2,k_n} + a_{1,k_n}} + 1)^2}} < \infty, \quad (97)$$

where the final inequality in (97) can be established via integral test for sum of series. This implies (77) via the first Borel-Cantelli lemma. To prove (78), note that

$$\begin{aligned} &\mathbb{E} \left[\left| D_{\mathcal{G}_k(\mathbf{a}_k)}(P, Q) - \hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n, Y^n) \right| \right] \\ &= \int_0^{\infty} \mathbb{P} \left(\left| D_{\mathcal{G}_k(\mathbf{a}_k)}(P, Q) - \hat{D}_{\mathcal{G}_k(\mathbf{a}_k)}(X^n, Y^n) \right| \geq \delta \right) d\delta \\ &\leq CE_{k,\mathbf{a}_k,n,\gamma} + \int_{CE_{k,\mathbf{a}_k,n,\gamma}}^{\infty} 2Ce^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k}^2k^2(e^{ka_{2,k} + a_{3,k}} + 1)^2}} d\delta \\ &= O \left(n^{-\frac{1}{2}}k^{\frac{3}{2}}a_{2,k}e^{ka_{2,k} + a_{3,k}} \right). \end{aligned} \quad (98)$$

B.2 Proof of Proposition 1

From proof of Corollary 1 (see (53)), there exists extensions $f_p^{(e)}, f_q^{(e)} \in \mathcal{B}_{b' \vee c'} \cap \mathcal{S}_{s, b'}(\mathbb{R}^d)$ of f, \bar{f} , respectively (see (55) and (56) for definitions of b' and c'). Define $f_{\text{KL}}^{(e)} := f_p^{(e)} - f_q^{(e)}$. Note that since $f_p^{(e)}, f_q^{(e)} \in \mathcal{L}^1(\mathbb{R}^d) \cup \mathcal{L}^2(\mathbb{R}^d)$, their Fourier transforms exists. Hence, we have

$$B\left(f_{\text{KL}}^{(e)}\right) \stackrel{(a)}{\leq} B\left(f_p^{(e)}\right) + B\left(f_q^{(e)}\right) \stackrel{(b)}{\leq} 2(b' \vee c'), \quad (99)$$

$$\max_{x \in \mathcal{X}} \left| f_{\text{KL}}^{(e)}(x) \right| \leq \max_{x \in \mathcal{X}} \left| f_p^{(e)}(x) \right| + \max_{x \in \mathcal{X}} \left| f_q^{(e)}(x) \right| \stackrel{(d)}{\leq} 2b, \quad (100)$$

where

- (a) follows from the definition in (4) and linearity of the Fourier transform;
- (b) (c) is since $f_p^{(e)}, f_q^{(e)} \in \mathcal{B}_{b' \vee c'}$;
- (d) is due to $(P, Q) \in \mathcal{L}_{\text{KL}}(b, c)$.

Hence, it follows from (99)-(100) that $f_{\text{KL}}^{(e)}|_{\mathcal{X}} \in \mathcal{I}(M)$ with $M = 2\bar{c}_{b, c, d}$ (since $b \leq b'$), where $\bar{c}_{b, c, d}$ is given in (54). The claim then follows from Theorem 4 since $f_{\text{KL}} = f_{\text{KL}}^{(e)}|_{\mathcal{X}}$.

C Appendix: χ^2 divergence

C.1 Proof of Theorem 5

Let $\chi_{\mathcal{G}_k(\mathbf{a}_k)}^2(P, Q) := H_{\gamma_{\chi^2}, \mathcal{G}_k(\mathbf{a}_k)}(P, Q)$. The proof of Theorem 5 is based on the lemma below (see Appendix C.1.1 for proof).

Lemma 3. *Let $P, Q \in \mathcal{P}(\mathcal{X})$ such that $P \ll Q$ and $\chi^2(P\|Q) < \infty$. For $X^n \sim P^{\otimes n}$ and $Y^n \sim Q^{\otimes n}$, the following holds for any $\alpha > 0$:*

$$(i) \text{ For } n, k_n, \mathbf{a}_{k_n} \text{ such that } k_n^{\frac{5}{2}} a_{2, k_n}^2 + k_n^{\frac{3}{2}} a_{2, k_n} a_{3, k_n} = O\left(n^{\frac{1-\alpha}{2}}\right),$$

$$\hat{\chi}_{\mathcal{G}_k(\mathbf{a}_{k_n})}^2(X^n, Y^n) \xrightarrow{n \rightarrow \infty} \chi_{\mathcal{G}_k(\mathbf{a}_{k_n})}^2(P, Q), \quad \mathbb{P} - a.s. \quad (101)$$

$$(ii) \text{ For } n, k, \mathbf{a}_k \text{ such that } k^{\frac{5}{2}} a_{2, k}^2 + k^{\frac{3}{2}} a_{2, k} a_{3, k} = O\left(n^{\frac{1-\alpha}{2}}\right),$$

$$\mathbb{E} \left[\left| \hat{\chi}_{\mathcal{G}_k(\mathbf{a}_k)}^2(X^n, Y^n) - \chi_{\mathcal{G}_k(\mathbf{a}_k)}^2(P, Q) \right| \right] = O\left(n^{-\frac{1}{2}} \left(k^{\frac{5}{2}} a_{2, k}^2 + k^{\frac{3}{2}} a_{2, k} a_{3, k} \right)\right). \quad (102)$$

The proof of (25) follows from (101), using similar arguments used to establish (20) and steps leading to (104) below. The details are omitted.

We proceed to prove (26). Since $f_{\chi^2} \in \mathcal{I}(M)$, we have similar to (88) that there exists $g_{\theta_k^*} \in \mathcal{G}_k^*(m_k)$

$$\|f_{\chi^2} - g_{\theta_k^*}\|_{\infty, P, Q} = D_{d, M, \mathbf{m}} k^{-\frac{1}{2}}, \quad (103)$$

where $D_{d, M, \mathbf{m}}$ is defined in (89). Also, $\chi^2(P\|Q) \geq \chi_{\mathcal{G}_k^*(m_k)}^2(P, Q)$ since $g_{\theta} \in \mathcal{G}_k^*(m_k)$ is bounded. Then, we have

$$\begin{aligned} & \left| \chi^2(P\|Q) - \chi_{\mathcal{G}_k^*(m_k)}^2(P, Q) \right| \\ &= \chi^2(P\|Q) - \chi_{\mathcal{G}_k^*(m_k)}^2(P, Q) \\ &\leq \chi^2(P\|Q) - \mathbb{E}_P[g_{\theta_k^*}(X)] - \mathbb{E}_Q \left[g_{\theta_k^*}(Y) + \frac{g_{\theta_k^*}^2(Y)}{4} \right] \end{aligned}$$

$$\begin{aligned}
 &\leq \mathbb{E}_P [|f_{\chi^2}(X) - g_{\theta_k^*}(X)|] + \mathbb{E}_Q \left[|f_{\chi^2}(Y) - g_{\theta_k^*}(Y)| + \frac{1}{4} |f_{\chi^2}^2(Y) - g_{\theta_k^*}^2(Y)| \right] \\
 &\leq 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \mathbb{E}_Q \left[\frac{1}{4} |f_{\chi^2}(Y) - g_{\theta_k^*}(Y)| |f_{\chi^2}(Y) + g_{\theta_k^*}(Y)| \right] \\
 &\leq 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \mathbb{E}_Q \left[\frac{1}{4} |f_{\chi^2}(Y) - g_{\theta_k^*}(Y)| |g_{\theta_k^*}(Y) - f_{\chi^2}(Y)| + \frac{1}{2} |f_{\chi^2}(Y) - g_{\theta_k^*}(Y)| |f_{\chi^2}(Y)| \right] \\
 &\leq 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \frac{D_{d,M,\mathbf{m}}^2}{4k} + \frac{D_{d,M,\mathbf{m}} M}{2\sqrt{k}}, \tag{104}
 \end{aligned}$$

where (104) is due to $f_{\chi^2} \in \mathcal{I}(M)$. Taking $a_{1,k} = \sqrt{k} \log k$, $ka_{2,k} = a_{3,k} = m_k$, and k, m_k satisfying $m_k^2 \sqrt{k} = O(n^{(1-\alpha)/2})$, we have

$$\begin{aligned}
 &\mathbb{E} \left[\left| \hat{\chi}^2 \mathcal{G}_{k(m_k)}^*(X^n, Y^n) - \chi^2(P\|Q) \right| \right] \\
 &\stackrel{(a)}{\leq} \left| \chi_{\mathcal{G}_{k(m_k)}^*}^2(P, Q) - \chi^2(P\|Q) \right| + \mathbb{E} \left[\left| \chi_{\mathcal{G}_{k(m_k)}^*}^2(P, Q) - \hat{\chi}_{\mathcal{G}_{k(m_k)}^*}^2(X^n, Y^n) \right| \right] \\
 &\stackrel{(b)}{\leq} 2D_{d,M,\mathbf{m}} k^{-\frac{1}{2}} + \frac{D_{d,M,\mathbf{m}}^2}{4k} + \frac{D_{d,M,\mathbf{m}} M}{2\sqrt{k}} + O\left(m_k^2 \sqrt{k} n^{-\frac{1}{2}}\right), \tag{105} \\
 &\stackrel{(c)}{=} O_{d,M}\left(\bar{m}(M, \mathbf{m}) k^{-\frac{1}{2}}\right) + O\left(m_k^2 \sqrt{k} n^{-\frac{1}{2}}\right),
 \end{aligned}$$

where

(a) is due to triangle inequality;

(b) follows from (102) and (104);

(c) is by the definition of $D_{d,M,\mathbf{m}}$ in (89) and since $\bar{m}(M, \mathbf{m}) \geq 1$.

Setting $\mathbf{m} = \{0.5 \log k\}_{k \in \mathbb{N}}$ in (105) yields (26), thus completing the proof.

C.1.1 Proof of Lemma 3

For $\gamma_{\chi^2}(x) = x + \frac{x^2}{4}$, we have

$$\begin{aligned}
 \gamma'_{\mathcal{G}_k(\mathbf{a}_k)} &= \sup_{\substack{x \in \mathcal{X}, \\ g_{\theta} \in \mathcal{G}_k(\mathbf{a}_k)}} \gamma'_{\chi^2}(g_{\theta}(x)) \leq 0.5(ka_{2,k} + a_{3,k}) + 1, \\
 R_{k,\mathbf{a}_k,\gamma} &\leq 2\sqrt{k} (0.5(ka_{2,k} + a_{3,k}) + 2), \tag{106}
 \end{aligned}$$

where $\gamma'_{\chi^2}(\cdot)$ denotes the derivative of γ_{χ^2} . Since

$$0 \leq E_{k,\mathbf{a}_k,n,\gamma} \leq 4\sqrt{2} n^{-\frac{1}{2}} k^{\frac{3}{2}} a_{2,k} (0.5(ka_{2,k} + a_{3,k}) + 2) \xrightarrow{n \rightarrow \infty} 0, \tag{107}$$

for k, \mathbf{a}_k such that $k^{\frac{5}{2}} a_{2,k}^2 + k^{\frac{3}{2}} a_{2,k} a_{3,k} = O\left(n^{\frac{1-\alpha}{2}}\right)$, it follows from (17) that for any $k \in \mathbb{N}$, $\delta > 0$, and n sufficiently large,

$$\mathbb{P} \left(\left| \hat{\chi}_{\mathcal{G}_k(\mathbf{a}_k)}^2(X^n, Y^n) - \chi_{\mathcal{G}_k(\mathbf{a}_k)}^2(P, Q) \right| \geq \delta \right) \leq 2Ce^{-\frac{n(\delta - CE_{k,\mathbf{a}_k,n,\gamma})^2}{16Ca_{2,k}^2 k^2 (0.5(ka_{2,k} + a_{3,k}) + 2)^2}}. \tag{108}$$

Then, (101) and (102) follows using similar steps used to prove (77) (see (97)) and (78) (see (98)) in Theorem 4, respectively. This completes the proof.

C.2 Proof of Proposition 2

It follows from (53) that there exists extensions $f_p^{(e)}, f_q^{(e)} \in \mathcal{B}_{b' \vee c'} \cap \tilde{\mathcal{S}}_{s, b'}(\mathbb{R}^d)$ of $f, \bar{f} \in \mathcal{H}_{b, c}^{s, \delta}(\mathcal{U})$, respectively, where $\tilde{\mathcal{S}}_{s, b'}(\mathbb{R}^d)$ is defined in (52). Let $f_{\chi^2}^{(e)} = 2(f_p^{(e)} \cdot f_q^{(e)} - 1)$. Recall the notation α_j for a multi-index of order j . We have from the chain rule for differentiation that $D^{\alpha_j} f_{\chi^2}^{(e)}(x)$ is the sum of 2^j terms of the form $D^{\alpha_{j_1}} f_p^{(e)}(x) \cdot D^{\alpha_{j_2}} f_q^{(e)}(x)$, where $\alpha_{j_1} + \alpha_{j_2} = \alpha_j$. Also, note from (50) and (51) that for $j = 0, \dots, s$, $f_p^{(e)}, f_q^{(e)}$ satisfies

$$\left| D^{\alpha_j} f_p^{(e)}(x) \right| \vee \left| D^{\alpha_j} f_q^{(e)}(x) \right| \leq \hat{b} \leq b', \quad \forall x \in \mathbb{R}^d, \quad (109a)$$

$$\left\| D^{\alpha_j} f_p^{(e)} \right\|_{L_2(\mathbb{R}^d)} \vee \left\| D^{\alpha_j} f_q^{(e)} \right\|_{L_2(\mathbb{R}^d)} \leq b'. \quad (109b)$$

Then, it follows that for $j = 0, \dots, s$,

$$\begin{aligned} \left\| D^{\alpha_j} f_{\chi^2}^{(e)} \right\|_{L_2(\mathbb{R}^d)} &\leq 2 + 2 \left\| \sum_{\substack{\alpha_{j_1}, \alpha_{j_2}: \\ \alpha_{j_1} + \alpha_{j_2} = \alpha_j}} D^{\alpha_{j_1}} f_p^{(e)} \cdot D^{\alpha_{j_2}} f_q^{(e)} \right\|_{L_2(\mathbb{R}^d)} \\ &\leq 2 + 2^{j+1} b' \max_{\alpha_{j_2}} \left\| D^{\alpha_{j_2}} f_q^{(e)} \right\|_{L_2(\mathbb{R}^d)} \\ &\leq 2 + 2^{j+1} b'^2. \end{aligned} \quad (110)$$

Hence, $f_{\chi^2}^{(e)} \in \tilde{\mathcal{S}}_{s, 2+2^{s+1}b'^2}(\mathbb{R}^d)$. From Lemma 1, it follows that $B(f_{\chi^2}^{(e)}) \leq (2 + 2^{s+1}b'^2)\kappa_d \sqrt{d}$. Moreover, we have

$$\sup_{x \in \mathcal{X}} \left| f_{\chi^2}^{(e)} \right| \leq 2 + 2 \sup_{x \in \mathcal{X}} \frac{p(x)}{q(x)} \leq 2 + 2b^2. \quad (111)$$

This implies that $f_{\chi^2}^{(e)}|_{\mathcal{X}} \in \mathcal{I}\left((2 + 2^{s+1}b'^2)(\kappa_d \sqrt{d} \vee 1)\right)$ since $b' \geq b$. The claim then follows from Theorem 5 by noting that $f_{\chi^2} = f_{\chi^2}^{(e)}|_{\mathcal{X}}$ and $b'^2 \leq \tilde{c}_{b, c, d}^2$.

D Appendix: Squared Hellinger distance

D.1 Proof of Theorem 6

Let $H_{\tilde{\mathcal{G}}_k(\mathbf{a}_k, t)}^2(P, Q) := H_{\gamma_{H^2}, \tilde{\mathcal{G}}_k(\mathbf{a}_k, t)}(P, Q)$. The proof of Theorem 6 hinges on the following lemma, whose proof is given in Appendix D.1.1.

Lemma 4. *Let $P, Q \in \mathcal{P}(\mathcal{X})$ such that $P \ll Q$ and $H^2(P, Q) < \infty$. For $X^n \sim P^{\otimes n}$ and $Y^n \sim Q^{\otimes n}$, the following holds for any $\alpha > 0$:*

$$(i) \text{ For } n, k, \mathbf{a}_{k_n} \text{ such that } k_n^{\frac{3}{2}} a_{2, k_n} t_{k_n}^{-2} = O\left(n^{\frac{1-\alpha}{2}}\right),$$

$$\hat{H}_{\tilde{\mathcal{G}}_{k_n}(\mathbf{a}_{k_n}, t_{k_n})}^2(X^n, Y^n) \xrightarrow{n \rightarrow \infty} H_{\tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)}^2(P, Q), \quad \mathbb{P} - a.s. \quad (112)$$

$$(ii) \text{ For } n, k, \mathbf{a}_k \text{ such that } k^{\frac{3}{2}} a_{2, k} t_k^{-2} = O\left(n^{\frac{1-\alpha}{2}}\right),$$

$$\mathbb{E} \left[\left| \hat{H}_{\tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)}^2(X^n, Y^n) - H_{\tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)}^2(P, Q) \right| \right] = O\left(n^{-\frac{1}{2}} k^{\frac{3}{2}} a_{2, k} t_k^{-2}\right). \quad (113)$$

We first prove (31). Since $f_{H^2} \in \mathcal{C}(\mathcal{X})$ for a compact set \mathcal{X} , its supremum is achieved at some $x^* \in \mathcal{X}$. Also, since $\left\| \frac{dP}{dQ} \right\|_{\infty} < \infty$ by definition of the Radon-Nikodym derivative, we have $\sup_{x \in \mathcal{X}} f_{H^2}(x) = f_{H^2}(x^*) < 1$.

Moreover, $t_k \leq 1 - f_{H^2}(x^*)$ for sufficiently large k since $t_k \rightarrow 0$. Then, it follows from Stinchcombe and White (1990, Theorem 2.8) that for any $\epsilon > 0$ and $k \geq k_0(\epsilon)$ (some integer), there exists a $g_{\theta^*} \in \tilde{\mathcal{G}}_{k,t_k}^{(1)}$ such that

$$\sup_{x \in \mathcal{X}} |f_{H^2}(x) - g_{\theta^*}(x)| \leq \epsilon. \quad (114)$$

This implies similar to (80) in Theorem 4 that

$$\lim_{k \rightarrow \infty} H_{\tilde{\mathcal{G}}_{k,t_k}^{(1)}}^2(P, Q) = H^2(P, Q). \quad (115)$$

Then, (31) follows from (112) and (115).

Next, we prove (32). Since $f_{H^2} \in \mathcal{I}_{H^2}(M)$, $1 - f_{H^2}(x) \geq \frac{1}{M}$ for all $x \in \mathcal{X}$. Using $t_k \rightarrow 0$, we have from (12) that for k such that $t_k \leq \frac{1}{M}$ and $m_k \geq M$, there exists $g_{\theta} \in \tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}$ such that

$$\|f_{H^2} - g_{\theta}\|_{\infty, P, Q} \leq \tilde{C}_{d,M} k^{-\frac{1}{2}}. \quad (116)$$

On the other hand, for k such that $t_k > \frac{1}{M}$ or $m_k < M$, taking $g_0 = 0$ yields $\|f_{H^2} - g_0\|_{\infty, P, Q} \leq M$ as $f_{H^2} \in \mathcal{I}(M)$. Then, denoting $\mathbf{t} = \{t_k\}_{k \in \mathbb{N}}$, it follows similar to (88) that for all k , there exists $g_{\theta_k^*} \in \tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}$ such that

$$\|f_{H^2} - g_{\theta_k^*}\|_{\infty, P, Q} \leq \tilde{C}_{d,M} k^{-\frac{1}{2}} \vee \left(\sqrt{\bar{t}(M^{-1}, \mathbf{t})} \vee \sqrt{\bar{m}(M, \mathbf{m})} \right) M k^{-\frac{1}{2}} =: \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}}, \quad (117)$$

where $\bar{t}(M^{-1}, \mathbf{t}) := \inf\{k : t_k \leq M^{-1}\}$. Moreover, note that by definition, $H^2(P, Q) \geq H_{\tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}}^2(P, Q)$. Then, we have

$$\begin{aligned} & \left| H^2(P, Q) - H_{\tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}}^2(P, Q) \right| \\ &= H^2(P, Q) - H_{\tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}}^2(P, Q) \\ &\leq \mathbb{E}_P[f_{H^2}(X)] - \mathbb{E}_Q\left[\frac{f_{H^2}(Y)}{1 - f_{H^2}(Y)}\right] - \mathbb{E}_P[g_{\theta_k^*}(X)] + \mathbb{E}_Q\left[\frac{g_{\theta_k^*}(Y)}{1 - g_{\theta_k^*}(Y)}\right] \\ &\leq \mathbb{E}_P[|f_{H^2}(X) - g_{\theta_k^*}(X)|] + \mathbb{E}_Q\left[\left|\frac{f_{H^2}(Y)}{1 - f_{H^2}(Y)} - \frac{g_{\theta_k^*}(Y)}{1 - g_{\theta_k^*}(Y)}\right|\right] \\ &\leq \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + \mathbb{E}_Q\left[\left|\frac{f_{H^2}(Y) - g_{\theta_k^*}(Y)}{(1 - f_{H^2}(Y))(1 - g_{\theta_k^*}(Y))}\right|\right] \\ &\leq \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + M t_k^{-1} \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}}, \end{aligned} \quad (118)$$

where (118) is due to $1 - g_{\theta^*}(x) \geq t_k$, $(1 - f_{H^2}(x))^{-1} \leq M$ for all $x \in \mathcal{X}$, and (117).

Then, it follows from (113) and (118) that by taking $a_{1,k} = \sqrt{k} \log k$, $ka_{2,k} = a_{3,k} = m_k$, and $\sqrt{k} m_k t_k^{-2} = O(n^{(1-\alpha)/2})$ for some $\alpha > 0$, we have

$$\begin{aligned} & \mathbb{E} \left[\left| \hat{H}_{\tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}}^2(X^n, Y^n) - H^2(P, Q) \right| \right] \\ &\leq \left| H^2(P, Q) - H_{\tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}}^2(P, Q) \right| + \mathbb{E} \left[\left| \hat{H}_{\tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}}^2(X^n, Y^n) - H_{\tilde{\mathcal{G}}_{k,m_k,t_k}^{(2)}}^2(P, Q) \right| \right] \\ &\leq \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + M t_k^{-1} \bar{D}_{d,M,\mathbf{t},\mathbf{m}} k^{-\frac{1}{2}} + O\left(m_k \sqrt{k} t_k^{-2} n^{-\frac{1}{2}}\right) \\ &= O_{d,M}\left(\sqrt{\bar{t}(M^{-1}, \mathbf{t})} \vee \sqrt{\bar{m}(M, \mathbf{m})} t_k^{-1} k^{-\frac{1}{2}}\right) + O\left(m_k \sqrt{k} t_k^{-2} n^{-\frac{1}{2}}\right). \end{aligned} \quad (119)$$

Setting $m_k = 0.5 \log k$ and $t_k = \log^{-1} k$ in (119) yields (32), thus completing the proof.

D.1.1 Proof of Lemma 4

Note that Theorem 3 continues to hold with $\mathcal{G}_k(\mathbf{a})$ in (16) and (17) replaced with $\tilde{\mathcal{G}}_k(\mathbf{a}, t)$, since for $\gamma_{H^2}(x) = \frac{x}{1-x}$,

$$\bar{\gamma}'_{\tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)} = \sup_{\substack{x \in \mathcal{X}, \\ g_\theta \in \tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)}} \gamma'_{H^2}(g_\theta(x)) = \sup_{\substack{x \in \mathcal{X}, \\ g_\theta \in \tilde{\mathcal{G}}_k(\mathbf{a}_k, t_k)}} \frac{1}{(1 - g_\theta)^2} \leq \frac{1}{t_k^2},$$

where $\gamma'_{H^2}(\cdot)$ denotes the derivative of γ_{H^2} . This implies that $R_{k, \mathbf{a}_k, \gamma} \leq 2\sqrt{k}(t_k^{-2} + 1)$, and

$$0 \leq E_{k, \mathbf{a}_k, n, \gamma} \leq 4\sqrt{2}n^{-\frac{1}{2}}k^{\frac{3}{2}}a_{2,k}(t_k^{-2} + 1) \xrightarrow{n \rightarrow \infty} 0,$$

for k, \mathbf{a}_k, t_k such that $k^{\frac{3}{2}}a_{2,k}t_k^{-2} = O\left(n^{\frac{1-\alpha}{2}}\right)$. It then follows from (17) that for any $k \in \mathbb{N}$, $\delta > 0$, and n sufficiently large,

$$\mathbb{P}\left(\left|\hat{H}^2_{\mathcal{G}_k(\mathbf{a}_k)}(X^n, Y^n) - H^2_{\mathcal{G}_k(\mathbf{a}_k)}(P, Q)\right| \geq \delta\right) \leq 2Ce^{-\frac{n(\delta - CE_{k, \mathbf{a}_k, n, \gamma})^2}{16Ca_{2,k}^2k^2(t_k^{-2} + 1)^2}}.$$

Then, (112) and (113) follows using similar steps used to prove (77) (see (97)) and (78) (see (98)) in Theorem 4, respectively. This completes the proof.