A Appendix: Remaining Proofs

In this section, we present the proofs of Theorems 3.1 and 3.4.

A.1 Proof of Theorem 3.1

In the following, we quote a version of Talagrand’s inequality due to Bousquet (2002) from Steinwart and Christmann (2008, Theorem 7.5) and a (slightly simplified) bound on the expected supremum of empirical processes indexed by Vapnik-Červonenkis (VC) classes of functions, from Giné and Guillou (2001, Proposition 2.1). Both will be used to prove Theorem 3.1.

Theorem A.1. Let $(Z, P)$ be a probability space and $\mathcal{G}$ be a set of measurable functions from $Z$ to $\mathbb{R}$. Furthermore, let $B \geq 0$ and $\sigma \geq 0$ be constants such that $\mathbb{E}_P g = 0$, $\mathbb{E}_P g^2 \leq \sigma^2$, and $\|g\|_\infty \leq B$ for all $g \in \mathcal{G}$. For $n \geq 1$, define $G : Z^n \to \mathbb{R}$ by

$$G(z) := \sup_{g \in \mathcal{G}} \left| \frac{1}{n} \sum_{j=1}^{n} g(z_j) \right|, \quad z = (z_1, \ldots, z_n) \in Z^n.$$

Then, for all $\tau > 0$, we have

$$P^n\left( \left\{ z \in Z^n : G(z) \geq 4\mathbb{E}_P G + \sqrt{\frac{2\tau\sigma^2}{n}} + \frac{\tau B}{n} \right\} \right) \leq e^{-\tau}.$$ 

Theorem A.2. Let $(Z, P)$ be a probability space and $\mathcal{G}$ be a set of measurable functions from $Z$ to $\mathbb{R}$. Furthermore, let $B \geq 0$ and $0 \leq \sigma \leq B$ be constants such that $\mathbb{E}_P g^2 \leq \sigma^2$, and $\|g\|_\infty \leq B$ for all $g \in \mathcal{G}$. Suppose $\mathcal{G}$ is a uniformly bounded VC-class, i.e., there exist positive numbers $A$ and $\nu$ such that, for every probability measure $P$ on $Z$ and every $0 < \epsilon \leq B$, the covering numbers satisfy

$$\mathcal{N}(\mathcal{G}, L_2(P), \epsilon) \leq \left( \frac{AB}{\epsilon} \right)^\nu.$$ 

Then there exists a universal constant $C$ such that $G$ defined as in Theorem A.1 satisfies

$$\mathbb{E}_P G \leq C \left( \frac{\nu B}{n} \log \frac{AB}{\sigma} + \sqrt{\frac{\nu \sigma^2}{n} \log \frac{AB}{\sigma}} \right). \quad (15)$$

Given a measurable $g : \mathbb{R}^d \to \mathbb{R}$ and a $\delta > 0$ we define the function $g_\delta : \mathbb{R}^d \to \mathbb{R}$ by $g_\delta(x) := g(x/\delta)$, $x \in \mathbb{R}^d$. The following lemma, which establishes a stability of covering number bounds under this operation, will also be needed in the proof of Theorem 3.1.

Lemma A.3. Let $\mathcal{G}$ be set of measurable functions $g : \mathbb{R}^d \to \mathbb{R}$ such that there exists a constant $B \geq 0$ with $\|g\|_\infty \leq B$ for all $g \in \mathcal{G}$. For $\delta > 0$, we write $\mathcal{G}_\delta := \{g_\delta : g \in \mathcal{G}\}$. Then, for all $\epsilon \in (0, B]$ and all $\delta > 0$, we have

$$\sup_P \mathcal{N}(\mathcal{G}, L_2(P), \epsilon) = \sup_P \mathcal{N}(\mathcal{G}_\delta, L_2(P), \epsilon),$$

where the suprema are taken over all probability measures $P$ on $\mathbb{R}^d$.

Proof. We only prove “$\leq$”, the converse inequality can be shown analogously. Let us fix $\epsilon, \delta > 0$ and a distribution $P$ on $\mathbb{R}^d$. We define a new distribution $P'$ on $\mathbb{R}^d$ by $P'(A) := \frac{1}{\delta} P(\frac{A}{\delta})$ for all measurable $A \subset \mathbb{R}^d$. Furthermore, let $\mathcal{G}'$ be an $\epsilon$-net of $\mathcal{G}_\delta$ with respect to $L_2(P')$. For $\mathcal{G}' := \mathcal{G}'_{1/\delta}$, we then have $|\mathcal{G}'| = |\mathcal{G}'|$, and hence it suffices to show that $\mathcal{G}'$ is an $\epsilon$-net of $\mathcal{G}$ with respect to $L_2(P)$. To this end, we fix a $g \in \mathcal{G}$. Then $g_\delta \in \mathcal{G}_\delta$, and hence there exists an $h' \in \mathcal{G}'$ with $\|g_\delta - h\|_{L_2(P')} \leq \epsilon$. Moreover, we have $h := h_{1/\delta} \in \mathcal{G}'$, and since the definition of $P'$ ensures $\mathbb{E}_{P} g_\delta = \mathbb{E}_{P'} h'$ for all measurable $f : \mathbb{R}^d \to [0, \infty)$, we obtain

$$\|g - h\|_{L_2(P)} = \|g_\delta - h\|_{L_2(P')} = \|g_\delta - h'\|_{L_2(P')} \leq \epsilon,$$

i.e. $\mathcal{G}'$ is an $\epsilon$-net of $\mathcal{G}$ with respect to $L_2(P)$.

We further need the following result, which is a reformulation of van der Vaart and Wellner (1996, Theorem 2.6.4).

Theorem A.4. Let $A$ be a set of subsets of $Z$ that has finite VC-dimension $V$. Then the corresponding set of indicator functions $\mathcal{G} := \{1_A : A \in A\}$ is a uniformly bounded VC-class and the corresponding VC-characteristics $A$ and $\nu$ only depend on $V$.

With these preparation we are now able to establish the following generalization of Theorem 3.1. Applying this generalization to $K$ of the form (3) immediately proves Theorem 3.1.

Proposition A.5. Let $P$ be a probability measure on $\mathbb{R}^d$ with a bounded Lebesgue density $h$ and $K$ be a real-valued function on $X$ such that $K \in L_\infty(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$. Suppose that

$$\mathcal{F} := \{ K(x - \cdot) : x \in X \}$$

is a uniformly bounded VC-class. Then, there exists a positive constant $C$ only depending on $K$, $h$ and VC-characteristics $A$ and $\nu$ of $\mathcal{F}$ such that, for all $n \geq 1$, $\delta > 0$, and $\tau > 0$ we have

$$P^n\left( \left\{ x \in X^n : \left\| \tilde{h}_{K, \epsilon, \tau} - h_{K, \epsilon, \tau} \right\|_\infty < \frac{C}{n^d} \log \frac{C}{\delta} \right. \right.$$ 

$$\left. + \frac{C}{n^d} \log \frac{C}{\delta} + \frac{C}{n^d} \log \frac{C}{\sqrt{n^d}} \right) \geq 1 - e^{-\tau}.$$
Proof. Let us assume without loss of generality that \( \|f\|_\infty \leq 1 \). We define \( k_{x,\delta} := \delta^{-d} K(x, y) \) and \( f_{x,\delta} := k_{x,\delta} - E P k_{x,\delta} \). Then it is easy to check that 

\[
E P f_{x,\delta}^2 = 0 \quad \text{and} \quad \|f_{x,\delta}\|_\infty \leq 2\delta^{-d} \quad \forall x \in X \quad \text{and} \quad \delta > 0.
\]

Moreover, we have

\[
E P f_{x,\delta}^2 \leq \delta^{-d} \|K\|_2^2 \int K^2(\frac{x - y}{\delta}) \, dy \leq \delta^{-d} \|h\|_\infty \|K\|_2^2
\]

for all \( x \in X \) and \( \delta > 0 \), where the norm \( \|K\|_2 \) is with respect to the Lebesgue measure on \( \mathbb{R}^d \). In addition, we have

\[
\frac{1}{n} \sum_{i=1}^{n} f_{x,\delta}(x_i) = \bar{h}_{D,\delta}(x) - \bar{h}_{P,\delta}(x),
\]

where \( \bar{h}_{P,\delta} \) and \( \bar{h}_{D,\delta} \) are defined in (1) and (2) respectively. Applying Theorem A.1 to \( \mathcal{G} := \{f_{x,\delta} : x \in \mathbb{R}^d\} \), we hence obtain, for all \( \delta > 0 \), \( \tau > 0 \), and \( n \geq 1 \), that

\[
\|\bar{h}_{D,\delta} - \bar{h}_{P,\delta}\|_\infty < 4E P^n \|h_{D,\delta} - h_{P,\delta}\|_\infty + \frac{2\sqrt{2\tau}}{n^d\delta} \leq 2\sqrt{\frac{2\tau}{n^d\delta}} + \frac{2\sqrt{2\tau}}{n^d\delta} \tag{16}
\]

holds with probability \( P^n \) not smaller than \( 1 - e^{-\tau} \). It thus remains to bound the term \( E P^n \|h_{D,\delta} - h_{P,\delta}\|_\infty \).

Note that since \( \mathcal{F} \) is a uniformly bounded VC-class, so is \( \mathcal{F} := \{f_{x,\delta} : f \in \mathcal{F}, \alpha \in [-1, 1]\} \); i.e. there exist positive numbers \( A \) and \( \nu \) such that

\[
\sup_{P} \mathcal{N}(\mathcal{F}, L_2(P), \epsilon) \leq \left( \frac{2A}{\nu} \right)^\nu
\]

for all \( 0 < \epsilon < 2 \). For \( \delta > 0 \), we further have \( \delta^d \mathcal{G} \subset \mathcal{F}_\delta \), and hence Lemma A.3 implies

\[
\mathcal{N}(\delta^d \mathcal{G}, L_2(P), \epsilon) \leq \mathcal{N}(\mathcal{F}_\delta, L_2(P), \epsilon) \leq \left( \frac{2A}{\nu} \right)^\nu
\]

for all probability measures \( P \) on \( \mathbb{R}^d \) and all \( 0 < \epsilon < 2 \).

Now, our very first estimates show that every \( g \in \mathcal{G} := \delta^d \mathcal{G} \) satisfies \( \|g\|_\infty \leq 2 \) and \( E P g^2 \leq \delta^d \|h\|_\infty \|K\|_2^2 \), and hence Theorem A.2 yields

\[
E P^n \sup_{g \in \mathcal{G}} \left\{ \frac{1}{n} \sum_{j=1}^{n} g(X_j) \right\} \leq C \left( \frac{2\nu}{n} \log \frac{2A}{\sqrt{\delta^d \|h\|_\infty \|K\|_2^2}} \right) \quad \text{and} \quad \left( \frac{2\nu}{n} \log \frac{2A}{\sqrt{\delta^d \|h\|_\infty \|K\|_2^2}} \right)
\]

Multiplying both sides by \( \delta^{-d} \), we obtain

\[
E P^n \|h_{D,\delta} - h_{P,\delta}\|_\infty \leq C \left( \frac{2\nu}{n\delta^d} \log \frac{2A}{\sqrt{\delta^d \|h\|_\infty \|K\|_2^2}} \right)
\]

which, when used in (16), yields the result.

\[\square\]
the stopping criterion of Algorithm 1 is satisfied, that is, $\rho^*(D) \leq \rho^* + \varepsilon^* + \eta^* + 2\varepsilon + 2\eta$.

ii). Theorem 3.3 shows that in its last loop Algorithm 1 identifies exactly the topologically connected components of $M_{\rho^*(D),\delta}$ that belong to the set $\zeta_\varepsilon(C(M_{\rho^*(D)+\varepsilon+\eta}))$, where $\zeta_\varepsilon : C(M_{\rho^*(D)+\varepsilon+\eta}) \to C(M_{\rho^*(D),\delta})$ is the top-CCRM. Moreover, since Algorithm 1 stops at $\rho^*(D)$, we have $|\zeta_\varepsilon(C(M_{\rho^*(D)+\varepsilon+\eta}))| \neq 1$ and thus $|C(M_{\rho^*(D)+\varepsilon+\eta})| \neq 1$. From $\rho^*(D) + \varepsilon + \eta \leq \rho^{**}$ and $(c_1)$ we thus conclude that $|C(M_{\rho^*(D)+\varepsilon+\eta})| = 2$. For later purposes, note that the latter implies the injectivity of $\zeta_\varepsilon$. In addition, since $|C(M_{\rho^*(D)+\varepsilon+\eta})| = 2$, $(c_3)$ yields $\zeta_{+,\rho^*(D)+\varepsilon+\eta} : C(M_{\rho^*}) \to C(M_{\rho^*(D)+\varepsilon+\eta})$ is bijective. Since $\rho^*(D) + 3\varepsilon + 3\eta > \rho^*$, it follows from $(c_1)$–$(c_3)$ that we have $\zeta_{+,\rho^*(D)+\varepsilon+\eta} : C(M_{\rho^*}) \to C(M_{\rho^*(D)+3\varepsilon+3\eta})$ is bijective. Using the composition property of top-CCRMs in $(b_2)$, we obtain that $\zeta_{+,\rho^*(D)+\varepsilon+\eta} : C(M_{\rho^*(D)+3\varepsilon+3\eta}) \to C(M_{\rho^*(D)+\varepsilon+\eta})$ is bijective, and hence $|C(M_{\rho^*(D)+3\varepsilon+3\eta})| = 2$. Let us now consider the following commutative diagram:

\[
\begin{array}{ccc}
C(M_{\rho^*(D)+3\varepsilon+3\eta}) & \xrightarrow{\zeta_{+,\rho^*(D)+\varepsilon+\eta}} & C(M_{\rho^*(D)+\varepsilon+\eta}) \\
\zeta_\varepsilon \downarrow & & \downarrow \zeta_\varepsilon \\
C(M_{\rho^*(D)+2\varepsilon+2\eta,\delta}) & \xrightarrow{\zeta_f} & C(M_{\rho^*(D),\delta})
\end{array}
\]

where again, all occurring maps are the top-CCRMs between the respective sets. Now we have already shown that $\zeta_\varepsilon$ is injective and that $\zeta_{+,\rho^*(D)+\varepsilon+\eta}$ is bijective. Consequently, $\zeta_\varepsilon$ is injective.

iii). Follows from Theorem 3.3 and $\rho^*(D) + 2\varepsilon + 2\eta \leq \rho^{**} - 3\varepsilon - 3\eta$.

iv). Since $\rho^*(D) + 3\varepsilon + 3\eta > \rho^*(D) + \varepsilon + \eta > \rho^*$, by $(c_1)$–$(c_3)$, we see that the maps $\zeta_{-,\rho^*}$ and $\zeta_{+,\rho^*}$ are bijective. Therefore $\zeta_{+,\rho^*(D)+\varepsilon+\eta}$ is bijective and the diagram follows. \[\square\]