Appendix: Upper Bound for the Clustering Rademacher Complexity

Let $\mathcal{F}_C$ be a family of $k$-valued functions with
\[
\mathcal{F}_C := \{ f_C = (f_{c_1}, \ldots, f_{c_k}) : C \in \mathcal{H}^k \}. \tag{1}
\]
Let $\varphi : \mathbb{R}^k \to \mathbb{R}$ be a minimum function:
\[
\forall \alpha \in \mathbb{R}^k, \varphi(\alpha) = \min_{i=1,\ldots,k} \alpha_i \tag{2}
\]
and $\mathcal{G}_C$ be a “minimum” family of the functions $\mathcal{F}_C$,
\[
\mathcal{G}_C := \{ g_C = \varphi \circ f_C \mid f_C \in \mathcal{F}_C, g_C(x) = \varphi(f_C(x)) \}. \tag{3}
\]

**Definition 1** (Clustering Rademacher Complexity). Let $\mathcal{G}_C$ be a family of functions defined in (3). $S = (x_1, \ldots, x_n)$ be a fixed sample of size $n$ with elements in $\mathcal{X}$, and $D = \{ \Phi_i = \psi(x_i) \}_{i=1}^n$. Then, the clustering empirical Rademacher complexity of $\mathcal{G}_C$ with respect to $D$ is defined by
\[
\mathcal{R}_n(\mathcal{G}_C) = \mathbb{E}_\sigma \left[ \sup_{g_C \in \mathcal{G}_C} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i g_C(x_i) \right| \right],
\]
where $\sigma_1, \ldots, \sigma_n$ are independent random variables with equal probability of taking values $+1$ or $-1$. Its expectation is $\mathcal{R}(\mathcal{G}_C) = \mathbb{E} [ \mathcal{R}_n(\mathcal{G}_C) ]$.

Based on the recently improvement of the upper bound of Rademacher complexity of $L$-Lipschitz with respect to the $L_\infty$ norm [5], we provide a refined bound of clustering Rademacher complexity:

**Lemma 1.** If $\forall x \in \mathcal{X}, \| \Phi(x) \| \leq 1$, then, for any $S = \{ x_1, \ldots, x_n \} \in \mathcal{X}^n$, there exists a constant $c > 0$ such that
\[
\mathcal{R}_n(\mathcal{G}_C) \leq c \sqrt{k} \max_i \hat{\mathcal{R}}_n(\mathcal{F}_{C_i}) \log^2(\sqrt{n}),
\]
where $\mathcal{G}_C$ is a family of clustering functions defined in (3), $\mathcal{F}_C$ is a family of $k$-valued functions associate with the clustering center $C = [c_1, \ldots, c_k]$ defined in [1], $\mathcal{F}_{C_i}$ is a family of the output coordinate $i$ of $\mathcal{F}_C$, and $\hat{\mathcal{R}}_n(\mathcal{F}_{C_i}) = \sup_{S \in \mathcal{X}^n} \mathcal{R}_n(\mathcal{F}_{C_i})$.

The above result shows that the upper bound of the clustering Rademacher complexity is linearly dependent on $\sqrt{k}$, which substantially improves the existing bounds linearly dependent on $k$.

**Remark.** The upper bound of the clustering Rademacher complexity involves a constant $c$ and a logarithmic term $\log(n)$. Thus, if one requires its absolute value to be smaller than the existing bounds defined, there may exist some cases which acquire a large $k$. However, from a statistical perspective, our bound with linear dependence on $\sqrt{k}$ substantially improves the existing ones with linear dependence on $k$.

In the following, we will show that Lemma 1 cannot be improved from a statistical view when ignoring the logarithmic terms.
Then, we have

Without loss of generality, we assume that

ϕ

which does not match the upper bound of

To prove Lemma 1, we first give the following two lemmas:

Proof of Lemma 1. We first show that the minimum function

being the associated norm. Let

ρ

∀

φ

has been given in [5], a lower bound linearly dependent on

Remark. A lower bound linearly dependent on

for a

k

Lemma 2. There exists a set

C

such that

\( R_n(\mathcal{G}_C) \geq \frac{\sqrt{k}}{3\sqrt{2}} \cdot \max_i \tilde{R}_n(\mathcal{F}_{C_i}) \).

Lemma 2 shows that the lower bound of \( R_n(\mathcal{G}_C) \) is \( \Omega(\sqrt{k} \max_i \tilde{R}_n(\mathcal{F}_{C_i})) \), which implies that the upper bound of order \( O(\sqrt{k} \max_i \tilde{R}_n(\mathcal{F}_{C_i})) \) in Lemma 1 is (nearly) optimal when ignoring the logarithmic terms.

Appendix: Proof of Lemma 1

To prove Lemma 1, we first give the following two lemmas:

Lemma 3 (L∞ Contraction Inequality, Theorem 1 in [5]). Let \( \mathcal{F} \subseteq \{ f : \mathcal{X} \to \mathbb{R}^k \} \), and let \( \phi : \mathbb{R}^k \to \mathbb{R} \) be L-Lipschitz with respect to the \( L_\infty \) norm, that is \( \|\phi(\nu) - \phi(\nu')\|_\infty \leq L \cdot \|\nu - \nu'\|_\infty \). For any \( a > 0 \), there exists a constant \( C > 0 \) such that if \( \max_i \{\|f(x_i)\|, \|f(x)\|_\infty\} \leq \rho \), then

\[
R_n(\phi \circ \mathcal{F}) \leq C \cdot L \sqrt{k} \max_i \tilde{R}_n(\mathcal{F}_i) \log^{3/4} \left( \frac{\rho m}{\max_i \tilde{R}_n(\mathcal{F}_i)} \right),
\]

where \( R_n(\phi \circ \mathcal{F}) = \mathbb{E}_{\sigma} \left[ \sup_{f \in \mathcal{F}} \left\{ \sum_{i=1}^n \sigma_i \phi(f(x_i)) \right\} \right], \tilde{R}_n(\mathcal{F}_i) = \sup_{S \subseteq \mathcal{X}} \mathbb{R}_n(\mathcal{F}_i) \).

Lemma 4 (Lemma 24(a) in [7] with \( p = 2 \)). Let \( \eta_1, \ldots, \eta_n \in \mathcal{H} \), where \( \mathcal{H} \) is a Hilbert space with \( \|\cdot\| \) being the associated norm. Let \( \sigma_1, \ldots, \sigma_n \) be a sequence of independent Rademacher variables. Then, we have

\[
\mathbb{E}_{\sigma} \left\| \sum_{i=1}^n \sigma_i \eta_i \right\|^2 \leq \sum_{i=1}^n \|\eta_i\|^2 \quad (4)
\]

and

\[
\mathbb{E} \left\| \sum_{i=1}^n \sigma_i \eta_i \right\| \geq \sqrt{\frac{\sqrt{2}}{2}} \sqrt{\sum_{i=1}^n \|\eta_i\|^2}. \quad (5)
\]

Proof of Lemma 1. We first show that the minimum function

\( \varphi(\nu) = \min(\nu_1, \ldots, \nu_k) \)

defined in [2] is 1-Lipschitz continuous with respect to the \( L_\infty \) norm, that is

\[
\forall \nu, \nu' : \mathbb{R}^k, |\varphi(\nu) - \varphi(\nu')| \leq \|\nu - \nu'\|_\infty. \quad (6)
\]

Without loss of generality, we assume that \( \varphi(\nu) \geq \varphi(\nu') \). Let

\[ j = \arg \min_{i=1, \ldots, k} \nu_i', \]

then from the definition of \( \varphi \), we know that \( \varphi(\nu') = \nu_j' \). Thus, we can obtain that

\[
|\varphi(\nu) - \varphi(\nu')| = |\varphi(\nu) - \nu_j'| \leq \nu_j - \nu_j' \leq \nu_j - \nu_j \]

(by the fact that \( \varphi(\nu) \leq \nu_j \))
We then show that \( \max\{\|\varphi(f_C(x))\|, \|f_C(x)\|_\infty\} \) is bounded by a constant. From the definition of \( f_C \) (see Eq.(1)), we know that

\[
f_C(x) = (f_{c_1}(x), \ldots, f_{c_k}(x)) \quad \text{and} \quad f_{c_j}(x) = \|\Phi_x - c_j\|^2.
\]

Note that \( \|\Phi_x\| \leq 1 \) and \( c_j \in H \), so we have

\[
\|c_j\| \leq 1 \quad \text{and} \quad f_{c_j}(x) \leq 2\|\Phi_x\| + 2\|c_j\| \leq 4, \forall x \in X.
\]

(7)

Thus, one can see that \( \|f_C(x)\|_\infty = \max_j |f_{c_j}(x)| \leq 4 \) and \( |\varphi(f_C(x))| = |\min_{j=1,\ldots,k} f_{c_j}(x)| \leq 4. \)

From the above analysis, we know that \( \varphi(\nu) \) is 1-continuous with respect to the \( L_\infty \)-norm, and \( \max\{\|f_C(x)\|, \|f_C(x)\|_\infty\} \leq 4. \) Thus, using Lemma 3 with \( L = 1, \rho = 4 \) and \( a = 1/2 \), we have

\[
R_n(G_C) \leq C \sqrt{k \max_i \tilde{R}_n(F_{C_i}) \log \frac{4n}{\max_i \tilde{R}_n(F_{C_i})}}.
\]

(8)

Let

\[
c_i := \sup_{x \in X} \sup_{f_{c} \in F_{C_i}} |f_{c}(x)| \quad \text{and} \quad c = \max\{c_i, i = 1, \ldots, k\}.
\]

(9)

From (7), we know that \( c \) is a constant and \( c \leq 4. \) By definition of \( \tilde{R}_n(F_{C_i}) \), we can obtain that

\[
\forall j, \tilde{R}_n(F_{C_j}) = \sup_{S \in X^n} E_{\sigma} \left[ \sup_{f_{c} \in F_{C_j}} \left| \sum_{i=1}^n \sigma_i f_{c}(x_i) \right| \right]
\]

\[
\geq \sup_{x \in X} E_{\sigma} \left[ \sup_{f_{c} \in F_{C_j}} \left| \sum_{i=1}^n \sigma_i f_{c}(x) \right| \right]
\]

\[
\geq \sup_{x \in X, f_{c} \in F_{C_j}} E_{\sigma} \left| \sum_{i=1}^n \sigma_i f_{c}(x) \right| \quad (\text{by Jensen’s inequality})
\]

(10)

\[
\geq \sqrt{\frac{2n}{2}} \sup_{x \in X, f_{c} \in F_{C_j}} \sqrt{|f_{c}(x)|} \quad (\text{by Eq.}(3) \text{ of Lemma 4})
\]

\[
= \sqrt{\frac{2nc_j}{2}} \quad (\text{by Eq.}(9)).
\]

Thus, one can see that \( \max_i \tilde{R}_n(F_{C_i}) \geq \sqrt{\frac{2nc}{2}} \), where \( c = \max\{c_i, i = 1, \ldots, k\}. \) So, we have \( \max_i \tilde{R}_n(F_{C_i}) \leq \sqrt{\frac{2n}{c}}. \) Plugging this into (8) proves the result.

\[\Box\]

**Appendix: Proof of Theorem 1**

To prove Theorem 1, we first give the following two lemmas:

**Lemma 5.** If \( \forall x \in X, \|\Phi_x\| \leq 1 \) then for all \( S \in X^n \) and \( C \in H^k \), we have

\[
\max_i \tilde{R}_n(F_{C_i}) \leq 3\sqrt{n}.
\]

To prove Theorem 1, we first give the following two lemmas:
Proof. \( \forall S \in \mathcal{X}^n, C \in \mathcal{H}^k \) and \( i \in \{1, \ldots, k\} \), we have

\[
\mathcal{R}_n(\mathcal{F}_{C_i}) = \mathbb{E}_\sigma \sup_{f_i \in \mathcal{F}_{C_i}} \left| \sum_{j=1}^n \sigma_j f_i(x_j) \right|
\]

\[
= \mathbb{E}_\sigma \sup_{c \in \mathcal{H}} \left| \sum_{j=1}^n \sigma_j \|\Phi_j - c\|^2 \right|
\]

\[
= \mathbb{E}_\sigma \sup_{c \in \mathcal{H}} \left| \sum_{j=1}^n \sigma_j \left[ -2\langle \Phi_j, c \rangle + \|c\|^2 + \|\Phi_j\|^2 \right] \right|
\]  \[ (11) \]

\[
\leq 2\mathbb{E}_\sigma \sup_{c \in \mathcal{H}} \left| \sum_{j=1}^n \sigma_j \langle \Phi_j, c \rangle \right| + \mathbb{E}_\sigma \sup_{c \in \mathcal{H}} \left| \sum_{j=1}^n \sigma_j \|c\|^2 \right|
\]

One can see that

\[
\mathbb{E}_\sigma \sup_{c \in \mathcal{H}} \left| \sum_{j=1}^n \sigma_j \|c\|^2 \right| \leq \mathbb{E}_\sigma \left| \sum_{j=1}^n \sigma_j \right| \quad \text{(since } \|c\| \leq 1) \]

\[
\leq \sqrt{\mathbb{E}_\sigma \left| \sum_{j=1}^n \sigma_j \right|^2} \leq \sqrt{n} \quad \text{(by Eq.\,(4) of Lemma\,4),} \quad \text{(12)}
\]

and

\[
\mathbb{E}_\sigma \sup_{c \in \mathcal{H}} \left| \sum_{j=1}^n \sigma_j \langle \Phi_j, c \rangle \right| = \mathbb{E}_\sigma \left| \sum_{j=1}^n \sigma_j \Phi_j \right| \quad \text{(by } \|c\| \leq 1) \]

\[
\leq \sqrt{\mathbb{E}_\sigma \left| \sum_{j=1}^n \sigma_j \Phi_j \right|^2} \leq \sqrt{\sum_{i=1}^n \|\Phi_i\|^2} \quad \text{(by Eq.\,(4) of Lemma\,4)} \quad \text{(13)}
\]

\[
\leq \sqrt{n} \quad \text{(since } \|\Phi_i\| \leq 1).
\]

Substituting (12) and (13) into (11), we can prove the result. \( \square \)

To prove Theorem 1, we first propose the following lemma:

**Lemma 6.** For any \( \delta \in (0, 1) \), with probability \( 1 - \delta \), there exists a constant \( c > 0 \), such that

\[
\mathcal{R}(\mathcal{G}_C) \leq c\sqrt{kn} \log^2 (\sqrt{n}) + \sqrt{2n \log \left( \frac{1}{\delta} \right)}.
\]

**Proof.** From (8) or (11), with probability \( 1 - \delta \), we have

\[
\mathcal{R}(\mathcal{G}_C) \leq \mathcal{R}_n(\mathcal{G}_C) + \sqrt{2n \log \left( \frac{1}{\delta} \right)}. \quad \text{(14)}
\]
Thus, we have
\[ \mathcal{R}(\mathcal{G}_c) \leq \mathcal{R}_n(\mathcal{G}_c) + \sqrt{2n \log \left( \frac{1}{\delta} \right)} \]
\[ \leq c \sqrt{k} \max_i \mathcal{R}_n(\mathcal{F}_c) \log^2 (\sqrt{n}) + \sqrt{2n \log \left( \frac{1}{\delta} \right)} \]  
(by Lemma 1)
\[ \leq 3c \sqrt{k} n \log^2 (\sqrt{n}) + \sqrt{2n \log \left( \frac{1}{\delta} \right)} . \]  
(by Lemma 5)

\[ \square \]

**Proof of Theorem 1.** The starting point of our analysis is the following elementary inequality (see Ch.8 in [4] or page 2 in [3]):
\[ \mathbb{E}[W(C_n, P)] - W^*(P) \]
\[ = \mathbb{E}[W(C_n, P) - W(C_n, P_n)] + \mathbb{E}[W(C_n, P_n)] - W^*(P) \]
\[ \leq \mathbb{E}[W(C_n, P) - W(C_n, P_n)] + \mathbb{E}[W(C^*, P_n)] - W^*(P) \]
\[ (W(C_n, P_n) \leq W(C^*, P_n) \text{ as } C_n \text{ is optimal w.r.t. } W(\cdot, P_n)) \]
\[ \leq \mathbb{E} \sup_{C \in \mathcal{H}^k} (W(C, P) - W(C, P_n)) + \mathbb{E} \sup_{C \in \mathcal{H}^k} |W(C, P_n) - W(C, P)| \]
\[ \leq 2 \mathbb{E} \sup_{C \in \mathcal{H}^k} |W(C, P_n) - W(C, P)|. \]  
(15)

Let \( x_1', \ldots, x_n' \) be a copy of \( x_1, \ldots, x_n \), independent of the \( \sigma_i \)'s. Then, by a standard symmetrization argument (11) (can also be seen in the proof of Lemma 4.3 of [3]), we can write
\[ \mathbb{E} \sup_{C \in \mathcal{H}^k} |W(C, P_n) - W(C, P)| \leq \mathbb{E} \sup_{g_C \in \mathcal{G}_C} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i [g_C(x) - g_C(x')] \right| \]
\[ \leq 2 \mathbb{E} \sup_{g_C \in \mathcal{G}_C} \left| \frac{1}{n} \sum_{i=1}^n \sigma_i g_C(x) \right| = \frac{2}{n} \mathcal{R}(\mathcal{G}_C). \]  
(16)

Thus, we can obtain that
\[ \mathbb{E}[W(C_n, P)] - W^*(P) \leq \frac{4}{n} \mathcal{R}(\mathcal{G}_C) \]  
(by Eq. (15) and Eq. (16))
\[ \leq 4c \sqrt{\frac{k}{n} \log^2 (\sqrt{n}) + 4 \sqrt{\frac{2 \log \frac{1}{\delta}}{n}}} \]  
(by Lemma 6).  

This proves the result.  
\[ \square \]

**Appendix: Proof of Theorem 2**

**Proof.** Note that
\[ \mathbb{E}[W(\tilde{C}_n, P)] - W^*(P) \]
\[ = \mathbb{E}_{A_1}[W(\tilde{C}_n, P) - W(\tilde{C}_n, P_n)] + \mathbb{E}_{A_2}[W(\tilde{C}_n, P_n) - W(C_n, P_n)] \]
\[ + \mathbb{E}_{A_3}[W(C_n, P_n) - W(C_n, P)] + \mathbb{E}_{A_4}[W(C_n, P) - W^*(P)]. \]
Also note that $A_2$ is bounded by $\zeta$, and $A_4$ can be obtained from Theorem 1. From Eq. (16), we know that $A_1$ and $A_3$ can be bounded by the Rademacher complexity:

\[ A_1 \leq E \sup_{C \in \mathcal{H}} |W(C, \mathbb{P}_n) - W(C, \mathbb{P})| \leq \frac{2}{n} R(G_C), \]

\[ A_3 \leq E \sup_{C \in \mathcal{H}} |W(C, \mathbb{P}_n) - W(C, \mathbb{P})| \leq \frac{2}{n} R(G_C). \]

Thus, we can obtain that

\[ \mathbb{E}[|W(C_n, \mathbb{P})|] - W^*(\mathbb{P}) \leq \frac{4}{n} R(G_C) + c \sqrt{\frac{k}{n} \log^2 (\sqrt{n})} + c \sqrt{\frac{\log \frac{1}{\delta}}{n}} + \zeta. \]  

(17)

Substituting Lemma 6 into Eq. (17), we can prove the result.

Appendix: Proof of Theorem 3

Proof. Note that

\[ \mathbb{E} \left[ E_A[W(C_n^A, \mathbb{P})] \right] = \mathbb{E} \left[ E_A[W(C_n^A, \mathbb{P}) - E_A[W(C_n^A, \mathbb{P}_n)]] \right] + \mathbb{E} \left[ E_A[W(C_n^A, \mathbb{P}_n)] \right]. \]

From Lemma 2, we can obtain that

\[ \mathbb{E} \left[ E_A[W(C_n^A, \mathbb{P}_n)] \right] \leq \beta \cdot \mathbb{E}[W(C_n, \mathbb{P}_n)] = \beta \cdot \mathbb{E} \left[ W(C_n, \mathbb{P}_n) - W(C_n, \mathbb{P}) \right] + \beta \cdot \mathbb{E}[W(C_n, \mathbb{P})]. \]

Thus, we can obtain that

\[ \mathbb{E} \left[ E_A[W(C_n^A, \mathbb{P})] \right] \leq \mathbb{E} \left[ E_A[W(C_n^A, \mathbb{P}) - E_A[W(C_n^A, \mathbb{P}_n)]] \right] + \beta \cdot \mathbb{E} \left[ W(C_n, \mathbb{P}_n) - W(C_n, \mathbb{P}) \right] + \beta \cdot \mathbb{E}[W(C_n, \mathbb{P})]. \]  

(18)

Note that

\[ A_1, A_2 \leq E \sup_{C \in \mathcal{H}} |W(C, \mathbb{P}_n) - W(C, \mathbb{P})| \leq \frac{2}{n} R(G_C) \leq \tilde{O} \left( \sqrt{\frac{k}{n}} \right). \]  

(19)

By Theorem 1, we can obtain that

\[ \mathbb{E}[W(C_n^A, \mathbb{P})] \leq W^*(\mathbb{P}) + c \sqrt{\frac{k}{n} \log^2 (\sqrt{n})} + c \sqrt{\frac{\log \frac{1}{\delta}}{n}}. \]

Substituting the above inequality and Eq. (19) into Eq. (18), we have

\[ \mathbb{E} \left[ E_A[W(C_n^A, \mathbb{P}_n)] \right] \leq \tilde{O} \left( \sqrt{\frac{k}{n}} + W^*(\mathbb{P}) \right). \]  

□
Appendix: Proof of Theorem 4

To prove Theorem 4, we first propose the following lemma:

**Lemma 7.** With probability at least $1 - \delta$, we have

$$\mathbb{E}[\mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P}_n) - \mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P})] \leq \tilde{O}\left(\sqrt{\frac{k}{n}}\right).$$

**Proof.** Note that

$$\mathbb{E}[\mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P}_n) - \mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P})] \leq \mathbb{E} \sup_{\mathcal{C} \in H_k} |\mathcal{W}(\mathcal{C}, \mathcal{P}_n) - \mathcal{W}(\mathcal{C}, \mathcal{P})|$$

$$\leq \frac{2}{n} R(\mathcal{G}_C) \quad \text{(by Eq. (16))}$$

$$= \tilde{O}\left(\sqrt{\frac{k}{n}}\right) \quad \text{(by Lemma 6)}.$$

This proves the result. □

**Lemma 8.** If constructing $\mathcal{I}$ by uniformly sampling

$$m \geq C \sqrt{n} \log(1/\delta) \min(k, \Xi)/\sqrt{k},$$

then for all $S \in \mathcal{X}^n$, with probability at least $1 - \delta$, we have

$$\mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P}_n) - \mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P}) \leq C \sqrt{\frac{k}{n}},$$

where $\Xi = \text{Tr}(K_n(K_n + \mathbb{I}_n)^{-1})$ is the effective dimension of $K_n$, and $C$ is a constant.

**Proof.** This can be directly proved by combining Lemma 1 and Lemma 2 of [2] by setting $\varepsilon = 1/2$. □

**Proof of Theorem 4.** Note that

$$\mathbb{E}[\mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P})] - \mathcal{W}^*(\mathcal{P})
= \mathbb{E}[\mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P}) - \mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P}_n)] + \mathbb{E}[\mathcal{W}(\mathcal{C}_{n,m}, \mathcal{P}_n) - \mathcal{W}(\mathcal{C}_n, \mathcal{P}_n)]
+ \mathbb{E}[\mathcal{W}(\mathcal{C}_n, \mathcal{P}_n) - \mathcal{W}(\mathcal{C}_n, \mathcal{P})] + \mathbb{E}[\mathcal{W}(\mathcal{C}_n, \mathcal{P}) - \mathcal{W}^*(\mathcal{P})].$$

Note that

$$A_3 \leq \mathbb{E} \sup_{\mathcal{C} \in H_k} |\mathcal{W}(\mathcal{C}, \mathcal{P}_n) - \mathcal{W}(\mathcal{C}, \mathcal{P})|$$

$$\leq \frac{2}{n} R(\mathcal{G}_C) \quad \text{(by Eq. (16))} \quad (20)$$

$$\leq \tilde{O}\left(\sqrt{\frac{k}{n}}\right). \quad \text{(by Lemma 6)}$$

One can see that $A_4$ can be bounded by $\tilde{O}(\sqrt{k/n})$ using Theorem 1. $A_1$ and $A_2$ can both be bounded as $\tilde{O}(\sqrt{k/n})$ using Lemma 7 and Lemma 8 respectively. □
Appendix: Proof of Theorem 5

Proof. From the definition of effective dimension, we have

\[ \Xi = \text{Tr}(K^T(K + I)^{-1}) = \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_i + 1} \]

\[ = \sum_{i=1}^{\sqrt{k}} \frac{\lambda_i}{\lambda_i + 1} + \sum_{i=\sqrt{k}+1}^{n} \frac{\lambda_i}{\lambda_i + 1} \leq \sum_{i=1}^{\sqrt{k}} 1 + \sum_{i=\sqrt{k}+1}^{n} \lambda_i \]

\[ \leq \sqrt{k} + \sum_{i=\sqrt{k}+1}^{n} \lambda_i \leq \sqrt{k} + \sum_{i=\sqrt{k}+1}^{n} \frac{c}{\alpha} \leq \sqrt{k} + c \int_{\sqrt{k}}^{\infty} x^{-\alpha} dx = \sqrt{k} + \frac{c}{\alpha - 1} \sqrt{k} - \alpha \]

\[ \leq \left(1 + \frac{c}{\alpha - 1}\right) \sqrt{k} . \]

Thus, we can obtain that

\[ \frac{\min(k, \Xi)}{\sqrt{k}} \leq \frac{\Xi}{\sqrt{k}} \leq 1 + \frac{c}{\alpha - 1} . \]

Substituting the above inequality into Theorem 4, we can prove this result. \qed

Appendix: Proof of Theorem 6

Proof. Note that

\[ \mathbb{E}[\mathbb{E}[W(C_{m,n}, P) - W^*(P)]] = \mathbb{E}[\mathbb{E}[W(C_{m,n}, P) - W(C_{m,n}, P_n)] + \mathbb{E}[W(C_{m,n}, P_n) - W(C_{m,n}, P_n)] + \mathbb{E}[W(C_{m,n}, P) - W^*(P)] . \]

Also note that \( A_2 \) is bounded by \( \zeta \), \( A_4 \) can be obtained from Theorem 5, and \( A_1 \) and \( A_3 \) can be bounded by the Rademacher complexity:

\[ A_1, A_3 \leq \mathbb{E} \sup_{C \in \mathcal{G}_c} |W(C, P_n) - W(C, P)| \leq \frac{2}{n} \mathcal{R}(\mathcal{G}_c) . \]

Thus, we can obtain that

\[ \mathbb{E}[W(C_{m,n}, P) - W^*(P)] = \mathcal{O} \left( \frac{\mathcal{R}(\mathcal{G}_c)}{n} + \sqrt{\frac{k}{n}} + \zeta \right) . \quad (21) \]

Substituting Lemma 6 into Eq. (21), we can prove this result. \qed

Appendix: Proof of Theorem 7

Proof. Note that

\[ \mathbb{E}[\mathbb{E}_A[W(C_{n,m}^{A}, P)]] = \mathbb{E}[\mathbb{E}_A[W(C_{n,m}^{A}, P) - \mathbb{E}_A[W(C_{n,m}^{A}, P_n)]] + \mathbb{E}[\mathbb{E}_A[W(C_{n,m}^{A}, P_n)]] . \]

By Lemma 2, we can obtain that

\[ \mathbb{E}[\mathbb{E}_A[W(C_{n,m}^{A}, P)]] \leq \beta \cdot \mathbb{E}[W(C_{n,m}, P)] \]

\[ = \beta \cdot \mathbb{E}[W(C_{n,m}, P_n) - W(C_{n,m}, P)] + \beta \cdot \mathbb{E}[W(C_{n,m}, P)] . \]
Thus, we can obtain that

\[
\mathbb{E} \left[ E_A \mathcal{W}(C_{n,m}, \mathbb{P}) \right] \\
\leq \mathbb{E} \left[ E_A \mathcal{W}(C_{n,m}, \mathbb{P}) - E_A \mathcal{W}(C_{n,m}, \mathbb{P}) \right] \\
+ \beta \cdot \mathbb{E} \left[ \mathcal{W}(C_{n,m}, \mathbb{P}) - \mathcal{W}(C_{n,m}, \mathbb{P}) \right] + \beta \cdot \mathbb{E} \left[ \mathcal{W}(C_{n,m}, \mathbb{P}) \right].
\]

Note that

\[
A_1, A_2 \leq \mathbb{E} \sup_{C \in \mathcal{H}} |\mathcal{W}(C, \mathbb{P}) - \mathcal{W}(C, \mathbb{P})| \\
\leq \frac{2}{n} \mathcal{R}(\mathcal{C}) \quad \text{(by Eq. (16))}
\]

\[
= \tilde{O} \left( \sqrt{\frac{k}{n}} \right) \quad \text{(by Lemma 6)}.
\]

By Corollary 5, \( A_3 \) can be bounded:

\[
A_3 = \mathbb{E} \left[ \mathcal{W}(C_{n,m}, \mathbb{P}) \right] \leq \mathcal{W}^* (\mathbb{P}) + c \sqrt{\frac{k}{n}} \log^2 (\sqrt{n}).
\]

This proves the result.

**Appendix: Proof of Lemma 2**

We first prove that the maximum Rademacher complexity can be bounded by \( 3\sqrt{n} \). Then, following the same idea as [5] and using the Khintchine inequality [6], we show that there exists a hypothesis function \( \mathcal{F}_C \) such that \( \mathcal{R}_n(\mathcal{C}) \geq \sqrt{\frac{2k}{n}} \).

**Lemma 9** (Khintchine inequality with \( p = 1 \) in [6]). Let \( \sigma_1, \ldots, \sigma_n \) be Rademacher variables with equal probability of taking values \( +1 \) or \( -1 \). Then, we have \( \mathbb{E}_{\sigma} |\sum_{i=1}^{n} \sigma_i| \geq \sqrt{\frac{k}{2}} \).

**Proof of Lemma 2** Let \( \epsilon_1, \ldots, \epsilon_k \) be independent random variables with equal probability of taking values \( +1 \) or \( -1 \). Let \( \mathcal{C} = (\epsilon_1 \nu_1, \ldots, \epsilon_k \nu_k) \), where \( \nu_i \) is the \( i \)th standard basis function in \( \mathcal{H} \), that is \( \langle \nu_i, \nu_j \rangle = 1 \) if \( i = j \), otherwise 0. We choose the hypothesis space

\[
\mathcal{F}_C = \left\{ f_C = (f_{\epsilon_1 \nu_1}, \ldots, f_{\epsilon_k \nu_k}) \mid f_{\epsilon_i \nu_i} (\mathbb{X}) = \| \Phi - \epsilon_i \nu_i \|^2, \epsilon_i \in \{ \pm 1 \}^k \right\}.
\]

Assume that \( n \) is divisible by \( k \). We set \( \Phi_1, \ldots, \Phi_{n/k} = \nu_1, \Phi_{(n+1)/k}, \ldots, \Phi_{2n/k} = \nu_2, \ldots \), and so on, and let \( i_t \) be the index such that \( \Phi_t = \nu_{i_t} \). Let \( \sigma^t \in \{ \pm 1 \}^n \) be Rademacher variables. From the
definition of clustering Rademacher complexity, we can obtain that
\[ R_n(G_C) = R_n(\varphi \circ F_C) \]
\[ = \mathbb{E}_{\sigma' \in \{\pm 1\}^n} \sup_{e \in \{\pm 1\}^k} \left| \sum_{t=1}^{n} \sigma'_t \min_{1 \leq i \leq k} \left\| \Phi_t - \epsilon_i \nu_i \right\|^2 \right| \]
\[ = \mathbb{E}_{\sigma' \in \{\pm 1\}^n} \sup_{e \in \{\pm 1\}^k} \left| \sum_{t=1}^{n} \sigma'_t \min_{1 \leq i \leq k} (2 - 2(\Phi_t, \epsilon_i \nu_i)) \right| \]
(since \( \Phi_t = \nu_i \) and \( \nu_i \) is the \( i \)-th standard basis function in \( H \))
\[ = 2 \mathbb{E}_{\sigma' \in \{\pm 1\}^n} \sup_{e \in \{\pm 1\}^k} \left| \sum_{t=1}^{n} \sigma'_t \max_{1 \leq i \leq k}(\Phi_t, \epsilon_i \nu_i) \right| \]
\[ = 2 \mathbb{E}_{\sigma' \in \{\pm 1\}^n} \sup_{e \in \{\pm 1\}^k} \left| \sum_{t=1}^{n} \sigma'_t \max\{\epsilon_i, 0\} \right| \]
\[ \geq 2 \mathbb{E}_{\sigma' \in \{\pm 1\}^n} \sup_{e \in \{\pm 1\}^k} \left| \sum_{t=1}^{n} \sigma'_t \max\{\epsilon, 0\} \right| \]
\[ = 2k \cdot \mathbb{E}_{\sigma' \in \{\pm 1\}^{n/k}} \sup_{e \in \{\pm 1\}^k} \left| \sum_{t=1}^{n/k} \sigma'_t \max\{\epsilon, 0\} \right| \]
\[ = 2k \cdot \mathbb{E}_{\sigma' \in \{\pm 1\}^{n/k}} \left| \sum_{t=1}^{n/k} \sigma'_t \max\{\epsilon, 0\} \right| \geq k \sqrt{\frac{n}{2k}} \quad \text{(by Lemma 9)} \]
\[ = \sqrt{\frac{n}{2}}. \]

From Lemma 5 we know that
\[ \max_i \mathcal{R}_n(F_{C_i}) \leq 3 \sqrt{n}. \]
Thus, by the above upper bounds the lower bound (Eq. (23)), we can prove that there exists a hypothesis space \( F_C \) defined in (22), such that
\[ R_n(G_C) \geq \frac{\sqrt{k}}{3 \sqrt{2}} \cdot \max_i \mathcal{R}_n(F_{C_i}). \]
This proves the result. \( \square \)

References


