Supplementary Materials for "Adversarial Learning with Local Coordinate Coding"

Lemma 2 *Let $(\boldsymbol{\gamma}, \mathcal{C})$ be an arbitrary coordinate coding on $\mathbb{R}^{d_B}$. Given an $(L_{\mathbf{h}}, L_G)$-Lipschitz smooth generator $G_u(\mathbf{h})$ and an $L_{\mathbf{x}}$-Lipschitz discriminator D_v, for all $\mathbf{h} \in \mathbb{R}^{d_B}$:*

$$\left| D_v(G_u(\mathbf{h})) - D_v\left(\sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v})\right)\right| \leq L_{\mathbf{x}} L_{\mathbf{h}} \|\mathbf{h} - \mathbf{r}(\mathbf{h})\|_2 + L_{\mathbf{x}} L_G \sum_{\mathbf{v} \in \mathcal{C}} |\gamma_{\mathbf{v}}(\mathbf{h})| \|\mathbf{v} - \mathbf{r}(\mathbf{h})\|_2^2.$$

Proof Given an $(L_{\mathbf{h}}, L_G)$-Lipschitz smooth generator $G_u(\mathbf{h})$, an $L_{\mathbf{x}}$-Lipschitz discriminator D_v, and let $\gamma_{\mathbf{v}} = \gamma_{\mathbf{v}}(\mathbf{h})$ and $\mathbf{h}' = \mathbf{r}(\mathbf{h}) = \sum_{\mathbf{v} \in \mathcal{C}} \gamma_{\mathbf{v}} \mathbf{v}$. We have

$$\begin{aligned}
&\left| \widetilde{D}_v(G_u(\mathbf{h})) - \widetilde{D}_v\left(\sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v})\right)\right| \\
=&\left| D_v(G_u(\mathbf{h})) - D_v\left(\sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v})\right)\right| \\
=&\left| D_v(G_u(\mathbf{h})) - D_v(G_u(\mathbf{h}')) - \left(D_v\left(\sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v})\right) - D_v(G_u(\mathbf{h}'))\right)\right| \\
\leq& \left| D_v\left(G_u(\mathbf{h})\right) - D_v\left(G_u(\mathbf{h}')\right)\right| + \left| D_v\left(\sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v})\right) - D_v\left(G_u(\mathbf{h}')\right)\right| \\
\leq& L_{\mathbf{x}} \left\| G_u(\mathbf{h}) - G_u(\mathbf{h}')\right\|_2 + L_{\mathbf{x}} \left\| \sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v}) - G_u(\mathbf{h}')\right\|_2 \\
\leq& L_{\mathbf{x}} \left\| G_u(\mathbf{h}) - G_u(\mathbf{h}')\right\|_2 + L_{\mathbf{x}} \left\| \sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) \left(G_u(\mathbf{v}) - G_u(\mathbf{h}') - \Delta G_u(\mathbf{h}')^{\mathsf{T}} \left(\mathbf{v} - \mathbf{h}'\right)\right)\right\|_2 \\
\leq& L_{\mathbf{x}} \left\| G_u(\mathbf{h}) - G_u(\mathbf{h}')\right\|_2 + L_{\mathbf{x}} \sum_{\mathbf{v} \in \mathcal{C}} |\gamma_{\mathbf{v}}| \left\| G_u(\mathbf{v}) - G_u(\mathbf{h}') - \Delta G_u(\mathbf{h}')^{\mathsf{T}} (\mathbf{v} - \mathbf{h}')\right\|_2 \\
\leq& L_{\mathbf{x}} L_{\mathbf{h}} \|\mathbf{h} - \mathbf{h}'\|_2 + L_{\mathbf{x}} L_G \sum_{\mathbf{v} \in \mathcal{C}} |\gamma_{\mathbf{v}}| \|\mathbf{v} - \mathbf{h}'\|_2^2 \\
=& L_{\mathbf{x}} L_{\mathbf{h}} \|\mathbf{h} - \mathbf{r}(\mathbf{h})\|_2 + L_{\mathbf{x}} L_G \sum_{\mathbf{v} \in \mathcal{C}} |\gamma_{\mathbf{v}}| \|\mathbf{v} - \mathbf{r}(\mathbf{h})\|_2^2,
\end{aligned}$$

where $\widetilde{D}_v(\cdot) = 1 - D_v(\cdot)$. In the above derivation, the first inequality holds by the triangle inequality. The second inequality uses an assumption that D_v is Lipschitz smooth w.r.t. the input. The third inequality uses the facts that $\sum_{\mathbf{v} \in \mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{x}) = 1$ and $\mathbf{h}' = \sum_{\mathbf{v} \in \mathcal{C}} \gamma_{\mathbf{v}} \mathbf{v}$. The last inequality uses the $(L_{\mathbf{h}}, L_G)$-Lipschitz smooth generator G_u, that is

$$\left\| G_u(\mathbf{v}) - G_u(\mathbf{h}') - \Delta G_u(\mathbf{h}')^{\mathsf{T}} (\mathbf{v} - \mathbf{h}')\right\|_2 \leq L_G \|\mathbf{v} - \mathbf{h}'\|_2^2.$$

□

8. Proof of Lemma 1

Lemma 1 (Generator Approximation) *Let $(\boldsymbol{\gamma}, \mathcal{C})$ be an arbitrary coordinate coding on $\mathbb{R}^{d_B}$. Given a Lipschitz smooth generator $G_u(\mathbf{h})$, for all $\mathbf{h} \in \mathbb{R}^{d_B}$:*

$$\left\| G_u\left(\sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h})\mathbf{v}\right) - \sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v}) \right\|_2 \leq 2L_{\mathbf{h}}\|\mathbf{h}-\mathbf{r}(\mathbf{h})\|_2 + L_G \sum_{\mathbf{v}\in\mathcal{C}} |\gamma_{\mathbf{v}}(\mathbf{h})| \|\mathbf{v}-\mathbf{r}(\mathbf{h})\|_2^2.$$

Proof From Lemma 2, when the discriminator is identity function: $D_v(t) = t$, that is

$$\begin{aligned}\left| D_v(G_u(\mathbf{h})) - D_v\left(\sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v})\right)\right| &= \left\| G_u(\mathbf{h}) - \sum_{\mathbf{v}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v})\right\|_2 \\ &\leq L_{\mathbf{h}}\|\mathbf{h}-\mathbf{r}(\mathbf{h})\|_2 + L_G \sum_{\mathbf{v}\in\mathcal{C}} |\gamma_{\mathbf{v}}| \|\mathbf{v}-\mathbf{r}(\mathbf{h})\|_2^2,\end{aligned}$$

then, we have

$$\begin{aligned}\left\| G_u\left(\sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h})\mathbf{v}\right) - \sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v}) \right\|_2 &= \left\| G_u\left(\sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h})\mathbf{v}\right) - G_u(\mathbf{h}) + G_u(\mathbf{h}) - \sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v}) \right\|_2 \\ &\leq \left\| G_u\left(\sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h})\mathbf{v}\right) - G_u(\mathbf{h}) \right\|_2 + \left\| G_u(\mathbf{h}) - \sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h}) G_u(\mathbf{v}) \right\|_2 \\ &\leq 2L_{\mathbf{h}}\|\mathbf{h}-\mathbf{r}(\mathbf{h})\|_2 + L_G \sum_{\mathbf{v}\in\mathcal{C}} |\gamma_{\mathbf{v}}(\mathbf{h})| \|\mathbf{v}-\mathbf{r}(\mathbf{h})\|_2^2,\end{aligned}$$

where $\mathbf{r}(\mathbf{h}) = \sum_{\mathbf{v}\in\mathcal{C}} \gamma_{\mathbf{v}}(\mathbf{h})\mathbf{v}$. □

9. Proof of Theorem 1

In order to provide a generalization bound w.r.t. the neural net distance, we first give some relevant lemmas and theorems. When the latent points lie on a latent manifold and the generator is Lipschitz smooth, $Q_{L_{\mathbf{h}}, L_G}(\boldsymbol{\gamma}, \mathcal{C})$ has a bound as follows.

Lemma 3 (Manifold Coding (Yu et al., 2009)) *If the latent points lie on a compact smooth manifold $\mathcal{M}$, given an $(L_{\mathbf{h}}, L_G)$-Lipschitz smooth generator $G_u(\mathbf{h})$ and any $\epsilon > 0$, then there exist anchor points $\mathcal{C} \subset \mathcal{M}$ and coding $\boldsymbol{\gamma}$ such that*

$$Q_{L_{\mathbf{h}}, L_G}(\boldsymbol{\gamma}, \mathcal{C}) \leq \left[L_{\mathbf{h}} c_{\mathcal{M}} + \left(1 + \sqrt{d_{\mathcal{M}}} + 4\sqrt{d_{\mathcal{M}}}\right) L_G\right] \epsilon^2.$$

Lemma 3 shows that the complexity of local coordinate coding depends on the intrinsic dimension of the manifold instead of the dimension of the basis.

Theorem 1 *Suppose measuring function $\phi(\cdot)$ is Lipschitz smooth: $|\phi'(\cdot)| \leq L_\phi$, and bounded in $[-\Delta, \Delta]$. Consider coordinate coding $(\boldsymbol{\gamma}, \mathcal{C})$, an example set $\mathcal{H}$ in latent space and the empirical distribution $\widehat{\mathcal{D}}_{real}$, if the generator is Lipschitz smooth , then the expected generalization error satisfies the inequality:*

$$\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}(\boldsymbol{\gamma}(\mathbf{h}))}, \widehat{\mathcal{D}}_{real}\right)\right] \leq \inf_{\mathcal{G}} \mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u(\mathbf{h})}, \widehat{\mathcal{D}}_{real}\right)\right] + \epsilon(d_{\mathcal{M}}),$$

where $\epsilon(d_{\mathcal{M}}) = L_\phi Q_{L_{\mathbf{h}}, L_G}(\boldsymbol{\gamma}, \mathcal{C}) + 2\Delta$, and generative quality $Q_{L_{\mathbf{h}}, L_G}(\boldsymbol{\gamma}, \mathcal{C})$ has an upper bound w.r.t. $d_{\mathcal{M}}$ in Lemma 3 of supplementary material.

Proof Let $\mathcal{H}^{(k)} = \left\{\mathbf{h}_1^{(k)}, \mathbf{h}_2^{(k)}, \ldots, \mathbf{h}_r^{(k)}\right\}$ be a set of r latent samples which lie on the latent distribution. Consider $n+1$ independent experiments over the latent distribution, we have $\mathcal{H}_{r,n+1} = \left\{\mathcal{H}^{(1)}, \mathcal{H}^{(2)}, \ldots, \mathcal{H}^{(n+1)}\right\}$. Recall the optimization problem, we consider an empirical version of the expected loss:

$$[\widetilde{w}] = \arg\min_{[w]} \left[\frac{1}{n}\sum_{i=1}^{n+1} d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_{w,\mathcal{H}^{(i)}}(\boldsymbol{\gamma}(\mathbf{h}))}, \widehat{\mathcal{D}}_{real}\right)\right]. \tag{5}$$

Let k be an integer randomly drawn from $\{1, 2, \ldots, n+1\}$. Let $\left[\widehat{w}^{(k)}\right]$ be the solution of

$$\left[\widehat{w}^{(k)}\right] = \arg\min_{[w]} \left[\frac{1}{n}\sum_{i\neq k}^{n+1} d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_{w,\mathcal{H}^{(i)}}(\boldsymbol{\gamma}(\mathbf{h}))}, \widehat{\mathcal{D}}_{real}\right)\right], \tag{6}$$

with the k-th example left-out.

Recall the definition of the neural net distance, we have

$$d_{\mathcal{F},\phi}(\mu,\nu) = \sup_{\mathcal{F}}\left|\mathop{\mathbb{E}}_{\mathbf{x}\sim\mu}[\phi(D_v(\mathbf{x}))] + \mathop{\mathbb{E}}_{\mathbf{x}\sim\nu}\left[\phi(\widetilde{D}_v(\mathbf{x}))\right]\right|,$$

where $\mathcal{F} = \{D_v, v \in \mathcal{V}\}$. Given the k-th sample experiment, the same real distribution $\widehat{\mathcal{D}}_{real}$ over the training samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$, and two different distributions generated by $G_{\widehat{w}^{(k)},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))$ and $G_{\widetilde{w},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))$, respectively, the difference value of the neural net distance between these two generated distributions is:

$$\begin{aligned}
&d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}^{(k)},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))}, \widehat{\mathcal{D}}_{real}\right) - d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widetilde{w},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))}, \widehat{\mathcal{D}}_{real}\right)\\
=&\sup\left|\mathbb{E}_{\mathbf{x}\in\widehat{\mathcal{D}}_{real}}[\phi(D_v(\mathbf{x}))] + \mathbb{E}_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v\left(G_{\widehat{w}^{(k)},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))\right)\right)\right]\right|\\
&-\sup\left|\mathbb{E}_{\mathbf{x}\in\widehat{\mathcal{D}}_{real}}[\phi(D_v(\mathbf{x}))] + \mathbb{E}_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v\left(G_{\widetilde{w},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))\right)\right)\right]\right|\\
\leq&\sup\left|\mathbb{E}_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v\left(G_{\widehat{w}^{(k)},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))\right)\right)\right] - \mathbb{E}_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v\left(G_{\widetilde{w},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))\right)\right)\right]\right|\\
=&\sup\left|\frac{1}{\left|\mathcal{H}^{(k)}\right|}\sum_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v\left(G_{\widehat{w}^{(k)},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))\right)\right) - \phi\left(\widetilde{D}_v\left(G_{\widetilde{w},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))\right)\right)\right]\right| \leq 2\Delta,
\end{aligned}$$

where $\widetilde{D}_v(\cdot) = 1 - D_v(\cdot)$. In the above derivation, the first equality uses the definition of the neural net distance. The last inequality holds by the assumption that $\phi(\cdot)$ is L_ϕ-Lipschitz and bounded in $[-\Delta, \Delta]$.

By summing over k, and consider any fixed $G_u \in \mathcal{G}$, we obtain:

$$\begin{aligned}
\sum_{k=1}^{n+1} d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}^{(k)},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))}, \widehat{\mathcal{D}}_{real}\right) &\leq \sum_{k=1}^{n+1} d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widetilde{w},\mathcal{H}^{(k)}}(\boldsymbol{\gamma}(\mathbf{h}))}, \widehat{\mathcal{D}}_{real}\right) + 2(n+1)\Delta\\
&\leq \sum_{\mathbf{h}\in\mathcal{H}^{(k)},k=1}^{n+1} d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{\sum_{\mathbf{v}\in\mathcal{C}}\gamma_{\mathbf{v}}(\mathbf{h})G_u(\mathbf{v})}, \widehat{\mathcal{D}}_{real}\right) + 2(n+1)\Delta\\
&\leq \sum_{\mathbf{h}\in\mathcal{H}^{(k)},k=1}^{n+1} d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_u(\mathbf{h})}, \widehat{\mathcal{D}}_{real}\right) + \sum_{k=1}^{n+1} L_\phi Q_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C}) + 2(n+1)\Delta,
\end{aligned}$$

where $Q_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C}) = \mathbb{E}_{\mathbf{h}}\left[L_{\mathbf{h}}\|\mathbf{h}-\mathbf{r}(\mathbf{h})\|_2 + L_G\sum_{\mathbf{v}\in\mathcal{C}}|\gamma_{\mathbf{v}}|\|\mathbf{v}-\mathbf{r}(\mathbf{h})\|_2^2\right]$. In the above derivation, the second inequality holds since $\widetilde{w}$ is the minimizer of Problem (5). The third inequality follows from the concavity of $\phi(\cdot)$ and Lemma 1:

$$\begin{aligned}
d_{\mathcal{F},\phi}\left(\mathcal{D}_{\sum_{\mathbf{v}\in\mathcal{C},\mathbf{h}\in\mathcal{H}^{(k)}}\gamma_{\mathbf{v}}(\mathbf{h})G_u(\mathbf{v})}, \widehat{\mathcal{D}}_{real}\right) =&\sup\left|\mathbb{E}_{\mathbf{x}\in\widehat{\mathcal{D}}_{real}}[\phi(D_v(\mathbf{x}))] + \mathbb{E}_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v\left(\sum\nolimits_{\mathbf{v}\in\mathcal{C}}\gamma_{\mathbf{v}}(\mathbf{h})\,G_u(\mathbf{v})\right)\right)\right]\right|\\
\leq&\sup\left|\mathbb{E}_{\mathbf{x}\in\widehat{\mathcal{D}}_{real}}[\phi(D_v(\mathbf{x}))] + \mathbb{E}_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v(G_u(\mathbf{h})) + \widehat{Q}_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C})\right)\right]\right|\\
\leq&\sup\left|\mathbb{E}_{\mathbf{x}\in\widehat{\mathcal{D}}_{real}}[\phi(D_v(\mathbf{x}))] + \mathbb{E}_{\mathbf{h}\in\mathcal{H}^{(k)}}\left[\phi\left(\widetilde{D}_v(G_u(\mathbf{h}))\right)\right]\right| + L_\phi Q_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C})\\
=&d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u(\mathbf{h})}, \widehat{\mathcal{D}}_{real}\right) + L_\phi Q_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C}),
\end{aligned}$$

where $\widehat{Q}_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C}) = L_{\mathbf{h}}\|\mathbf{h}-\mathbf{r}(\mathbf{h})\|_2 + L_G\sum_{\mathbf{v}\in\mathcal{C}}|\gamma_{\mathbf{v}}|\|\mathbf{v}-\mathbf{r}(\mathbf{h})\|_2^2$ and $\mathbb{E}_{\mathbf{h}}\left[\widehat{Q}_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C})\right] = Q_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C})$. In the above derivation, the firth equality holds by the definition of the neural net distance. The first inequality because of Lemma 1 and the fact that $\phi(\cdot)$ is a concave measuring function in Definition 4. Here, we suppose $\phi(\cdot)$ is a monotonically increasing function. The second inequality holds by the following derivation:

$$
\begin{aligned}
&\left|\phi\left(\widetilde{D}_v\left(G_u(\mathbf{h})\right)+\widehat{Q}_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C})\right)-\phi\left(\widetilde{D}_v\left(G_u(\mathbf{h})\right)\right)\right| \\
\leq&\left|\phi'\left(\widetilde{D}_v\left(G_u(\mathbf{h})\right)\right)\left[\left(\widetilde{D}_v\left(G_u(\mathbf{h})\right)+\widehat{Q}_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C})\right)-\widetilde{D}_v\left(G_u(\mathbf{h})\right)\right]\right| \\
=&\left|\phi'\left(\widetilde{D}_v\left(G_u(\mathbf{h})\right)\right)\right|\widehat{Q}_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C}) \\
\leq& L_\phi\widehat{Q}_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C}),
\end{aligned}
$$

In the above derivation, the first inequality uses the concavity of measuring function $\phi(\cdot)$. The last inequality follows from that $|\phi'|\leq L_\phi$. Now by taking expectation w.r.t. $\mathcal{H}_{r,n+1}$, we obtain

$$
\begin{aligned}
&\mathbb{E}_{\mathcal{H}\subseteq\mathcal{H}_{r,n+1}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w},\mathcal{H}}(\boldsymbol{\gamma}(\mathbf{h}))},\widehat{\mathcal{D}}_{real}\right)\right] \\
\leq&\mathbb{E}_{\mathcal{H}\subseteq\mathcal{H}_{r,n+1}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{u,\mathbf{h}\in\mathcal{H}}(\mathbf{h})},\widehat{\mathcal{D}}_{real}\right)\right]+L_\phi Q_{L_{\mathbf{h}},L_G}(\boldsymbol{\gamma},\mathcal{C})+2\Delta.
\end{aligned}
$$

□

10. Proof of Theorem 2

Theorem 2 *Under the condition of Theorem 1, and given an empirical distribution $\widehat{\mathcal{D}}_{real}$ drawn from $\mathcal{D}_{real}$, then the following holds with probability at least $1-\delta$,*

$$
\left|\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\mathcal{D}_{real}\right)\right]-\inf_{\mathcal{G}}\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\mathcal{D}_{real}\right)\right]\right|\leq 2R_{\mathcal{X}}(\mathcal{F})+2\Delta\sqrt{\frac{2}{N}\log(\frac{1}{\delta})}+2\epsilon(d_{\mathcal{M}}),
$$

where $R_{\mathcal{X}}(\mathcal{F}) = \mathop{\mathbb{E}}_{\sigma,\mathcal{X}}\left[\sup_{\mathcal{F}}\frac{1}{N}\sum_{i=1}^{N}\sigma_i\phi\left(D_v(\mathbf{x}_i)\right)\right]$ and $\sigma_i\in\{-1,1\}, i=1,2,\ldots,m$ are independent uniform random variables.

Proof For the real distribution $\mathcal{D}_{real}$, we are interested in the generalization error in term of the following neural net distance:

$$
\begin{aligned}
&\left|\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\mathcal{D}_{real}\right)\right]-\inf_{\mathcal{G}}\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\mathcal{D}_{real}\right)\right]\right| \\
\leq&\left|\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\mathcal{D}_{real}\right)\right]-\mathbb{E}_{\mathcal{H}}\left[\inf_{\mathcal{G}}d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\mathcal{D}_{real}\right)\right]\right| \\
=&\left|\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\mathcal{D}_{real}\right)-d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\widehat{\mathcal{D}}_{real}\right)+d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\widehat{\mathcal{D}}_{real}\right)-\inf_{\mathcal{G}}d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\mathcal{D}_{real}\right)\right]\right| \\
\leq&\left|\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\mathcal{D}_{real}\right)-d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}},\widehat{\mathcal{D}}_{real}\right)+\inf_{\mathcal{G}}d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\widehat{\mathcal{D}}_{real}\right)-\inf_{\mathcal{G}}d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\mathcal{D}_{real}\right)+\epsilon(d_{\mathcal{M}})\right]\right| \\
\leq&2\mathbb{E}_{\mathcal{H}}\left[\sup_{\mathcal{G}}\left|d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\mathcal{D}_{real}\right)-d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u},\widehat{\mathcal{D}}_{real}\right)\right|+\epsilon(d_{\mathcal{M}})\right] \\
=&2\mathbb{E}_{\mathcal{H}}\left[\sup_{\mathcal{G}}\left|\sup_{D_v\in\mathcal{F}}\left|\mathop{\mathbb{E}}_{\mathbf{x}\in\mathcal{D}_{real}}\left[\phi\left(D_v(\mathbf{x})\right)\right]+\mathop{\mathbb{E}}_{\mathbf{x}\in\mathcal{D}_{G_u}}\left[\phi\left(\widetilde{D}_v(\mathbf{x})\right)\right]\right|-\sup_{D_v\in\mathcal{F}}\left|\mathop{\mathbb{E}}_{\mathbf{x}\in\widehat{\mathcal{D}}_{real}}\left[\phi\left(D_v(\mathbf{x})\right)\right]+\mathop{\mathbb{E}}_{\mathbf{x}\in\mathcal{D}_{G_u}}\left[\phi\left(\widetilde{D}_v(\mathbf{x})\right)\right]\right|\right|+\epsilon(d_{\mathcal{M}})\right] \\
\leq&2\sup_{D_v\in\mathcal{F}}\left|\mathop{\mathbb{E}}_{\mathbf{x}\in\mathcal{D}_{real}}\left[\phi\left(D_v(\mathbf{x})\right)\right]-\mathop{\mathbb{E}}_{\mathbf{x}\in\widehat{\mathcal{D}}_{real}}\left[\phi\left(D_v(\mathbf{x})\right)\right]\right|+2\epsilon(d_{\mathcal{M}}).
\end{aligned}
\tag{7}
$$

In the above derivation, the first inequality holds by by Jensen's inequality and the concavity of the infimum function. The second inequality holds by Theorem 1. The third inequality satisfies when we take supremum w.r.t. $G_u\in\mathcal{G}$. The last inequality uses the definition of the neural net distance and holds by triangle inequality. This reduces the problem to bounding the distance

$$d'_{\mathcal{F}}\left(\mathcal{D}_{real}, \widehat{\mathcal{D}}_{real}\right) := \sup_{D_v \in \mathcal{F}} \left| \mathbb{E}_{\mathbf{x} \in \mathcal{D}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] - \mathbb{E}_{\mathbf{x} \in \widehat{\mathcal{D}}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] \right|,$$

between the true distribution and its empirical distribution. This can be achieved by the uniform concentration bounds developed in statistical learning theory, and thus the distance $d'_{\mathcal{F}}\left(\mathcal{D}_{real}, \widehat{\mathcal{D}}_{real}\right)$ can be achieved by the Rademacher complexity. Let $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N \in \mathcal{X}$ be a set of N independent random samples in data space. We introduce a function

$$h\left(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\right) = \sup_{D_v \in \mathcal{F}} \left| \mathbb{E}_{\mathbf{x} \in \mathcal{D}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] - \mathbb{E}_{\mathbf{x} \in \widehat{\mathcal{D}}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] \right|.$$

Since measuring function ϕ is Lipschitz and bounded in $[-\Delta, \Delta]$, changing $\mathbf{x}_i$ to another independent sample $\mathbf{x}'_i$ can change the function h by no more than $\frac{4\Delta}{N}$, that is,

$$h\left(\mathbf{x}_1, \ldots, \mathbf{x}_i \ldots, \mathbf{x}_N\right) - h\left(\mathbf{x}_1, \ldots, \mathbf{x}'_i, \ldots, \mathbf{x}_N\right) \leq \frac{4\Delta}{N},$$

for all $i \in [1, N]$ and any points $\mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{x}'_i \in \mathcal{X}$. McDiarmid's inequality implies that with probability at least $1 - \delta$, the following inequality holds:

$$\begin{aligned} &\sup_{D_v \in \mathcal{F}} \left| \mathbb{E}_{\mathbf{x} \in \mathcal{D}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] - \mathbb{E}_{\mathbf{x} \in \widehat{\mathcal{D}}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] \right| \\ \leq &\mathbb{E}\left[\sup_{D_v \in \mathcal{F}} \left| \mathbb{E}_{\mathbf{x} \in \mathcal{D}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] - \mathbb{E}_{\mathbf{x} \in \widehat{\mathcal{D}}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] \right| \right] + 2\Delta \sqrt{\frac{2\log\left(\frac{1}{\delta}\right)}{N}}. \end{aligned} \tag{8}$$

From the bound on Rademacher complexity, we have

$$\begin{aligned} &\mathbb{E}\left[\sup_{D_v \in \mathcal{F}} \left| \mathbb{E}_{\mathbf{x} \in \mathcal{D}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] - \mathbb{E}_{\mathbf{x} \in \widehat{\mathcal{D}}_{real}} \left[\phi\left(D_v(\mathbf{x})\right)\right] \right| \right] \\ \leq &2\mathbb{E}_{\sigma, \mathcal{X}} \left[\sup_{D_v \in \mathcal{F}} \frac{1}{N} \sum_{i=1}^{N} \sigma_i \phi\left(D_v(\mathbf{x}_i)\right) \right] = 2R_{\mathcal{X}}(\mathcal{F}). \end{aligned} \tag{9}$$

Combining the inequalities (7), (8) and (9), we have

$$\mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\widehat{\mathcal{D}}_{G_{\widehat{w}}}, \mathcal{D}_{real}\right)\right] - \inf_{G_u} \mathbb{E}_{\mathcal{H}}\left[d_{\mathcal{F},\phi}\left(\mathcal{D}_{G_u}, \mathcal{D}_{real}\right)\right] \leq 2R_{\mathcal{X}}(\mathcal{F}) + 2\Delta\sqrt{\frac{2\log(\frac{1}{\delta})}{N}} + 2\epsilon(d_{\mathcal{M}}).$$

□