A APPENDIX

A.1 PROPERTIES OF EQUIVARIANT MAPS

Theorem 15 (Degree of Freedom of Equivariant Maps). *Let a group G act on sets $\mathcal{S}$ and $\mathcal{T}$, and $\mathcal{B} \subset \mathcal{T}$ a base space. Then, a G-equivariant map $F : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$ can be represented using its generator as*

$$F[x](t) = F_{\mathcal{B}}[g_t^{-1} \cdot x](P_{\mathcal{B}}(t)), \tag{9}$$

where $g_t \in G$ is an arbitrary element which satisfies $g_t \cdot P_{\mathcal{B}}(t) = t$. Conversely, for an arbitrary map $F_{\mathcal{B}} : \mathcal{C}(\mathcal{S}) \to \mathcal{C}(\mathcal{B})$, a map F defined by (9) is an equivariant map whose generator equals $F_{\mathcal{B}}$.

[Proof] . For any $x \in \mathbb{R}^{\mathcal{S}}$ and $t \in \mathcal{T}$, the following holds:

$$\begin{aligned} F_{\mathcal{B}}[g_t^{-1} \cdot x](P_{\mathcal{B}}(t)) &= F_{\mathcal{B}}[g_t^{-1} \cdot x](g_t^{-1} \cdot t) \\ &= F[g_t^{-1} \cdot x](g_t^{-1} \cdot t) \\ &= (g_t^{-1} \cdot F[x])(g_t^{-1} \cdot t) \\ &= F[x](t), \end{aligned} \tag{10}$$

where $g_t \in G$ is an arbitrary element which satisfies $g_t \cdot P_{\mathcal{B}}(t) = t$ and the third equality follows from the equivariance of F.

Conversely, for an arbitrary map $F_{\mathcal{B}} : \mathcal{C}(\mathcal{S}) \to \mathcal{C}(\mathcal{B})$, a map F defined by (9) is an equivariant map whose generator equals $F_{\mathcal{B}}$ as follows:

$$\begin{aligned} (g \cdot F[x])(t) &= F[x](g^{-1} \cdot t) \\ &= F_{\mathcal{B}}[g_{g^{-1} \cdot t}^{-1} \cdot x](P_{\mathcal{B}}(g^{-1} \cdot t)) \\ &= F_{\mathcal{B}}[g_{g^{-1} \cdot t}^{-1} \cdot x](P_{\mathcal{B}}(t)) \\ &= F_{\mathcal{B}}[(g g_t^{-1}) \cdot x](P_{\mathcal{B}}(t)) \\ &= F_{\mathcal{B}}[g_t^{-1} \cdot (g \cdot x)](P_{\mathcal{B}}(t)) \\ &= F[g \cdot x](t), \end{aligned}$$

where we used $g_{g^{-1} \cdot t} = g g_t^{-1}$ in the forth equality because $(g^{-1} g_t) \cdot P_{\mathcal{B}}(g^{-1} \cdot t) = g^{-1} \cdot t$. ■

Theorem 15 clarifies the rigidity and flexibility of the class of equivariant maps. That is, equivariant maps are completely rigid given generators in the sense that the generator determines those. On the other hand, the generators of equivariant maps are entirely flexible because they have no restrictions on constructing equivariant maps.

From the following proposition, the distance between equivariant maps is calculated from their generators.

Proposition 16 (Isometric Restriction). *Let a group G act on sets $\mathcal{S}$ and $\mathcal{T}$, and $\mathcal{B} \subset \mathcal{T}$ an arbitrary base space. The restriction $R_{\mathcal{B}}$ onto $\mathcal{B}$ is isometry from equivariant maps. That is, for two G-equivariant maps F and $\tilde{F} : \mathbb{R}^{\mathcal{S}} \to \mathbb{R}^{\mathcal{T}}$,*

$$\|F - \tilde{F}\|_\infty = \|F_{\mathcal{B}} - \tilde{F}_{\mathcal{B}}\|_\infty. \tag{11}$$

[Proof] . We note that, for any base space $\mathcal{B} \subset \mathcal{T}$, $g \in G$ and $\tau \in \mathcal{B}$,

$$F[x](g \cdot \tau) = F_{\mathcal{B}}[g^{-1} \cdot x] \circ P_{\mathcal{B}}(\tau) = F_{\mathcal{B}}[g^{-1} \cdot x](\tau).$$

Thus,

$$\begin{aligned}\left\|F-\tilde{F}\right\|_{\infty} &:= \sup_{x\in\mathbb{R}^{\mathcal{S}}}\sup_{t\in\mathcal{T}}\left|F[x](t)-\tilde{F}[x](t)\right| \\ &= \sup_{x\in\mathbb{R}^{\mathcal{S}}}\sup_{\tau\in\mathcal{B},g\in G}\left|F[x](g\cdot\tau)-\tilde{F}[x](g\cdot\tau)\right| \\ &= \sup_{x\in\mathbb{R}^{\mathcal{S}}}\sup_{\tau\in\mathcal{B},g\in G}\left|F_{\mathcal{B}}[g^{-1}\cdot x](\tau)-\tilde{F}_{\mathcal{B}}[g^{-1}\cdot x](\tau)\right| \\ &= \sup_{x\in\mathbb{R}^{\mathcal{S}}}\sup_{\tau\in\mathcal{B}}\left|F_{\mathcal{B}}[x](\tau)-\tilde{F}_{\mathcal{B}}[x](\tau)\right| \\ &= \left\|F_{\mathcal{B}}-\tilde{F}_{\mathcal{B}}\right\|_{\infty}.\end{aligned}$$

This completes the proof of Proposition 16. ■

We immediately obtain the following corollary from Proposition 16.

Corollary 17 (Identity Condition)**.** *Let a group G act on sets $\mathcal{S}$ and $\mathcal{T}$, and $\mathcal{B}\subset\mathcal{T}$ an arbitrary base space. Let F and $\tilde{F}:\mathbb{R}^{\mathcal{S}}\to\mathbb{R}^{\mathcal{T}}$ be G-equivariant maps. Then, $F=\tilde{F}$ if and only if $F_{\mathcal{B}}=\tilde{F}_{\mathcal{B}}$.*

A.2 Proof of Theorem 13

Guss & Salakhutdinov (2019) derived the following theorem in infinite-dimensional settings.

Theorem 18 (Universal Approximation for Continuous Maps by FNNs, Guss & Salakhutdinov (2019))**.** *Let an activation function $\rho:\mathbb{R}\to\mathbb{R}$ be continuous and non-polynomial. Let $\mathcal{S}\subset\mathbb{R}^d$ and $\mathcal{T}\subset\mathbb{R}^{d'}$ be compact domains. Let $F:\mathcal{C}(\mathcal{S})\to\mathcal{C}(\mathcal{T})$ be a continuous map. Then, for any compact $E\subset\mathcal{C}(\mathcal{S})$ and $\epsilon>0$, there exist $N\in\mathbb{N}$ and a two-layer fully connected neural network $\phi_E=A_2\circ\rho\circ A_1\in\mathcal{N}_{\mathsf{FNN}}(\rho,2;\mathcal{S},\mathcal{T})$ such that $A_1[\cdot]=W^{(1)}[\cdot]+b^{(1)}:E\to\mathcal{C}([N])=\mathbb{R}^N$, $A_2[\cdot]=W^{(2)}[\cdot]+b^{(2)}:\mathbb{R}^N\to\mathcal{C}(\mathcal{T})$, μ_{ϕ_E} is the Lebesgue measure, and $\|F|_E-\phi_E\|_\infty<\epsilon$.*

Thus, any continuous function can be approximated by an FNN whose neurons in the hidden layer is finite. Krukowski (2018) derived the following theorem[10].

Theorem 19 (Arzelà-Ascoli Theorem for $C_0(\mathcal{S})$, Krukowski (2018))**.** *Let $\mathcal{S}$ be a locally compact Hausdorff space. A subset $E\subset C_0(\mathcal{S})$ is relatively compact if and only if the following three conditions hold:*

- *(A1) E is point-wise bounded, i.e. for any $s\in\mathcal{S}$, the inequality $\sup_{x\in E}|x(s)|<\infty$ holds,*
- *(A2) E is equicontinuous, i.e., for any $\varepsilon>0$ and $s\in\mathcal{S}$, there exists a neighborhood U around s, for any $\tilde{s}\in U$, the inequality $\sup_{x\in E}|x(\tilde{s})-x(s)|<\varepsilon$ holds,*
- *(A3) E is equivanishing, i.e., for any $\varepsilon>0$, there exists a compact set $K\subset\mathcal{S}$, for any $s\notin K$, the inequality $\sup_{x\in E}|x(s)|<\varepsilon$ holds.*

Note that E in Theorem 13 is relatively compact since any compact set is relatively compact in an arbitrary metric space, and thus, we can use Theorem 19.

Let $L>0$ be the Lipschitz constant of F. By (A3) of Theorem 19, for any $\epsilon>0$, there exist a large $r>0$ and a small $\delta>0$, for any $\bar{r}\geq r-\delta$, a compact ball $B_{\bar{r}}:=\{x\in\mathcal{S}\,|\,\|x\|_{\mathbb{R}^d}\leq\bar{r}\}\subset\mathcal{S}$ centered at $0\in\mathbb{R}^d$ with radius $\bar{r}$ satisfies

$$\sup_{x\in E}\|x-x\cdot\mathbf{1}_{B_{\bar{r}}}\|_\infty<\frac{\epsilon}{4L}, \tag{12}$$

where $\mathbf{1}_{B_{\bar{r}}}$ is the indicator function on $B_{\bar{r}}$. Here, although $x\cdot\mathbf{1}_{B_{\bar{r}}}$ approximates x, it may not be included in $\mathcal{C}_0(\mathcal{S})$. Then, we can take a continuous approximation function $\tilde{\mathbf{1}}_{B_r}\in\mathcal{C}_0(\mathcal{S})$ of the

[10] Although Arzelà-Ascoli Theorem for functions on *compact* Hausdorff spaces is well-known, here we require its non-compact version.

indicator function $\mathbf{1}_{B_r}$ such that the support equals B_r, $0 \leq \tilde{\mathbf{1}}_{B_r} \leq 1$, $\tilde{\mathbf{1}}_{B_r} = 1$ on $B_{r-\delta}$ and it satisfies

$$\sup_{x \in E} \|x \cdot \mathbf{1}_{B_r} - x \cdot \tilde{\mathbf{1}}_{B_r}\|_\infty < \frac{\epsilon}{4L}. \tag{13}$$

From (12) and (13), we obtain

$$\sup_{x \in E} \|x - x \cdot \tilde{\mathbf{1}}_{B_r}\|_\infty < \frac{\epsilon}{2L}. \tag{14}$$

Since E is assumed to be compact and F is continuous, the image $F(E)$ is also compact in $\mathcal{C}_0(\mathcal{T})$. Thus, using (A3) of Theorem 19 again, for any $\epsilon > 0$, there exist $r' > 0$ and δ', for any $\bar{r}'$, a compact ball $B_{\bar{r}'} := \{x \in \mathcal{T} | \|x\|_{\mathbb{R}^{d'}} \leq \bar{r}'\} \subset \mathcal{T}$ centered at $0 \in \mathbb{R}^{d'}$ with radius $\bar{r}'$ satisfies

$$\sup_{x \in E} \|F[x] - F[x] \cdot \mathbf{1}_{B_{\bar{r}'}}\|_\infty < \frac{\epsilon}{4}. \tag{15}$$

Then, we can take an continuous approximation function $\tilde{\mathbf{1}}_{B_{r'}} \in \mathcal{C}_0(\mathcal{T})$ of the indicator function $\mathbf{1}_{B_{r'}} \in \mathcal{C}(\mathcal{T})$ such that the support equals $B_{r'}$, $0 \leq \tilde{\mathbf{1}}_{B_{r'}} \leq 1$, $\tilde{\mathbf{1}}_{B_{r'}} = 1$ on $B_{r'-\delta'}$ and it satisfies

$$\sup_{x \in E} \|F[x] \cdot \mathbf{1}_{B_{r'}} - F[x] \cdot \tilde{\mathbf{1}}_{B_{r'}}\|_\infty < \frac{\epsilon}{4}. \tag{16}$$

From (15) and (16), we obtain

$$\sup_{x \in E} \|F[x] - F[x] \cdot \tilde{\mathbf{1}}_{B_{r'}}\|_\infty < \frac{\epsilon}{2}. \tag{17}$$

We define the smoothed restriction function $\tilde{R}_{B_r} : \mathcal{C}_0(\mathcal{S}) \to \mathcal{C}_0(\mathcal{S})$ as $\tilde{R}_{B_r}(x) := x \cdot \tilde{\mathbf{1}}_{B_r}$ and $\tilde{R}_{B_{r'}} : \mathcal{C}_0(\mathcal{T}) \to \mathcal{C}_0(\mathcal{T})$ as $\tilde{R}_{B_{r'}}(x') := x' \cdot \tilde{\mathbf{1}}_{B_{r'}}$. Then, for any $x \in E$, we obtain

$$\begin{aligned}
& \|F|_E[x] - \tilde{R}_{B_{r'}} \circ F \circ \tilde{R}_{B_r}[x]\|_\infty \\
\leq\ & \|F[x] - F[x \cdot \tilde{\mathbf{1}}_{B_r}] \cdot \tilde{\mathbf{1}}_{B_{r'}}\|_\infty \\
\leq\ & \|F[x] - F[x] \cdot \tilde{\mathbf{1}}_{B_{r'}}\|_\infty + \|F[x] \cdot \tilde{\mathbf{1}}_{B_{r'}} - F[x \cdot \tilde{\mathbf{1}}_{B_r}] \cdot \tilde{\mathbf{1}}_{B_{r'}}\|_\infty \\
\leq\ & \|F[x] - F[x] \cdot \tilde{\mathbf{1}}_{B_{r'}}\|_\infty + \|F[x] - F[x \cdot \tilde{\mathbf{1}}_{B_r}]\|_\infty \\
\leq\ & \|F[x] - F[x] \cdot \tilde{\mathbf{1}}_{B_{r'}}\|_\infty + L\|x - x \cdot \tilde{\mathbf{1}}_{B_r}\|_\infty \\
<\ & \frac{\epsilon}{2} + L \cdot \frac{\epsilon}{2L} = \epsilon.
\end{aligned}$$

From the above discussion, we can approximate $F|_E$ by $\tilde{R}_{B_{r'}} \circ F \circ \tilde{R}_{B_r}$. Thus, it is enough to show that $\tilde{R}_{B_{r'}} \circ F \circ \tilde{R}_{B_r}$ can be approximated by an FNN.

For a compact set $K \subset \mathcal{S}$, let $\mathcal{C}(K)|_{\partial K=0} := \{x \in \mathcal{C}(K) | x|_{\partial K} \equiv 0\}$, where ∂K is the boundary set of K. Then, we define the inclusion $\iota_K : \mathcal{C}(K)|_{\partial K=0} \to \mathcal{C}_0(\mathbb{R}^d)$ as

$$\iota_K(x)(s) = \begin{cases} x(s) & (s \in K) \\ 0 & (s \notin K). \end{cases}$$

We can verify that ι_K is a bounded affine map. Moreover, we define the restriction function $R_K : \mathcal{C}_0(\mathcal{S}) \to \mathcal{C}(K)$ for a subset $K \subset \mathcal{S}$ as $R_K(x) := x|_K$. Using the above notions, we have

$$\tilde{R}_{B_{r'}} \circ F \circ \tilde{R}_{B_r} = \iota_{B_{r'}} \circ (R_{B_{r'}} \circ \tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}) \circ (R_{B_r} \circ \tilde{R}_{B_r}). \tag{18}$$

Thus, in order to approximate $\tilde{R}_{B_{r'}} \circ F \circ \tilde{R}_{B_r}$ by an FNN, we show that both $R_{B_{r'}} \circ \tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}|_{E'}$ and $R_{B_r} \circ \tilde{R}_{B_r}$ can be approximated by FNNs, where $E' := R_{B_r} \circ \tilde{R}_{B_r}(E)$.

First, we prove that $R_{B_{r'}} \circ \tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}|_{E'}$ can be approximated by a two-layer FNN. Since $R_{B_r} \circ \tilde{R}_{B_r} : E \to \mathcal{C}(B_r)|_{\partial B_r = 0}$ is continuous, the image E' is compact in $\mathcal{C}(B_r)|_{\partial B_r=0} \subset \mathcal{C}(B_r)$ because of the compactness of E. Then, using Theorem 18, $R_{B_{r'}} \circ \tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}|_{E'} : E' \to \mathcal{C}(B_{r'})$ is approximated by a two-layer FNN with any precision.

Next, we prove that $R_{B_r} \circ \tilde{R}_{B_r}$ can be approximated by a bounded affine map. We denote by δ_t the Dirac delta function at $t \in \mathcal{S}$. Let $w(t, s) := \tilde{\mathbf{1}}_{B_{r'}}(s)\delta_t(s)$ and $b(t) \equiv 0$ in (2). Then, the following holds:

$$A[x](t) = \int_{\mathbb{R}^d} x(s)\tilde{\mathbf{1}}_{B_{r'}}(s)\delta_t(s)d\mu(s) = x(t)\tilde{\mathbf{1}}_{B_{r'}}(t) = R_{B_r} \circ \tilde{R}_{B_r}[x](t).$$

Thus, $R_{B_r} \circ \tilde{R}_{B_r}$ is exactly represented by a bounded affine map if the Dirac delta function is allowed. However, the Dirac delta function is not a function but a generalized function. Here, the Dirac delta δ_t can be approximated by a smooth function $\tilde{\delta}_t$ called a mollifier with any precision. Thus, instead of $\tilde{\mathbf{1}}_{B_{r'}}(s)\delta_t(s)$, taking $\tilde{\mathbf{1}}_{B_{r'}}(s)\tilde{\delta}_t(s)$ as $w(t, s)$, we can verify that $R_{B_r} \circ \tilde{R}_{B_r}$ is approximated by a bounded affine map with any precision.

From the above discussion, for any $\epsilon > 0$, there exist a bounded affine map A and a two-layer FNN ϕ such that $\|R_{B_r} \circ \tilde{R}_{B_r} - A\|_\infty \leq \frac{\epsilon}{2L}$ and $\|\tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}|_{E'} - \phi\|_\infty \leq \frac{\epsilon}{2}$. Thus, we have

$$\begin{aligned}
&\|\tilde{R}_{B_{r'}} \circ F \circ \tilde{R}_{B_r} - \iota_{B_{r'}} \circ \phi \circ A\|_\infty \\
=\ &\|\iota_{B_{r'}} \circ (R_{B_{r'}} \circ \tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}) \circ (R_{B_r} \circ \tilde{R}_{B_r}) - \iota_{B_{r'}} \circ \phi \circ A\|_\infty \\
\leq\ &\|(R_{B_{r'}} \circ \tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}) \circ (R_{B_r} \circ \tilde{R}_{B_r}) - \phi \circ A\|_\infty \\
\leq\ &\|(\tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}) \circ (R_{B_r} \circ \tilde{R}_{B_r}) - (\tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}) \circ A\|_\infty \\
&+ \|(\tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r}) \circ A - \phi \circ A\|_\infty \\
\leq\ &L\|R_{B_r} \circ \tilde{R}_{B_r} - A\|_\infty + \|\tilde{R}_{B_{r'}} \circ F \circ \iota_{B_r} - \phi\|_\infty \\
\leq\ &\epsilon.
\end{aligned}$$

Since A and $\iota_{B_{r'}}$ are affine and ϕ is a two-layer FNN, the map $\phi_E := \iota_{B_{r'}} \circ \phi \circ A$ is also a two-layer FNN. Thus, this concludes the proof. ■

B Proof of Conversion Theorem

In this section, we prove Theorem 9. Since $\phi : E \to \mathcal{C}_0(\mathcal{B}_\mathcal{T})$ is a fully-connected neural network, there exist topological spaces $\mathcal{B}_\ell$ for $\ell = 1, \ldots, L-1$ and affine maps $A_1 : E \to \mathcal{C}_0(\mathcal{B}_1)$, $A_\ell : \mathcal{C}_0(\mathcal{B}_{\ell-1}) \to \mathcal{C}_0(\mathcal{B}_\ell)$ for $\ell = 1, \ldots, L-1$, and $A_L : \mathcal{C}_0(\mathcal{B}_{L-1}) \to \mathcal{C}_0(\mathcal{B}_\mathcal{T})$ such that the FNN $\phi = A_L \circ \rho \circ A_{L-1} \circ \cdots \circ \rho \circ A_1 : E \to \mathcal{C}_0(\mathcal{B}_\mathcal{T})$. Here, we note that the sets $\mathcal{B}_\ell$ for $\ell = 1, \ldots, L-1$ does not relate to the action of the group G while $\mathcal{B}_\mathcal{S}$ and $\mathcal{B}_\mathcal{T}$ are defined via the action of a group G. When we define as $\mathcal{S}_\ell := G/H_\mathcal{T} \times \mathcal{B}_\ell$ for $\ell = 1, \ldots, L-1$, the action of G on $\mathcal{S}_\ell$ is naturally defined by the action of G on $G/H_\mathcal{T}$. Then, the sets $\mathcal{B}_\ell$ for $\ell = 1, \ldots, L-1$ become the base space by the definition of $\mathcal{S}_\ell$. For brevity, we denote $\mathcal{B}_\mathcal{T}$ by $\mathcal{B}_L$ and $\mathcal{T}$ by $\mathcal{S}_L$.

In the following, for the fully-connected neural network ϕ, we show the existence of a group-convolutional neural network $\Phi := C_L \circ \rho \circ C_{L-1} \circ \cdots \circ \rho \circ C_1$ such that $C_1 : E \to \mathcal{C}_0(\mathcal{S}_1)$ and $C_\ell : \mathcal{C}_0(\mathcal{S}_{\ell-1}) \to \mathcal{C}_0(\mathcal{S}_\ell)$ for $\ell = 2, \ldots, L$ are biased G-convolutions and Φ satisfies (7).

First, we construct C_1. Since $A_1 : E \to \mathcal{C}_0(\mathcal{B}_1)$ is affine, there are $w^{(1)}(\tau, \cdot) \in \mathcal{C}(\mathcal{S})$ and $b^{(1)}(\tau) \in \mathbb{R}$ for each $\tau \in \mathcal{B}_1$ such that $A_1[\cdot] = W^{(1)}[\cdot] + b^{(1)}$, where $W^{(1)} : E \to \mathcal{C}_0(\mathcal{B}_1)$ satisfies

$$W^{(1)}[x](\tau) = \int_\mathcal{S} w^{(1)}(\tau, s)x(s)d\mu^{(1)}(s).$$

From the assumption (C1) of Theorem 9, there exists a G-invariant measure ν_1 such that $\mu^{(1)}$ is absolute continuous with respect to ν_1. Thus, we can set in (6) as

$$\begin{aligned}
v_1((g, \tau), s) &:= w^{(1)}(\tau, g^{-1} \cdot s)\frac{d\mu^{(1)}}{d\nu_1}(g^{-1} \cdot s), \\
b_1(g, \tau) &:= b^{(1)}(\tau),
\end{aligned}$$

where $g \in G/H_\mathcal{T}$, and $t \in \mathcal{B}_1$. Then, one can easily verify that these functions are G-invariant. Then, $C_1 : E \to \mathcal{C}(\mathcal{S}_1)$ is given by

$$C_1[x](g, \tau) := \int_\mathcal{S} v_1((g, \tau), s)x(s)d\nu_1(s) + b_1(g, \tau),$$

where $x \in E$. Moreover, the following holds for arbitrary $x \in E$ and $\tau \in \mathcal{B}_1$:

$$\begin{aligned} R_{\mathcal{B}_1} \circ C_1[x](\tau) &= \int_{\mathcal{S}} x(s) w^{(1)}(\tau, 1^{-1} \cdot s) \frac{d\mu^{(1)}}{d\nu_1}(1^{-1} \cdot s) d\nu(s) + b^{(1)}(\tau) \\ &= \int_{\mathcal{S}} w^{(1)}(\tau, s) x(s) d\mu^{(1)}(s) + b^{(1)}(\tau) \\ &= A_1[x](t). \end{aligned} \tag{19}$$

Thus, we obtain

$$R_{\mathcal{B}_1} \circ C_1 = A_1. \tag{20}$$

Next, we construct C_ℓ for $\ell \in \{2, \dots, L\}$. Since $A_\ell : \mathcal{C}(\mathcal{B}_{\ell-1}) \to \mathcal{C}(\mathcal{B}_\ell)$ is affine, there are $w^{(\ell)}(\tau, \varsigma) \in \mathbb{R}$ and $b^{(\ell)}(\tau) \in \mathbb{R}$ for each $\tau \in \mathcal{B}_\ell$ and $\varsigma \in \mathcal{B}_{\ell-1}$ such that $A_\ell[\cdot] = W^{(\ell)}[\cdot] + b^{(\ell)}$, where $W^{(\ell)} : \mathcal{C}(\mathcal{B}_{\ell-1}) \to \mathcal{C}(\mathcal{B}_\ell)$ satisfies

$$W^{(\ell)}[x](\tau) = \int_{\mathcal{B}_{\ell-1}} x(\varsigma) w^{(\ell)}(\tau, \varsigma) d\mu^{(\ell)}(\varsigma),$$

where $x \in \mathcal{C}(\mathcal{B}_{\ell-1})$ and $\tau \in \mathcal{B}_\ell$. For $\ell \in \{2, \dots, L\}$, we set as $\nu_\ell := \nu_{G/H_\mathcal{T}} \times \mu^{(\ell)}$, $v_\ell((g, \tau), (h, \varsigma)) := \delta(h, g) w^{(\ell)}(\tau, \varsigma)$ and $b_\ell(g, \tau) := b^{(\ell)}(\tau)$ in (6), where $h, g \in G/H_\mathcal{T}, \varsigma \in \mathcal{B}_{\ell-1}$, $\tau \in \mathcal{B}_\ell$, and $\delta(h, \cdot)$ is the Dirac delta function at $h \in G/H_\mathcal{T}$. Then, $C_\ell : \mathcal{C}(\mathcal{S}_{\ell-1}) \to \mathcal{C}(\mathcal{S}_\ell)$ is given by

$$\begin{aligned} C_\ell[x](g, \tau) &:= \int_{G/H_\mathcal{T} \times \mathcal{B}_{\ell-1}} x(h, \varsigma) v_\ell((g, \tau), (h, \varsigma)) d\nu_{G/H_\mathcal{T}}(h) d\mu^{(\ell)}(\varsigma) + b_\ell(\tau) \\ &= \int_{G/H_\mathcal{T} \times \mathcal{B}_{\ell-1}} x(h, \varsigma) \delta(h, g) w^{(\ell)}(\tau, \varsigma) d\nu_{G/H_\mathcal{T}}(h) d\mu^{(\ell)}(\varsigma) + b^{(\ell)}(\tau) \\ &= \int_{\mathcal{B}_{\ell-1}} x(g, \varsigma) w^{(\ell)}(\tau, \varsigma) \mu^{(\ell)}(\varsigma) + b^{(\ell)}(\tau) \\ &= W^{(\ell)}[x_g](\tau) + b^{(\ell)}(\tau) \\ &= A_\ell[x_g](t), \end{aligned}$$

where $x \in \mathcal{C}(\mathcal{S}_{\ell-1})$ and $x_g(\varsigma) := x(g, \varsigma)$. Then, the following equation holds for $\ell \in \{2, \dots, L-1\}$ by the definition of $R_{\mathcal{B}_\ell}$:

$$R_{\mathcal{B}_\ell} \circ C_\ell - A_\ell \circ R_{\mathcal{B}_{\ell-1}} = 0.$$

We note that the Dirac delta function δ used above is not a function but a generalized function. Here, it can be approximated by a smooth function $\tilde{\delta}$ called a mollifier with any precision.

Thus, replacing δ by $\tilde{\delta}$ in C_ℓ, we obtain the following inequality for $\ell \in \{2, \dots, L\}$:

$$\| R_{\mathcal{B}_\ell} \circ C_\ell - A_\ell \circ R_{\mathcal{B}_{\ell-1}} \|_\infty < \epsilon'. \tag{21}$$

Since ρ acts component-wise, the following equation hold:

$$R_{\mathcal{B}_1} \circ \rho = \rho \circ R_{\mathcal{B}_1}. \tag{22}$$

By (20), (21), and (22), we can see the diagram in Figure 2 is "approximately" commutative.

We note that every A_ℓ is Lipschitz because $W^{(\ell)}$ is a bounded linear operator. Using (20), (21), (22), and the fact that A_ℓ is Lipschitz, and taking a small ϵ', we obtain

$$\| R_{\mathcal{B}_L} \circ \Phi - \phi \|_\infty < \epsilon. \tag{23}$$

Lastly, we show the inequality (8):

$$\begin{aligned} \| F|_E - \Phi \|_\infty &= \| R_{\mathcal{B}_L} \circ F|_E - R_{\mathcal{B}_L} \circ \Phi \|_\infty \qquad (24) \\ &= \| (F_{\mathcal{B}_\mathcal{T}}|_E - \phi) + (\phi - R_{\mathcal{B}_L} \circ \Phi) \|_\infty \\ &\leq \| F_{\mathcal{B}_\mathcal{T}}|_E - \phi \|_\infty + \| \phi - R_{\mathcal{B}_L} \circ \Phi \|_\infty \\ &\leq \| F_{\mathcal{B}_\mathcal{T}}|_E - \phi \|_\infty + \epsilon, \qquad (25) \end{aligned}$$

where we used Proposition 16 in (24) and (23) in (24). ■

$$
\begin{tikzcd}
E \arrow[r, "\rho\circ C_1"] \arrow[dr, "\rho\circ A_1"'] & \mathcal{C}_0(\mathcal{S}_1) \arrow[r, "\rho\circ C_2"] \arrow[d, "R_{\mathcal{B}_1}"] & \cdots \arrow[r, "\rho\circ C_{L-1}"] \arrow[d] & \mathcal{C}_0(\mathcal{S}_{L-1}) \arrow[r, "C_L"] \arrow[d, "R_{\mathcal{B}_{L-1}}"] & \mathcal{C}_0(\mathcal{T}) \arrow[d, "R_{\mathcal{B}_L}"] \\
& \mathcal{C}_0(\mathcal{B}_1) \arrow[r, "\rho\circ A_2"'] & \cdots \arrow[r, "\rho\circ A_{L-1}"'] & \mathcal{C}_0(\mathcal{B}_{L-1}) \arrow[r, "A_L"'] & \mathcal{C}_0(\mathcal{B}_{\mathcal{T}})
\end{tikzcd}
$$

Figure 2: Approximately commutative diagram.

C Proof of Universality of DeepSets

We set $G = S_n$, $H = \text{Stab}(1) := \{s \in S_n \mid s(1) = 1\}$ and $B = \{*\}$, where $\{*\}$ is a singleton. Then we can see that Stab(1) is a subgroup of G and its left cosets $G/H = [n]$.

Lemma 20. *As a set, $S_n/Stab(1)$ is equal to $[n]$, and the canonical S_n-action on $S_n/Stab(1)$ is equivalent to the permutation action on $[n]$.*

[Proof] . Firstly, we can see that Stab(1) is isomorphic to S_{n-1} as a group, since Stab(1) can freely permute any element other than 1. Therefore $|S_n/\text{Stab}(1)| = |S_n/S_{n-1}| = n!/(n-1)! = n$. Next, we confirm that the action on $S_n/\text{Stab}(1)$ is equal to permutation on $[n]$ as a representation. To see this, we consider a complete system of representatives of $S_n/\text{Stab}(1)$. We can take $\overline{(1\,1)}, \overline{(1\,2)}, \ldots, \overline{(1\,n)}$ as a complete system of representatives. This is because for any $s \in S_n$ there is a decomposition $s = (1\ s(1)) \cdot t$ for some $t \in$ Stab(1). Here, we note that $\overline{s} = \overline{(1\ s(1))}$ by this formula. Finally, we see that the S_n-action on $\{\overline{(1\,1)}, \overline{(1\,2)}, \ldots, \overline{(1\,n)}\} = [n]$ ($\overline{(1\,j)} \mapsto j$) coincide with the permutation action. When we take $s \in S_n$ and $\overline{(1\,j)} \in \{\overline{(1\,1)}, \overline{(1\,2)}, \ldots, \overline{(1\,n)}\}$, we have $s \cdot (1\ j) = (1\ s(j)) \cdot t'$ for some $t' \in$ Stab(1) . This implies that $s \cdot \overline{(1\,j)} = \overline{(1\ s(j))}$ and this is equivalent to the permutation action $s \cdot j = s(j)$ by the correspondence above. ■

Therefore, $\mathcal{C}(G/H \times B) = \mathcal{C}([n]) = \mathbb{R}^n$ holds, and the equivariant model of our paper is equal to the one of DeepSets.

Theorem 12. *For any permutation equivariant function $F : \mathbb{R}^n \to \mathbb{R}^n$, a compact set $E \subset \mathbb{R}^n$ and $\epsilon > 0$, there is an equivariant model of DeepSets (or equivalently, our model) $\Phi_E : E \to \mathbb{R}^n$ such that $\|\Phi_E(x) - F|_E(x)\|_\infty < \epsilon$.*

[Proof] . Firstly, we see that our model is equal to the equivariant model of DeepSets when $\mathcal{S}$ is $[n]$. Our group convolution is defined by

$$C_{\nu,v,b}[x](t) := \int_{\mathcal{S}} v(t,s)x(s)d\nu(s) + b(t).$$

Since $\mathcal{S} = \mathcal{T} = [n]$, we have

$$\int_{\mathcal{S}} v(t,s)x(s)d\nu(s) + b(t) = \sum_{s\in\mathcal{S}} v(t,s)x(s) + b(t).$$

Therefore, the map $C_{\nu,v,b} : \mathbb{R}^n \to \mathbb{R}^n$ is induced by the matrix $W = (v(i,j))_{i,j\in[n]}$ and bias $b(t)$. Here, since $v : \mathcal{S}\times\mathcal{T} \to \mathbb{R}$ is G-invariant, $v(i,j)$ satisfies the condition $v((k,l)\cdot i, (k,l)\cdot j) = v(i,j)$ for any transition (k,l). This implies the parametrization $W = \lambda E + \gamma \mathbf{1}\mathbf{1}^\top$ by direct calculation. ■