On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups — Supplementary Material

Risi Kondor [1] **Shubhendu Trivedi** [2]

1. Background from group and representation theory

For a more detailed background on representation theory, we point the reader to (Serre, 1977).

Groups. A **group** is a set G endowed with an operation $G\times G\to G$ (usually denoted multiplicatively) obeying the following axioms:

G1. for any $g_1,g_2\in G$, $g_1g_2\in G$ (closure);

G2. for any $g_1,g_2,g_3\in G$, $g_1(g_2g_3)=(g_1g_2)g_3$ (associativity);

G3. there is a unique $e\in G$, called the **identity** of G, such that $eg=ge=g$ for any $u\in G$;

G4. for any $g\in G$, there is a corresponding element $g^{-1}\in G$ called the **inverse** of g, such that $g\,g^{-1}=g^{-1}g=e$.

We do *not* require that the group operation be commutative, i.e., in general, $g_1g_2\neq g_2g_1$. Groups can be finite or infinite, countable or uncountable, compact or non-compact. While most of the results in this paper would generalize to any compact group, to keep the exposition as simple as possible, throughout we assume that G is finite or countably infinite. As usual, $|G|$ will denote the size (cardinality) of G, sometimes also called the **order** of the group. A subset H of G is called a **subgroup** of G, denoted $H\leq G$, if H itself forms a group under the same operation as G, i.e., if for any $g_1,g_2\in H$, $g_1g_2\in H$.

Homogeneous Spaces. Let G be a group acting on a set $\mathcal{X}$. We say that $\mathcal{X}$ is a **homogeneous space** of G if for any $x,y\in\mathcal{X}$, there is a $g\in G$ such that $y=g(x)$. The significance of homogeneous spaces for our purposes is that once we fix the "origin" x_0, the above correspondence between points in $\mathcal{X}$ and the group elements that map x_0 to them allows to lift various operations on the homogeneous space to the group. Because expressions like $g(x_0)$ appear so often in the following, we introduce the shorthand $[g]_{\mathcal{X}}:=g(x_0)$. Note that this hides the dependency on the (arbitrary) choice of x_0.

As an example, $\mathbb{Z}^2$ is a homogeneous space of itself with respect to the trivial action $(i,j)\mapsto(g_1+i,g_2+j)$. The sphere S^2 is a homogeneous space of the rotation group $\mathrm{SO}(3)$ with respect to the action

$$x\mapsto R(x)\qquad R(x)=Rx\qquad x\in S^2. \tag{1}$$

Representations. A (finite dimensional) **representation** of a group G over a field $\mathbb{F}$ is a matrix-valued function $\rho\colon G\to\mathbb{F}^{d_\rho\times d_\rho}$ such that $\rho(g_1)\,\rho(g_2)=\rho(g_1g_2)$ for any $g_1,g_2\in G$. In this paper, unless stated otherwise, we always assume that $\mathbb{F}=\mathbb{C}$. A representation ρ is said to be **unitary** if $\rho(g^{-1})=\rho(g)^\dagger$ for any $g\in G$. One representation shared by every group is the **trivial representation** ρ_{tr} that simply evaluates to the one dimensional matrix $\rho_{\mathrm{tr}}(g)=(1)$ on every group element.

Equivalence, reducibility and irreps. Two representations ρ and ρ' of the same dimensionality d are said to be **equivalent** if for some invertible matrix $Q\in\mathbb{C}^{d\times d}$, $\rho(g)=Q^{-1}\rho'(g)\,Q$ for any $g\in G$. A representation ρ is said to be **reducible** if it decomposes into a direct sum of smaller representations in the form

$$\begin{aligned}&\rho(g)\\ &\quad=Q^{-1}\left(\rho_1(g)\oplus\rho_2(g)\right)\,Q\\ &\quad=Q^{-1}\left(\begin{array}{c|c}\rho_1(g)&0\\\hline 0&\rho_2(g)\end{array}\right)Q\qquad\forall\,g\in G\end{aligned}$$

for some invertible matrix $Q\in\mathbb{C}^{d_\rho\times d_\rho}$. We use $\mathcal{R}_G$ to denote a complete set of inequivalent irreducible representations of G. However, since this is quite a mouthful, in this paper we also use the alternative term **system of irreps** to refer to $\mathcal{R}_G$. Note that the choice of irreps in $\mathcal{R}_G$ is far from unique, since each $\rho\in\mathcal{R}_G$ can be replaced by an equivalent irrep $Q^\top\rho(g)\,Q$, where Q is any orthogonal matrix of the appropriate size.

Complete reducibility and irreps. Representation theory takes on its simplest form when G is compact (and

[1]Departments of Statistics and Computer Science, The University of Chicago [2]Toyota Technological Institute at Chicago. Correspondence to: Risi Kondor <risi@cs.uchicago.edu>, Shubhendu Trivedi <shubhendu@ttic.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.

$\mathbb{F}=\mathbb{C}$). One of the reasons for this is that it is possible to prove ("theorem of complete reducibility") that any representation ρ of a compact group can be reduced into a direct sum of irreducible ones, i.e.,

$$\rho(g)=Q^{-1}\big(\rho_{(1)}(g)\oplus\rho_{(2)}(g)\oplus\ldots\oplus\rho_{(k)}(g)\big)\,Q, \quad (2)$$

for some sequence $\rho_{(1)},\rho_{(2)},\ldots,\rho_{(k)}$ of irreducible representations of G and some $Q\in\mathbb{C}^{d\times d}$. In this sense, for compact groups, $\mathcal{R}_G$ plays a role very similar to the primes in arithmetic. Fixing $\mathcal{R}_G$, the number of times that a particular $\rho'\in\mathcal{R}_G$ appears in (2) is a well-defined quantity called the **multiplicity** of ρ' in ρ, denoted $m_\rho(\rho')$. Compactness also has a number of other advantages:

1. When G is compact, $\mathcal{R}_G$ is a countable set, therefore we can refer to the individual irreps as $\rho_1,\rho_2,\ldots$. (When G is finite, $\mathcal{R}_G$ is not only countable but finite.)
2. The system of irreps of a compact group is essentially unique in the sense that if $\mathcal{R}'_G$ is any other system of irreps, then there is a bijection $\phi\colon \mathcal{R}_G\to\mathcal{R}'_G$ mapping each irrep $\rho\in\mathcal{R}_G$ to an equivalent irrep $\phi(\rho)\in\mathcal{R}'_G$.
3. When G is compact, $\mathcal{R}_G$ can be chosen in such a way that each $\rho\in\mathcal{R}$ is unitary.

Restricted representations. Given any representation ρ of G and subgroup $H\leq G$, the *restriction* of ρ to H is defined as the function $\rho|_H\colon H\to\mathbb{C}^{d_\rho\times d_\rho}$, where $\rho|_H(h)=\rho(h)$ for all $h\in H$. It is trivial to check that $\rho|_H$ is a representation of H, but, in general, it is not irreducible (even when ρ itself is irreducible).

Fourier Transforms. In the Euclidean domain convolution and cross-correlation have close relationships with the Fourier transform

$$\widehat{f}(k)=\int e^{-2\pi\iota kx}\,f(x)\,dx, \quad (3)$$

where ι is the imaginary unit, $\sqrt{-1}$. In particular, the Fourier transform of $f*g$ is just the pointwise product of the Fourier transforms of f and g,

$$\widehat{f*g}(k)=\widehat{f}(k)\,\widehat{g}(k), \quad (4)$$

while cross-correlation is

$$\widehat{f\star g}(k)=\widehat{f}(k)^*\,\widehat{g}(k). \quad (5)$$

The concept of *group representations* (see Section 1) allows generalizing the Fourier transform to any compact group. The **Fourier transform** of $f\colon G\to\mathbb{C}$ is defined as:

$$\widehat{f}(\rho_i)=\int_G \rho_i(u)\,f(u)\,d\mu(u), \qquad i=1,2,\ldots, \quad (6)$$

which, in the countable (or finite) case simplifies to

$$\widehat{f}(\rho_i)=\sum_{u\in G} f(u)\,\rho(u), \qquad i=1,2,\ldots. \quad (7)$$

Despite $\mathbb{R}$ not being a compact group, (3) can be seen as a special case of (6), since $e^{-2\pi\iota kx}$ trivially obeys $e^{-2\pi\iota k(x_1+x_2)}=e^{-2\pi\iota kx_1}e^{-2\pi\iota kx_2}$, and the functions $\rho_k(x)=e^{-2\pi\iota kx}$ are, in fact, the irreducible representations of $\mathbb{R}$. The fundamental novelty in (6) and (7) compared to (3), however, is that since, in general (in particular, when G is not commutative), irreducible representations are matrix valued functions, each "Fourier component" $\widehat{f}(\rho)$ is now a matrix. In other respects, Fourier transforms on groups behave very similarly to classical Fourier transforms. For example, we have an inverse Fourier transform

$$f(u)=\frac{1}{|G|}\sum_{\rho\in\mathcal{R}} d_\rho\,\mathrm{tr}\big[f(\rho)\,\rho(u)^{-1}\big],$$

and also an analog of the convolution theorem, which is stated in the main body of the paper.

2. Convolution of vector valued functions

Since neural nets have multiple channels, we need to further extend equations 6-12 to vector/matrix valued functions. Once again, there are multiple cases to consider.

Definition 1. *Let G be a finite or countable group, and $\mathcal{X}$ and $\mathcal{Y}$ be (left or right) quotient spaces of G.*

1. *If $f\colon\mathcal{X}\to\mathbb{C}^m$, and $g\colon\mathcal{Y}\to\mathbb{C}^m$, we define $f*g\colon G\to\mathbb{C}$ with*
$$(f*g)(u)=\sum_{v\in G} f{\uparrow}^G(uv^{-1})\cdot g{\uparrow}^G(v), \quad (8)$$
where $\cdot$ denotes the dot product.
2. *If $f\colon\mathcal{X}\to\mathbb{C}^{n\times m}$, and $g\colon\mathcal{Y}\to\mathbb{C}^m$, we define $f*g\colon G\to\mathbb{C}^n$ with*
$$(f*g)(u)=\sum_{v\in G} f{\uparrow}^G(uv^{-1})\,\times\,g{\uparrow}^G(v), \quad (9)$$
where $\times$ denotes the matrix/vector product.
3. *If $f\colon\mathcal{X}\to\mathbb{C}^m$, and $g\colon\mathcal{Y}\to\mathbb{C}^{n\times m}$, we define $f*g\colon G\to\mathbb{C}^m$ with*
$$(f*g)(u)=\sum_{v\in G} f{\uparrow}^G(uv^{-1})\,\tilde{\times}\,g{\uparrow}^G(v), \quad (10)$$
where $\boldsymbol{v}\tilde{\times}A$ denotes the "reverse matrix/vector product" $A\boldsymbol{v}$.

Since in cases 2 and 3 the nature of the product is clear from the definition of f and g, we will omit the $\times$ and $\tilde{\times}$ symbols. The specializations of these formulae to the cases of Equations 6-12 are as to be expected.

3. Proof of Proposition 1.

Proposition 1 has three parts. To proceed with the proof, we introduce two simple lemmas.

Recall that if H is a subgroup of G, a function $f\colon G\to\mathbb{C}$ is called **right** H**–invariant** if $f(uh)=f(u)$ for all $h\in H$ and all $u\in G$, and it is called **left** H**–invariant** if $f(hu)=f(u)$ for all $h\in H$ and all $u\in G$.

Lemma 1. *Let H and K be two subgroups of a group G. Then*

1. *If $f\colon G/H\to\mathbb{C}$, then $f{\uparrow}^G\colon G\to\mathbb{C}$ is right H–invariant.*
2. *If $f\colon H\backslash G\to\mathbb{C}$, then $f{\uparrow}^G\colon G\to\mathbb{C}$ is left H–invariant.*
3. *If $f\colon K\backslash G/H\to\mathbb{C}$, then $f{\uparrow}^G\colon G\to\mathbb{C}$ is right H invariant and left K–invariant.*

Lemma 2. *Let ρ be an irreducible representation of a countable group G. Then $\sum_{u\in G}\rho(u)=0$ unless ρ is the trivial representation, $\rho_{\rm tr}(u)=(1)$.*

Proof. Let us define the functions $r^{\rho}_{i,j}(u)=[\rho(u)]_{i,j}$. Recall that for $f,g\colon G\to\mathbb{C}$, the inner product $\langle f,g\rangle$ is defined $\langle f,g\rangle=\sum_{u\in G}f(u)^*g(u)$. The Fourier transform of a function f can then be written element-wise as $[\widehat{f}(\rho)]_{i,j}=\langle {r^{\rho}_{i,j}}^*,f\rangle$. However, since the Fourier transform is a unitary transformation, for any $\rho,\rho'\in\mathcal{R}_G$, unless $\rho=\rho'$, $i=i'$ and $j=j'$, we must have $\langle r^{\rho}_{i,j},r^{\rho'}_{i',j'}\rangle=0$. In particular, $\big[\sum_{u\in G}\rho(u)\big]_{i,j}=\langle r^{\rho_{\rm tr}}_{1,1},r^{\rho}_{i,j}\rangle=0$, unless $\rho=\rho_{\rm tr}$ (and $i=j=1$). ■

Now recall that given an irrep ρ of G, the *restriction* of ρ to H is $\rho|_H\colon H\to\mathbb{C}^{d_\rho\times d_\rho}$, where $\rho|_H(h)=\rho(h)$ for all $h\in H$. It is trivial to check that $\rho|_H$ is a representation of H, but, in general, it is not irreducible. Thus, by the Theorem of Complete Decomposability (see section 1), it must decompose in the form $\rho|_H(h)=Q(\mu_1(h)\oplus\mu_2(h)\oplus\ldots\oplus\mu_k(h))Q^\dagger$ for some sequence $\mu_1,\ldots,\mu_k$ of irreps of H and some unitary martrix Q. In the special case when the irreps of G and H are adapted to $H\le G$, however, Q is just the unity.

Proof of Proposition 1, Part 1. The fact that any $u\in G$ can be written uniquely as $u=gh$ where g is the representative of one of the gH cosets and $h\in H$ immediately tells us that $\widehat{f}(\rho)$ factors as

$$\begin{aligned}\widehat{f}(\rho)&=\sum_{u\in G}f{\uparrow}^G(u)\,\rho(u)\\&=\sum_{x\in G/H}\sum_{h\in H}f{\uparrow}^G(\overline{x}h)\,\rho(\overline{x}h)\\&=\sum_{x\in G/H}\sum_{h\in H}f(x)\,\rho(\overline{x}h)\\&=\sum_{x\in G/H}\sum_{h\in H}f(x)\,\rho(\overline{x})\rho(h)\\&=\sum_{x\in G/H}f(x)\,\rho(\overline{x})\Big[\sum_{h\in H}\rho(h)\Big].\end{aligned}$$

However, $\rho(h)=\mu_1(h)\oplus\mu_2(h)\oplus\ldots\oplus\mu_k(h)$ for some sequence of irreps $\mu_1,\ldots,\mu_k$ of H, so

$$\sum_{h\in H}\rho(h)=\Big[\sum_{h\in H}\mu_1(h)\Big]\oplus\Big[\sum_{h\in H}\mu_2(h)\Big]\oplus\ldots\oplus\Big[\sum_{h\in H}\mu_k(h)\Big],$$

and by Lemma 2 each of the terms in this sum where μ_i is *not* the trivial representation (on H) is a zero matrix, zeroing out all the corresponding columns in $\widehat{f}(\rho)$. ■

Proof of Proposition 1, Part 2. Analogous to the proof of part 1, using $u=hg$ except that $\sum_{h\in H}\rho(h)$ will now multiply $\sum_{x\in H\backslash G}f(x)\,\rho(\overline{x})$ from the left. ■

Proof of Proposition 1, Part 3. Immediate from combining case 3 of Lemma 1 with Parts 1 and 2. ■

4. Proof of Proposition 2.

Proof. Let us assume that G is countable. Then

$$\begin{aligned}\widehat{f*g}(\rho_i)&=\sum_{u\in G}\Big[\sum_{v\in G}f(uv^{-1})\,g(v)\Big]\rho_i(u)\\&=\sum_{u\in G}\sum_{v\in G}f(uv^{-1})\,g(v)\rho_i(uv^{-1})\rho_i(v)\\&=\sum_{v\in G}\sum_{u\in G}f(uv^{-1})\,g(v)\rho_i(uv^{-1})\rho_i(v)\\&=\sum_{v\in G}\Big[\sum_{u\in G}f(uv^{-1})\,\rho_i(uv^{-1})\Big]g(v)\rho_i(v)\\&=\sum_{v\in G}\Big[\sum_{w\in G}f(w)\,\rho_i(w)\Big]g(v)\rho_i(v)\\&=\Big[\sum_{w\in G}f(w)\,\rho_i(w)\Big]\Big[\sum_{v\in G}g(v)\rho_i(v)\Big]\\&=\widehat{f}(\rho_i)\,\widehat{g}(\rho_i).\end{aligned}$$

The continuous case is proved similarly but with integrals with respect Haar measure instead of sums. ■

5. Proof of Theorem 1.

The proof of the Theorem for the forward direction, i.e., that convolution implies equivariance, is in the main text. Here we provide the proof for the reverse direction, i.e., that a network is equivariant *only* if it is convolutional. We start with two versions of Schur's Lemma.

Lemma 3. ***(Schur's lemma I)*** *Let $\{\rho(g)\colon U \to U\}_{g\in G}$ and $\{\rho'(g)\colon V \to V\}_{g\in G}$ be two irreducible representations of a compact group G. Let $\phi\colon U \to V$ be a linear (not necessarily invertible) mapping that is equivariant with these representations in the sense that $\phi(\rho(g)(u)) = \rho'(g)(\phi(u))$ for any $u \in U$. Then, unless ϕ is the zero map, ρ and ρ' are equivalent representations.*

Lemma 4. ***(Schur's lemma II)*** *Let $\{\rho(g)\colon U \to U\}_{g\in G}$ be an irreducible representation of a compact group G on a space U, and $\phi\colon U \to U$ a linear map that commutes with each $\rho(g)$ (i.e., $\rho(g) \circ \phi = \phi \circ \rho(g)$ for any $g \in G$). Then ϕ is a multiple of the identity.*

We build up the proof through a sequence of lemmas.

Lemma 5. *Let U and V be two vector spaces on which a compact group G acts by the linear actions $\{T_g\colon U \to U\}_{g\in G}$ and $\{T'_g\colon V \to V\}_{g\in G}$, respectively. Let $\phi\colon U \to V$ be a linear map that is equivariant with the $\{T_g\}$ and $\{T'_g\}$ actions, and W be an irreducible subspace of U (with respect to $\{T_g\}$). Then $Z = \phi(W)$ is an irreducible subspace of V, and the restriction of $\{T_g\}$ to W, as a representation, is equivalent with the restriction of $\{T'_g\}$ to Z.*

Proof. Assume for contradiction that Z is reducible, i.e., that it has a proper subspace $\mathcal{Z} \subset Z$ that is fixed by $\{T'_g\}$ (in other words, $T'_g(v) \in \mathcal{Z}$ for all $v \in \mathcal{Z}$ and $g \in G$). Let v be any nonzero vector in $\mathcal{Z}$, $u \in U$ be such that $\phi(u) = v$, and $\mathcal{W} = \operatorname{span}\{\, T_g(u) \mid g \in G \,\}$. Since W is irreducible, $\mathcal{W}$ cannot be a proper subspace of W, so $\mathcal{W} = W$. Thus,

$$\begin{aligned} Z &= \phi(\operatorname{span}\{\, T_g(u) \mid g \in G \,\}) \\ &= \operatorname{span}\{T'_g(\phi(u)) | g \in G\} = \operatorname{span}\{T'_g(v) | g \in G\} \subseteq \mathcal{Z}, \end{aligned} \tag{11}$$

contradicting our assumption. Thus, the restriction $\{T_g|_W\}$ of $\{T_g\}$ to W and the restriction $\{T'_g|_Z\}$ of $\{T'_g\}$ to Z are both irreducible representations, and $\phi\colon W \to Z$ is a linear map that is equivariant with them. By Schur's lemma it follows that $\{T_g|_W\}$ and $\{T'_g|_Z\}$ are equivalent representations. ■

Lemma 6. *Let U and V be two vector spaces on which a compact group G acts by the linear actions $\{T_g\colon U \to U\}_{g\in G}$ and $\{T'_g\colon V \to V\}_{g\in G}$, and let $U = U_1 \oplus U_2 \oplus \ldots$ and $V = V_1 \oplus V_2 \oplus \ldots$ be the corresponding isotypic decompositions. Let $\phi\colon U \to V$ be a linear map that is equivariant with the $\{T_g\}$ and $\{T'_g\}$ actions. Then $\phi(U_i) \subseteq V_i$ for any i.*

Proof. Let $U_i = U_i^1 \oplus U_i^2 \oplus \ldots$ be the decomposition of U_i into irreducible G–modules, and $V_i^j = \phi(U_i^j)$. By Lemma 5, each V_i^j is an irreducible G–module that is equivalent with U_i^j, hence $V_i^j \subseteq V_i$. Consequently, $\phi(U_i) = \phi(U_i^1 \oplus U_i^2 \oplus \ldots) \subseteq V_i$. ■

Lemma 7. *Let $\mathcal{X} = G/H$ and $\mathcal{X}' = G/K$ be two homogeneous spaces of a compact group G, let $\{\mathbb{T}_g\colon L(\mathcal{X}) \to L(\mathcal{X})\}_{g\in G}$ and $\{\mathbb{T}'_g\colon L(\mathcal{X}') \to L(\mathcal{X}')\}_{g\in G}$ be the corresponding translation actions, and let $\phi\colon L(\mathcal{X}) \to L(\mathcal{X}')$ be a linear map that is equivariant with these actions. Given $f \in L(\mathcal{X})$ let $\widehat{f}$ denote its Fourier transform with respect to a specific choice of origin $x_0 \in \mathcal{X}$ and system or irreps $\mathcal{R}_G = \{\rho_1, \rho_2, \ldots\}$. Similarly, $\widehat{f'}$ is the Fourier transform of $f' \in L(\mathcal{X}')$, with respect to some $x'_0 \in \mathcal{X}'$ and the same system of irreps.*

Now if $f' = \phi(f)$, then each Fourier component of f' is a linear function of the corresponding Fourier component of f, i.e., there is a sequence of linear maps $\{\Phi_i\}$ such that $\widehat{f'}(\rho_i) = \Phi_i(\widehat{f}(\rho_i))$.

Proof. Let $U_1 \oplus U_2 \oplus \ldots$ and $V_1 \oplus V_2 \oplus \ldots$ be the isotypic decompositions of $L(\mathcal{X})$ and $L(\mathcal{X}')$ with respect to the $\{\mathbb{T}_g\}$ and $\{\mathbb{T}'_g\}$ actions. Each Fourier component $\widehat{f}(\rho_i)$ captures the part of f falling in the corresponding isotypic subspace U_i. Similarly, $\widehat{f'}(\rho_j)$ captures the part of f' falling in V_j. Lemma 6 tells us that because ϕ is equivariant with the translation actions, it maps each U_i to the corresponding isotypic V_i. Therefore, $\widehat{f'}(\rho_i) = \Phi_i(\widehat{f}(\rho_i))$ for some function Φ_i. By the linearity of ϕ, each Φ_i must be linear. ■

Lemma 7 is a big step towards describing what form equivariant mappings take in Fourier space, but it doesn't yet fully pin down the individual Φ_i maps. We now focus on a single pair of isotypics (U_i, V_i) and the corresponding map Φ_i taking $\widehat{f}(\rho_i) \mapsto \widehat{f'}(\rho_i)$. We will say that Φ_i is an *allowable* map if there is some equivariant ϕ such that $\widehat{\phi(f)}(\rho_i) = \Phi_i(\widehat{f}(\rho_i))$. Clearly, if $\Phi_1, \Phi_2, \ldots$ are individually allowable, then they are also jointly allowable.

Lemma 8. *All linear maps of the form $\Phi_i\colon M \mapsto MB$ where $B \in \mathbb{C}^{\delta\times\delta}$ are allowable.*

Proof. Recall that the $\{\mathbb{T}_g\}$ action takes $f \mapsto f^g$, where $f^g(x) = f(g^{-1}x)$. In Fourier space,

$$\widehat{f^g}(\rho_i) = \sum_{u \in G} \rho_i(u)\, f^g{\uparrow}^G(u) = \sum_{u \in G} \rho_i(u)\, f{\uparrow}^G(g^{-1}u)$$
$$= \sum_{w \in G} \rho_i(gw)\, f{\uparrow}^G(w) = \rho_i(g) \sum_{w \in G} \rho_i(w)\, f{\uparrow}^G(w)$$
$$= \rho_i(g)\, \widehat{f}(\rho_i). \quad (12)$$

(This is actually a general result called the (left) translation theorem.) Thus,

$$\Phi_i\big(\widehat{\mathbb{T}_g(f)}(\rho_i)\big) = \Phi_i\big(\rho_i(g)\widehat{f}(\rho_i)\big) = \rho_i(g)\, \widehat{f}(\rho_i)\, B.$$

Similarly, the $\{\mathbb{T}'_g\}$ action maps $\widehat{f'}(\rho_i) \mapsto g(\rho_i)\widehat{f'}(\rho_i)$, so

$$\mathbb{T}'_g\big(\Phi_i(\widehat{f}(\rho_i))\big) = \mathbb{T}'_g\big(\widehat{f}(\rho_i)\, B\big) = \rho_i(g)\, \widehat{f}(\rho_i)\, B.$$

Therefore, Φ_i is equivariant with the $\{\mathbb{T}\}$ and $\{\mathbb{T}'\}$ actions. ■

Lemma 9. *Let $\Phi_i \colon M \mapsto BM$ for some $B \in \mathbb{C}^{\delta \times \delta}$. Then Φ_i is not allowable unless B is a multiple of the identity. Moreover, this theorem also hold in the columnwise sense that if $\Phi_i \colon M \to M'$ such that $[M']_{*,j} = B_j\, [M]_{*,j}$ for some sequence of matrices $B_1, \ldots, B_d$, then Φ_i is not allowable unless each B_j is a multiple of the identity.*

Proof. Following the same steps as in the proof of Lemma 8, we now have

$$\Phi_i\big(\widehat{\mathbb{T}_g(f)}(\rho_i)\big) = B\, \rho_i(g)\, \widehat{f}(\rho_i),$$
$$\mathbb{T}'_g\big(\Phi_i(\widehat{f}(\rho_i))\big) = \rho_i(g)\, B\, \widehat{f}(\rho_i).$$

However, by the second form of Schur's Lemma, we cannot have $B\, \rho_i(g) = \rho_i(g) B$ for all $g \in G$, unless B is a multiple of the identity. ■

Lemma 10. *Φ_i is allowable if and only if it is of the form $M \mapsto MB$ for some $B \in \mathbb{C}^{\delta \times \delta}$.*

Proof. For the "if" part of this lemma, see Lemma 8. For the "only if" part, note that the set of allowable Φ_i form a subspace of all linear maps $\mathbb{C}^{\delta \times \delta} \to \mathbb{C}^{\delta \times \delta}$, and any allowable Φ_i can be expressed in the form

$$[\Phi_i(M)]_{a,b} = \sum_{c,d} \alpha_{a,b,c,d} M_{c,d}.$$

By Lemma 9, if $a \neq c$ but $b = d$, then $\alpha_{a,b,c,d} = 0$. On the other hand, by Lemma 8 if $a = c$, then $\alpha_{a,b,c,d}$ can take on any value, regardless of the values of b and d, as long as $\alpha_{a,b,a,d}$ is constant across varying a.

Now consider the remaining case $a \neq c$ and $b \neq d$, and assume that $\alpha_{a,b,c,d} \neq 0$ while Φ_i is still allowable. Then, by Lemma 8, it is possible to construct a second allowable map Φ'_i (namely one in which $\alpha'_{a,d,a,b} = 1$ and $\alpha'_{a,d,x,y} = 0$ for all $(x, y) \neq (c, d)$) such that in the composite map $\Phi''_i = \Phi'_i \circ \Phi_i$, $\alpha''_{a,d,c,d} \neq 0$. Thus, Φ''_i is not allowable. However, the composition of one allowable map with another allowable map is allowable, contradicting our assumption that Φ_i is allowable.

Thus, we have established that if Φ_i is allowable, then $\alpha_{a,b,c,d} = 0$, unless $a = c$. To show that any allowable Φ_i of the form $M \mapsto MB$, it remains to prove that additionally $\alpha_{a,b,a,d}$ is constant across a. Assume for contradiction that Φ_i is allowable, but for some (a, e, b, d) indices $\alpha_{a,b,a,d} \neq \alpha_{e,b,e,d}$. Now let Φ_0 be the allowable map that zeros out every column except column d (i.e., $\alpha^0_{x,d,x,d} = 1$ for all x, but all other coefficients are zero), and let Φ' be the allowable map that moves column b to column d (i.e., $\alpha'_{x,d,x,b} = 1$ for any x, but all other coeffcients are zero). Since the composition of allowable maps is allowable, we expect $\Phi'' = \Phi' \circ \Phi \circ \Phi^0$ to be allowable. However Φ'' is a map that falls under the purview of Lemma 9, yet $\alpha''_{a,d,a,d} \neq \alpha''_{e,d,e,d}$ (i.e., M_j is not a multiple of the identity) creating a contradiction. ■

Proof of Theorem 1 (Reverse direction). For simplicty we first prove the theorem assuming $\mathcal{Y}_\ell = \mathbb{C}$ for each ℓ.

Since $\mathcal{N}$ is a G-CNN, each of the mappings $(\xi_\ell \circ \phi_\ell) \colon L(\mathcal{X}_{\ell-1}) \to L(\mathcal{X}_\ell)$ is equivariant with the corresponding translation actions $\{\mathbb{T}^{\ell-1}_g\}_{g \in G}$ and $\{\mathbb{T}^{\ell}_g\}_{g \in G}$. Since ξ_ℓ is a pointwise operator, this is equivalent to asserting that ϕ_ℓ is equivariant with $\{\mathbb{T}^{\ell-1}_g\}_{g \in G}$ and $\{\mathbb{T}^{\ell}_g\}_{g \in G}$.

Letting $\mathcal{X} = \mathcal{X}_{\ell-1}$ and $\mathcal{X}' = \mathcal{X}_\ell$, Lemma 8 then tells us the the Fourier transforms of $f_{\ell-1}$ and $\phi_\ell(f_{\ell-1})$ are related by

$$\widehat{\phi_\ell(f_{\ell-1})}(\rho_i) = \Phi\big(\widehat{f_{\ell-1}}(\rho_i)\big)$$

for some fixed set of linear maps $\Phi_1, \Phi_2, \ldots$. Furthermore, by Lemma 10, each Φ_i must be of the form $M \mapsto MB_i$ for some appropriate matrix $B_i \in \mathbb{C}^{d_\rho \times d_\rho}$. If we then define χ_ℓ as the inverse Fourier transform of $(B_1, B_2, \ldots)$, then by the convolution theorem (Proposition 2), $\phi_\ell(f_{\ell-1}) = f_{\ell-1} * \chi$, confirming that $\mathcal{N}$ is a G-CNN. The extension of this result to the vector valued case, $f_\ell \colon \mathcal{X}_\ell \to V_\ell$, is straightforward. ■

References

Serre, Jean-Pierre. *Linear Representations of Finite Groups*, volume 42 of *Graduate Texts in Mathematics*. Springer, 1977.