Lecture 1: Foundations of Group Equivariant Deep Learning

Lecturer: Renjie Liao *Date: Jan 13* *Scribes: Renjie Liao*

Groups

A **Group** $(G, \cdot)$ is a set G equipped with a binary operation $\cdot$ (called product) that satisfyies four axioms:

- Closure: $\forall g, h \in G, g \cdot h \in G$.
- Associativity: $(g \cdot h) \cdot k = g \cdot (h \cdot k)$.
- Identity: $\exists e \in G$ s.t. $g \cdot e = e \cdot g = g$.
- Inverse: $\forall g \in G, \exists g^{-1} \in G$ s.t. $g \cdot g^{-1} = e$.

Examples: *The Cyclic Group of Order 4 C_4:* This group represents the symmetries of a square that can be reached via discrete rotations with angles $\{0°, 90°, 180°, 270°\}$.

- Group Elements: $\{0°, 90°, 180°, 270°\}$
- Group Operation: The composition of rotations. Because we are dealing with a circle, the result is calculated modulo 360°. For example, $90° \cdot 180° = 270°$, $270° \cdot 180° = 450°$. Since $450° \equiv 90°$ (mod 360), the result is 90°.
- Identity (e): 0°. For any angle $g \in C_4$: $g \cdot 0° = 0° \cdot g = g$.
- Inverse (g^{-1}): The inverse is the operation that "undoes" a rotation, returning the object to its original orientation. For example, inverse of 0°: 0°, inverse of 90°: 270° since $90° \cdot 270° = 360° \equiv 0°$, and inverse of 180°: 180° since $180° \cdot 180° = 360° \equiv 0°$.

The Translation Group: For a d-dimensional space (usually $d = 2$ for images), the translation group is the set of all vectors $\mathbb{R}^d$ (continuous) or $\mathbb{Z}^d$ (discrete) under the operation of vector addition.

- Group Elements: A vector $\mathbf{v} \in \mathbb{R}^d$ (continuous) or $\mathbf{v} \in \mathbb{Z}^d$ (discrete) representing a shift. For example, $g = (3, -2)$ means "shift 3 units right and 2 units down."
- Group Operation ($+$): The "product" of two translations g_1 and g_2 is their vector sum: $g_1 + g_2$.
- Identity (e): The zero vector $\mathbf{0} = (0, 0)$. Shifting by zero changes nothing.
- Inverse (g^{-1}): For a shift $\mathbf{v}$, the inverse is $-\mathbf{v}$. If you shift right by 3, you undo it by shifting left by 3.

Subgroups

Let $(G, \cdot)$ be a group. A subset $H \subseteq G$ is a **subgroup** if it satisfies three conditions:

- Identity: The identity element e of G is in H.
- Closure: If $h_1, h_2 \in H$, then $h_1 \cdot h_2 \in H$.
- Inverses: If $h \in H$, then $h^{-1} \in H$.

Examples: *The Trivial Subgroup:* For any group G, the set containing only the identity $\{e\}$ is the smallest possible subgroup.

Integers in Real Numbers: $(\mathbb{Z}, +)$ is a subgroup of $(\mathbb{R}, +)$. If you add two integers, you get an integer; the negative of an integer is an integer; and 0 is an integer.

Group Homomorphism

Let $(G, \cdot)$ and $(H, *)$ be two groups. A function $\phi : G \to H$ is a **group homomorphism** if, for all $a, b \in G$, the following identity holds:

$$\phi(a \cdot b) = \phi(a) * \phi(b).$$

In other words, the result is the same whether you perform the group operation before or after applying the map. A homomorphism is a map that respects the group operation. It allows us to relate the structure of one group to another, even if the groups have different sizes.

Examples: *Determinants:* Consider the group of invertible $n \times n$ matrices under multiplication, $GL(n, \mathbb{R})$, and the group of non-zero real numbers under multiplication, $(\mathbb{R} \setminus \{0\}, \times)$.

The determinant map $\det : GL(n, \mathbb{R}) \to \mathbb{R} \setminus \{0\}$ is a homomorphism because:

$$\det(A \cdot B) = \det(A) \times \det(B).$$

This homomorphism collapses a complex, high-dimensional space (matrices) into a simpler one (scalars) while preserving the "logic" of multiplication.

Group Isomorphism

A homomorphism $\phi : G \to H$ is an **isomorphism** if ϕ is a bijection (it is both injective/one-to-one and surjective/onto). If such a map exists, we say G and H are isomorphic, denoted as $G \cong H$.

An isomorphism implies that G and H are identical in structure; they only differ in the "names" of their elements.

Examples: *Logarithms:* The group of positive real numbers under multiplication $(\mathbb{R}^+, \times)$ is isomorphic to the group of all real numbers under addition $(\mathbb{R}, +)$ via $\phi(x) = \ln(x)$, since

$$\ln(a \cdot b) = \ln(a) + \ln(b).$$

Rotations and Integers: The group C_4 of discrete $90°$ rotations is isomorphic to the group of integers modulo 4, $(\mathbb{Z}_4, + \pmod 4)$, where we construct the mapping as $0° \to 0$, $90° \to 1$, $180° \to 2$, and $270° \to 3$. Rotating $90°$ twice results in $180°$, just as $1 + 1 = 2 \pmod 4$.

Normal Subgroups

A subgroup H of a group G is called a **Normal Subgroup** (denoted $H \trianglelefteq G$) if it is invariant under conjugation by any element of G. Formally, H is normal if for every $g \in G$ and every $h \in H$:

$$ghg^{-1} \in H.$$

Intuition: If you take an element h from the subgroup, "translate" it using g, and then "translate it back" using g^{-1}, you must land back inside the subgroup. If H is normal, the "structure" of H looks the same from the perspective of any element in G.

Examples: In $(\mathbb{Z}, +)$, the subgroup of even numbers $2\mathbb{Z}$ is normal since for any $g \in \mathbb{Z}$ and $h \in 2\mathbb{Z}$, we have $g + h - g = h$, *i.e.*, $ghg^{-1} = h$.

Semi-direct Product

The Semi-direct Product (denoted by the symbol $\rtimes$) is a way of "gluing" two groups together to form a larger group where one group acts on the other.

Let N and H be groups, and let $\varphi : H \to \mathrm{Aut}(N)$ be a homomorphism that describes how H acts on N. Here, $\mathrm{Aut}(N)$ is the *automorphism group* of N, *i.e.*, the set of all group isomorphisms $\alpha : N \to N$ with composition as the group operation. The semi-direct product $G = N \rtimes_\varphi H$ is the set of pairs (n, h) with the following multiplication law:

$$(n_1, h_1) \cdot (n_2, h_2) = (n_1 \cdot \varphi_{h_1}(n_2), \quad h_1 \cdot h_2).$$

Examples: *Special Euclidean Groups $SE(2)$:* The Special Euclidean group $SE(2)$ is the semi-direct product of the continuous translation group $\mathbb{R}^2$ and the rotation group $SO(2)$.

- N (The Normal Subgroup): $\mathbb{R}^2$ (Translations).
- H (The Acting Subgroup): $SO(2)$ (Rotations R_θ).
- The Action: $\varphi_\theta(\mathbf{v}) = R_\theta \mathbf{v}$ (The rotation matrix R_θ rotates the vector $\mathbf{v}$).
- The Product in $SE(2)$: Suppose you have $g_1 = (\mathbf{u}, \theta)$ and $g_2 = (\mathbf{v}, \phi)$, then

$$g_1 \cdot g_2 = (\mathbf{u} + R_\theta \mathbf{v}, \quad \theta + \phi).$$

Plane Symmetry Groups p_4: The $p4$ group is the semi-direct product of the discrete grid $\mathbb{Z}^2$ and the four 90° rotations C_4. Elements: $g = ((x, y), r)$, where $x, y \in \mathbb{Z}$ and $r \in \{0, 1, 2, 3\}$.

The "Twist": When $r = 1$ (90° rotation), the translation $(1, 0)$ (one step right) becomes $(0, 1)$ (one step up). For example, let $g_1 = ((0, 0), 1)$ (a 90° rotation at origin) and $g_2 = ((1, 0), 0)$ (a shift 1 unit right).

$$g_1 \cdot g_2 = ((0, 0) + R_{90^\circ}(1, 0), \quad 1 + 0) = ((0, 1), 1).$$

Even though g_2 was a shift to the right, the total operation results in a shift up because the rotation was applied first.

Group Actions

A **Group Action** describes how a group G transforms a set X. Formally, a map $G \times X \to X$ is an action if:

- Identity: $e \cdot x = x$
- Composition: $g \cdot (h \cdot x) = (gh) \cdot x$

Example: The translation group $(\mathbb{Z}^2, +)$ acts on the pixel grid X. A group element $g = (u, v)$ acts on a coordinate (x, y) by addition: $(x + u, y + v)$.

Left Regular Representations

Let G be a group, X be a set upon which G acts. Let $\mathcal{F}$ be a function space, where each element $f \in \mathcal{F}$ is a function defined on X. The **Left Regular Representation** L is a map that assigns to each group element $g \in G$ a linear operator $L_g : \mathcal{F} \to \mathcal{F}$, defined by:

$$[L_g f](x) = f(g^{-1}x).$$

Note that sometimes we simplify $[L_g f](x)$ as $L_g f(x)$. But you should understand it as "take the function f, transform it by g to get a new function $L_g f$, and then evaluate that new function at point x."

To make sense of the abstract notations, we can think of

- f (The Signal): The input "state" of the layer, *e.g.*, an image. $f(x)$ is the value observed at position x.
- g (The Transformation): An element of the symmetry group we wish to respect, *e.g.*, a 90° rotation.
- $L_g f$ (The Transformed Signal): The entire function f after it has been moved by g.
- g^{-1} (The Inverse): Ensures that the value of the transformed function at point x is taken from the "source" point that would land on x after being moved by g.

Why the Inverse (g^{-1})? This is the most common point of confusion. We use the inverse to ensure that the representation is "covariant" with the group composition. If you want to shift a signal "forward" by g, you must look at what the signal was "backward" at $g^{-1}x$. It satisfies

$$[L_g L_h f](x) = [L_h f](g^{-1}x) = f(h^{-1}g^{-1}x) = [L_{gh}]f(x).$$

If we used g instead of g^{-1}, the order of operations would flip ($L_h L_g$), making it a "right" representation.

Example: The "Shift" Operator in 1D convolution.

- Domain X: The set of integers $\mathbb{Z}$ (pixel indices).
- Function f: The input vector where $f(i)$ is the intensity at pixel i.
- Group G: The translation group $(\mathbb{Z}, +)$.
- Action: Let $g = 2$ (a "right shift by 2").
- The Result: $[L_2 f](i) = f(i - 2)$.

If the original signal had a peak at $i = 0$ ($f(0) = 1$), the new signal $L_2 f$ will have its peak at $i = 2$, because $[L_2 f](2) = f(2 - 2) = f(0) = 1$. This confirms that L_g successfully "shifts" the feature map.

Quotient Spaces

Let G be a group and H be a subgroup of G. The **Quotient Space** (or set of cosets), denoted G/H (read "G mod H"), is the set of all left cosets of H in G:

$$G/H = \{gH \mid g \in G\}$$

where a coset gH is the set $\{gh \mid h \in H\}$.

Intuition: The quotient space G/H treats all elements in the same "cluster" (coset) as the same point. If G is a set of transformations, H represents a "sub-transformation" that we want to consider redundant.

Example: The Circle as a Quotient ($G = \mathbb{R}$, $H = \mathbb{Z}$). Consider the group of real numbers under addition $(\mathbb{R}, +)$ and its subgroup of integers $\mathbb{Z}$. The quotient space $\mathbb{R}/\mathbb{Z}$ essentially says: "Two numbers are the same if they differ by an integer." For example, 0.2, 1.2, and 2.2 all belong to the same coset $0.2 + \mathbb{Z}$. The result of this "collapsing" is the Circle (S^1). Every number is mapped to its fractional part in $[0, 1)$, where 1 wraps back to 0.

Homogeneous Spaces

A set X is a **homogeneous space** of a group G if G acts **transitively** on X. Transitivity means that for any two points $x_1, x_2 \in X$, there exists at least one group element $g \in G$ such that:

$$g \cdot x_1 = x_2$$

Because of this property, a homogeneous space X can always be identified with a Quotient Space G/H, where H is the *stabilizer subgroup* of a chosen origin point $x_0 \in X$. The stabilizer is defined as:

$$H = \{g \in G \mid g \cdot x_0 = x_0\}$$

Examples: *The 2D Euclidean Plane* $\mathbb{R}^2$*:* The plane is a homogeneous space for the Special Euclidean Group $SE(2)$ (2D translations and 2D rotations).

- Transitivity: Given any two coordinates on a map, you can move from one to the other by translating (sliding) and potentially rotating the map.
- Stabilizer (H): If you pick the origin $(0, 0)$, the set of all transformations that leave $(0, 0)$ fixed are the rotations around the origin ($SO(2)$).
- Relation: Therefore, $\mathbb{R}^2 \cong SE(2)/SO(2)$.

The 2-Sphere S^2*:* The surface of the Earth (idealized as a sphere) is a homogeneous space for the Special Orthogonal Group $SO(3)$ (3D rotations).

- Transitivity: You can reach any city on Earth from any other city simply by rotating the globe around some axis.
- Stabilizer (H): If you pick the North Pole as your point, the rotations that leave the North Pole fixed are the "spinning" motions around the vertical axis, *i.e.*, $SO(2)$.
- Relation: Therefore, $S^2 \cong SO(3)/SO(2)$.

The Discrete Grid $\mathbb{Z}^2$*:* In standard Image Processing, the pixel grid is a homogeneous space for the translation group $(\mathbb{Z}^2, +)$.

- Transitivity: Any pixel can be reached from any other pixel by a discrete shift. This is why we can use the same convolution kernel at every pixel—the "neighborhood" of every pixel is identical.
- Stabilizer (H): If you pick the origin $(0, 0)$, the set of all transformations that leave the entire grid unchanged are rotations by multiples of $90°$ around the point $(0, 0)$, *i.e.*, C_4.
- Relation: Therefore, $\mathbb{Z}^2 \cong p4/C_4$.