

Choices of Invariants for SE(3) Equivariant Transformers

Niklas Mather

July 31, 2023

1 Introduction

2 Mathematical Background

Functions on point clouds of homogeneous spaces

In this paper, we are concerned with constructing neural networks that can leverage data that exists as a point cloud on a homogeneous space X , corresponding to some group G_X . Such a space is a manifold, defined as the set of points on which G acts transitively. That is:

$$X = \{x \in X \mid x = g \cdot y, y \in X, g \in G_X\}$$

The notation $g \cdot x$ denotes a group element ‘acting’ on x . Formally, such a group action is a bijective function, from X to itself, also known as an automorphism of X . We also denote:

$$\text{Aut}(X) = \{T : X \rightarrow X \mid X \text{ is a bijection}\}$$

Where by definition we require that the map $T : G \rightarrow \text{Aut}(X)$ is an injective group homomorphism, that is $T(gh) = T(g)T(h) \forall g, h \in G_X$. Because not all groups are abelian, we must technically distinguish between ‘left’ and ‘right’ actions, which differ in whether h or g acts on x first during the action of the product gh . However, in this paper I will always refer to left actions, where $T_{gh}(x) = T_g(T_h(x))$.

A **point cloud** is a discrete subset of the manifold X , which I will denote \tilde{X} . It is common to also consider the points as the vertices of a graph that is embedded in X . That is, given a point $x \in \tilde{X}$, there is a set $N(x) \subset \tilde{X}$ which ‘share an edge’ with x . This graph rarely may or may not have any true semantic meaning, but in either case gives us a convenient notation with which to describe pairwise relationships between specific points.

Throughout this paper, I will reference molecular chemistry as a running example of the type of problem I am trying to solve. A molecule is a set of points (atoms) which are distributed through three-dimensional space. That is, one molecule’s position differs from another by the action of a translation on their coordinate vectors. Moreover, while atoms may be rotationally symmetric, *pairs* of atoms may differ by a 3d rotation. Thus, a molecule is a discrete set of points distributed on a manifold corresponding the group generated by all 3d rotations and translations, also known as $SE(3)$.

2.1 Feature fields

Of course, molecules are defined by other types of properties than their raw location. For example, each atom has a particular number of protons and neutrons (corresponding to elemental and isotope classes). Moreover, each atom may have geometric quantities with it (e.g. the sum of all force vectors acting on it). We refer to these as **features** defined on the point cloud.

While it is tempting to define features as a vector-valued function on a point cloud, it is more natural and useful to define them as functions on the underlying manifold. Then, for any discrete subset of the manifold $\tilde{X} \subset X$ we can define a function which agrees with the global function. That is, given a function $f : X \mapsto Y$ we can define a function $f' : \tilde{X} \mapsto Y$ which by construction has $f'(x) = f(x) \forall x \in \tilde{X}$.

$$f'(x) := \sum_j f(x_j) \delta(x - x_j)$$

What is Y , the codomain of the functions that we defined above? While we could consider the domain as \mathbb{R}^d , this ignores that there is a natural structure to the features, and so they should not be treated as a simple ‘list of numbers’. In particular, we noted that many features have a natu-

ral structure: for example, a force vector has a geometric structure which implies that it should transform under the action of the underlying group.

Problem statement: learning functions on point clouds

Equivariance

In this paper, we are particularly interested in functions

Features, Functions and Integral Transforms

Given a manifold X and Y , we can define an associated space of functions:

$$F(X) : -\{f : X \rightarrow \mathbb{R}^d\}$$

As stated above, we are interested in constructing equivariant maps between such spaces. Three natural questions are:

- How do we ensure that a given transform is equivariant?
- Is this process the same for attention layers?
- Can we generalise this construction beyond the case where $\text{Aut}(X)$ are group actions?

The first of these two have been answered in the literature before, but I took a slightly different route to establishing them, using a construction of integral transforms from category theory. I will now describe this construction here and show how it can be used to answer the questions above.

We are given two spaces: X and Y . For each of these we may define a space of functions $F(X)$ and $F(Y)$, and our goal is to construct a map $K : F(X) \rightarrow F(Y)$

First, consider the product space and the associated projection maps:

$$p(x, y) = x, \quad q(x, y) = y$$

$$p_X^X \times q_Y^Y$$

Now, we can use p to construct a map from $F(X)$ to $F(X \times Y)$ via precomposition. That is, given a function $f \in F(X)$, there is a unique function $\tilde{f}(x, y) = f \circ p(x, y) \in F(X \times Y)$. Denoting precomposition as p^* we have obtained a map $p^* : F(X) \rightarrow F(X \times Y)$. In category theory, this map is referred to as the 'pull-back'.

How can we construct a map from $F(X \times Y)$ to $F(Y)$? Our goal is to 'remove' the dependence on x for a given function $f(x, y)$. Thus we must aggregate over the fibres of the projection map $q^{-1}(y)$. There are many possible aggregation functions, but in keeping with the standard definition of the integral transform I will here use integration. Thus we have defined the 'push-forward' $q_* : F(X \times Y) \rightarrow F(Y)$.

$$(q_* f)(y) = \int_{q^{-1}(y)} f(x, y) dx$$

Finally, we note that multiplication by a kernel is a transformation $k \cdot : F(X) \times Y \rightarrow F(X \times Y)$. Thus, we have the following diagram:

$$\begin{array}{ccc} F(X) \times Y & \xrightarrow{k \cdot} & F(X \times Y) \\ \downarrow p^* & & \downarrow q_* \\ F(X) & \xrightarrow{k \cdot p^*} & F(Y) \end{array}$$

Constructing Equivariant Integral Transforms

This construction allows us to explore the questions we raised at the beginning of the section. First, I will answer how we ensure that K is equivariant. Suppose we have group actions $T_X : X \mapsto X$ and $T_Y : Y \mapsto Y$. Each of these lift to actions $\tilde{T}_X : F(X) \mapsto F(X)$ and $\tilde{T}_Y : F(Y) \mapsto F(Y)$. Under what circumstances does $K\tilde{T}_X f = \tilde{T}_Y Kf$ for every choice of f ?

Choose any $f \in F(X)$, $T_X : x \mapsto g \cdot x$, $T_Y : y \mapsto h \cdot y$. Then, our commutativity condition gives:

$$\begin{aligned} K\tilde{T}_X f &= \tilde{T}_Y Kf \\ q_* k \cdot p^* T_g f(y) &= q_* k \cdot p^* f(h^{-1}y) \end{aligned}$$

Recall that q_* maps from the fibres $q^{-1}(y)$, therefore, consider the following illustration of each step of the integral transform:

Now, in what sense are group actions special in the above construction? We did not rely on The answer is that groups allow us to construct the correspondences $T_g \mapsto T_y$ very efficiently using orbits.

Attention Mechanisms as Integral Transforms

Tensor Products

Choices of invariants

3 Design of the Transformer

4 Ablation Study