Choices of Invariants for SE(3) Equivariant Attention Mechanisms

Niklas Mather

August 18, 2023

1 Introduction

Attention mechanisms are a neural network layer which parameterise the interactions between different inputs to a neural network by the features associated with those inputs. One of their benefits is that they can , and so provide a recipe that generalises fairly easily to new domains, including (REF), on problems ranging from X (REF) to Y (REF) Z (REF). While this generality is useful, it is at odds with a competing objective of neural network design: to leverage prior knowledge of the domain structure for better performance.

For example, group-equivariant neural networks use knowledge of the symmetries of the input domain to reduce generalisation error (REF) and reduce sample complexity (REF). This approach relies on constraining the forms of the layers in a given neural network so that they are 'equivariant' with respect to symmetry transformations, and has been extended to include equivariant attention mechanisms, which achieve competitive performance with standard convolutional group-equivariant networks on a number of benchmarks.

These group-equivariant attention mechanisms are the focus of this paper. Specifically, we build on previous work that solved for the equivariance constraints in the case of the standard Euclidean group in 3 dimensions (SE(3)) (REF). Within this framework, there are still a number of design choices available to the user, especially in how to paramaterise the group invariants. The focus of our paper will be on the impact of these design choices.

The work is divided into three sections:

- A summary of the theory of group-equivariant attention mechanisms and a precise statement of the problem.

- A demonstration of how equivariance arises from integral transforms, from the perspective of category theory.

- An ablation study of the impact of different choices of group invariants on the MD17 dataset.

2   Mathematical Background

Functions on point clouds of homogeneous spaces

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. A homogeneous space is defined as the set of points on which its corresponding group acts transitively. That is:

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

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:

$$\mathrm{Aut}(X) = \{T : X \to X | X \text{ is a bijection }\}$$

Where by definition we require that the map $T : G \to \mathrm{Aut}(X)$ 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 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 $N(x) \subset \tilde{X}$ of points which share an edge with x - called the neighbourhood of X - which share an edge with x. This graph 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.

These features could be described as a vector-valued function on a point cloud, but it is more natural and useful to define them as functions over the underlying space if we expect them to be continuous on that domain. The two perspectives can be made compatible in the following way: given a function $f : X \mapsto Y$ we can define a function $f' : \tilde{X} \mapsto Y$ which by construction has $f'(x) = f(x) \forall x \in \tilde{X}$.

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

Steerability

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

3   Equivariance from a categorical perspective

Features, Functions and Integral Transforms

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

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

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 $\mathrm{Aut}(X)$ 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 consuct a map $K : F(X) - > F(Y)$

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

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

$$X \times Y$$

$$p \qquad q$$

$$X \qquad\qquad Y$$
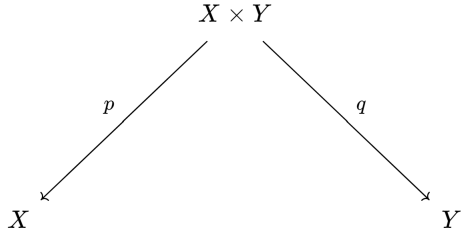
Now, we can use p to construct a map from $F(X)$ to $F(X \times Y)$ via precomposition. That is, given a function $f \in F(X)$, there is a unique function $\tilde{f}(x,y) = f \circ p(x,y) \in F(X \times Y)$. Denoting precomposition as p^* we have obtained a map $p^* : F(X) \to F(X) \times Y$. In category 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 definiton of the integral transform I will here use integration. Thus we have defined the 'push-forward' $q_* : F(X \times Y)$.

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

Finally, we note that multiplication by a kernel is a transformation

$$k \cdot : F(X) \times Y \to F(X \times Y)$$

Thus, we have the following diagram:

The diagram shows:
$$F(X) \times Y \xrightarrow{\;k\cdot\;} F(X \times Y)$$
with vertical arrows p^* (left) and q_* (right), and bottom:
$$F(X) \dashrightarrow F(Y), \qquad K = q_*[k \cdot p*]$$

Constructing Equivariant Integral Transforms

The above diagram characterises all possible linear transformations from $F(X)$ to $F(Y)$, but we are explicitly considered in those that that commute with morphisms of the category $F(X)$. If X is a homogeneous space and the morphisms are group actions, this corresponds to solving for the equivariance constraints.

Suppose we have two transformations $T_X : X \mapsto Y$ and $T_Y : Y \mapsto Y$. Each of these lift to actions $\tilde{T}_X : F(X) \mapsto F(X)$ and $\tilde{T}_Y : F(Y) \mapsto F(Y)$, given by $\tilde{T}_X f(x) := f(T(x))$. Under what circumstances does $K T_X f = T_Y K f$ for every choice of f?

Choose any $f \in F(X)$. Then, the commutativity condition gives:

$$K \tilde{T}_X f = \tilde{T}_Y K f$$
$$q_* k \cdot p^* \tilde{T}_X f(y) f = q_* k \cdot p^* f(T_Y(y))$$
$$\int_{q^{-1}(y)} k(x,y) f(T_X(x)) dx = \int_{q^{-1}(T_Y(y))} k(x, T_Y(y)) f(x) dx$$
$$\int_{q^{-1}(y)} k(T_X^{-1}(x), y) f(x) dx = \int_{q^{-1}(T_Y(y))} k(x, T_Y(y)) f(x) dx$$

Now, under the condition that the fibres (considered as sets) are invariant under the group action:

$$q^{-1}(y) = q^{-1}(T_Y(y))$$

We can equate the terms under the integral signs using the fundamental theorem of calculus:

$$k(T_X^{-1}(x), y)f(x) = k(x, T_Y(y))f(x)$$

Choosing in particular the case where f is a constant function, we have:

$$k(T_X^{-1}(x), y) = k(x, T_Y(y))$$
$$k(x, y) = k(T_X(x), T_Y(y))$$

Therefore, we see that there is a correspondence between $\tilde{T}_X$ and $\tilde{T}_Y$ as long as k is constant on the curves in the product space $X \times Y$ induced by T_X and T_Y.

Visual Intuition

This construction also affords us some interesting visual intuition of how equivariance is achieved. Consider the following illustration of $F(X) \times Y$ and $F(X \times Y)$.

Attention Mechanisms as Integral Transforms

Our original goal was to understand how attention mechanisms could be made equivariant. This can be acheived straightforwardly using the compositional construction. We only have to modify the definition of q so that it maps from the set of neighbours of a point to the point:

$$q^{-1}(y) := N(y)$$

Then, the fibres $q^{-1}(y)$ are once again invariant under the group structure, and we can repeat the argument above.

Moreover, the coefficients α_{ij} define a kernel $k(x_i, y_j)$

All parts of this diagram are by definition invariant under group actions except for the computation of the key, value, and query operations.

Discussion: does this generalise beyond groups?

In the above discussion, I deliberately avoided referencing the transformations as group actions. It's worth pausing to consider wha

4 Ablation Study on SE(3)

Equivariant maps in the SE(3) context

4.1 Architecture design

4.2 Results

5 Conclusion