# A Geometric Insight into Equivariant Message Passing Neural Networks on Riemannian Manifolds

**Ilyes Batatia** [1 2]

## Abstract

This work proposes a geometric insight into equivariant message passing on Riemannian manifolds. As previously proposed, numerical features on Riemannian manifolds are represented as coordinate-independent feature fields on the manifold. To any coordinate-independent feature field on a manifold comes attached an equivariant embedding of the principal bundle to the space of numerical features. We argue that the metric this embedding induces on the numerical feature space should optimally preserve the principal bundle's original metric. This optimality criterion leads to the minimization of a twisted form of the Polyakov action with respect to the graph of this embedding, yielding an equivariant diffusion process on the associated vector bundle. We obtain a message passing scheme on the manifold by discretizing the diffusion equation flow for a fixed time step. We propose a higher-order equivariant diffusion process equivalent to diffusion on the cartesian product of the base manifold. The discretization of the higher-order diffusion process on a graph yields a new general class of equivariant GNN, generalizing the ACE and MACE formalism to data on Riemannian manifolds.

## 1. Introduction

In the last decade, deep learning has emerged as the predominant paradigm for a wide range of applications in machine learning. Exploiting the underlying structure of the data is at the core of the success of deep learning. Convolutional Neural Networks (CNNs) (LeCun et al., 1989) exploit the spatial structure of images via the locality of the filters and the weight sharing that enhances its generalizability. The success of CNNs in computer vision has made clear the benefits of explicitly using translation equivariant networks. However, many relevant problems exhibit more complex symmetries than images. Such problems benefit from using latent representations related to the specific underlying symmetry group theory. Equivariant neural networks have emerged as the general class of models that internally transform according to a given symmetry group including permutation invariance for graphs (Gilmer et al., 2017; Thiede et al., 2021); spatial rotations for 2D (Cohen & Welling, 2016; Esteves et al., 2017) and 3D (Bartók et al., 2013; Anderson et al., 2019; Drautz, 2020; Batzner et al., 2022; Batatia et al., 2022b; Toshev et al., 2023) data; or Lorentz boost (Bogatskiy et al., 2022; Li et al., 2022; Munoz et al., 2022) or more generally Lie groups (Batatia et al., 2023). The concept of spaces with symmetries is inherent to all physical phenomena, for instance, the SO(3) group in molecular interaction and the SO(3,1) group in particle physics. Beyond the symmetries, it is crucial to generalize equivariant neural networks to processes on more general domains than the Euclidean space, as physics exhibits many spatial domains, including the Minkowski space in special relativity. Previous study (Weiler et al., 2021; Cohen et al., 2019) generalized CNNs to Riemannian manifolds by defining a notion of coordinate-independent feature fields.

Machine learning on graphs has shown to be successful in a broad range of applications, including physical science (Carleo et al., 2019; Bronstein et al., 2021). Large amounts of data are available as graphs; therefore, developing efficient algorithms over graphs is crucial. Graph Neural Networks (GNNs) emerged as one of the most promising methods to apply Deep Learning principles to data over graphs. The relationship between GNNs and diffusion processes has been studied in the attentional flavour (Chamberlain et al., 2021). The more general message-passing scheme can be understood as a partial differential equation (Gilmer et al., 2017).

This study aims to extend the concept of equivariant message passing to data on Riemannian Manifolds. Our approach demonstrates how, without non-linear activation,

[1]ENS Paris-Saclay, Université Paris-Saclay, 91190 Gif-sur-Yvette, France [2]University of Cambridge, Cambridge, CB2 1PZ UK. Correspondence to: Ilyes Batatia <ilyes.batatia@ens-paris-saclay.fr>.

*Proceedings of the $2^{nd}$ Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML) at the $40^{th}$ International Conference on Machine Learning, Honolulu, Hawaii, USA. 2023.* 

one can derive equivariant message passing from the minimization of a specific functional: the twisted Polyakov action. The specialization to Euclidean spaces recovers well known equivariant message passing architectures. This functional can be seen as measuring how accurate the feature field is at embedding the geometry of the manifold along with the action of a given group on it into a vector space. The update rule of the architecture comes from the discretization of the diffusion process resulting from the minimization of a twisted form of this functional. We propose a higher-order equivariant diffusion process equivalent to diffusion on the cartesian product of the base manifold. The discretization of the higher-order diffusion process on a graph yields a new general class of equivariant GNN closely related to the MACE (multi- Atomic Cluster Expansion) formalism. While the ideas developed in this paper could interest the community, they are far from mature and complete. This paper is intended to spark interest and hopefully lead to a complete understanding of the interplay between modern geometric machine learning architectures and the associated functional they minimize.

In the first chapter, we present Message Passing Neural Networks (MPNNs) formalism and make the equivariance constraint explicit. We explain the connection between MPNNs and diffusion through the discretization of Beltrami flow. In the second chapter, we construct an equivariant feature diffusion to apply the analogy between MPNNs and diffusion processes to equivariant MPNNs. We first represent coordinate independent feature fields on Riemannian manifolds as sections of associated vector bundles. To any coordinate-independent feature field comes a canonically attached equivariant embedding of the principal bundle to the space of numerical features. Minimizing the Polyakov action with respect to the graph of this embedding yields a flow on the associated bundle that is equivalent to a diffusion process. By discretizing this flow for a fixed time step, one obtains a message passing on the manifold. To realize a higher-order message passing on the manifold, we considered Equivariant feature diffusion on the Cartesian product of the manifold allows us to address this question. Finally, we obtain a discrete version of the message passing on the graph by embedding a graph into a Riemannian manifold. We show that such formulation is very closely related to the MACE architecture proposed for the representation of molecules in GNNs in the case of manifolds being $\mathbb{R}^3$ and the group $SE(3)$.

## 2. Background

### 2.1. Equivariant Message Passing Neural Networks

Message Passing Neural Networks (MPNN) (Gilmer et al., 2017) are a general class of Graph Neural Networks that parametrize a mapping from labelled graphs to a vector space. In MPNN frameworks, each node is labelled with a state updated via successive message passing between neighbours. After a fixed number of iterations, a readout function maps the state to the space of real numbers. Equivariant MPNNs emerge if the states of the nodes have additional vector features along with an action of a group on them. In this case, one seeks mappings that preserve a given group's action.

**Nodes states** Let $\Gamma = (\mathcal{V} = \{1, ..., n\}, \mathcal{E})$ be a graph with $\mathcal{V}$ and $\mathcal{E}$ denoting the nodes and edges respectively. The state of node $i$, $\sigma_i^{(t)}$ is composed of two properties :

$$\sigma_i^{(t)} = (r_i, h_i^{(t)}) \tag{1}$$

with $r_i$ the positional attribute of the node and $h_i^{(t)}$ is a learnable feature of node $i$. These learnable features are updated after each iteration of message passing, with an iteration index by $t$.

**Message passing and update** During each round of message passing, features on each node $h_i^{(t)}$ are updated based on aggregated messages, $m_i^{(t)}$ derived from the states of the atoms in the neighbourhood of $i$, denoted by $\mathcal{N}(i)$ :

$$m_i^{(t)} = \bigoplus_{j \in \mathcal{N}(i)} M_t(\sigma_j^{(t)}, \sigma_i^{(t)}) \tag{2}$$

where $\bigoplus_{j \in \mathcal{N}(i)}$ is any permutation invariant operation over the neighbours of node i and $M_t$ is a learnable function acting on states of nodes $i$ and $j$. The messages are used to update the features of node $i$ with a learnable update function $U_t$:

$$\sigma_i^{(t+1)} = (r_i, h_i^{(t+1)}) = (r_i, U_t(m_i^{(t)})) \tag{3}$$

**Readout** After $T$ iterations of message passing and update, in the readout phase, the states of the nodes are mapped to the output $y_i$ by a learnable function $R$:

$$y_i = R(\sigma_i^{(T)}) \tag{4}$$

**Equivariance** One can ask for an MPNN to be equivariant with respect to the action of a group $G$. We will consider the case where $G$ is a reductive Lie group, and V is a representation of $G$ with action $\rho$. Formally, a message as a generic function of the states $(\sigma_{i_1}, ..., \sigma_{i_n})$ is said to be $(\rho, G)$ equivariant if it respects the constraint:

$$m^{(t)}(g \cdot (\sigma_{i_1}^{(t)}, ..., \sigma_{i_n}^{(t)})) = \rho(g) \cdot m^{(t)}(\sigma_{i_1}^{(t)}, ..., \sigma_{i_n}^{(t)}) \quad \forall g \in G \tag{5}$$

where $g \cdot (\sigma_{i_1}^{(t)}, ..., \sigma_{i_n}^{(t)})$ denotes an action of $G$ over the states such that

$$g \cdot \sigma_i^{(t)} = (g \cdot r_i, \rho_h(g) h_i^{(t)}) \tag{6}$$

with $\rho_h(g)$ denoting the action of g on $h$. In practice from equivariance constraint of equation (5) results constraints on the type of operations $M_t, \bigoplus, U_t, R_t$.

### 2.2. Diffusion process and message passing

Recently a graph Beltrami flow has been introduced (Chamberlain et al., 2021) by analogy with diffusion processes on images and related it to a subcase of MPNNs referred to as Attentional graph neural networks. The graph Beltrami flow evolves the features as :

$$\frac{\partial h_i^{(t)}}{\partial t} = \sum_{j:i\to j\in\mathcal{E}} a(\sigma_j^{(t)}, \sigma_i^{(t)})(h_j^{(t)} - h_i^{(t)}) \tag{7}$$

The solution of this equation for long times minimizes a discretized version of the Polyakov action, which is equivalent to finding an optimal embedding. With a discretization with a time step of 1, one obtains:

$$h_i^{(t+1)} = h_i^{(t)} + \sum_{j:i\to j\in\mathcal{E}} a(\sigma_j^{(t)}, \sigma_i^{(t)})(h_j^{(t)} - h_i^{(t)}) \tag{8}$$

This update formula corresponds precisely to the attention flavour of MPNNs, and the attention weights $a(\sigma_j^{(t)}, \sigma_i^{(t)})$ can be understood as anisotropic diffusion coefficients. More general flavours are possible by studying the non-linear equation of evolution of the type $\frac{\partial h_i^{(t)}}{\partial t} = \Phi(\{\sigma_j\}_{j\in\mathcal{N}(i)})$. In the next section, we will show that equivariant message passing can be achieved as an equivariant diffusion process resulting from a twisted form of the Polyakov action. Other analogies of GNNs with energy minimization have been proposed in (Giovanni et al., 2022).

## 3. A geometric insight into equivariant message passing

To apply the rich analogy between diffusion processes and equivariant message passing, one must represent message passing as a diffusion that respects the underlying symmetry imposed by the structure group. For this purpose, we will define an equivariant features diffusion process on a Riemannian Manifold. This will rely on defining equivariant features as sections of an associated bundle as proposed in (Cohen et al., 2019; Weiler et al., 2021). A general Laplacian will carry the diffusion process that updates the features on the associated bundle. This diffusion process results from minimizing a twisted form of the Polyakov action that preserves equivariance. The equivariant features diffusion process is thus related to finding an optimal mapping between the principal bundle of a manifold and a given irreducible representation of the structure group.

### 3.1. Equivarient features fields as sections of associated bundles

We aim to construct a geometric notion of equivariant features on the manifold as outlined in (Weiler et al., 2021; Cohen et al., 2019). We will review some fundamentals of differential geometry for this construction. For more mathematical details, see (Kobayashi & Nomizu, 1963). As manifolds do not come with a preferential reference frame (gauge), features must be defined up to an arbitrary choice of coordinates. It is natural to ask for the network to be independent of the coordination as proposed in (Weiler et al., 2021). It results naturally in asking the network to be equivariant under gauge transformations, i.e. local change of reference frame. Let $(\mathcal{M}, \eta_{\mathcal{M}})$ be a Riemannian manifold of dimension $\mathfrak{m}$. We will start by making more specific the notion of coordinate independence. To each point $x \in \mathcal{M}$, one can attach a tangent space $T_x\mathcal{M}$ which locally looks like $\mathbb{R}^{\mathfrak{m}}$. Each tangent space $T_x\mathcal{M}$ is isomorphic to $\mathbb{R}^{\mathfrak{m}}$; however, there is no canonical isomorphism. A preferred choice of local isomorphism is called a gauge (Naber, 1997).

**Definition 3.1** (Gauge)**.** Let $x \in \mathcal{M}$ and $U^A$ be a neighborhood of $x$. A gauge is defined as a smooth and invertible map:

$$\psi_x^A : T_x\mathcal{M} \to \mathbb{R}^{\mathfrak{m}} \tag{9}$$

that specifies a preferred choice of isomorphism between $T_x\mathcal{M}$ and $\mathbb{R}^{\mathfrak{m}}$.

Gauges coordinatize tangent spaces only in local neighbourhoods, and in all generality, they can not be extended to the full manifold while preserving their smoothness requirement. One can consider a collection of local neighbourhoods and their corresponding gauges called an atlas :

$$\mathcal{A} = \{(U^X, \psi^X)\}_{X\in\mathcal{X}} \tag{10}$$

such that $\bigcup_{X\in\mathcal{X}} U^X = \mathcal{M}$. On intersections $U^A \cap U^B \neq \varnothing$, the different gauges $\psi_x^A, \psi_x^B$ are stitched together using smooth transition functions defined on the structure group $G$. From now on, we will consider $G$ to be a reductive Lie group.

**Definition 3.2** (Transition functions)**.** A transition function (see 3.1) between intersections $U^A \cap U^B \neq \varnothing$, is a map :

$$g^{BA} = U^A \cap U^B \to G, x \mapsto \psi_x^B \circ (\psi_x^A)^{-1} \tag{11}$$

Therefore one observes that a single coordinate free tangent vector $v \in T_xM$ is represented by two vectors in $v^A, v^B \in \mathbb{R}^{\mathfrak{m}}$ :

$$v^A = g^{AB} v^B, g^{BA} \in G \tag{12}$$

depending on the gauges $\psi^A, \psi^B$. We see that coordinate independence is closely related to the action of the structure group on the manifold that defines the transition maps

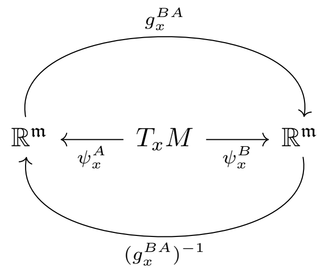


*Figure 1.* Relationship between gauge maps and transition functions.

between gauges. In practical implementations, we aim for equivariant message passing to diffuse feature fields relative to some local reference frame that take values in some vector space $V$ of dimension $c$. Let $\psi^A$ be a local gauge on a neighborhood $U^A$ of $M$. Relative to this gauge, a local feature field is defined as :

$$f^A : U^A \to V \tag{13}$$

given by a vector with $c$-dimensional (corresponding to $c$ channels) feature vector $f^A(x)$ at each point $x \in M$. According to some other gauge $\psi^B$ on $U^B$, one can measure $f^B$. As we require the global feature field $h$ to be coordinate independent, the different local feature fields must transform principally. Because gauge transformations are elements of $G$, the local features fields will transform according to some linear representations of $G$ on $V$ :

$$\rho : G \to Aut(V) = GL(c) \tag{14}$$

and we have the local features fields transforming as :

$$f^B(x) = \rho(g_x^{BA}) f^A(x), x \in U^A \cap U^B \tag{15}$$

The type of representations models different types of features. The scalar fields transform according to the trivial representation $\rho(g) = \mathbf{1}, \forall g \in G$ whose local numerical values are invariant under gauge transformations. More generally, feature fields that transform according to irreducible representations of $G$ are of great practical importance in physics. As $G$ is a reductive Lie group, any finite representation can be decomposed as a sum of irreducible representations. We will therefore restrict ourselves to the case where $V$ is a finite-dimensional representation.

To properly define the diffusion of equivariant feature fields, we must give a geometric interpretation of coordinate-independent feature fields. Global coordinate independent feature fields are sections of an associated vector bundle. Let $G$ be a reductive Lie group and $V$ a finite-dimensional representation of $G$ of dimension c and denote by $\rho$ the linear representation of $G$ on $V$. Fiber bundles allow for a global description of fields over a manifold and can thus represent feature fields.

**Definition 3.3** (Fiber bundle)**.** A fiber bundle is a quadruplet $(E, \pi, M, F)$ with E the total space, $M$ the base space and $F$ the typical fiber and a smooth surjective map $\pi : E \to M$. A fiber bundle is locally trivial as $\forall x \in M$ there exists a neighbourhood $U$ of $x$ and a diffeomorphism $\phi : U \times F \to \pi^{-1}(U)$, such that $\pi \circ \phi(x, f) = x, \forall f \in F$. The tuple $(\psi, U)$ is called a local trivialization.

To include more specific mathematical structures on the typical fiber, one can define subtypes of bundles. An important example is vector bundles, where F is a vector space, and this is crucial as it formalizes the concept of parameterizing a vector space by a manifold.

**Definition 3.4** (Vector bundle)**.** A vector bundle is a triplet $(E, \pi, M)$ with $E, M$ two manifolds, and $\pi : E \to M$ such that the preimage $\pi^{-1}(x)$ of $x \in M$ has the structure of a vector space.

To combine geometric properties with differential geometry, one needs to construct a principal bundle $P$, that locally looks like the product of $M$ with a structure group $G$ and such that transition maps are isomorphism (Warner, 1983).

**Definition 3.5** (Principal bundle)**.** A principal bundle is a quadruplet $(P, \pi, M, G)$ where $P$ and $M$ are two manifolds, $G$ a Lie group, $\pi : P \to M$ is a surjective map such that $\pi^{-1}(x)$ is diffeomorphic to $G$ and there is an action $\cdot$ of $G$ on $P$ such that :

- $\pi(p \cdot g) = \pi(p)$ for $p \in \pi^{-1}(x)$ and $g \in G$
- the restriction $G \times \pi^{-1}(x) \to \pi^{-1}(x)$ is free and transitive.

We call $M$ the base manifold, $P$ the total space and $G$ the structure group of the principal bundle. When no confusion can be made, $(P, \pi, M, G)$ is called $P$.

The disjoint union of bases of $T_xM, x \in M$ that are equivalent by the action of $G$ forms the total space of a principal bundle $PG(TM)$ over $M$ called the bundle of G-frames of $TM$.For now we will refer by $P$ to $PG(TM)$.

In the following, we will construct associated feature vector bundles with feature coefficients in $V$ as typical fibers. Under gauge transformations, these fibers are acted on by the group linear representation $\rho : G \to Aut(V)$. These features vector bundles are constructed as quotients.

**Definition 3.6** (Associated bundle)**.** Let $(P, \pi, M, G)$ be a principal bundle and $\rho$ a linear representation of $G$ on $V$. We define $E = (P \times V)/G$, a point of $E$ is of the form

$$[p, h] = \{(p \cdot g, \rho(g^{-1})h), g \in G\} \tag{16}$$

where $p \in P$ and $h \in V$. Let $\pi_E$ : $E$ be given by

$\pi_E[p, h] = \pi(p)$. Then $(E, \pi_E, M)$ forms a vector bundle called an associated vector bundle to P, denoted $P \times_{(\rho,G)} V$.

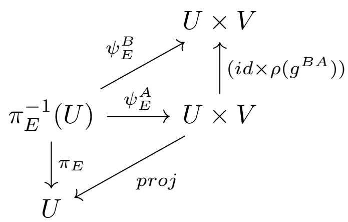


*Figure 2.* Trivialization of $E \xrightarrow{\pi_E} M$

The construction of $E$ is equivalent to coordinate independent feature vectors on M : features $f(x) \in E$ are expressed relative to arbitrary frames in $P$.

**Definition 3.7** (Coordinate free features field)**.** Coordinate free features fields are defined as smooth global sections $f \in \Gamma(E)$ that is a smooth map $f : M \to E$ such that $\pi_E \circ f = id_M$.

**Lemma 3.1.** The coordinate free feature fields $f \in \Gamma(E)$ and $(\rho, G)$-equivariant functions $h$ on $P$ are in one-to-one correspondence such that one can attach to any $h$ a canonical coordinate free feature field $f_h$.

*Proof.* Define a $(\rho, G)$-equivariant function $h : P \to V$ from the principal bundle to the representation $V$. Let $p_1, p_2 \in \pi^{-1}(x)$ and define $f_h^1(x) = [p_1, h(p_1)]$ and $f_h^2(x) = [p_2, h(p_2)]$. As $p_1, p_2 \in \pi^{-1}(x)$, there exists a $g \in G$ such that $p_1 = p_2 \cdot g$. By $(\rho, G)$-equivariance of $h$ and by the definition of a point of the associated bundle (16),

$$f_h^1(x) = [p_1, h(p_1)] = [p_2 \cdot g, h(p_2 \cdot g)] = \quad (17)$$

$$[p_2 \cdot g, \rho(g^{-1})h(p_2)] = [p_1, h(p_1)] \quad (18)$$

we have $f_h(x)$ invariant by the choice of $p \in \pi^{-1}(x)$. Thus for any $p \in \pi^{-1}(x)$, $\pi_E \circ f_h(x) = x$, and $f_h(x) \in \Gamma(E)$. Conversely, for $f \in \Gamma(E)$, we put $h_f(p) = v$ such that $f \circ \pi(p) = [p, v]$. Then $h_f$ is $(\rho, G)$-equivariant. Therefore the coordinate-free features fields naturally generate a corresponding equivariant map. We will denote $f_h$ the coordinate-free feature map canonically associated to the equivariant function $h$. □

**Definition 3.8** (Canonical association function)**.** We will write $\gamma$ the map associating to an equivariant function $h \in C^\infty(P, V)^{(\rho,G)}$ the section of $E$, $f_h$ such that $f_h = [p, h(p)]$ :

$$\gamma : C^\infty(P, V)^{(\rho,G)} \to \Gamma(E), h \mapsto f_h \quad (19)$$

On local neighborhood $U^A$ of $M$, the coordinate-free feature field is trivializable on an arbitrary local frame by the action of a gauge $\psi_{E,x}^A$ into numerical features $f_h^A$:

$$f_h^A(x) = \psi_{E,x}^A(f_h(x)), x \in U^A \quad (20)$$

A different choice of trivialization on different neighbourhoods $U^B$ yields different numerical features related by the action of the linear representation of $V$

$$f_h^B(x) = \rho(g^{BA})\psi_{E,x}^A(f_h(x)) \quad (21)$$

In all generality, the features consist of multiple independent feature fields on the same base space transforming according to different representations. The whole space is defined as the Whitney sum $\bigoplus_i E_i$. A common practice in equivariant message passing the construction of feature fields from a set of finite dimensional irreducible representations with the highest weight $\lambda$.

### 3.2. Diffusion of features fields and Polyakov Action

This section considers a coordinate-free feature field $f_h \in \Gamma(E)$ introduced in the previous section over a single finite-dimensional representation $V$ of $G$. We argue that an implicit regularization for the feature field is to realize an optimal embedding of $M$ in $E$, finding the optimal featurization of $M$ into the representation $V$ that respects the underlying structure imposed by the group $G$. This optimality constraint applies to its corresponding equivariant function $h$. In order to make the notion of optimality more precise, we will introduce the Polyakov action that measures the energy of an embedding and will be our optimality criterion.

We will assume that $G$ is a simply connected compact Lie group with Lie algebra $\mathfrak{g}$. For locally compact reductive Lie groups, most of the following discussion can be applied identically by considering the exposition on the maximal compact subgroup of the universal cover of the complexification of $G$ as they share the same finite-dimensional representations. One can induce a Riemannian metric $u$ on $P$ using the the Riemannian metric on $\mathcal{M}$, $\eta_{\mathcal{M}}$ and the invariant metric on $G$, $\eta_G$. To do so, consider the tangent space of $P$ at the point $p \in P$, $T_pP$. It can be decomposed into a sum of two subspaces, $V_pP$ called the vertical space, which is the kernel of the pushforward $\pi^* : T_pP \to T_{\pi(p)M}$ and the horizontal space $H_pP$ which is the complementary subspace. One can define naturally a connection on $P$ by looking at the following map $\phi_p : \mathfrak{g} \to V_pP$:

$$\phi_p(X) = \frac{d}{dt}(p \cdot \exp(tX))|_{t=0}, \quad X \in \mathfrak{g} \quad (22)$$

In the case where $G$ is a compact simply connected Lie groups, this map is invertible, and we call $\phi_p^{-1}$ the inverse. A similar map can be formed if the group has a finite number of disconnected components by multiplying the map by a discrete group $\mathbf{H} \in G$ containing a representative from each connected component. From this map, one can define an inner product on $P$. Let $dp_1 = v_1 + h_1$ and $dp_2 = v_2 + h_2$ be two points on $T_pP$, the inner product is defined as,

$$(dp_1, dp_2) = \eta_{\mathcal{M}}(\pi(h_1), \pi(h_2)) + \eta_G(\phi_p^{-1}(v_1), \phi_p^{-1}(v_2)) \tag{23}$$

The metric $u$ on $P$ is the metric induced by this inner product.

Moreover, we assume that $P \times V$ is equipped with a Riemannian metric $v = u \oplus \kappa \mathbf{I}_d$, for an arbitrary positive number $\kappa$ multiplying the identity of $V$. The graph of $h$ given by $\varphi_h : P \to P \times V$ realizes an embedding of $P$ into $P \times V$. This embedding induces a natural metric on $P \times V$ given by $\gamma_{\mu,\nu} = \frac{\partial \varphi_h^i}{\partial x_\mu} \frac{\partial \varphi_h^j}{\partial x_\nu} v_{i,j}$. We will now make precise our optimality criterion.

**Definition 3.9** (Optimality of the feature fields)**.** A feature field $f_h \in \Gamma(E)$ is optimal if the induced metric associated with the embedding $\varphi_h$ of the principal bundle to the numerical features space $V$ preserves optimally the metric on the Principal bundle.

A natural way to think about this optimal condition is that one wishes to construct features in a vector space that preserves as much as possible the geometry of the initial manifold. To measure the energy of $\varphi_h$ and correlatively of $f_h$, we study its Polyakov Action.

**Definition 3.10** (Polyakov action)**.** The Polyakov action of the embedding $\varphi$ is defined as,

$$S[\varphi_h, u, v] = \int_P u^{\mu,\nu} \frac{\partial \varphi_h^i}{\partial x_\mu} \frac{\partial \varphi_h^j}{\partial x_\nu} v_{i,j} dP \tag{24}$$

Minimizing the Polyakov with respect to embedding metrics $u$ and $v$ is equivalent to finding the optimal embedding $\varphi_{h,opt}$ of $P$ in $P \times V$.

The minimization of the precedent action yields the resolution of the Euler-Lagrange equation. One obtains a gradient descent flow in the form of a heat equation on $h : P \to V$ :

$$\frac{\partial h_t}{\partial t} = \Delta^P h \tag{25}$$

where $\Delta^P$ corresponds to the Laplace Beltrami operator on the principle bundle $P$. However, this flow equation is not guaranteed to preserve equivariance. To preserve equivariance of the initial condition $h^{(0)}$, (Batard & Sochen, 2012) proposed introducing a twisted version of the Polyakov action resulting in a gradient descent flow. The additional term in the action is expressed as a scalar product $\langle , \rangle$ on $V$.

**Definition 3.11** (Casimir Operator)**.** Let $\mathfrak{g}$ be the Lie algebra of $G$. Let $(\mathfrak{g}_1, ..., \mathfrak{g}_n)$ be an orthonormal basis of $\mathfrak{g}$ and $d\rho : \mathfrak{g} \to GL(V)$ the representation of $\mathfrak{g}$ induced by $\rho$. Let $(\mathfrak{g}^1, ..., \mathfrak{g}^n)$ be the dual basis on $\mathfrak{g}$ with respect to the Killing form. The Casimir operator $Cas \in GL(V)$ is defined as :

$$Cas = \sum_i^n d\rho(\mathfrak{g}^i) d\rho(\mathfrak{g}_i) \tag{26}$$

The twisted Polyakov action is given by adding a term to the Polyakov action :

$$S[\varphi_h, u, v] = \int_P u^{\mu,\nu} \frac{\partial \varphi_h^i}{\partial x_\mu} \frac{\partial \varphi_h^i}{\partial x_\nu} v_{i,j} + \frac{1}{2} \langle \text{Cas} \cdot h, h \rangle dP \tag{27}$$

Minimizing the equation with respect to $\varphi_h$, one obtains the following heat equation :

$$\frac{\partial h^t}{\partial t} = -(\Delta^P \otimes \mathbb{I} + \mathbb{I} \otimes Cas) h^t \tag{28}$$

which one can rewrite :

$$\frac{\partial h^t}{\partial t} = \Delta^E h^t \tag{29}$$

where we define $\Delta^E = -(\Delta^P \otimes \mathbb{I} + \mathbb{I} \otimes Cas)$ the generalized Laplacian on the associated bundle $E$. In the special case where $V$ is an irreducible representation, $Cas = \lambda \mathbb{I}$ and $\Delta^E = -(\Delta^P \otimes \mathbb{I} + \lambda \mathbb{I})$.

**Definition 3.12** (Equivariant Feature Diffusion Process)**.** Let $h^{(0)} : P \to V$ be an initial equivariant feature map. Let $\Delta^E$ be the generalized Laplacian on the associated bundle $E$. We define an Equivariant Diffusion Process as follows:

$$h^{(T)} = h^{(0)} + \int_0^T \Delta^E h^{(t)} dt \quad y = \mathbf{R}(h^{(T)}) \tag{30}$$

The feature map $h^{(T)}$ will be the optimal equivariant map from $P$ to the irreducible representation $V$ of $G$ for $T$ sufficiently long. Furthermore, $\mathbf{R}$ is a learnable readout function.

From the equivariant diffusion process, we can construct a smooth coordinate independent feature field using the feature field canonically attached to any equivariant function on $P$ :

$$f_h^{(t)}(x) = [p, h^{(t)}(p)], p \in \pi^{-1}(x) \tag{31}$$

### 3.3. Equivariant features propagator on reductive groups

The diffusion process on the associated bundle, $E$, can be understood in terms of the scalar heat kernel on $P$ as detailed in (Berline et al., 1992). First, we observe the following remark :

**Remark.** Let $\Delta^E = -(\Delta^P \otimes \mathbb{I} + \mathbb{I} \otimes Cas)$, then for some parameter $t$,

$$e^{-t\Delta^E} = e^{-tCas} e^{-t\Delta^P}$$

We have the following equality for $p_1 \in P$

$$(e^{-t\Delta^E} h)(p_1) = \int_E \langle p_1 | e^{-t\Delta^E} | p_2 \rangle h(p_2) dp_2 \tag{32}$$

where $\langle p_1 | e^{-t\Delta^E} | p_2 \rangle$ is the Schwartz Kernel of $e^{-t\Delta^E}$ in the Dirac notation.

**Proposition 3.1** (Getzler - Vergne - Berline)**.** If $p_1$ is a point of the principle bundle $P$ of a locally compact group $G$ and $\Delta^E$ the generalized Laplacian over the associated bundle $E$. Set $p_2$ is a chosen representative in $\pi^{-1}(x)$, for $x \in \mathcal{M}$. Then for any function $h : P \to V$, and $|dx|$ a Riemannian density,

$$(e^{-t\Delta^E} h)(p_1) = e^{-tCas} \int_G \int_{\mathcal{M}} \langle p_1 | e^{-t\Delta^P} | p_2 g \rangle \rho(g)^{-1} h(p_2) dg dx \tag{33}$$

In the case of $G$ is a non-compact reductive Lie group, the first integral taken over the maximal compact subgroup of the complexification of $G$ referred to as $K_{\mathbb{C}}$, following the Weyl unitary trick.

**Definition 3.13.** Let's $e^{-t\Delta^E}$ be the equivariant propagator of the diffusion such that :

$$h^{(t')} = e^{-(t'-t)\Delta^E} h^{(t)} \tag{34}$$

**Definition 3.14.** For any $p_1, p_2 \in P$, denote

$$\langle p_1 | e^{-(t'-t)\Delta^P} | p_2 \rangle = k_{t'-t}(p_1, p_2) \tag{35}$$

the heat kernel of $\Delta^P$ such that one can rewrite the propagator of equation (34) as

$$h^{(t')} = e^{-(t'-t)\Delta^E} h^{(t)} = \tag{36}$$

$$e^{-tCas} \int_{\mathcal{M}} \int_G k_{t'-t}(p_1, p_2 g) \rho(g)^{-1} h^{(t)}(p_2) dg dx \tag{37}$$

**Definition 3.15** (GM - Message passing)**.** Let $h^{(0)} : P \to V$ be an initial equivariant feature map. Let $\Delta^E$ be the generalized Laplacian on the associated bundle $E$. We define the message operation in a GM - Message passing as :

$$m^{(T)}(p_1) = e^{-TCas} \int_{\mathcal{M}} \int_G k_T(p_1, p_2 g) \rho(g)^{-1} h^{(0)}(p_2) dg dx \tag{38}$$

$$h^{(T)}(p_1) = U^{(T)}(m^{(T)}(p_1)) \tag{39}$$

with the function $U^{(T)}$ a learnable equivariant function. The usual equivariant updates are linear combinations from the vector space of $C^\infty(P, V)^{(\rho, G)}$.

**Remark.** GM-Message passing networks are gauge equivariant. A weight-sharing constraint needs to be preserved for the method to be also equivariant to the diffeomorphism of the manifold. GM-Message passing networks are indeed diffeomorphism-preserving maps as the heat kernel is shared across the manifold.

As this diffusion process has low correlation order (only pairwise interaction), it can be inefficient at modeling highly correlated interactions. Therefore, it is convenient to construct a higher-order GM-Message passing on a Cartesian product of manifolds. First, define $P^n$ the principal bundle on the base manifold $M^n$.

**Definition 3.16** (Higher order GM - Message passing)**.** Let $\tilde{h}^{(0)} : P^n \to V^n$, be an initial equivariant feature map. Let $E^n$ be the associated vector bundle to $P^n$ and $\Delta^{E^n}$ be the generalized Laplacian on the associated bundle $E^n$. Let $dx^n = \prod_\xi^n dx_\xi$ be a volume element on the manifold $\mathcal{M}^n$. We define a Higher order GM - Message passing as :

$$m^{(T)}(p_1) = \tag{40}$$

$$e^{-TCas} \int_{\mathcal{M}^n} \int_G \tilde{k}_T^N(p_1, p_2 g, .., p_n g) \rho(g)^{-1} \tilde{h}^{(0)}(p_2, ..., p_n) dg dx^n$$

Assume the following structure on the diffusion kernel and feature function,

$$\tilde{k}_T(p_1, p_2 g, .., p_n g) = \prod_{\xi=1}^n k_T(p_1, p_\xi g) \tag{41}$$

$$\tilde{h}^{(0)}(p_2 g, ..., p_n g) = \rho(g)^{n-1} \prod_{\xi=1}^n h^{(0)}(p_\xi g) \tag{42}$$

We can rewrite the Higher-order GM-Message Passing of equation (40) as :

$$m^{(T)}(p_1) = \tag{43}$$

$$e^{-TCas} \int_G \rho(g)^{-1} dg \prod_{\xi=1}^n \int_{\mathcal{M}} k_T(p_1, p_\xi g) h^{(0)}(p_\xi) dx_\xi$$

$$h^{(T)}(p_1) = U^{(T)}(m^{(T)}(p_1)) \tag{44}$$

The propagated equivariant map is projected back to numerical features using the canonically attached feature field and the gauge,

$$f_h^{A,(T)}(x) = \psi_E^A \circ \gamma(h^{(T)}(p)), p \in \pi^{-1}(x) \tag{45}$$

### 3.4. Propagation beyond Diffusion

Equivariant diffusion allows for an insightful interpretation regarding the minimization of the Polyakov action. It also gives theorems and techniques to understand geometrically the diffusion kernel. However, diffusion poses severe limitations. The update function should be linear to conserve the original heat equation, which is not always valid in message-passing neural networks. However, this corresponds to the case of MACE (Batatia et al., 2022b), which uses linear updates. We allow GM - Message passing in all generality to have a non-linear update function. The nonlinearity should be of a gate form to avoid breaking equivariance (Weiler et al., 2021). It is for future work to analyze the impact of nonlinearities update on the process.

### 3.5. Equivariant Message Passing over graphs

In the previous section, we introduced a method to update the coordinate-free features field by applying an equivariant diffusion process to the canonically attached equivariant map. In this section, we will apply this formalism to discretized manifolds. We consider now $\mathcal{G} = (\mathcal{V} = \{1, ..., n\}, \mathcal{E})$ a graph with $\mathcal{V}$ and $\mathcal{E}$ denoting nodes and edges respectively. Let $(\mathcal{M}, \eta_{\mathcal{M}})$ be a Riemannian manifold of dimension $\mathfrak{m}$. The induced topology on the graph by the metric $\eta_{\mathcal{M}}$ is defined by the set $\mathcal{E} = \{(i,j) \in \mathcal{V} | \eta_{\mathcal{M}}(r_i, r_j) \leq r_c\}$ of neighboring points. By fixing one point $i$, one can construct from the set $\mathcal{E}$, neighbors of $i$ defined as $\mathcal{N}(i) = \{j \in \mathcal{V} | \eta_{\mathcal{M}}(r_i, r_j) \leq r_c, j \neq i\}$. We find the discrete analogy to the GM-Message passing, equivalent to finding the optimal equivariant map $h : P \to V$ with information on a discretized manifold represented by the graph $\mathcal{G}$.

**Definition 3.17** (State function)**.** We define $(P, \pi, \mathcal{M}, G)$ the principal bundle of $\mathcal{M}$. Let $\sigma : \mathcal{V} \to \mathcal{M} \times V$ be the state function of the nodes (atoms) $\mathcal{V}$ such that $\sigma_i = (r_i, f_{h,i})$ with $r_i$ the positions and $f_{h,i}$ the canonically attached coordinate independent feature field attached to $h : P \to V$. We call $\sigma^A$ the state relative to a gauge $\psi^A$ on $\mathcal{M}$.

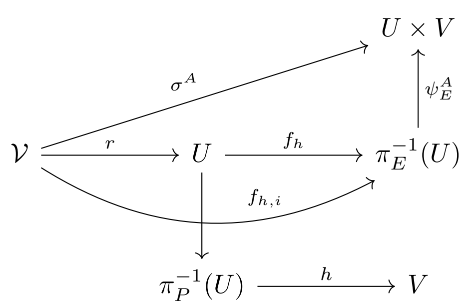


*Figure 3.* The state function labels the set of nodes $\mathcal{V}$ by a tuple of positional features in $\mathcal{M}$ and a the numerical feature field in $V$ a vector space.

**Definition 3.18** (Equivariant Message Passing)**.** Let $h^{(0)} : P \to V$ be an initial equivariant feature map. Let $\Delta^E$ be the generalized Laplacian on the associated bundle $E$ and $k_T$ the heat kernel of $\Delta^E$. We define the message operation of Equivariant - Message passing on $\mathcal{G}$ :

$$m^{(t)}(p_i) = e^{-tCas} \int_G \rho(g)^{-1} dg \prod_{\xi=1}^{n} \sum_{j \in \mathcal{N}(i)} k_t(p_i, p_j g) h^{(t-1)}(p_j) \tag{46}$$

The feature map $h^{(T)}$ will be the optimal equivariant map from $P$ to the irreducible representation $V$ of $G$ for $T$ sufficiently long over the graph $\mathcal{G}$. We identify this operation as the pooling operation of message passing,

$$\bigoplus_{j \in \mathcal{N}(i)} M_t(\sigma_j^{(t)}, \sigma_i^{(t)}) =$$
$$e^{-tCas} \int_G \rho(g)^{-1} dg \prod_{\xi=1}^{n} \sum_{j \in \mathcal{N}(i)} k_t(p_i, p_j g) h^{(t-1)}(p_j) \tag{47}$$

The kernel function can be learnable in all generality depending on the iteration $t$, such as a neural network. The update function $U_t$ can be a linear or non-linear neural network, shifting from classical diffusion to a partial differential equation evolution.

Let's assume that, $\mathcal{M} = \mathbb{R}^3$, $G = SE(3)$ and $V$ is the irreducible representation of invariant scalar to $SE(3)$, such that $\rho(g) = I, \forall g \in G$. If one expands the heat kernel into a spherical series, we recover the equation of the MACE architecture (Batatia et al., 2022b;a).

Let $k(p,q)$ be a translationally invariant kernel. Then $\forall t \in \mathbb{R}, k(p+t, q+t) = k(p,q) = k(0, q-p) = \tilde{k}(q-p)$. Thus we see that $\tilde{k}$ and $k$ contain as much information.

**Spherical expension of the kernel** First consider the heat kernel on the Sphere $S^2$. By Mercer's theorem, any kernel $k(p-q) \in L(S^2)$ can be expended in the form of spherical harmonics. By extending it to $\mathbb{R}^3$ any kernel on $k(p-q) \in L(R^3)$ can be expended into,

$$k(p-q) = \sum_{n=0}^{+\infty} \sum_{l=0}^{+\infty} \sum_{m=-l}^{m=l} R^{(n)}(\boldsymbol{p} - \boldsymbol{q}) c_{lm} Y_m^l(\hat{p} - \hat{q}) \tag{48}$$

where $Y_m^l$ are spherical harmonics of order $lm$ By truncating the expansion to a maximal $l$ value and injecting it into the previous message passing equation, one recovers the exact equation for the MACE messages if the time step is constant as the Casimir term will just become a constant re-scaling.

## 4. Discussion

In this work, we have introduced a new geometric interpretation of message passing based on differential geometry and diffusion that led to the formulation of a general class of networks on Riemannian manifolds. Implementing these models on test manifolds is still needed, and numerical experiments are required to validate the proposed approach. Fast Fourier transform on the group could simplify the computation of the integral over the group arising in the formulation of $GM$-message. The connection with equivariant message passing could also be fruitful for the interpretability of these methods and to help create new numerical schemes based on our proposed understanding. A more general discussion is needed on the connection between message passing and non-linear partial differential equations on manifolds. Extending this work beyond point clouds using non-commutative geometry would represent a challenging task but might be a fruitful endeavour.

## 5. Acknowledgements

The author would like to express heartfelt gratitude towards the reviewers, denoted as $uudT$ and $8mvL$. The invaluable feedback and constructive critiques provided have been instrumental in shaping and refining this work.

## References

Anderson, B., Hy, T. S., and Kondor, R. Cormorant: Covariant molecular neural networks. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/file/03573b32b2746e6e8ca98b9123f2249b-Paper.pdf.

Bartók, A. P., Kondor, R., and Csányi, G. On representing chemical environments. *Physical Review B*, 87(18): 184115, 2013.

Batard, T. and Sochen, N. A Class of generalized Laplacians on vector bundles devoted to multi-channels image processing. working paper or preprint, February 2012. URL https://hal.archives-ouvertes.fr/hal-00683953.

Batatia, I., Batzner, S., Kovács, D. P., Musaelian, A., Simm, G. N. C., Drautz, R., Ortner, C., Kozinsky, B., and Csányi, G. The design space of e(3)-equivariant atom-centered interatomic potentials, 2022a. URL https://arxiv.org/abs/2205.06643.

Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csanyi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In *Advances in Neural Information Processing Systems*, volume 35, 2022b.

Batatia, I., Geiger, M., Munoz, J., Smidt, T., Silberman, L., and Ortner, C. A general framework for equivariant neural networks on reductive lie groups, 2023.

Batzner, S., Musaelian, A., Sun, L., Geiger, M., Mailoa, J. P., Kornbluth, M., Molinari, N., Smidt, T. E., and Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13(1):2453, 2022.

Berline, N., Getzler, E., and Vergne, M. Heat kernels and dirac operators. 1992.

Bogatskiy, A., Hoffman, T., Miller, D. W., and Offermann, J. T. PELICAN: Permutation Equivariant and Lorentz Invariant or Covariant Aggregator Network for Particle Physics. 11 2022.

Bronstein, M. M., Bruna, J., Cohen, T., and Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges, 2021.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., and Zdeborová, L. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019. doi: 10.1103/RevModPhys.91.045002. URL https://link.aps.org/doi/10.1103/RevModPhys.91.045002.

Chamberlain, B. P., Rowbottom, J., Eynard, D., Giovanni, F. D., Dong, X., and Bronstein, M. M. Beltrami flow and neural diffusion on graphs. *CoRR*, abs/2110.09443, 2021. URL https://arxiv.org/abs/2110.09443.

Cohen, T. S. and Welling, M. Steerable cnns. 2016. doi: 10.48550/ARXIV.1612.08498. URL https://arxiv.org/abs/1612.08498.

Cohen, T. S., Weiler, M., Kicanaoglu, B., and Welling, M. Gauge equivariant convolutional networks and the icosahedral cnn, 2019.

Drautz, R. Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer. *Phys. Rev. B*, 102:024104, Jul 2020. doi: 10.1103/PhysRevB.102.024104. URL https://link.aps.org/doi/10.1103/PhysRevB.102.024104.

Esteves, C., Allen-Blanchette, C., Zhou, X., and Daniilidis, K. Polar transformer networks, 2017. URL https://arxiv.org/abs/1709.01889.

Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. Neural message passing for quantum chemistry. *CoRR*, abs/1704.01212, 2017. URL http://arxiv.org/abs/1704.01212.

Giovanni, F. D., Rowbottom, J., Chamberlain, B. P., Markovich, T., and Bronstein, M. M. Graph neural networks as gradient flows: understanding graph convolutions via energy, 2022.

Kobayashi, S. and Nomizu, K. *Foundations of differential geometry. I*, volume 15 of *Intersci. Tracts Pure Appl. Math.* Interscience Publishers, New York, NY, 1963.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1(4):541–551, 1989. doi: 10.1162/neco.1989.1.4.541.

Li, C., Qu, H., Qian, S., Meng, Q., Gong, S., Zhang, J., Liu, T.-Y., and Li, Q. Does Lorentz-symmetric design boost network performance in jet physics? 8 2022.

Munoz, J. M., Batatia, I., and Ortner, C. Bip: Boost invariant polynomials for efficient jet tagging, 2022. URL https://arxiv.org/abs/2207.08272.

Naber, G. L. *Topology, geometry, and gauge fields: Foundations*. Springer, New York, USA, 1997. doi: 10.1007/978-1-4419-7254-5.

Thiede, E. H., Zhou, W., and Kondor, R. Autobahn: Automorphism-based graph neural nets. In Beygelzimer, A., Dauphin, Y., Liang, P., and Vaughan, J. W. (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=NU69dglcsS.

Toshev, A. P., Galletti, G., Brandstetter, J., Adami, S., and Adams, N. A. E(3) equivariant graph neural networks for particle-based fluid mechanics, 2023.

Warner, F. *Foundations of differentiable manifolds and Lie Groups*. Graduate Texts in Mathematics, Vol. 94, Springer, 1983.

Weiler, M., Forré, P., Verlinde, E., and Welling, M. Coordinate independent convolutional networks – isometry and gauge equivariant convolutions on riemannian manifolds, 2021. URL https://arxiv.org/abs/2106.06020.