On Non-Linear operators for Geometric Deep Learning

Anonymous Author(s)
Affiliation
Address
email

Abstract

This work studies operators mapping vector and scalar fields defined over a manifold $\mathcal{M}$, and which commute with its group of diffeomorphisms Diff$(\mathcal{M})$. We prove that in the case of scalar fields $L^p_\omega(\mathcal{M}, \mathbb{R})$, those operators correspond to point-wise non-linearities, recovering and extending known results on $\mathbb{R}^d$. In the context of Neural Networks defined over $\mathcal{M}$, it indicates that point-wise non-linear operators are the only universal family that commutes with any group of symmetries, and justifies their systematic use in combination with dedicated linear operators commuting with specific symmetries. In the case of vector fields $L^p_\omega(\mathcal{M}, T\mathcal{M})$, we show that those operators are solely the scalar multiplication. It indicates that Diff$(\mathcal{M})$ is too rich and that there is no universal class of non-linear operators to motivate the design of Neural Networks over the symmetries of $\mathcal{M}$.

1 Introduction

Given a physical domain $\mathcal{M}$ and measurements $f : \mathcal{M} \to \mathcal{Y}$ observed over it, one is often interested in processing *intrinsic* information from f, i.e. consistent with the *symmetries* of the domain. In words, if two measurements f, $\tilde{f} = g.f$ are related by a symmetry g of the domain, like a rigid motion on an observed molecular compound, we would like our processed data $M(f)$ and $M(\tilde{f})$ to be related by the same symmetry — thus that $M(g.f) = g.M(f)$ or equivalently that M commutes with the symmetry transformation of the domain. The study of operators that satisfy such symmetry constraints has played a long and central role in the history of physics and mathematics, motivated by the inherent symmetries of physical laws. More recently, such importance has also extended to the design of machine learning systems, where symmetries improve the sample complexity [22, 3]. For instance, Convolutional Neural Networks build translation symmetry, whereas Graph Neural Networks build permutation symmetry, amongst other examples coined under the 'Geometric Deep Learning' umbrella [5, 4].

Lie groups of transformations are of particular interest, because there exists a precise and systematic framework to build such intrinsic operators. Indeed, for a locally compact group G, it is possible to define a Haar measure which is invariant to the action of G [2]; then a simple filtering along the orbit of G allows to define a class of *linear* operators that commute with the group action. Examples of locally compact groups are given by specific Lie groups acting on $\mathbb{R}^d$, such as the translations or the rotations $O_d(\mathbb{R})$. Often these Lie groups G only act on a manifold $\mathcal{M}$, and one tries to average along the orbit induced by G. Note that it is possible, beyond invariance, to linearize more complex groups of variability like diffeomorphisms Diff$(\mathcal{M})$ [7].

While the description of such linear intrinsic structures is of central mathematical importance and forms the basis of Representation theory [27], in itself is not sufficient to bear fruit in the context of Representation *learning* using Neural Networks [11]. Indeed, linear operators do not have the

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

capacity to extract rich information needed to solve challenging high-dimensional learning problems. It is therefore necessary to extend the systematic construction and classification of intrinsic operators to the non-linear case.

With that purpose in mind, our work aims at studying the class of (*non-linear*) operators M which commute with the action of the group Diff($\mathcal{M}$), the diffeomorphisms over $\mathcal{M}$. This approach will lead to a natural class of non-linear intrinsic operators. Indeed, any group G of symmetries is, by definition, a subgroup of Diff($\mathcal{M}$), and thus commutes with such M [21]. Consequently, obtaining a non-linear invariant to a symmetry group G could be done by using a cascade of interlacing non-linear operators which commute with Diff($\mathcal{M}$) and linear operators which commute with G.
A notable example of linear operators that are covariant to the Lie group of translations is a given by the convolutions along the orbit of the group. These can be constructed thanks to the canonical Haar measure [28]. However, such an approach fails for infinite dimensional groups, like our object of interest: contrary to Lie groups, Diff($\mathcal{M}$) is not locally compact and it is thus not possible to define a Haar measure on this group.

Our first contribution is to demonstrate that the *non-linear* operators which act on vector fields (elements of $L^p_\omega(\mathcal{M}, T\mathcal{M})$) and which commute with the group of diffeomorphisms, are actually just scalar multiplications. This implies that Diff($\mathcal{M}$) is too rich to obtain non-trivial operators. Our second contribution is to demonstrate that *non-linear* operators acting on signals in $L^p_\omega(\mathcal{M}, \mathbb{R})$ are pointwise non-linearities. This fills a gap in the results of [7], and *a fortiori* justifies the use of point-wise non-linearities in geometric Deep Learning [4].

Our paper is structured as follows: Sec. 2 introduces the necessary formalism, that we use through this paper: in particular, we formally define the action of diffeomorphism. Then, we state and discuss our theorems in Sec. 3.1 and sketch their proofs in Sec. 3.2. Rigorous proofs of each statement can be found in the Appendix.

2 Problem Setup

2.1 Related work and motivation

In this section, we discuss the notion of intrinsic operators, invariant and covariant non-linear operators and linear representation over standard symetry groups. Then, we formally state our objective.

Intrinsic Operators As discussed above, in this work we are interested in *intrinsic* operators $M : L^p(\mathcal{M}, E) \to L^p(\mathcal{M}, E)$, where $\mathcal{M}$ is a Riemannian manifold, and $E = \mathbb{R}$ or $E = T\mathcal{M}$, capturing respectively the setting of scalar signals and vector fields over $\mathcal{M}$. Here the notion of 'intrinsic' means that M is consistent with an equivalence class induced by a symmetry group G in $L^p(\mathcal{M}, E)$: if $f, \tilde{f} \in L^p(\mathcal{M}, E)$ are related by a transformation $g \in G$ (in which case we write $f = g.\tilde{f}$), then $M(f) = g.M(\tilde{f})$. Naturally, a stronger equivalence class imposes a stronger requirement towards M, and consequently restrains the complexity of M. We now describe the plausible techniques used to design such operators M.

GM-Convolutions The notion of GM-convolutions [30] is an example of linear covariant operators which commute with the reparametrization of a manifold. In practice, this implies that the weights of a GM-convolution are shared and the action of GM-convolutions is local – two properties that facilitate implementation and point out the similarity with Lie groups. Another example of symmetry group corresponds to the isometry group of a Riemaniann manifold, whose pushforward preserves the tensor metric. In this case, it is well known that isometries [29] are the only diffeomomorphism which commute with a manifold Laplacian. Thus, any *linear* operators which commute with isometries is stabilized by Laplacian's eigenspaces. However, little is known on the *non-linear* counterpart of the symmetry-covariant operators. In this work, we characterize *non-linear* operators which commute with Diff($\mathcal{M}$). We will see that such operators are intrinsically defined by Diff($\mathcal{M}$) and could be combined with any linear operators covariant with a symmetry group G.

Non-linear operators It has been shown that Convolutional Neural Networks are dense in the set of *non-linear* covariant operators [31]. The recipe of the corresponding proof is an extension of the proof of the universal approximation theorem [13]. The Scattering Transform [6, 20] is also an example of a well-understood non-linear operator which corresponds to a cascade of complex wavelet transforms followed by a point-wise modulus non-linearity. This representation provably linearizes small deformations.

Compact Lie Groups In the context of geometric Machine Learning [5], there are several relevant notions of equivalence. For instance, we can consider a compact Lie Group G acting on $\mathcal{M}$, and an associated representation in $\mathcal{F} = \{f : \mathcal{M} \to \mathbb{R}\}$: Given $g \in G$ and $f \in \mathcal{F}$, then $g.f(x) \triangleq f(g^{-1}.x)$ for $x \in \mathcal{M}$. We then consider $f \sim \tilde{f}$, related by this group action: $\tilde{f} = g.f$ for some $g \in G$. The operators M which are compatible with such group action are referred as being G-equivariant (or covariant to the action of G) in the ML literature [12, 4]. Such groups are typically of finite and small dimension, e.g. the Euclidean transformations of $\mathcal{M} = \mathbb{R}^d$, with $d = 2$ for computer vision applications, or $d = 3$ for computational biology/chemistry applications. In this case, it is possible to characterize all *linear* intrinsic operators M as group convolutions [18], leading to a rich family of non-linear intrinsic operators by composing such group convolutions with element-wise non-linear operators, as implemented in modern Neural Networks. We highlight that stability to symetries via non-linear operators finds useful application, in particular for flat manifolds [7].

Isometries Riemanian manifolds $\mathcal{M}$ come with a default equivalence class, which is given by isometries. If $m_u : T_u\mathcal{M} \times T_u\mathcal{M} \to \mathbb{R}$ denotes the Riemannian metric tensor at point $u \in \mathcal{M}$, a diffeomorphism $\psi : \mathcal{M} \to \mathcal{M}$ is an isometry if $g_u(v, w) = g_{\psi(u)}(d\psi_u(v), d\psi_u(w))$ for any $u \in \mathcal{M}$ and $v, w \in T_u\mathcal{M}$. In words, isometries are changes of variables that preserve the local distances in the domain. The ensemble of all isometries forms a Lie Group which is locally compact [24]. In this case, one can also build a rich class of intrinsic operators by following the previously explained 'blueprint', namely composing linear intrinsic operators with element-wise non-linearities. As a representative example, the Laplace-Beltrami operator of $\mathcal{M}$ only depends on intrinsic metric properties [29]: as said above, isometries preserve the invariant subspaces of a Laplacian.

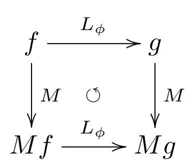
Beyond Isometries While isometries are the 'natural' transformations of the geometric domain, they cannot express high-dimensional sources of variability; indeed, if $\mathcal{M}$ is a d-dimensional complete connected Riemannian manifold, its isometry group has dimension at most $d(d + 1)/2$ [9]. This raises the question whether one can characterize intrinsic operators relative to a broader class of transformations. Another class of important symmetries corresponds to the ones which are gauge invariant, i.e. which leads to transformations which preserve the change of parametrization and which are used in [10, 30] through the notion of G-structure.

In this work, we consider the class of transformations given by Diff($\mathcal{M}$), the diffeomorphisms over $\mathcal{M}$. As shown in the Appendix, compactly supported deformations $\psi : \mathcal{M} \to \mathcal{M}$ define bounded linear operators L_ψ acting on $L^p(\mathcal{M}, E) \to L^p(\mathcal{M}, E)$, and constitute a far broader class of transformations than isometries. Our proof is mainly based on the use of compactly supported diffeomorphisms.

Our objective is to characterize the (non-linear) operators M such that

$$\forall \phi \in \text{Diff}(\mathcal{M}), L_\phi M = M L_\phi \,.$$

In other words, we aim to understand continuous operators M that commute with deformations. We will show that such operators are act locally and that they can be descriped explicitly, with simple formula. The commutation condition is visualized in the following diagram:



2.2 Notations

We will now formally introduce the mathematical objects of interest in this document. Let $(\mathcal{M}, g)$ be an orientable, connected, Riemannian manifold, of finite dimension $d \in \mathbb{N}^*$, with $g \in \Gamma(T^*M \otimes T^*M)$ a section of symmetric definite positive bilinear forms on the tangent bundle of M. Fix $p \in [1, +\infty[$. For any volume form $\omega \in \Gamma(\bigwedge^d T^*M)$ let us define $L^p_\omega(\mathcal{M}, T\mathcal{M})$, the space of L^p

vector fields, defined as the subspace of measurable functions $f : \mathcal{M} \to T\mathcal{M}$ such that $f(x) \in T_x M$ almost surely and

$$\|f\|_p^p \triangleq \int_{x\in\mathcal{M}} g_x(f(x), f(x))^{\frac{p}{2}}\, d\omega(x) < +\infty\,. \tag{1}$$

We will also consider $L^p_\omega(\mathcal{M}, \mathbb{R})$ the space of measurable scalar functions $f : \mathcal{M} \to \mathbb{R}$ that fulfill

$$\|f\|_p^p \triangleq \int_{x\in\mathcal{M}} |f(x)|^p\, d\omega(x) < +\infty\,. \tag{2}$$

We may write $\|\cdot\|$ instead of $\|\cdot\|_p$ when there is no ambiguity. For a C^∞ diffeomorphism $\phi \in \mathrm{Diff}(\mathcal{M})$, we will consider the action of $L_\phi : L^p_\omega(\mathcal{M}, T\mathcal{M}) \to L^p_\omega(\mathcal{M}, T\mathcal{M})$ which we define for for any $f \in L^p_\omega(\mathcal{M}, \mathbb{R})$ as

$$L_\phi f(u) \triangleq d\phi(u)^{-1}.f(\phi(u))\,.$$

Note that this action is contravariant:

$$L_{\psi\circ\phi} f(u) = d(\psi\circ\phi)^{-1}.f(\psi\circ\phi(u)) = L_\phi L_\psi f(u)$$

For scalar function $f \in L^p_\omega(\mathcal{M}, \mathbb{R})$, we define the action of ϕ via

$$L_\phi f(u) \triangleq f(\phi(u))\,.$$

This latter operator is also contravariant. If there is no ambiguity, we will use the same notation L_ϕ, whether we apply it to $L^p_\omega(\mathcal{M}, \mathbb{R})$ or $L^p_\omega(\mathcal{M}, T\mathcal{M})$. Throughout the article we restrict ourselves to ϕ such that L_ϕ is a bounded operator. Write $\mathrm{supp}(\phi) = \{x, \phi(x) \neq x\}$ for the support of ϕ and say that ϕ has a compact support if $\mathrm{supp}(\phi)$ is compact. We denote by $\mathrm{Diff}_c(\mathcal{M}) \subset \mathrm{Diff}(\mathcal{M})$ the set of compactly supported diffeomorphisms. Recall that since a $\mathcal{M}$ is second-countable, $\mathcal{C}^\infty_c(\mathcal{M})$ is dense in $L^p_\omega(\mathcal{M}, \mathbb{R})$ and $\mathcal{C}^\infty_c(\mathcal{M}, T\mathcal{M})$ is dense in $L^p_\omega(\mathcal{M}, T\mathcal{M})$. Finally, denote by $O_d(\mathbb{R})$ the set of unitary operators on $\mathbb{R}^d$. Throughout the article, we might not write explicitly that equalities hold almost surely, since this is the default in L^p spaces.
As mentioned earlier, compactly supported diffeomorphisms lead to continuous operators, which is made rigorous by the following lemma whose proof is in the appendix.

Lemma 1. *If $\mathrm{supp}(\phi)$ is compact, then L_ϕ is bounded.*

3 Main theorems

In this section we present our main results. We first show that any (non-linear) deformation-equivariant operator acting on scalar fields must be point-wise (Theorem 1), and then establish that any deformation-equivariant operator acting on vector fields corresponds to a multiplication by a scalar (Theorem 2).

3.1 Theorem statements

Now, we are ready to state our two main theorems:

Theorem 1 (Scalar case)**.** *Let $\mathcal{M}$ be a connected and orientable manifold of dimension $d \geq 1$. We consider a Lipschitz continuous operator $M : L^p_\omega(\mathcal{M}, \mathbb{R}) \to L^p_\omega(\mathcal{M}, \mathbb{R})$, where $1 \leq p < \infty$. Then,*

$$\forall\, \phi \in \mathit{Diff}(\mathcal{M}) :\ ML_\phi = L_\phi M$$

is equivalent to the existence of a Lipschitz continuous function $\rho : \mathbb{R} \to \mathbb{R}$ that fulfills

$$M[f](m) = \rho(f(m)) \quad \text{for almost all } m \in \mathcal{M}.$$

In that case, we have $\rho(0) = 0$ if $\omega(\mathcal{M}) = \infty$.

Theorem 2 (Vector case)**.** *Let $\mathcal{M}$ be a connected and orientable manifold of dimension $d \geq 1$. We consider a continuous operator $M : L^p_\omega(\mathcal{M}, T\mathcal{M}) \to L^p_\omega(\mathcal{M}, T\mathcal{M})$, where $1 \leq p < \infty$. Then,*

$$\forall\, \phi \in \mathit{Diff}(\mathcal{M}) :\ ML_\phi = L_\phi M$$

is equivalent to the existence of a scalar $\lambda \in \mathbb{R}$ such that

$$\forall f \in L^p_\omega(\mathcal{M}, T\mathcal{M}) :\ M[f](m) = \lambda f(m) \quad \text{for almost all } m \in \mathcal{M}.$$

We highlight that our theorems are quite generic in the sense that they apply to the manifolds usually used in applications or theory, $\mathbb{R}^d$ in particular.

Remark 1. *The scalar case allows to recover standard operators which are exploited for Deep Neural Networks architectures. However, Theorem 2 indicates that the group of diffeomorphism is too rich to obtain non-trivial non-linear operators.*

Remark 2. *The case $p = \infty$ leads to different results. For instance, in the scalar case we may consider the operator $Mf(x) = \sup_y |f(y)|$ which fulfills $L_\phi M f = M L_\phi f$ but is not pointwise.*

Remark 3. *The condition "$\omega(\mathcal{M}) = \infty \implies \rho(0) = 0$" in Theorem 1 is necessary, since in the case $\mathcal{M} = \mathbb{R}$, the operator $Mf(x) \triangleq e^{if(x)}$ is not in $L^p_\omega(\mathcal{M}, \mathbb{R})$.*

Remark 4. *The Lipschitz condition in Theorem 1 is crucial, otherwise, $Mf(x) = \rho(f(x))$ might not be an operator of $L^p_\omega(\mathcal{M}, \mathbb{R})$. For instance, if $p = 2$, $\mathcal{M} = [0, 1]$ and $Mf(x) = \sqrt{f(x)}$, we see that in this case, let $f(x) = x$, then $f \in L^p_\omega(\mathcal{M}, \mathbb{R})$ and $Mf \notin L^p_\omega(\mathcal{M}, \mathbb{R})$*

Remark 5. *If M is not Lipschitz, we can find an example which is not even continuous. The following example holds in both cases, the scalar case and the vector case. In both cases $f \in L^p(M, \mathbb{R})$, the only thing that changes is the action of L_ϕ on f. $\mathcal{M} = \mathbb{R}$, let for all $f \in L^p(M, \mathbb{R})$:*

$$Mf(x) = 1_{\{z, \lim_{y\to z} f(y) = f(z)\}}(x) f(x).$$

It is a measurable function. Let us show that this M is a counterexample to the vector case: for any $\phi \in$ Diff$(\mathcal{M})$ and $x \in \mathbb{R}$, one has

$$ML_\phi f(x) = 1_{\{z, \lim_{y\to z} f(\phi(y)) = f(\phi(z))\}}(x) \quad d\phi(x)^{-1} f(\phi(x)) \tag{3}$$

$$= 1_{\{z, \lim_{y\to \phi(z)} f(y) = f(\phi(z))\}}(x) \quad d\phi(x)^{-1} f(\phi(x)) \tag{4}$$

$$= 1_{\{z, \lim_{y\to z} f(y) = f(z)\}}(\phi(x)) \quad d\phi(x)^{-1} f(\phi(x)) \tag{5}$$

$$= L_\phi M f(x)\,. \tag{6}$$

However, M is not continuous as changing any function to 0 on $\mathbb{Q}$ does not change its norm but changes the set where the limits exists. More precisely let $c > 0$ be a strictly positive scalar, $M[c] = c$; let $f = c1[x \notin \mathbb{Q}]$, $M[f] = 0$ as $\{z, \exists \lim_{y\to z} f(\phi(y))\} = \emptyset$. However $c = f$ almost surely but $M[c] \neq M[f]$ therefore M is not continuous.

3.2 Proof Sketch

We now describe the main ideas for proving the Theorems 1 and 2. The appendix contains complete formal arguments and technical lemmata which we omit here due to lack of space. The two proofs share quite some similarities despite substantially different final results. Three ideas guide our proofs: First, we prove that it is possible to localize M on a certain class of open sets which behaves nicely with the manifold structure, the strongly convex sets which we denote as $\mathcal{O}_1$. This is closely related to the notion of pre-sheaves [14]. Secondly, we characterize M on small open-sets. In the scalar case, we will study the representation of locally constant functions. In the vector case, we will show that locally, the image $M(1_U c)$ of a vector field c is co-linear to c provided that U is small enough. We will also show that those local properties are independent of the position on the manifold $\mathcal{M}$ via a connectedness argument. Thirdly and finally, we combine a compacity and a density argument to extend this characterization to $\mathcal{M}$, which is developed in Sec. 3.3. Throughout the presentation, we will use the following definitions and theorems obtained from other works:

Definition 1 (Strong convexity, from [17])**.** *Let $\mathcal{O}_1$ be the collection of open sets which are bounded and strongly convex, meaning that any points p, q in such a set can be joined by a geodesic contained in the set. Furthermore let $\dot{\mathcal{O}}_1 = \{A \in \mathcal{O}_1 : \exists B \in \mathcal{O}_1, \bar{A} \subset B$ and $\omega(\bar{A}\backslash A) = 0\}$.*

The intuition behind the definition of $\dot{\mathcal{O}}_1$ is that all of its elements are contained in a 'security' open set,which avoids degenerated effects on the manifold. In particular, this allows to control the boundary of a given open set.

Theorem 3 (theorem adapted from [16, 17])**.** ***(1)** $\dot{\mathcal{O}}_1$ is a system of neighborhoods. **(2)** Any element of $\mathcal{O}_1$ is diffeomorph to $\mathbb{R}^d$. **(3)** Both $\mathcal{O}_1$ and $\dot{\mathcal{O}}_1$ are stable by intersection.*

Theorem 4 (Flowbox theorem, as stated in [8])**.** *Let $f, g \in \mathcal{C}^\infty_c(\mathcal{M}, T\mathcal{M})$. For any $m \in \mathcal{M}$ with $f(m) \neq 0$ and $g(m) \neq 0$, there exists an open set $U \subset \mathcal{M}$ and $\phi \in$ Diff$(\mathcal{M})$ such that $\phi(m) = m$ and $L_\phi(1_U f) = 1_{\phi(U)} g$.*

We will now present some lemmata that are necessary the proofs of theorems 1 and 2. As a first step, we argue that one may assume $M(0) = 0$ where 0 denotes the constant 0-function. This is because in the appendix we show that $M(0)$ is a constant function C, with $C = 0$ if $\omega(\mathcal{M}) = \infty$. Therefore, we may substract C from ρ and λ, leaving us with having to show the theorems only for $M(0) = 0$. Next, a key idea of the proof is to exploit the flexibility of the deformation equivariance to *localise* the input, i.e. to show that the image of compactly supported functions is also compactly supported. To do so, the following lemma provides a way of collapsing an open ball into a singleton while maintaining a good control on the support of the diffeomorphism.

Lemma 2 (Key lemma). *Let $\epsilon > 0$. There exists a sequence of diffeomorphisms $\phi_n : \mathbb{R}^d \to \mathbb{R}^d$, compactly supported in $\mathcal{B}(0, 1 + \epsilon)$ such that:*

$$\phi_n(\mathcal{B}(0,1)) = \mathcal{B}(0, \frac{1}{n}),$$

and

$$\sup_{u \in \mathcal{B}(0,1)} \|d\phi_n(u)\| \leq \frac{1}{n}.$$

Proof. Set $\phi_n(u) = f_n(\|u\|)u$, where

$$f_n(r) = \begin{cases} \frac{1}{n} & \text{, if } |r| \leq 1 \\ 1 & \text{, if } |r| \geq 1 + \epsilon, \end{cases}$$

and f_n is smoothly interpolated for $|r| \in [1, 1 + \epsilon]$ in a way that it remains nondecreasing. It is then clear that ϕ_n fulfills the desired properties. □

We will often use that if the support of $\phi \in \text{Diff}(\mathcal{M})$ is such that $\text{supp}(\phi) \cap U = \emptyset$, then for any $f \in L^p_\omega(\mathcal{M}, \mathbb{R})$ one has $1_U f = L_\phi(1_U f)$. This implies the following important lemma, for which a rigorous proof can be found in the appendix:

Lemma 3. *Let $U \in \dot{\mathcal{O}}_1$ and M as in Theorem 1 or Theorem 2. Then, for any $f \in E$, where $E = L^p_\omega(\mathcal{M}, \mathbb{R})$ or $E = L^p_\omega(\mathcal{M}, T\mathcal{M})$ respectively, we have:*

$$M[f 1_U] = 1_U M[f].$$

Furthermore, if U is any closed set, the same conclusion applies.

Equipped with this result, our proof will characterize the image of functions of the type $1_U c$ where either $c \in \mathbb{R}$, or c is a vector field which can be straightened, via the following Lemma. In the Vector case:

Lemma 4 (Image of localized vector field). *For M as in Theorem 2 there is $U \in \dot{\mathcal{O}}_1$, such that for any $f \in L^p_\omega(M, TM)$ there is $\lambda(U)$ such that:*

$$M[f 1_U] = 1_U \lambda(U) f. \tag{7}$$

Here is the scalar case:

Lemma 5 (Image of constant functions, scalar case). *Let M as in Theorem 1. For any $U \in \dot{\mathcal{O}}_1$ and $c \in \mathbb{R}$, then: $M(c 1_U) = h(c, U) 1_U$. Furthermore, $c \to h(c, U)$ is Lipschitz for any $U \in \dot{\mathcal{O}}_1$.*

At this stage, we note that both representations are point-wise, and the next steps of the proofs will be identical both for the scalar and vector cases. The extension to $L^p_\omega(\mathcal{M}, \mathbb{R})$ or $L^p_\omega(\mathcal{M}, T\mathcal{M})$ will be done thanks to:

Lemma 6 (Image of a disjoint union of opensets). *Let $U_1, ..., U_n \in \mathcal{O}_1$ and M as in Theorem 2 or Theorem 1, s.t. $\forall i \neq j, \overline{U_i} \cap \overline{U_j} = \emptyset$. Then for any $f \in L^p_\omega(\mathcal{M}, T\mathcal{M})$:*

$$M[\sum_{i=1}^n 1_{U_i} f] = \sum_{i=1}^n M[1_{U_i} f].$$

This lemma states that we can completely characterize M on disjoint union of simple sets. We will then need a Vitali-covering-type argument in order to "glue" those open sets together, which shows that simple functions with disjoint support can approximate any elements of $L^p_\omega(\mathcal{M}, \mathbb{R})$ or $L^p_\omega(\mathcal{M}, T\mathcal{M})$ (we only state the lemma for $L^p_\omega(\mathcal{M}, \mathbb{R})$ as our proof on $L^p_\omega(\mathcal{M}, T\mathcal{M})$ does not necessarily need this result):

Lemma 7 (Local Vitali). *For $f \in \mathcal{C}_c^\infty(\mathcal{M})$ and $m \in \mathcal{M}$, there exists $U \in \dot{\mathcal{O}}_1$ with $m \in U$, such that for any $\epsilon > 0$, there exist subsets $U_1, ..., U_n \in \dot{\mathcal{O}}_1$ with $U_i \subset U$ and numbers $c_1, ..., c_n \in \mathbb{R}$ fulfilling*

$$\|\sum_n 1_{U_n} c_n - 1_U f\| < \epsilon\,.$$

Note that this type of covering is not possible on any open set without further assumptions on the manifold, such as bounds on its Ricci curvature [19]. Fortunately, we will only need a local version which is true because charts are locally bi-Lipschitz. Both Lemma 6 and Lemma 7 imply that:

Proposition 1. *Consider M from either Theorem 1 or 2. Assume that there exists $U \in \dot{\mathcal{O}}_1$ such that $M(c1_V) = h(c, V)1_V$ for any $V \subset U$, with $V \in \dot{\mathcal{O}}_1$, where c is either a vector field in the case $E = L^p_\omega(\mathcal{M}, T\mathcal{M})$ or a constant scalar in the case $E = L^p_\omega(\mathcal{M}, \mathbb{R})$. If we further assume that $c \to h(c, U)$ is L-Lipschitz, then*

$$\forall f \in E, \forall m \in \mathcal{M}, M[1_U f](m) = 1_U h(f(m), U)\,.$$

Furthermore, it does not depend on U, meaning that for any other such $\tilde{U}$, we have:

$$\forall f \in E, \forall m \in U \cap \tilde{U}, M[1_{\tilde{U}} f](m) = 1_U h(f(m), U)\,.$$

We briefly discuss the intuition behind Theorem 2. It is linked to the idea that the operators M at hand have to commute with local rotations, and this even for locally constant vector fields. We reduce the characterisation of deformation-equivariant vector operators using an invariance to symmetry argument: functions which are invariant to rotations are multiples of a scalar. The reason is contained in the following lemma, which is commonly used in physics:

Lemma 8 (Invariance to rotation). *Let $f : \mathbb{R}^d \to \mathbb{R}^d$ such that for any $W \in O_d(\mathbb{R})$ and $x \in \mathbb{R}^d$, one has $f(Wx) = Wf(x)$. Then, there is $\lambda : \mathbb{R}^d \to \mathbb{R}, f(x) = \lambda(\|x\|)x$.*

Proof. We write $f(x) = \lambda(x)x + x^\perp$, with $x^\perp(m) \neq 0$ and $x^\perp \perp x$. Then, we introduce $W \in O_d(\mathbb{R})$ such that $Wx^\perp(m) = -x^\perp(m)$ and $Wx(m) = x(m)$. From $f(x) = f(Wx) = Wf(x)$ we deduce that $x^\perp = 0$. Next, $\lambda(Wx) = \lambda(x)$ thus $\lambda(x) = \lambda(x')$ for any $\|x\| = \|x'\|$. □

Distinction between scalar and vector case The scalar case is simpler to handle than the vector case: there several more steps for the proof of Theorem 2. We also highlight that the non-linearity is fully defined by its image on locally constant functions.

Finally, we conclude the proof of the theorem by appealing to a common density argument of the functions smooth with compact support, combing all the lemmata we have just presented in Sec. 3.3.

3.3 Proofs conclusions (common to the scalar and vector case)

In this section, we prove that the local properties of M can be extended globally on $\mathcal{M}$. The main idea is to exploit the well-known Poincaré's formula, which states that:

$$1_{\cup_i U_i} = \sum_{k=1}^n (-1)^k \sum_{i_1 < ... < i_k} 1_{U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k}}\,,$$

and to localize the action of M on each $U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k} \in \dot{\mathcal{O}}_1$ thanks to Lemma 3.

Proof of Theorem 1 and Theorem 2. Let f be a smooth and compactly supported function. Further consider $\cup_{i \leq n} U_i$ a finite covering of its support with $U_i \in \dot{\mathcal{O}}_1$. Using an inclusion-exclusion formula together with Lemma 3, we obtain

$$\begin{aligned}1_{\cup_i U_i} M[f] &= \sum_{k=1}^n (-1)^k \sum_{i_1 < ... < i_k} 1_{U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k}} M[f] \\ &= \sum_{k=1}^n (-1)^k \sum_{i_1 < ... < i_k} M[f 1_{U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k}}]\,,\end{aligned}$$

where we used that $U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k} \in \dot{\mathcal{O}}_1$. Now, the support of f is closed and included in $\cup_i U_i$. Thus using Lemma 3:

$$M[f] = \sum_{k=1}^{n} (-1)^k \sum_{i_1 < ... < i_k} M[f 1_{U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k}}],$$

Note that if ρ is a pointwise operator with $\rho(0) = 0$, then $\rho(1_U f) = 1_U \rho(f)$ and

$$\sum_{k=1}^{n} (-1)^k \sum_{i_1 < ... < i_k} \rho(f 1_{U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k}}) = \sum_{k=1}^{n} (-1)^k \sum_{i_1 < ... < i_k} 1_{U_{i_1} \cap U_{i_2} \cap ... \cap U_{i_k}} \rho(f) \tag{8}$$

$$= 1_{\cup_i U_i} \rho(f) = \rho(f)\,. \tag{9}$$

Thus, $Mf = \rho(f)$ where ρ is obtained from Lemma 4 or 5 combined with Prop 1. We conclude by density in $L^p_\omega(\mathcal{M}, \mathbb{R})$ or $L^p_\omega(\mathcal{M}, T\mathcal{M})$ respectively. This finishes the proof. □

4 Remarks and conclusion

In this work, we have fully characterized non-linear operators which commute under the action of smooth deformations. In some sense, it settles the intuitive fact that commutation with the whole diffeomorphism group is too strong a property, leading to a small, nearly trivial family of *non-linear* intrinsic operators. While on their own they have limited interest for geometric deep representation learning, they can 'upgrade' any family of linear operators associated with any group $G \subset \text{Diff}(\mathcal{M})$ into a powerful non-linear class — the so-called GDL Blueprint in [4]. Also, this result is a first step towards characterizing the non-linear operators which commute with Gauge transformations and could give useful insights for specifying novel Gauge invariant architectures. We now state a couple of unsolved questions and future work.

On the commutativity assumption: For $\mathcal{M} = \mathbb{R}^d$, it is unclear which type of non-linear operators commute with smaller groups of symmetry such as the Euclidean group. In fact, a generic question holds for manifolds: for a given symmetry group G, what is elementary non-linear building block of a Neural Network? This could be, for instance, useful to design Neural Networks which are Gauge invariant. It is an open question for future work which would be relevant many applications in physics [15].

Example of vector operators for L^∞ It is slightly unclear how the vector case $p = \infty$ can be handled in our framework, yet [1] seems to have interesting insights toward this direction.

Linearization of Diff$(\mathcal{M})$ In this work, we considered an exact commutation between operators and a symmetries: however, it is unclear which operators approximatively commute with a given symmetry group. Such operators would be better to linearize a high-dimensional symmetry group like Diff$(\mathcal{M})$. An important instance of non-linear operators that are non-local and that 'nearly' commute with diffeomorphisms is the Wavelet Scattering representation [20, 7, 25].

References

[1] John C Baez. Diffeomorphism invariant generalized measures on the space of connections modulo gauge transformations. In *Proceedings of the Quantum Topology Conference*. World Scientific, Singapore, 1994.

[2] Christoph Bandt. Metric invariance of haar measure. *Proceedings of the American Mathematical Society*, pages 65–69, 1983.

[3] Alberto Bietti, Luca Venturi, and Joan Bruna. On the sample complexity of learning under geometric stability. *Advances in Neural Information Processing Systems*, 34, 2021.

[4] Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

[5] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[6] Joan Bruna. *Scattering representations for recognition*. PhD thesis, Ecole Polytechnique X, 2013.

[7] Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1872–1886, 2013.

[8] Craig Calcaterra and Axel Boldt. Lipschitz flow-box theorem. *Journal of mathematical analysis and applications*, 338(2):1108–1115, 2008.

[9] Zhi Chen, Yiqian Shi, and Bin Xu. The riemannian manifolds with boundary and large symmetry. *Chinese Annals of Mathematics, Series B*, 31(3):347–360, 2010.

[10] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral cnn. In *International conference on Machine learning*, pages 1321–1330. PMLR, 2019.

[11] Taco Cohen and Max Welling. Learning the irreducible representations of commutative lie groups. In *International Conference on Machine Learning*, pages 1755–1763. PMLR, 2014.

[12] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.

[13] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.

[14] Pedro Boavida de Brito and Michael S Weiss. Manifold calculus and homotopy sheaves. *arXiv preprint arXiv:1202.1305*, 2012.

[15] Michael Eickenberg, Erwan Allys, Azadeh Moradinezhad Dizgah, Pablo Lemos, Elena Massara, Muntazir Abidi, ChangHoon Hahn, Sultan Hassan, Bruno Regaldo-Saint Blancard, Shirley Ho, et al. Wavelet moments for cosmological parameter estimation. *arXiv preprint arXiv:2204.07646*, 2022.

[16] Stéphane Gonnord and Nicolas Tosel. *Calcul différentiel: thèmes d'analyse pour l'agrégation*. Ellipses, 1998.

[17] Sigmundur Gudmundsson. An introduction to riemannian geometry. *Lecture Notes version*, pages 1–235, 2004.

[18] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International Conference on Machine Learning*, pages 2747–2755. PMLR, 2018.

[19] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.

[20] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

[21] Stéphane Mallat. Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150203, 2016.

[22] Song Mei, Theodor Misiakiewicz, and Andrea Montanari. Learning with invariances in random features and kernel models. In *Conference on Learning Theory*, pages 3351–3418. PMLR, 2021.

[23] Peter W Michor and Cornelia Vizman. n-transitivity of certain diffeomorphism groups. *arXiv preprint dg-ga/9406005*, 1994.

[24] Sumner B Myers and Norman Earl Steenrod. The group of isometries of a riemannian manifold. *Annals of Mathematics*, pages 400–416, 1939.

[25] Edouard Oyallon. *Analyzing and introducing structures in deep convolutional neural networks*. PhD thesis, Paris Sciences et Lettres, 2017.

[26] Richard S Palais. Extending diffeomorphisms. *Proceedings of the American Mathematical Society*, 11(2):274–277, 1960.

[27] Jialun Ping, Fan Wang, and Jin-Quan Chen. *Group representation theory for physicists*. World Scientific Publishing Company, 2002.

[28] Mitsuo Sugiura. *Unitary representations and harmonic analysis: an introduction*. Elsevier, 1990.

[29] Bill Watson. Manifold maps commuting with the laplacian. *Journal of Differential Geometry*, 8(1):85–94, 1973.

[30] Maurice Weiler, Patrick Forré, Erik Verlinde, and Max Welling. Coordinate independent convolutional networks–isometry and gauge equivariant convolutions on riemannian manifolds. *arXiv preprint arXiv:2106.06020*, 2021.

[31] Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55(1):407–474, 2022.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [N/A]
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

A Technical Lemmata

Proof of Lemma 1. We simply exihibit the proof for $E = L^2_\omega(\mathcal{M}, T\mathcal{M})$. Indeed, let $f \in L^2_\omega(\mathcal{M}, T\mathcal{M})$, then:

$$\begin{aligned}
\|L_\phi f\|^2 &= \int g(L_\phi f, L_\phi f) d\omega && (10)\\
&= \int_{\text{supp}(\phi)} g(L_\phi f, L_\phi f) d\omega + \int_{\mathcal{M}\setminus\text{supp}(\phi)} g(L_\phi f, L_\phi f) d\omega && (11)\\
&= \int_{\phi(\text{supp}(\phi))} g(d\phi^{-1}.f, d\phi^{-1}.f) \det(J\phi^{-1}) d\omega' + \int_{\mathcal{M}\setminus\text{supp}(\phi)} g(f,f) d\omega && (12)\\
&\leq \int_{\text{supp}(\phi)} g(f,f)\|d\phi^{-1}\|^2 \det(J\phi^{-1}) d\omega' + \|f\|^2 && (13)\\
&\leq (\sup_{\omega\in\text{supp}(\phi)} \|d\phi^{-1}(\omega)\|^{2(d+1)} + 1)\|f\|^2 < \infty && (14)\\
& && (15)
\end{aligned}$$

Thus, L_ϕ is bounded. □

A.1 A remark on the Flowbox theorem

Usually, the Flowbox Theorem (here Theorem 4) is stated for a (often local) diffeomorphism. If $c(m) \neq 0, \tilde{c}(m) \neq 0$, then there exists U, V and $\phi : U \to V$ a diffeomorphism such that $m \in U \cap V$ and $L_\phi(1_U c) = 1_V \tilde{c}$. However, we note that thanks to Theorem 4 of [26], it is possible to find $\tilde{U}$ smaller such that there exists $\tilde{\phi} : \mathcal{M} \to \mathcal{M}$ which is a global diffeomorphism and $\forall m \in \tilde{U}, \tilde{\phi}(m) = \phi(m)$. In this case, $\tilde{\phi}, \tilde{U}$ and $\tilde{V} = \tilde{\phi}(\tilde{U})$ are the candidates of our statement in Theorem 4. As this is quite technical and rather intuitive, we skipped this remark in the main paper.

A.2 Spatial localization (common to the scalar and vector case)

We now explain how to localize our operator M. Equipped with Lemma 2, we can extend our contraction result on $\mathbb{R}^d$ to $\mathcal{M}$ as follow:

Corollary 1 (Contraction of an openset)**.** *For any $U \in \mathcal{O}_1$ and W openset such that $\bar{U} \subset W \subset \mathcal{M}$, there exists ϕ_n supported on W such that for any $f \in L^p_\omega(\mathcal{M}, T\mathcal{M})$:*

$$L_{\phi_n}(1_U f) \to 0\,.$$

Proof. We prove first the result for $U = \mathcal{B}(0,1)$ and $\bar{U} \subset W$. In this case, it is possible to find $\epsilon > 0$ such that $\mathcal{B}(0, 1+\epsilon) \subset W$. Now, taking ϕ_n^{-1} as in Lemma 2, we get:

$$\begin{aligned}
\int_{\mathbb{R}^d} \|L_{\phi_n}(1_\mathcal{B}(0,1) f)(u)\|^p\, du &= \int_{\mathbb{R}^d} \|1_\mathcal{B}(0,1)(\phi_n^{-1}(u)) d\phi_n(u).f(\phi_n^{-1}(u))\|^p\, du && (16)\\
&= \int_{\mathbb{R}^d} \|1_{\mathcal{B}(0,\frac{1}{n})}(u) d\phi_n(u).f(nu)\|^p\, du && (17)\\
&= \frac{1}{n^d}\int_{\mathbb{R}^d} 1_{\mathcal{B}(0,1)}(u)\|d\phi_n(\frac{u}{n}).f(u)\|^p\, du && (18)\\
&\leq \frac{1}{n^{d+1}}\|1_{\mathcal{B}(0,1)} f\|^p \to 0 && (19)
\end{aligned}$$

Next, getting back to the manifold, we know that if $U \in \dot{\mathcal{O}}_1$, there is $V \in \mathcal{O}_1$ such that $\bar{U} \subset V$. We can thus find an openset $\mathcal{B} \subset V$, such that in the chart of V, $\mathcal{B}$ is an open ball, and $U \subset \mathcal{B} \subset W$. We can thus apply the technique derived above to get $\phi_n : V \to V$, compactly supported, which contracts $\mathcal{B}$(and thus U) to 0 and supported in W. Since it is smooth, compactly supported on W, we can extend it on $\mathcal{M}$ and we get the result. □

Next, this technique can be used to build a sequence of contraction, which allows to explicitly localize the image of a compactly supported function, as follow:

Lemma 9 (Lemma 3 restated for closed sets)**.** *Let $F \subset \mathcal{M}$ a closed set. Then, for any $f \in L^p_\omega(\mathcal{M}, \mathbb{R})$, we have:*

$$M[f1_F] = 1_F M[f]$$

Proof. Because $\mathcal{M}$ is a manifold, it is second countable and thus there is a countable collection of opens such that $\mathcal{M}\backslash F = \cup_{i\geq 0} U_i$ with $U_i \in \mathcal{O}_1$. We use Lemma 12 and, we apply the dominated convergence theorem to $f_n = 1_{\cup_{i\leq n} U_i} f$ to conclude. □

Proof of Lemma 3. We note that if $U \in \dot{\mathcal{O}}_1$, then $\omega(\bar{U}\backslash U) = 0$ and we can thus use the Corollary 1 to conclude. □

A.3 Action on locally constant functions, for the scalar and vector cases

We now prove the part specific to the vector field setting, i.e., that the action of M is locally a multiplication by a scalar.

Proof of Lemma 4. **Step 1:**$M(1_U c)(m) = 1_V \lambda(m, U, c)c$ such that $c(m) \neq 0, \forall m \in U$.

Let $c \in C^\infty_c(\mathcal{M}, T\mathcal{M})$. For $U \in \dot{\mathcal{O}}_1$, $m_0 \in U$, fix a chart $\psi : U \to \mathbb{R}^d$, $\psi(m_0) = 0$ and c is constant in ψ denoted $c^\psi \in \mathbb{R}^d$, which is possible thanks to the Theorem 4. This can also be written as for m in a neighborhood of m_0:

$$d\psi(m).c(m) = c^\psi\,.$$

Following the strategy in Lemma 8,there is $W \in \mathcal{O}_d$ such that $Wc^\psi = c^\psi$ and $Wv = -v$ for any vector v orthogonal to c^ψ. By compacity, we can find A an open set small enough, with boundary of measure 0, such that $0 \in A$, and $\mathcal{W}A \subset \psi(U)$ for any $\mathcal{W} \in \mathcal{O}_d$. Now, setting $\tilde{\phi} = \psi^{-1} \circ W \circ \psi$, which is well defined on the open $\cup_{\mathcal{W}\in\mathcal{O}_d} \mathcal{W}A$, using Theorem 4 of [26](see remark Sec. A.1 of the appendix), we can can extend ϕ globally such that on a local neighborhood, $\forall m \in \tilde{U}, \phi(m) = \tilde{\phi}(m)$. Now, up to taking A even smaller, we can use: $V = \overline{\psi^{-1}(\cup_{n\in\mathbb{Z}} W^n A)} \subset U$, which is closed with a measure 0 boundary(we have a countable union). We get:

$$\begin{aligned} L_\phi(1_V c)(m_0) &= [d\psi^{-1}(m_0) \circ W \circ d\psi(m_0)]c(m_0)1_V & (20)\\ &= 1_V c(m_0)\,. & (21) \end{aligned}$$

Let us denote $p^\perp_{c^\psi}$ the orthogonal projection (with respect to the Euclidean scalar product) on the orthogonal plane to c^ψ.

As $V \subset U$, V is closed and $U \in \dot{\mathcal{O}}_1$ from Lemma 9, we know that:

$$M(c)(m_0) = M(1_U c)(m_0) = M(1_V c)(m_0) = \lambda(m_0, c, U)d\psi^{-1}(0)c^\psi + d\psi^{-1}(0)p^\perp_{c^\psi} M(1_V c)(m_0)$$

Yet, on the other hand:

$$L_\phi M(1_V c)(m_0) = \lambda(m_0, c, U)d\psi^{-1}(0)c_\psi - d\psi^{-1}(0)p^\perp_{c^\psi} M(1_V c)(m_0) \quad (22)$$

As this is true for any m_0, we thus proved that:

$$M(1_U c)(m) = 1_U \lambda(m, U, c)c$$

Step 2: In fact, $\lambda(m, c, U) = \lambda(m, U)$ if c does not cancel on U and $m \in U$.

Let $c, \tilde{c}$ be two vector fields as above and defined on U both not equal to 0, and $m \in U$. Using the Theorem 4 combined with the remark of Sec. A.1 of the appendix, there exists $\phi : \mathcal{M} \to \mathcal{M}$ a diffeomorphism and $\tilde{V}, V \subset U$ and $m \in \tilde{V} \cap V$, such that $L_\phi(1_V c(m)) = 1_{\tilde{V}}\tilde{c}(m)$ and $\phi(m) = m$ Now, we could take a smaller closed set $V \subset U$ with measure 0 boundary, so that $M[1_V c](m) = M[1_U c](m) = M[c](m)$, which would lead to, following a similar argument to above:

$$\lambda(m, \tilde{c}, U)\tilde{c}(m) = M[1_{\tilde{V}}\tilde{c}(m)] = L_\phi M[1_V c](m) = L_\phi(\lambda(., c, U)c)(m) = \lambda(m, c, U)\tilde{c}(m)$$

and then locally λ is independent of the choice of a vector field, which implies the desired property.

Step 3: In fact, $\lambda(m, U) = \lambda(U)$. Indeed, let $m, m_0 \in V$ and $\phi \in \text{Diff}(\mathcal{M})$ such that $\phi(m) = m_0$ (as V is connex, by using Lemma 11). Now, along the same line as above:

$$\lambda(m, U) = \lambda(m_0, U)$$

The previous results hold when the vector field can be locally straightened, however the vector fields that take value 0 on some points of U can not be straightened. We will now show that vector fields that can be straightened on $U \in \dot{\mathcal{O}}_1$ are dense dense in $C^\infty(U, TU)$ for the L^p_ω norm. Let $f \in C^\infty(U, TU)$, let $A = \{x \in U | f(x) = 0\}$, and $A^\epsilon = \{x \in U | \|f(x)\| \leq \epsilon\}$ for $\epsilon > 0$. By Urysohn's lemma there is $\chi^\epsilon : U \to \mathbb{R}$ be such that $\chi|_{A^\epsilon} = 1$ and $\chi|_{U \setminus A^{2\epsilon}} = 0$. Let,

$$f^\epsilon = f + 2\epsilon\chi^\epsilon$$

For any $x \in U$,

$$\|f^\epsilon(x)\| \geq |\|f(x)\| - 2\epsilon\chi^\epsilon(x)| \tag{23}$$

and by construction $|\|f(x)\| - 2\epsilon\chi^\epsilon(x)| > 0$.

Therefore,

$$M[f^\epsilon 1_U] = \lambda(U) f^\epsilon \tag{24}$$

Furthermore for all $0 < \epsilon \leq 1$, $\|f^\epsilon\|$ is bounded by $\|f\| + 2$ that is integrable, so by dominated convergence theorem, $f^\epsilon \xrightarrow[\epsilon \to 0]{L^p_\omega} f$. So, $M[f 1_U] = \lambda(U) f$.

To end the proof, one remarks that $C^\infty_c(\mathcal{M}, T\mathcal{M})$ is dense in $L^p_\omega(\mathcal{M}, T\mathcal{M})$.

□

The next Lemma shows that, in the scalar case, we can consider $\tilde{M}f \triangleq Mf - M(0)$ for $f \in L^p_\omega(\mathcal{M}, \mathbb{R})$ without losing in generality.

Lemma 10. *Under the assumptions of Theorem 1,* $M(0)$ *is constant, and if* $\omega(\mathcal{M}) = \infty$*, then* $M(0) = 0$.

Proof. Following the Theorem 1 of [23], for any $m, m_0 \in \mathcal{M}$, we can find ϕ global diffeomorphism such that $\phi(m) = m_0$. We note that $L_\phi(0) = 0$ and thus for any $m \in \mathcal{M}$:

$$M(0)(m) = M[L_\phi(0)] = L_\phi M(0)(m) = M(0)(m_0)$$

Thus, $M(0)$ is constant, and if $\omega(\mathcal{M}) = \infty$, it is necessary that $M(0) = 0$. □

The corresponding Lemma in the scalar case is substantially simpler, as strongly convex sets are connex:

Proof of Lemma 5. Fix $m_0 \in V$, and let $m \in V$, using Lemma 11(because $V \in \dot{\mathcal{O}}_1$ is connex, we can apply a connexity argument or the transitivity argument of Theorem 1 of [23] for compactly supported diffeomorphisms), we can find ϕ supported in V such that $\phi(m_0) = m$. Thus, $L_\phi f = f$ and $Mf(m_0) = ML_\phi f(m_0) = L_\phi Mf(m_0) = Mf(m)$. Thus, $M(c1_V) = h(c, V)1_V$. The Lipschitz aspect is inherited from the fact that M is Lipschitz. □

A.4 Extrapolation to any good open sets (common to the scalar and vector case)

In this section, we use the fact that we want to prove that both scalar and vector operators correspond to point-wise non-linearity, which are locally Lipschitz due to the regularity assumptions that we used.

Proof of Proposition 1. **Step 1:Fix c, for any $m \in U$ such that $V \subset U$, then $h(c, U)(m) = h(c, V)(m)$**

Indeed, we note that for $m \in U$, where we used Lemma 3:

$$M(1_V f)(m) = 1_V(m)M(f)(m) = 1_U(m)M(f)(m) = M(1_U f)(m)$$

Thus, $h(c, V)|_V = h(c, U)|_V$ for any $V \subset U$.

Step 2: extension by density, for any f, $M(f1_U) = 1_U h(f, U)$ for any $f \in L^p_\omega(\mathcal{M}, \mathbb{R})$. Using Lemma 7, consider $f \in \mathcal{C}^\infty_c(E)$, $f_n = \sum_n 1_{U_n} c_n$, where c_n is either a constant scalar, either a vector field, with disjoint support such that $\|1_U f - 1_U f_n\| < \epsilon$.

We know that, from Lemma 6 that:

$$M(1_U f_n) = M[\sum_n 1_{U_n} c_n] = \sum_n 1_{U_n} M[1_{U_n} c_n] = \sum_n 1_{U_n} h(c_n, U)$$

Next, we note that:

$$\begin{aligned} \|M1_U f - 1_U h(f, U)\| &\leq \|1_U(Mf_n - Mf)\| + \|1_U Mf_n - 1_U h(f_n, U)\| && (25)\\ &+ \|1_U(h(f_n, U) - h(f, U))\| && (26)\\ &\leq 2L\|1_U(f_n - f)\| && (27) \end{aligned}$$

and from this, given that $h(., U)$ is L-Lipschitz, we conclude by density of $\mathcal{C}^\infty_c(\mathcal{M})$ in $L^p_\omega(\mathcal{M}, \mathbb{R})$.

Step 3: Independence from U

Step 1 allows for the following definition of a global h from local h_U: let $m \in \mathcal{M}$, pose,

$$\forall U \in \dot{\mathcal{O}}_1 \quad h(f(m)) := h(f(m), U) \tag{28}$$

In the scalar case and in the vector case, one can build a scalar function and vector function such that, $f(m) = \mu \in \mathbb{R}$ or $f(m) = c \in T_x M$ (as shown in Step 3 of proof of 4). Therefore in the scalar case h is a function from $\mathbb{R}$ to $\mathbb{R}$ and in the vector case for any $x \in \mathcal{M}$ and $v \in T_x\mathcal{M}$, $h(x) = \lambda x$.

□

We only prove the Vitali version for $L^p_\omega(\mathcal{M}, \mathbb{R})$, as the proof for $L^p_\omega(\mathcal{M}, T\mathcal{M})$ would be identical, replacing solely the scalar by constant vector fields in their local parametrization.

Proof of Lemma 7. We consider U small enough such that $U \in \dot{\mathcal{O}}_1$, $m \in U$ and $\exp_m : \mathcal{B} \to U$ is locally a diffeomorphism from $\mathcal{B} \subset T\mathcal{M}_m$, and let $U_i = \exp_m(\mathcal{B}_i)$ with $\mathcal{B}(x_i, r_i) \subset \mathcal{B}$, which is strongly convex and thus $U_i \in \dot{\mathcal{O}}_1$. We remind that $\exp_m$ is bi-Lipschitz on the bounded set U. In this case, there is $C_1, C_2 > 0$ such that for any x_i, r_i with $\mathcal{B}(x_i, r_i) \subset \mathcal{B}$, we have $r_i^d \leq \lambda(\mathcal{B}(x_i, r_i)) \leq C_1\omega(U_i) \leq C_2\lambda(\mathcal{B}(x_i, r_i)) \leq C_d r^d$. By Vitali's lemma, we have for any $\epsilon > 0$ and $r > 0$, the existence of some $x_i, r_i < r$:

$$\|1_{\mathcal{B}} - \sum_{i=1}^n 1_{\mathcal{B}(x_i, r_i)}\|^p \leq \epsilon^p$$

For f smooth, let:

$$\|f(x)1_U - \sum_{i=1}^n f(x_i)1_{U_i}\|^p \leq \|\sum_{i=1}^n (f(x) - f(x_i))1_{U_i}\|^p + \|1_{U\setminus(\cup_i U_i)} f(x)\|^p \tag{29}$$

Now, as $\exp_m$ is bi-Lipschitz, we get a r small enough such that $|f(x) - f(x_i)| < \epsilon$. Next, because the sets are disjoint:

$$\| \sum_{i=1}^{n} (f(x) - f(x_i)) 1_{U_i} \|^p = \sum_{i=1}^{n} \int_{U_i} |f(x) - f(x_i)|^p \tag{30}$$

$$\leq \sum_{i=1}^{n} \omega(U_i) \epsilon^p \tag{31}$$

$$\leq \epsilon^p \omega(U)) \,. \tag{32}$$

Now, using $|f(x)| \leq \|f\|_\infty$, we get:

$$\|1_{U \setminus (\cup_i U_i)} f(x)\|^p \leq \|f\|_\infty \epsilon^p$$

And:

$$\|f - \sum_{i=1}^{n} f(x_i) 1_{U_i}\| < (1 + \omega(U))^{1/p} \epsilon \,.$$

□

The following Lemma allows to build diffeomorphism with compact support - we give this proof for the sake of completeness, at it is proved in [23].

Lemma 11. *Fix $\rho > 0$, and $x_0, x_1 \in \mathcal{B}(0, \rho)$, there exists ϕ diffeomorphism, such that $\phi(x_0) = x_1$ and supp(ϕ) $\subset \mathcal{B}(0, \rho)$.*

Proof. Consider f, smooth, supported in $[-1, 1]$ and such that $f(0) = 1$. We will use a connexity argument: let us fix $x_0 \in \mathcal{B}(0, \rho)$. Let's consider $\Gamma = \{x \in \mathcal{B}(0, \rho) : \exists \phi \text{ diffeomorphism } \phi(x) = x_0,\ \text{supp}(\phi) \subset \mathcal{B}(0, \rho)\}$. Let $x_1 \in \Gamma$, then there is $\eta < \frac{1}{2}$, $\mathcal{B}(x_1, \eta) \subset \mathcal{B}(0, \rho)$. For x_2 such that $\|x_1 - x_2\| \leq \frac{\eta}{4 \sup |f'|}$, we introduce:

$$\tau(x) = (x_2 - x_1) f(\frac{\|x - x_1\|^2}{\eta^2}) \,.$$

We have that supp($\mathbf{I} - \tau$) $\subset \mathcal{B}(x_1, \eta)$, and:

$$\frac{\partial \tau}{\partial x}(x) = 2 \frac{(x_2 - x_1)\langle x - x_1, x_1 \rangle}{\eta^2} f'(\frac{\|x - x_1\|^2}{\eta^2})$$

leading to:

$$\|\frac{\partial \tau}{\partial x}(x)\| < \frac{1}{2}$$

This implies that the spectrum of $\partial \tau$ is in $[0, 1[$ and thus, $\mathbf{I} - \partial \tau$ is invertible. Now, by assumption, we know there is ϕ such that $\phi(x_1) = x_0$, compactly supported in Ω. Introducing $\phi' = \phi \circ (\mathbf{I} - \tau)$, then ϕ' is a diffeomorphism, compactly supported in Ω and $\phi'(x_2) = \phi(x_1) = x$, thus $x_2 \in \Gamma$. This shows Γ is open. But also Γ is closed (otherwiwe, we can make a path ...). Thus, by connexity $\Gamma = \Omega$. □

The next Lemma is crucial in our proof, and allows to characterize union of well behaving opensets:

Lemma 12. *Let $n \geq 0$, $\{U_i\}_{i \leq n} \subset \dot{\mathcal{O}}_1$ and F a closed set such that $\bar{U}_i \cap F = \emptyset, \forall i$. Then for any $f \in L^p_\omega(\mathcal{M}, T\mathcal{M})$:*

$$1_F M[(1_F + 1_{\cup_{i \leq n} U_i}) f] = 1_F M[1_F f]$$

Proof. We work by induction on n. For $n = 0$, the result is true. Then, let's write $U^\epsilon_{n+1} = \{x, d(U_{n+1}, x) < \epsilon\}$. It's an openset which contains $\bar{U}_{n+1}$, and by assumption we can pick ϵ small enough such that $U^\epsilon_{n+1} \cap F = \emptyset$. Next, let's apply Lemma 1 to U_{n+1} and $W = U^\epsilon_{n+1}$. Then:

$$1_F M[(1_F + 1_{(\cup_{i \leq n} U_i \setminus U^\epsilon_{n+1}) \cup U_{n+1}}) f] = L_{\phi_n^{-1}} 1_F M[(1_F + 1_{(\cup_{i \leq n} U_i \setminus U^\epsilon_{n+1}) \cup U_{n+1}}) f] \tag{33}$$

$$= 1_F M[L_{\phi_n}(1_F f + 1_{(\cup_{i \leq n} U_i \setminus U^\epsilon_{n+1}) \cup U_{n+1}} f)] \tag{34}$$

$$\to 1_F M[1_F f + 1_{(\cup_{i \leq n} U_i \setminus U^\epsilon_{n+1})} f] \tag{35}$$

Now, we remark that:

$$1_F M[1_F f + 1_{(\cup_{i\leq n} U_i \setminus U^{\epsilon}_{n+1})} f] = 1_F M[1_F f + 1_{\cup_{i\leq n} U_i}(1_{\mathcal{M}\setminus U^{\epsilon}_{n+1}} f)] \tag{36}$$

And we apply the induction hypothesis to $(1_{\mathcal{M}\setminus U^{\epsilon}_{n+1}} f)$. □

The next Lemma is crucial in our proof, and allows to characterize disjoint union of well behaving opensets:

Proof of lemma 6. We note that $\cup_{i=1}^{n} \overline{U_i} = \overline{\cup_{i=1}^{n} U_i}$. Thus, using Lemma 9, given this union is closed and disjoint; as for any closed set,

$$M[f 1_F] 1_{F^c} = M[0] 1_{F^c} = 0 \tag{37}$$

the following linearity property holds,

$$M[\sum_{i=1}^{n} 1_{\bar{U}_i} f] = \sum_{i=1}^{n} 1_{\bar{U}_i} M[f] = \sum_{i=1}^{n} M[1_{\bar{U}_i} f]$$

Now, we conclude as the boundaries have measure 0. □