A Deep Neural Network for Manifold-valued Data with Applications to Neuroimaging

Rudrasis Chakraborty[†], Jose Bouza[−], Jonathan Manton[+] and Baba C. Vemuri[− ⋆]

[†] University of California, Berkeley, USA [−] University of Florida, USA
[+] University of Melbourne, Australia.

Abstract. Developing deep neural networks (DNNs) for manifold-valued data sets has gained significant interest of late in the deep learning research community. Examples of manifold-valued data in the medical imaging domain include (but are not limited to) diffusion magnetic resonance imaging, tensor-based morphometry, shape analysis and more. In this paper we present a novel theoretical framework for DNNs to cope with manifold-valued data inputs, taking inspiration from the convolutional neural network (CNN) architecture. We call our network the ManifoldNet.

Analogous to vector spaces where convolutions are equivalent to computing weighted means, manifold-valued data convolutions can be defined using the weighted Fréchet Mean (wFM). To this end, we present a provably convergent recursive algorithm for computation of the wFM of the given data, where the weights are to be learned. Further, we prove that the proposed wFM layer achieves a contraction mapping and hence the ManifoldNet need not have additional non-linear ReLU units used in standard CNNs to achieve a contraction mapping. Analogous to the equivariance of convolution in Euclidean space to translations, we prove that the wFM is equivariant to the action of the group of isometries admitted by the Riemannian manifold on which the data reside. This equivariance property facilitates weight sharing within the network. We present experiments using the ManifoldNet framework to achieve regression between diffusion MRI scans of Parkinson Disease (PD) patients and clinical information such as their Movement Disorder Society's Unified Parkinson's Disease Rating Scale (MDS-UPDRS) scores. In another experiment, we present results of finding group differences based on brain connectivity at the fiber bundle level between PD and controls.

1 Introduction

CNNs pioneered by [16] have gained much popularity since their significant success on Imagenet data reported in [15]. In the past few years there has been a growing interest in generalizing CNNs and deep networks in general to data that reside on smooth non-Euclidean spaces. In this context, at the outset, it would be useful to categorize problems into (1) those that involve data as samples of real-valued functions defined on

⋆ This research was in part funded by the NSF grants IIS-1525431 and IIS-1724174 to Baba C. Vemuri. email:vemuri@ufl.edu

This paper has appeared in the Proceedings of the International Conf. on Information Processing in Medical Imaging, 2019, vol 11492

a manifold and (2) those that are simply manifold-valued and hence are sample points on a manifold. In this paper we will consider the second problem, namely, when the input data are sample points on known Riemannian manifolds for example, the manifold of symmetric positive definite (SPD) matrices, $SPD(n)$, the special orthogonal group, $\mathsf{SO}(n)$, the n-sphere, $\mathbf{S}^n$, the Grassmannian, $\mathsf{Gr}(p,n)$, and others. To be precise, the domain of interest is an n-dimensional field of points sampled from a Riemannian manifold. There is very little prior work that we are aware of on DNNs that can cope with input data samples residing on these manifolds with the exception of [13,12]. In [13], authors presented a deep network architecture for classification of hand-crafted features residing on a Grassmann manifold that form the input to the network. In [12], authors presented a deep network architecture for data on $SPD(n)$. In both of these works, the architecture does not involve the use of any convolution or equivalent operations on $\mathsf{Gr}(p,n)$ or $SPD(n)$. Further, it does not use the natural invariant metric or intrinsic operations on the Grassmannian or the $SPD(n)$ in the network blocks. Using intrinsic operations within the layers guarantees that the result remains on the manifold and hence does not require any projection (extrinsic) operations to ensure the result lies in the same space. Further, using extrinsic operations can yield results that are susceptible to significant inaccuracies when the data variance is large [22]. Moreover, since there are no convolution type operations defined for data on these manifolds in their network, it can not be considered a generalization to the CNN and as a consequence does not consider the equivariance property to the action of the group of isometries denoted by $I(\mathcal{M})$, admitted by the manifold $\mathcal{M}$.

In this paper, we present a novel DNN framework called the ManifoldNet. This is a potential analog of a CNN that can cope with input data sampled from a Riemannian manifold. The intuition in defining the analog relies on the equivariance property. Note that convolution of functions in vector spaces are equivariant to translations in the input domain. Further, it is easy to show that traditional convolutions of functions are equivalent to computing the weighted mean [10]. Hence, for the case of manifold-valued data, we can define the analogous operation of a weighted Fréchet mean (wFM) and prove that it is equivariant to the action of $I(\mathcal{M})$. This will be presented in a subsequent section. Our key contributions in this work (**presented in section 2**) are (i) we define the analog of convolution operations for manifold-valued data to be one of estimating the wFM for which we present a provably convergent, efficient and recursive estimator. (ii) A proof of equivariance of wFM to the action of $I(\mathcal{M})$. This equivariance allows the network to share weights within the layers. (iii) A novel deep architecture involving the Riemannian counterparts to the conventional CNN units (**presented in section 3**). (iv) Two real data experiments, (a) regression between changes in diffusional structure – captured in the Cauchy deformation tensor obtained via nonrigid registration of the ensemble average propagator (EAP) field computed from the patient scan to the EAP control atlas – and function in movement disorder patients. (b) An experiment on finding group differences based on brain connectivity at the fiber bundle level specifically, the motor sensory area (M1) tract in both the brain hemispheres.

2 Group action equivariant network for manifold-valued data

In this section, we will define the equivalent of a convolution operation on Riemannian manifolds. As mentioned in the introduction, the domain of interest is an n-dimensional field of manifold valued points. Before formally defining such an operation and building the ManifoldNet, we first present some relevant concepts from differential geometry that will be used in the rest of the paper.

Preliminaries. Let $(\mathcal{M}, g^{\mathcal{M}})$ be a orientable complete Riemannian manifold with a Riemannian metric $g^{\mathcal{M}}$, i.e., $(\forall x \in \mathcal{M})\ g^{\mathcal{M}}_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbf{R}$ is a bi-linear symmetric positive definite map, where $T_x\mathcal{M}$ is the tangent space of $\mathcal{M}$ at $x \in \mathcal{M}$. Let $d : \mathcal{M} \times \mathcal{M} \to [0, \infty)$ be the metric (distance) induced by the Riemannian metric $g^{\mathcal{M}}$. With a slight abuse of notation we will denote a Riemannian manifold $(\mathcal{M}, g^{\mathcal{M}})$ by $\mathcal{M}$ unless specified otherwise. Let Δ be the supremum of the sectional curvatures of $\mathcal{M}$.

Definition 1. *Let $p \in \mathcal{M}$, $r > 0$. Define $\mathcal{B}_r(p) = \{q \in \mathcal{M} | d(p, q) < r\}$ to be a open ball at p of radius r.*

Definition 2. *[11] The local injectivity radius at $p \in \mathcal{M}$, $r_{inj}(p)$, is defined as $r_{inj}(p) = \sup\left\{r | Exp_p : (\mathcal{B}_r(\mathbf{0}) \subset T_p\mathcal{M}) \to \mathcal{M}$ is defined and is a diffeomorphism onto its image$\right\}$. The injectivity radius [19] of $\mathcal{M}$ is defined as $r_{inj}(\mathcal{M}) = \inf_{p\in\mathcal{M}}\{r_{inj}(p)\}$.*

Within $\mathcal{B}_r(p)$, where $r \leq r_{\text{inj}}(\mathcal{M})$, the mapping $\text{Exp}_p^{-1} : \mathcal{B}_r(p) \to \mathcal{U} \subset T_p\mathcal{M}$, is called the inverse Exponential/ Log map.

Definition 3. *[14] An open ball $\mathcal{B}_r(p)$ is a regular geodesic ball if $r < r_{inj}(p)$ and $r < \pi/\left(2\Delta^{1/2}\right)$.*

In Definition 3 and below, we interpret $1/\Delta^{1/2}$ as ∞ if $\Delta \leq 0$. It is well known that, if p and q are two points in a regular geodesic ball $\mathcal{B}_r(p)$, then they are joined by a unique geodesic within $\mathcal{B}_r(p)$ [14].

Definition 4. *[7] $\mathcal{U} \subset \mathcal{M}$ is strongly convex if for all $p, q \in \mathcal{U}$, there exists a unique length minimizing geodesic segment between p and q and the geodesic segment lies entirely in $\mathcal{U}$.*

Definition 5. *[11] Let $p \in \mathcal{M}$. The local convexity radius at p, $r_{cvx}(p)$, is defined as $r_{cvx}(p) = \sup\{r \leq r_{inj}(p) | \mathcal{B}_r(p)$ is strongly convex$\}$. The convexity radius of $\mathcal{M}$ is defined as $r_{cvx}(\mathcal{M}) = \inf_{p\in\mathcal{M}}\{r_{cvx}(p)\}$.*

For the rest of the paper, we will assume that the samples on $\mathcal{M}$ lie inside an open ball $U = \mathcal{B}_r(p)$ where $r = \min\{r_{\text{cvx}}(\mathcal{M}), r_{\text{inj}}(\mathcal{M})\}$, for some $p \in \mathcal{M}$, unless mentioned otherwise. Now, we are ready to define the operations necessary to develop the ManifoldNet architecture.

2.1 wFM on $\mathcal{M}$ as a generalization of convolution

We will begin by defining a convolution type operation on points sampled from $\mathcal{M}$. This convolution operation will perform an averaging over a moving window, where weighted

sums are replaced with weighted intrinsic averages. Let $\{X_i\}_{i=1}^N$ be the manifold-valued samples on $\mathcal{M}$. We define the convolution type operation on $\mathcal{M}$ as the weighted Fréchet mean (wFM) [20] of the samples $\{X_i\}_{i=1}^N$. Also, by the aforementioned condition on the samples, the existence and uniqueness of the FM is guaranteed [1]. As mentioned earlier, it is easy to show (see [10]) that convolution $\psi^* = b \star a$ of two functions $a : X \subset \mathbf{R}^n \to \mathbf{R}$ and $b : X \subset \mathbf{R}^n \to \mathbf{R}$ can be formulated as computation of the weighted mean $\psi^* = argmin_\psi \int a(\mathbf{u})(\psi - \widetilde{b}_\mathbf{u})^2 d\mathbf{u}$, where, $\forall \mathbf{x} \in X, \widetilde{b}_\mathbf{u}(\mathbf{x}) = b(\mathbf{u} + \mathbf{x})$ and $\int a(\mathbf{x}) d\mathbf{x} = 1$. Here, f^2 for any function f is defined pointwise. Further, the defining property of convolutions in vector spaces is the linear translation equivariance in both the domain and the range of the image. Since weighted mean in vector spaces can be generalized to wFM on manifolds and further, wFM can be shown (see below) to be equivariant to group actions admitted by the manifold, we claim that wFM is a generalization of convolution operations to manifold-valued data.

Let $\{w_i\}_{i=1}^N$ be the weights such that they satisfy the convexity constraint, i.e., $\forall i, w_i > 0$ and $\sum_i w_i = 1$, then wFM, $\mathsf{wFM}(\{X_i\}, \{w_i\})$ is defined as:

$$\mathsf{wFM}(\{X_i\}, \{w_i\}) = \underset{M \in \mathcal{M}}{\operatorname{argmin}} \sum_{i=1}^{N} w_i d^2(X_i, M) \tag{1}$$

Analogous to the equivariance property of convolution to translations in vector spaces, we will now show that the wFM is equivariant under the action of the group of isometries of $\mathcal{M}$. We will first formally define the group of isometries of $\mathcal{M}$ (let us denote it by G) and then define the equivariance property and show that wFM is G-equivariant.

Definition 6 (Group of isometries of $\mathcal{M}$ ($I(\mathcal{M})$)). *A diffeomorphism $\phi : \mathcal{M} \to \mathcal{M}$ is an isometry if it preserves distance, i.e., $d(\phi(x), \phi(y)) = d(x, y)$. The set $I(\mathcal{M})$ of all isometries of $\mathcal{M}$ forms a group with respect to function composition. Rather than write an isometry as a function ϕ, we will write it as a group action. Henceforth, let G denote the group $I(\mathcal{M})$, and for $g \in G$, and $x \in \mathcal{M}$, let $g.x$ denote the result of applying the isometry g to point x.*

Clearly $\mathcal{M}$ is a G set (see [9] for the definition of a G set). We will now define equivariance and show that wFM is G-equivariant.

Definition 7 (Equivariance). *Let X and Y be G sets. Then, $F : X \to Y$ is said to be G-equivariant if $\forall g \in G$, $\forall x \in X$, $F(g.x) = g.F(x)$.*

Let $U \subset \mathcal{M}$ be an open ball inside which FM exists and is unique, let P be the set of all possible finite subsets of U.

Theorem 1. *Given $\{w_i\}$ satisfying the convex constraint, let $F : P \to U$ be a function defined by $\{X_i\} \mapsto \mathsf{wFM}(\{X_i\}, \{w_i\})$. Then, F is G-equivariant.*

Proof. Let $g \in G$ and $\{X_i\}_{i=1}^N \in P$, now, let $M^* = \mathsf{wFM}(\{X_i\}, \{w_i\})$, as $g.F(\{X_i\}) = g.M^*$, it suffices to show $g.M^*$ is $\mathsf{wFM}(\{g.X_i\}, \{w_i\})$ (assuming the existence and uniqueness of $\mathsf{wFM}(\{g.X_i\}, \{w_i\})$ which is stated in the following claim).

Claim: Let $U = \mathcal{B}_r(p)$ for some $r > 0$ and $p \in \mathcal{M}$. Then, $\{g.X_i\} \subset \mathcal{B}_r(g.p)$ and hence $\mathsf{wFM}(\{g.X_i\}, \{w_i\})$ exists and is unique.

Let $\widetilde{M}$ be $\mathsf{wFM}(\{g.X_i\}, \{w_i\})$. Then, $\sum_{i=1}^N w_i d^2\left(g.X_i, \widetilde{M}\right) = \sum_{i=1}^N w_i d^2\left(X_i, g^{-1}.\widetilde{M}\right)$. Since, $M^* = \mathsf{wFM}(\{X_i\}, \{w_i\})$, hence, $M^* = g^{-1}.\widetilde{M}$, i.e., $\widetilde{M} = g.M^*$. Thus, $g.M^* = \mathsf{wFM}(\{g.X_i\}, \{w_i\})$, which implies F is G-equivariant.

A class of Riemannian manifolds on which G acts transitively are called Riemannian homogeneous spaces. We can see that on a Riemannian homogeneous space $\mathcal{M}$, wFM is G-equivariant. Equipped with a G-equivariant operator on $\mathcal{M}$, we can claim that the wFM (defined above) is a valid convolution operator since group equivariance is a unique defining property of a convolution operator.

The rest of this subsection will be devoted to developing an efficient way to compute wFM. The strategy is to cast the weighted FM computation as an unweighted FM computation, and then use efficient FM estimators. Let $\omega^{\mathcal{M}} > 0$ be the Riemannian volume form.

Let $p_{\mathbf{X}}$ be the probability density of a U-valued random variable $\mathbf{X}$ with respect to $\omega^{\mathcal{M}}$ on $U \subset \mathcal{M}$, so that $\mathsf{Pr}(X \in \mathfrak{A}) = \int_{\mathfrak{A}} p_X(Y)\omega^{\mathcal{M}}(Y)$ for any Borel-measurable subset $\mathfrak{A}$ of U. Let $Y \in U$, we can define the expectation of the real valued random variable $d^2(,Y) : U \to \mathbf{R}$ by $E\left[d^2(,Y)\right] = \int_U d^2(X,Y)p_{\mathbf{X}}(X)\omega^{\mathcal{M}}(X)$. Now, let $w : U \to (0,\infty)$ be an integrable function where $\int_U w(X)\,\omega^{\mathcal{M}}(X) = 1$.

Let $\widetilde{p}_X$ be the probability density corresponding to the probability measure $\widetilde{\mathsf{Pr}}$ defined by $\widetilde{\mathsf{Pr}}(X \in \mathfrak{X}) = \int_{\mathfrak{X}} \widetilde{p}_X(Y)\omega^{\mathcal{M}}(Y) := \int_{\mathfrak{X}} \frac{1}{C}\, p_X(Y)w(Y)\omega^{\mathcal{M}}(Y)$, where, $\mathfrak{X}$ lies in the Borel σ-algebra over U and let $C = \int_U p_X(Y)w(Y)\omega^{\mathcal{M}}(Y)$. Note that the constant $C > 0$, since p_X is a probability density, $w > 0$ and $\mathcal{M}$ is orientable.

Now, we will state and prove the following proposition.

Proposition 1. *Using the notation from above we have:* ***(i)*** $\mathsf{supp}(p_X) = \mathsf{supp}(\widetilde{p}_X)$. ***(ii)*** $\mathsf{wFE}(\mathsf{X}, w) = \mathsf{FE}\left(\widetilde{\mathsf{X}}\right)$.

Proof. Let $X \in \mathsf{supp}(p_X)$, then, $p_X(X) > 0$. Since, $w(X) > 0$, hence, $\widetilde{p}_X(X) > 0$ and thus, $X \in \mathsf{supp}(\widetilde{p}_X)$. On the other hand, assume $\widetilde{X}$ to be a sample drawn from $\widetilde{p}_X$. Then, either $p_X\left(\widetilde{X}\right) = 0$ or $p_X\left(\widetilde{X}\right) > 0$. If, $p_X\left(\widetilde{X}\right) = 0$, then, $\widetilde{p}_X\left(\widetilde{X}\right) = 0$ which contradicts our assumption. Hence, $p_X\left(\widetilde{X}\right) > 0$, i.e., $\widetilde{X} \in \mathsf{supp}(p_X)$. This concludes the proof of part (i).

Let X and $\widetilde{\mathsf{X}}$ be the $\mathcal{M}$ valued random variable following p_X and $\widetilde{p}_X$ respectively. We define the weighted Fréchet expectation (wFE) of X as $\mathsf{wFE}(\mathsf{X}, w) = \mathrm{argmin}_{Y\in\mathcal{M}} \int_{\mathcal{M}} w(X)d^2(X,Y)p_X(X)\omega^{\mathcal{M}}(X)$.

Observe, $E_w\left[d^2(,Y)\right] := \int_U w(X)d^2(X,Y)p_X(X)\,\omega^{\mathcal{M}}(X) = C\int_U d^2(X,Y)\widetilde{p}_X(X)\,\omega^{\mathcal{M}}(X) = C\,\widetilde{E}\left[d^2(,Y)\right]$ .. Hence, we get $\mathsf{FE}\left(\widetilde{\mathsf{X}}\right) = \mathsf{wFE}(\mathsf{X}, w)$, as C is independent of the choice of Y, which concludes the proof of part (ii).

Now let $\{X_i\}_{i=1}^N$ be samples drawn from p_X and $\left\{\widetilde{X}_i\right\}_{i=1}^N$ be samples drawn from $\widetilde{p}_X$. In order to compute wFM, we will now present an online algorithm (inductive FM

Estimator – dubbed iFME). Note that in [22,6,17], authors present recursive algorithms for FM computation on the hyper-sphere, Stiefel and $SPD(n)$ manifolds respectively. These specific algorithms are distinct from our work here since the wFM approach is applicable to any Riemannian manifold.

iFME wFM Estimator: Given, $\{X_i\}_{i=1}^N \subset U$ and $\{w_i := w(X_i)\}_{i=1}^N$ such that $\forall i, w_i > 0$, the n^{th} estimate, M_n of wFM $(\{X_i\}, \{w_i\})$ is given by the following recursion:

$$M_1 = X_1 \qquad\qquad M_n = \Gamma_{M_{n-1}}^{X_n}\left(\frac{w_n}{\sum_{j=1}^n w_j}\right). \tag{2}$$

In the above equation, $\Gamma_X^Y : [0,1] \to U$ is the shortest geodesic curve from X to Y. Observe that, in general wFM is defined with $\sum_{i=1}^N w_i = 1$, but in above definition, $\sum_{i=1}^N w_i \neq 1$. We can normalize $\{w_i\}$ to get $\{\widetilde{w}_i\}$ by $\widetilde{w}_i = w_i/\left(\sum_i w_i\right)$, but then Eq. 2 will not change as $\widetilde{w}_n/\left(\sum_{j=1}^n \widetilde{w}_j\right) = w_n/\left(\sum_{j=1}^n w_j\right)$. This gives us an efficient inductive/recursive way to define convolution operation on $\mathcal{M}$. We now show that the proposed wFM estimator is consistent in the following proposition (the proof is in supplementary section).

Proposition 2. *Using the above notations and assumptions, let $\{X_i\}_{i=1}^N$ be i.i.d. samples drawn from p_X on $\mathcal{M}$. Let the wFE be finite. Then, M_N converges a.s. to wFE as $N \to \infty$.*

2.2 Nonlinear operation between wFM-layers for $\mathcal{M}$-valued Data

In the traditional CNN model, we need an intermediate nonlinear function between convolutional layers (e.g. ReLU). As argued in [18], any nonlinear function used in CNNs is basically a contraction mapping. Formally, let F be a nonlinear mapping from U to V. Assume U and V are metric spaces equipped with metrics d_U and d_V respectively. Then F is a contraction mapping *iff* $\exists c < 1$ such that $d_V(F(x), F(y)) \leq c\, d_U(x,y)$. F is a non-expansive mapping [18] *iff* $d_V(F(x), F(y)) \leq d_U(x,y)$.

One can easily see that the popular choices for nonlinear operations like ReLU, sigmoid are indeed non-expansive mappings. We will now show that the function wFM as defined in 1, is a contraction mapping for any non-trivial choice of weights. Let $\{X_i\}_{i=1}^N$ and $\{Y_j\}_{j=1}^M$ be the two set of samples on $\mathcal{M}$. Without loss of generality assume $N \leq M$. We consider the set $\mathcal{U}^M = \underbrace{U \times \cdots \times U}_{M \text{ times}}$. Clearly $\{Y_j\}_{j=1}^M \in \mathcal{U}^M$ and we embed $\{X_i\}_{i=1}^N$ in $\mathcal{U}^M$ as follows: we construct $\left\{\widetilde{X}_i\right\}_{i=1}^M$ from $\{X_i\}_{i=1}^N$ by defining $\widetilde{X}_i = X_{(i-1) \bmod N+1}$. Let us denote the embedding by ι. Now, define the distance on $\mathcal{U}^M$ as $d\left(\left\{\widetilde{X}_i\right\}_{i=1}^M, \{Y_j\}_{j=1}^M\right) = \max_{i,j} d(X_i, Y_j)$. We say the choice of weights for wFM is trivial if one of the weights is 1 (hence all others are 0).

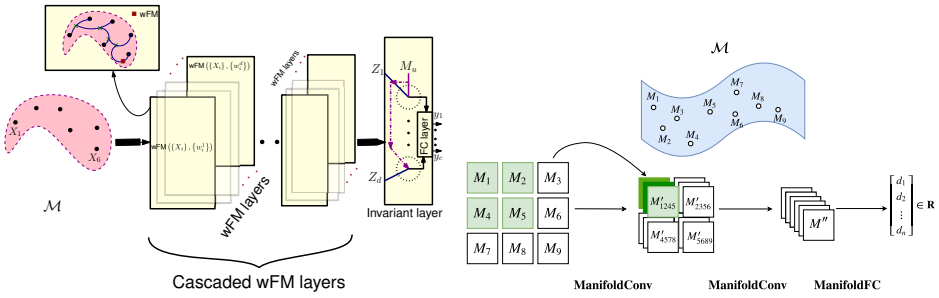

Fig. 1: *Left-Right* (a) Schematic diagram of ManifoldNet (b) 2×2 ManifoldNet conv. example

Proposition 3. *For all nontrivial choices of* $\{\alpha_i\}_{i=1}^N$ *and* $\{\beta_j\}_{j=1}^M$ *satisfying the convexity constraint ,* $\exists c < 1$ *such that,*

$$d\left(\mathit{wFM}\left(\{X_i\}_{i=1}^N, \{\alpha_i\}_{i=1}^N\right), \mathit{wFM}\left(\{Y_j\}_{i=1}^M, \{\beta_j\}_{i=1}^M\right)\right) \leq c\, d\left(\iota\left(\{X_i\}_{i=1}^N\right), \{Y_j\}_{j=1}^M\right) \tag{3}$$

2.3 The invariant (last) layer

We will form a deep network by cascading multiple sliding wFM windows each of which acts as a convolution-type layer. Each convolutional-type layer is equivariant to the group action, and hence at the end of the cascaded convolutional layers, the output is equivariant to the group action applied to the input of the network. Let d be the number of output channels each of which outputs a wFM, hence each of the channels is equivariant to the group action. However, in order to build a network that yields an output which is *invariant* to the group action we would like the last layer (i.e., the analogue to a linear classifier) to be invariant to the group action. The last layer is thus constructed as follows: Let $\{Z_1, \cdots, Z_d\} \subset \mathcal{M}$ be the output of d channels and $M_u = \mathsf{FM}\left(\{Z_i\}_{i=1}^d\right) = \mathsf{wFM}\left(\{Z_i\}_{i=1}^d, \{1/d\}_1^d\right)$ be the unweighted FM of the outputs $\{Z_i\}_{i=1}^d$. Then, we construct a layer with d outputs whose i^{th} output $o_i = d\left(M_u, Z_i\right)$. Let c be the number of classes for the classification task, then, a fully connected (FC) layer with inputs $\{o_i\}$ and c output nodes is used. Finally, a softmax operation is then used at the c output nodes to obtain the outputs $\{y_i\}_{i=1}^c$. In the following proposition we claim that this last layer with $\{Z_i\}_{i=1}^d$ inputs and $\{y_i\}_{i=1}^c$ outputs is group invariant.

Proposition 4. *The last layer with* $\{Z_i\}_{i=1}^d$ *inputs and* $\{y_i\}_{i=1}^c$ *outputs is group invariant.*

Proof. Using the above construction, let $W \in \mathbf{R}^{c \times d}$ and $\mathbf{b} \in \mathbf{R}^c$ be the weight matrix and bias respectively of the FC layer. Then,

$$\mathbf{y} = F\left(W^T \mathbf{o} + \mathbf{b}\right) = F\left(W^T d\left(M_u, Z\right) + \mathbf{b}\right), \tag{4}$$

where, F is the softmax function. In the above equation, we treat $d\left(M_u, Z\right)$ as the vector $\left[d\left(M_u, Z_1\right), \cdots, d\left(M_u, Z_d\right)\right]^t$. Observe that, $g.M_u = \mathsf{FM}\left(\{g.Z_i\}_{i=1}^d\right)$. As each of the d channels is group equivariant, Z_i becomes $g.Z_i$. Because of the property of the distance under group action, $d\left(g.M_u, g.Z_i\right) = d\left(M_u, Z_i\right)$. Hence, one can see that if we change the inputs $\{Z_i\}$ to $\{g.Z_i\}$, the output $\mathbf{y}$ will remain invariant.

In Fig. 1 we present a schematic of ManifoldNet depicting the different layers of processing the manifold-valued data as described above in Sections 2.1-2.3.

Unlike standard Euclidean CNN, note that, here we do not need a nonlinearity between two convolution layers as argued in subsection 2.2. Note that, in standard CNN, without the presence of non-linearity one can collapse a deep network into a shallow one. This raises the following question *Can ManifoldNet be collapsed to it's shallow counterpart as there is no non-linearity between layers ?* In order to answer this question, we will show that for manifolds with non-constant sectional curvatures, we can not collapse the ManifoldNet into a shallow network.

Here we present a proof (via a counter example) to show that a multi-layer ManifoldNet can not be collapsed to a single layer ManifoldNet when $\mathcal{M} = SPD(n)$, which is a manifold with non-constant curvature.

Theorem 2. *The multi-layer ManifoldNet is not equivalent to the single layer ManifoldNet for data on Riemannian manifolds with non-constant sectional curvature.*

Proof. This is a proof by counter example. Let us suppose we are given 4 SPD matrices, $A = \begin{bmatrix}0.9593 & 0.3429\\ 0.3429 & 0.1493\end{bmatrix}$, $B = \begin{bmatrix}1.2575 & 0.5475\\ 0.5475 & 1.8143\end{bmatrix}$, $C = \begin{bmatrix}1.2435 & 0.6396\\ 0.6396 & 1.1966\end{bmatrix}$, $D = \begin{bmatrix}1.2511 & 0.5446\\ 0.5447 & 1.3517\end{bmatrix}$, whose wFM we want to compute. Let us consider two sequences $S1 = \{A, B, C, D\}$ and $S2 = \{A, C, B, D\}$. Consider a one layer ManifoldNet for computing the wFM of these four matrices. For simplicity of exposition, suppose this one layer network learns equal weights $(= 0.25)$ for all matrices and hence yields the wFM $M = \begin{bmatrix}1.1640 & 0.4667\\ 0.4667 & 0.6388\end{bmatrix}$ as the solution for both sequences $S1$ and $S2$ respectively. To compute the wFM, we use a gradient descent applied to the weighted sum of square geodesic distances between the unknown wFM and the sample points.

Now, let us consider a two layer wFM. For $S1$, the first layer computes wFM of $\{A, B\}$ and $\{C, D\}$ respectively and returns $M1$ and $M2$ as the wFMs. Then, the second layer takes $M1$ and $M2$ as inputs and returns their wFM say, $M3$. Analogously for the sequence $S2$, the first layer computes wFM of $\{A, C\}$ and $\{B, D\}$ and returns $\bar{M}1$ and $\bar{M}2$. Then the second layer takes as input, $\bar{M}1$ and $\bar{M}2$ and returns $\bar{M}3$ as their wFM.

It can be verified that for the first layer if we use equal weights, we need the weights for the second layer to be 0.4980 and 0.5050 for $S1$ and $S2$ respectively such that both $M3 = M$ and $\bar{M}3 = M$. This counter example shows that the weights are dependent on the input data matrices, which means that in general a multi-layer ManifoldNet can not be collapsed to a single layer ManifoldNet.

3 Experiments

We now present the basic ManifoldNet architecture and evaluate its performance on two medical imaging tasks: (1) Diffusion Tensor field hypothesis testing and (2) nonlinear regression. We remark that although the ManifoldNet architecture is perfectly capable of tackling the classification of manifold-valued data, in the medical imaging domain, procuring a very large population of such data is either prohibitively expensive or unavailable via public access repositories. Hence, we chose to demonstrate the efficacy of the ManifoldNet on other challenging applications such as regression and group testing that do not require tens of thousands of images for training and testing.

We now describe in detail how to use the basic layers defined and analyzed in section 2 to create multilayer ManifoldNet architectures comparable to deep CNNs. Note that the input to ManifoldNet is always an N-dimensional field of points sampled from a Riemannian manifold. For expository purposes we will consider the case of a *manifold-valued image*, i.e., the input to any layer of the ManifoldNet architecture is, $(B \times W \times H \times C)$ points on a Riemannian manifold. Here, B the batch size, W the image width, H is the image height and C is the number of input channels.

Analogous to traditional convolution layers, we now define the **ManifoldConv** layer, which slides a small window of weights along the spatial dimensions (W and H) of the input, but instead of weighted sums we now compute the weighted FM in each window as described in section 2.1. As in the traditional CNNs, we can do this for several different weight tensors to generate multiple output channels. So the **ManifoldConv** layer transforms the original tensor of manifold valued points to a size $(B \times W' \times H' \times C')$ tensor of manifold valued points. Important properties of the **ManifoldConv** layers are:

- Traditional convolution layers are **ManifoldConv** with $\mathcal{M} = \mathbf{R}$, so **ManifoldConv** generalizes traditional convolution layers.
- The **ManifoldConv** generalization preserves the key property that has made convolutional layers among the most successful building blocks of deep network architectures: equivariance to isometric transformations of the input data. This property is preserved by wFM, and therefore also by the **ManifoldConv** layer.
- Since we have shown that the wFM operation is both a contraction mapping and, atleast in the case of non-constant curvature manifolds, non-collapsible, it follows that the same properties hold for the **ManifoldConv** layer. Therefore the stacking of such layers without an intermediate non-linearity is justified for non-constant curvature manifolds.
- The weights for a wFM computation should satisfy the convexity constraint, which can be imposed by squaring the weight matrix to be positive and then normalizing to sum to 1. This allows us to use regular backpropagation through the **ManifoldConv** layer.

After stacking several **ManifoldConv** layers we may require a vector valued output for classification purposes, and we would like this vector to be *invariant* to the natural group isometries admitted by the manifold on which the input data reside. To generate this vector we use the invariant final layer from section 2.3. We henceforth call this layer the **ManifoldFC** layer, since it is an analogue to the traditional fully connected layers.

Using these two layers we can build general classifiers of non-constant curvature manifold valued data using an architecture of the form

$$\textbf{ManifoldConv} \rightarrow \textbf{ManifoldConv} \rightarrow \cdots \rightarrow \textbf{ManifoldConv} \rightarrow \textbf{ManifoldFC}$$

And for other tasks such as manifold regression we can remove the **ManifoldFC** layer. This architecture can be trained end-to-end using traditional backpropagation since the weights are real valued.

Group testing on movement disorder patients: Diffusion MRI is a commonly used modality for diagnosing and studying neurological disorders. In this experiment we use a dataset consisting of scans from a control group of 44 patients and a group of 50 Parkinsons patients (see 2). All scans were performed using a $3.0T$ Philips Achieva scanner with a 32-channel volume head coil. The parameters of the diffusion imaging acquisition sequence were as follows: gradient directions = 64, b-values = $0/1000$ s/mm2, repetition time = $7748ms$, echo time = $86ms$, flip angle = 90, field of view = 224×224 mm, matrix size = 112112, number of contiguous axial slices = 60 and SENSE factor P = 2. for parameters of image acquisition. Sensory motor area tracts called M1 fiber tracts are first extracted from the scans using FSL software [2] from both the left ('LM1') and right hemispheres ('RM1') respectively. We then use the FSL software [4] to extract $SPD(3)$ diffusion tensors from each tract. Note that the space $SPD(3)$ of (3×3) SPD matrices is a homogeneous Riemannian manifold with the isometry group being $GL(n)$.

Following [5] closely, we fit a 3 layer ManifoldNet model to both the control group and the Parkinsons group data. Using the method from [23] we compute the distance between these two models, denoted by $\mathbf{d}$. Now we permute the class labels between the classes, retrain the two models and compute the network distance d_j. If there are significant differences between the classes we should expect that $\mathbf{d} > d_j$. We repeat this experiment for $j = 1, \ldots, 1000$ and let p be the proportion of experiments for which $\mathbf{d} \leq d_j$. This is a permutation test of the null hypothesis "there is no significant difference between the tract models learned from the two different classes." We can compare this to the performance of the similar dMRI architecture SPD-SRU and the baseline methods from [5]. We found that our ManifoldNet architecture achieved an 'LM1' p-value of 0.029 and an 'RM1' p-value of 0.024. The baseline method gave an 'LM1' and 'RM1' p-value of 0.17 and 0.34 respectively, while the SPD-SRU architecture gave p-values of 0.01 and 0.032 respectively. We can conclude that using ManifoldNet we can reject the null hypothesis with 95% confidence, which is competitive with SPD-SRU.

Nonlinear Regression between Structure and Function: This dataset contains high angular resolution diffusion image (HARDI) scans from, (1) healthy controls, (2) patients with essential tremor (ET) and (3) Parkinson's disease (PD) patients. This data pool contains scans from 25 controls, 15 ET and 26 PD patients. This HARDI data was acquired using the same parameters as before. The dimension of each image is $(112 \times 112 \times 60)$. From each of these images, we identify the region of interest (ROI) (40 voxels in size) containing the Substantia Nigra (a neuroanatomical structure known to be affected most by PD and ET).

In morphometric analysis, it is common to use the Cauchy deformation tensor (CDT) field to capture changes in a patient scan with respect to a reference template/atlas.

Thus, in order to capture changes in patient HARDI scans with respect to the control atlas, we first nonrigidly register each of the EAP (ensemble average propagator) fields estimated from the input HARDI scan to the computed EAP atlas and obtain the CDT at each voxel in the ROI, given by $\sqrt{J^T J}$, where, J is the Jacobian of the non-rigid transformation [8]. The CDT is an SPD matrix of dimension (3×3) in this case. Hence, for each patient we extract a CDT field of dimension $(3 \times 3 \times 40)$. In this experiment, we seek to find the relationship between structural information in the form of CDT and clinical measures such as the MDS-UPDRS (Movement Disorder Society's revision of the Unified Parkinson's Disease Rating Scale) [21]. The MDS-UPDRS score is widely used to follow the longitudinal course of PD. These scores are obtained via interviews and clinical observations by an expert. In this experiment, available to us are the MDS-UPDRS scores for all the 58 subjects in the population under consideration. This score is a nonnegative natural number, with smaller values indicating normality.

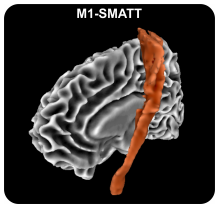

Fig. 2: M1-SMATT template

For these 58 patients, we used a 3 layer ManifoldNet to find the relation between CDT field and MDS-UPDRS scores. We used an MSE loss and obtained an R^2 statistic of 0.93. This R^2 statistic value is similar to the one reported in [3].

4 Conclusions

In this paper, we presented a novel deep network called ManifoldNet suited for processing manifold-valued data sets. Inputs to the ManifoldNet are manifold-valued and not real or complex-valued functions defined on non-Euclidean domains. Our key contributions are: (i) A novel deep network to be perceived as a generalization of the CNN to manifold-valued data inputs using purely intrinsic operations on the data manifold. (ii) Analogous to convolutions in vector spaces – which can be computed using the weighted mean – we present wFM operations on the manifold and prove the equivariance of the wFM to natural group actions admitted by the manifold. This equivariance allows us to share the learned weights within a layer of the ManifoldNet. (iii) An efficient recursive wFM estimator that is provably convergent. (iv) Experimental results demonstrating the efficacy of the ManifoldNet for, (a) regression between dMRI scans of PD patients and clinical MDS-UPDRS scores and (b) finding group differences between PD and Controls based on brain connectivity at the fiber bundle level.

References

1. Afsari, B.: Riemannian L^p center of mass: Existence, uniqueness, and convexity. Proceedings of the American Mathematical Society **139**(02), 655–655 (2011). https://doi.org/10.1090/S0002-9939-2010-10541-5, `http://www.ams.org/jourcgi/jour-getitem?pii=S0002-9939-2010-10541-5`
2. Archer, D., Vaillancourt, D., Coombes, S.: A template and probabilistic atlas of the human sensorimotor tracts using diffusion mri. Cerebral Cortex **28**, 1–15 (03 2017). https://doi.org/10.1093/cercor/bhx066

3. Banerjee, M., Chakraborty, R., Ofori, E., Okun, M.S., Viallancourt, D.E., Vemuri, B.C.: A nonlinear regression technique for manifold valued data with applications to medical image analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4424–4432 (2016)
4. Behrens, T., Berg, H.J., Jbabdi, S., Rushworth, M., Woolrich, M.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? NeuroImage **34**(1), 144 – 155 (2007). https://doi.org/https://doi.org/10.1016/j.neuroimage.2006.09.018, http://www.sciencedirect.com/science/article/pii/S1053811906009360
5. Chakraborty, R., Yang, C.H., Zhen, X., Banerjee, M., Archer, D., Vaillancourt, D., Singh, V., Vemuri, B.C.: Statistical Recurrent Models on Manifold valued Data. ArXiv e-prints (May 2018)
6. Chakraborty, R., Vemuri, B.C., et al.: Statistics on the stiefel manifold: Theory and applications. The Annals of Statistics **47**(1), 415–438 (2019)
7. Chavel, I.: Riemannian geometry: a modern introduction, vol. 98. Cambridge university press (2006)
8. Cheng, G., Vemuri, B.C., Carney, P.R., Mareci, T.H.: Non-rigid registration of high angular resolution diffusion images represented by gaussian mixture fields. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 190–197. Springer (2009)
9. Dummit, D.S., Foote, R.M.: Abstract algebra, vol. 3. Wiley Hoboken (2004)
10. Goh, A., Lenglet, C., Thompson, P.M., Vidal, R.: A nonparametric riemannian framework for processing high angular resolution diffusion images and its applications to odf-based morphometry. NeuroImage **56**(3), 1181–1201 (2011)
11. Groisser, D.: Newton's method, zeroes of vector fields, and the Riemannian center of mass. Advances in Applied Mathematics **33**(1), 95–135 (2004). https://doi.org/10.1016/j.aam.2003.08.003
12. Huang, Z., Van Gool, L.J.: A Riemannian Network for SPD Matrix Learning. In: AAAI. vol. 1, p. 3 (2017)
13. Huang, Z., Wu, J., Van Gool, L.: Building deep networks on Grassmann manifolds. arXiv preprint arXiv:1611.05742 (2016)
14. Kendall, W.S.: Probability, convexity, and harmonic maps with small image. I. Uniqueness and finite existence. Proc. London Math. Soc. (3) **61**(2), 371–406 (1990)
15. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet Classification with Deep Convolutional Neural Networks. Advances In Neural Information Processing Systems (2012). https://doi.org/http://dx.doi.org/10.1016/j.protcy.2014.09.007
16. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. of the IEEE (1998). https://doi.org/10.1109/5.726791
17. Lim, Y., Pálfia, M.: Weighted inductive means. Linear Algebra and its Applications **453**, 59–83 (2014)
18. Mallat, S.: Understanding Deep Convolutional Networks. Philosophical Transactions A **374**, 20150203 (2016). https://doi.org/10.1098/rsta.2015.0203, http://arxiv.org/abs/1601.04920
19. Manton, J.H.: A globally convergent numerical algorithm for computing the centre of mass on compact lie groups. In: Control, Automation, Robotics and Vision Conference, 2004. ICARCV 2004 8th. vol. 3, pp. 2211–2216. IEEE (2004)
20. Maurice Fréchet: Les éléments aléatoires de nature quelconque dans un espace distancié. Annales de l'I. H. P., **10**(4), 215–310 (1948)
21. Ramaker, C., Marinus, J., Stiggelbout, A.M., Van Hilten, B.J.: Systematic evaluation of rating scales for impairment and disability in parkinson's disease. Movement disorders: official journal of the Movement Disorder Society **17**(5), 867–876 (2002)

22. Salehian, H., Chakraborty, R., Ofori, E., Vaillancourt, D., Vemuri, B.C.: An efficient recursive estimator of the Fréchet mean on a hypersphere with applications to Medical Image Analysis. Mathematical Foundations of Computational Anatomy (2015)
23. Triacca, U.: Measuring the distance between sets of arma models. Econometrics **4**, 32 (07 2016). https://doi.org/10.3390/econometrics4030032