Neural Ordinary Differential Equations on Manifolds

Luca Falorsi [1 2] **Patrick Forré** [1]

Abstract

Normalizing flows are a powerful technique for obtaining reparameterizable samples from complex multimodal distributions. Unfortunately current approaches fall short when the underlying space has a non trivial topology, and are only available for the most basic geometries. Recently normalizing flows in Euclidean space based on Neural ODEs show great promise, yet suffer the same limitations. Using ideas from differential geometry and geometric control theory, we describe how neural ODEs can be extended to smooth manifolds. We show how vector fields provide a general framework for parameterizing a flexible class of invertible mapping on these spaces and we illustrate how gradient based learning can be performed. As a result we define a general methodology for building normalizing flows on manifolds.

1. Introduction

Recently Chen et al. (2018) showed how to effectively integrate Ordinary Differential Equations (ODE) with Deep Learning frameworks. Ubiquitous in all fields of science, differential equations are the main modelling tools for physical processes. In deep learning their introduction was initially motivated from the observation that the popular residual network (ResNet) architecture can be interpreted as an Euler discretization step of a differential equation (Haber & Ruthotto, 2017). Instead of relying on discretized maps, Chen et al. (2018) proposed to directly model the continuous dynamics using *vector fields* in $\mathbb{R}^n$. A vector field, through its associated ODE, indicates for every point an infinitesimal displacement change, and therefore implicitly describes a map from the space to itself called a *flow*. While the flow can be practically computed using numerical ODE solvers, the key observation of Chen et al. (2018) is that we can threat the ODE solution as a black-box. This means that in the backward pass we do not have to differentiate through the operations performed by the numerical solver, instead Chen et al. (2018) propose to use the adjoint sensitivity method (Pontryagin et al., 1962). Closely related with the Pontryagin Maximum Principle, one of the most prominent results in control theory, the adjoint sensitivity method allows to compute vector-Jacobian product of the ODE solutions with respect to its inputs. This is done by simulating the dynamics given by the initial ODE backwards, augmenting it with a linear differential equation, which intuitively can be thought of as a continuous version of the usual chain rule.

Neural ODEs found one of their major applications in the context of *normalizing flows* (Grathwohl et al., 2019; Finlay et al., 2020). Normalizing flows (NF) are a general methodology for defining complex reparameterizable densities by applying a series of diffeomorphism to samples from a (simple) base distribution (Rezende & Mohamed, 2015). See Papamakarios et al. (2019) for a general review of NF. The resulting density at the transformed points can be computed using the change of variable formula. Flows defined by vector fields are particularly amenable for this task, as for every time interval the ode solution defines a diffeomorphism. In this case, the change change of density is given simply by integrating the divergence of the vector field along integral curves.

As many real word problems are naturally defined on spaces with a non-trivial topology, recently there has been a great interest in building probabilistic deep learning frameworks that can works on manifolds different from the Euclidean space (Davidson et al., 2018; Falorsi et al., 2018; 2019; Pérez Rey et al., 2019; Nagano et al., 2019). For this objective the possibility of defining complex reparameterizable densities on manifolds through normalizing flows is of central importance. However as of today there exist few alternatives, mostly limited to the most basic and simple topologies.

The main obstacle for defining normalizing flows on manifolds is that there is no general methodology for parameterizing maps $F : M \to N$ between two manifolds. Neural networks can only accomplish this for the the Euclidean space, $\mathbb{R}^n$. In this work we propose to use vector fields on a manifold M as a flexible way to parameterize diffeomor-

[1]University of Amsterdam [2]Aiconic. Correspondence to: Luca Falorsi <luca.falorsi@gmail.com>.

Second workshop on Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models (ICML 2020), Virtual Conference

phic maps from the manifold to itself. As a vector fields defines an infinitesimal displacement on the manifold for every point, they naturally give rise to diffeomorphisms without needing to impose further constraints. In addition, there exist decades old research on how to numerically integrate ODEs on manifolds. See Hairer (2011) for a review of the main methods.

We start in Section 2 by delineating how vector fields and ODEs on a manifold M can be defined in the context of differential geometry. We then explain how vector fields naturally give rise, through their associated flow, to diffeomorphisms on M. In Section 3 we describe how the adjoint sensitivity method can be generalized to vector fields on manifolds in the context of geometric control theory (Agrachev & Sachkov, 2013). This highlights important connections with symplectic geometry and the Hamiltonian formalism. Similarly as in the adjoint method in the Euclidean space, to backpropagate through the flow defined by a vector field we have to solve an ODE in an augmented space. In this case the ODE is given by a vector field on the cotangent space T^*M, called *cotangent lift*, which is a *lift* of the original vector field on M. In Section 4 we demonstrate how flows defined by vector fields allow to define *continuous normalizing flows* on manifold. As a proof of concept we show how the defined framework can be used to model complex multimodal densities on the hypersphere in Appendix A and provide practical advice on how to parameterize vector fields on a manifold using neural networks in Appendix C.

2. Vector fields and flows on manifolds

Throughout the paper M will be a n-dimensional smooth manifold. Vector fields are the mathematical objects that allows us to generalize the concept of ODEs to manifolds. A **smooth vector field** X is defined as a smooth section of the tangent bundle $X \in \Gamma^\infty(M, TM)$. A smooth time dependent vector field is a smooth function $X : \mathbb{R} \times M \to TM$ such that $\forall t \in \mathbb{R}$, $X_t := X(t, \cdot) \in \Gamma^\infty(M, TM)$ is a smooth vector field.

Definition 1. *Let $X : \mathbb{R} \times M \to TM$ be a smooth time dependent vector field and $J \subseteq \mathbb{R}$ an interval. A curve $\gamma : J \to M$ is an* **integral curve** *of X if:*

$$\dot{\gamma}(t) = X(t, \gamma(t)), \quad \forall t \in J. \tag{1}$$

We call **maximal integral curve** *an integral curve that cannot be extended to a larger interval J.*

Writing Equation (1) in a local smooth chart we find that it is equivalent to (locally) solving a system of ODEs. We can then apply the existence and uniqueness theorem to show that every smooth vector field always admits integral curves:

Theorem 1 (Theorem 2.15 in (Agrachev et al., 2019)). *Let X be a smooth time dependent vector field, then for any point $(t_0, p_0) \in \mathbb{R} \times M$ there exists a unique maximal integral curve $\gamma : t_0 \in J \to M$ with starting point q_0 at starting time t_0 denoted by $\gamma(t; t_0, q_0)$. We call γ a solution of the Cauchy problem:*

$$\begin{cases} \dot{q}(t) = X(t, q(t)), \\ q(t_0) = q_0. \end{cases} \tag{2}$$

Moreover the map $(t_0, q_0) \to \gamma(t; t_0, q_0)$ is smooth in a neighborhood of (t_0, q_0).

A time-dependent vector field is **complete** if for every $(t_0, q_0) \in \mathbb{R} \times M$, the maximal solution $\gamma(t; t_0, q_0)$ of the Cauchy problem is defined on all $\mathbb{R}$.

Through integral curves, vector fields on manifolds give us a flexible and convenient way of defining maps from M to itself. Restricting to the time independent case, this means considering the family of maps $\phi^t_X : M \to M, \quad \phi^t_X(q) = \gamma(t; 0, q)$ where $t \in \mathbb{R}$. In this case we say that the vector field generates a **flow**[1].

Definition 2. *A smooth flow is a smooth left $\mathbb{R}$-action on a manifold M; that is, a family of smooth diffeomorphisms $\phi^t : M \to M, \forall t \in \mathbb{R}$, satisfying the following properties:*

$$\phi^0 = Id, \qquad \phi^t \circ \phi^s = \phi^{t+s} \quad \forall t, s \in \mathbb{R} \tag{3}$$

Every smooth flow ϕ uniquely generates a smooth complete vector field X by $X_q = \frac{d}{dt}\big|_{t=0} \phi^t(q), \quad \forall q \in M$. Conversely every complete smooth vector field generates smooth flow $\{\phi^t_X\}_{t \in \mathbb{R}}$ through its integral curves. This result is known as the **Fundamental Theorem of Flows**.

For time dependent vector field we have to take into account the additional time dependence. We therefore have **time dependent flows**.

Definition 3. *A time dependent[2] smooth flow on a smooth manifold is a two parameter family of diffeomorphism $\{\phi^{t_1, t_0}\}_{t_0, t_1 \in \mathbb{R}}$, with $\phi^{t_1, t_0} : M \to M$, $\forall t_0, t_1 \in \mathbb{R}$ that satisfy the following conditions:*

$$\phi^{t,t} = Id, \quad \phi^{r,s} \circ \phi^{s,t} = \phi^{r,t}, \quad \forall t, s, r \in \mathbb{R}. \tag{4}$$

Similarly, as in the time independent case, we have that a time dependent complete smooth vector field uniquely generates a time dependent smooth flow and vice versa. Summing up, we have seen how vector fields naturally allow us define maps on a generic smooth manifold M. We refer to (Lee, 2013) for proofs and additional results on vector fields and flows on manifolds.

[1] Not to be confused with NF, a **flow** is only defined for a subset $\mathcal{D} \subseteq \mathbb{R} \times M$ in general, since not all vector fields are complete. A flow defined on all $\mathbb{R} \times M$ is often called a **global flow**. For simplicity we restrict our attention to complete vector fields and global flows. In the rest of the paper a flow will denote a globally defined flow.

[2] A time independent flow can be considered a time dependent flow simply by defining $\phi^{s,t} := \phi^{s-t}$

3. Cotangent lift

Let $X \in \Gamma^\infty(M, TM)$ be a complete smooth vector field [3] and $\{\phi_X^t\}_{t\in\mathbb{R}}$ its generated flow. We are interested in differentiating through $\phi_X^t : M \to M$. In differential geometry this corresponds to computing the pullback map [4] $(\phi_X^t)^* : T^*M \to T^*M$, where T^*M is the cotangent bundle of M. The key observations of the adjoint sensitivity method is that in $M = \mathbb{R}^n$ this quantity can be computed simulating the reverse dynamics of X, augmenting it with an additional linear ODE called the **adjoint equation**. We will show how to generalize this method for vector fields on an arbitrary smooth manifold M. Consider the family of maps[5]:

$$\left(\phi_{-X}^t\right)^* : T^*M \to T^*M, \qquad \forall t \in \mathbb{R}, \tag{5}$$

$$p_q \mapsto \left(\left(\phi_{-X}^t\right)^* p\right)_{\phi_X^t(q)}. \tag{6}$$

Using the properties of the pullback and the fact that $\{\phi_{-X}^t\}_{t\in\mathbb{R}}$ is a flow, it is easy to show that the family $\{\left(\phi_{-X}^t\right)^*\}_{t\in\mathbb{R}}$ defines a flow on T^*M. Following Section 2 there exists a unique vector field $X^{T^*} \in \Gamma^\infty(T^*M, TT^*M)$ on the cotangent bundle that generates this flow. This vector field is called the **cotangent lift** of X (Bullo & Lewis, 2004):

$$X_{p_q}^{T^*} = \frac{d}{dt}\bigg|_{t=0} \left(\left(\phi_{-X}^t\right)^* p\right)_{\phi_X^t(q)}, \quad \forall p_q \in M. \tag{7}$$

This means that given the cotangent vector $p_q \in T_q^*M$, to compute the pullback of p_q by ϕ_X^t we can solve the Cauchy problem defined by $-X^{T^*}$ with starting point p_q. The cotangent lift has the following important properties (See Remark S1.11 in (Bullo & Lewis, 2004)):

Property 1. *X^{T^*} is a Hamiltonian vector field with respect to the canonical symplectic structure of the cotangent bundle T^*M. The Hamiltonian that generates X^{T^*} is $H_X(p_q) = p_q(X_q)$. We therefore have:*

$$X^{T^*} \lrcorner\, \omega = dH_X, \tag{8}$$

where ω is the canonical symplectic form on T^*M

Property 2. *X^{T^*} is a linear vector field on the fibers of T^*M. That is, given $p_q, p'_q \in T_q^*M$, and $a, b \in \mathbb{R}$, it holds that $X^{T^*}(a\cdot p_q + b\cdot p'_q) = a\cdot X^{T^*}(p_q) + b\cdot X^{T^*}(p'_q)$.*

This fundamentally descends because the pullback map is fiberwise linear.

Property 3. *X^{T^*} is a lift of X. That is, $d\pi_M(X_{p_q}^{T^*}) = X_q, \forall p_q \in T^*M$, where $\pi_M : T^*M \to M$ is the projection of the cotangent bundle onto its base space M.*

3.1. Cotangent lift in local coordinates

Let's compute a local expression for the cotangent lift X^{T^*} on a local coordinate chart $(U; x_i)$, $U \subseteq M$. In this chart the vector field will have an expression $X|_U = \sum_{i=1}^n f_i \partial_{x_i}$ where $f_i \in C^\infty(U)$. Since X^{T^*} is a vector field on T^*M we can find its components with respect to the frame $\{\partial_{x_i}, \partial_{\xi_i}\}_{i=1}^n$ adapted to cotangent coordinates $(T^*U; x_i, \xi_i)$ (See Section 2.1 in (Da Silva, 2001)). Since the field is Hamiltonian we can leverage the fact that an Hamiltonian vector fields Y with Hamiltonian H in local cotangent coordinates can be written using Hamilton equations:

$$Y\big|_{T^*U} = \sum_{i=1} \left(\frac{\partial H}{\partial \xi_i}\partial_{x_i} - \frac{\partial H}{\partial x_i}\partial_{\xi_i}\right). \tag{9}$$

In our specific case we have that the local expression for H_X is:

$$H_X\big|_{T^*U} = \sum_{i=1}^n \xi_i dx_i \left(\sum_{j=1}^n f_j \partial_{x^j}\right) = \sum_{i=1}^n \xi_i f_i. \tag{10}$$

Therefore:

$$X^{T^*}\big|_{T^*U} = \sum_{i=1}^n f_i \partial_{x_i} - \sum_{i=1}^n \left(\sum_{j=1}^n \frac{\partial f_i}{\partial x_j}\xi_j\right)\partial_{\xi_i}. \tag{11}$$

Notice that, as we expected from Property 2 and 3, the cotangent lift (11) is linear on the components ∂_{ξ_i} and coincides with X if projected onto the components ∂_{x_i}. Notice that this expression is the same as the adjoint equation. Therefore, for $M = \mathbb{R}^n$, the cotangent lift coincides with the adjoint equation.

3.2. Cotangent lift on embedded submanifolds

Let M be a properly embedded smooth submanifold of $\mathbb{R}^m$, and let $\iota : M \hookrightarrow \mathbb{R}^m$ denote the inclusion map. Consider a smooth vector field $X \in \Gamma^\infty(M, TM)$ and $\overline{X} \in \Gamma^\infty(\mathbb{R}^n, T\mathbb{R}^n)$ a smooth tangent vector field that extends X. We first observe that, since the $\overline{X}$ and X are ι-related, their flows commute with the inclusion map (Proposition 9.6 (Lee, 2013)). We therefore have that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{R}^m & \xrightarrow{\phi_{\overline{X}}^t} & \mathbb{R}^m \\ \uparrow \iota & & \uparrow \iota \\ M & \xrightarrow{\phi_X^t} & M. \end{array} \tag{12}$$

[3] We can consider a time dependent vector field $Y : \mathbb{R} \times M \to TM$ as a vector field $\check{Y} \in \Gamma^\infty(\mathbb{R} \times M, T\mathbb{R} \times TM)$ on the augmented space $\mathbb{R} \times M$: $\check{Y}(s, q) := \left(\frac{\partial}{\partial t}|_{t=s}, Y(s, q)\right)$.

[4] The pullback generalizes the VJP operation to manifolds

[5] Given $p \in T^*M$ we use the notation p_q to stress that p has base point $q \in M$, this means $p \in T_q^*M$.

Suppose we are interested in computing the differential of the function $f \circ \iota \circ \phi_X^t = f \circ \phi_X^t : M \to \mathbb{R}$, where $f : \mathbb{R}^m \to \mathbb{R}$. We can then both write:

$$\begin{aligned} d\left(f \circ \iota \circ \phi_X^t\right) &= \left(\phi_X^t\right)^* \circ \iota^* \circ df = \phi^t_{-X^{T^*}} \circ \iota^* \circ df \\ &= \iota^* \circ \left(\phi_{\overline{X}}^t\right)^* \circ df = \iota^* \circ \phi^t_{-\overline{X}^{T^*}} \circ df. \end{aligned}$$

This means that we can use the cotangent lift of $\overline{X}$ to compute the pullback of cotangent vectors by the flow of X.

4. Continuous normalizing flows on manifolds

Let (M, g) be an orientable Riemannian manifold and $\mu_g \in \Gamma^\infty(M, \Lambda^n T^* M)$ its Riemannian volume form (Lee (2013), Proposition 15.29). Additionally, let X be a complete smooth time dependent vector fields on M and $\{\phi_X^{t,s}\}_{t,s\in\mathbb{R}}$ its smooth time dependent flow. In Section 2 we saw that the maps $\phi_X^{t,s} : M \to M$ define diffeomorphisms on M. We can then use the flow induced by X to define continuous normalizing flows on M.

We represent our initial probability density using the volume form $\rho_0\mu_g \in \Gamma^\infty(M, \Lambda^n T^* M)$ where $\rho_0 \in C^\infty(M)$ is a smooth non-negative function on M that integrates to one: $\int \rho_0\mu_g = 1$. To describe our reparameterized density we define $\rho_t \in C^\infty(M)$ as the smooth function such that:

$$\rho_t\mu_g = \left(\phi_X^{0,t}\right)^* (\rho_0\mu_g), \tag{13}$$

where $(\phi_X^{0,t})^* : \Gamma^\infty(M, \Lambda^n T^* M) \to \Gamma^\infty(M, \Lambda^n T^* M)$ is the pullback of volume forms induced by $\phi_X^{0,t}$ (Lee (2013), Chapter 14). The evolution of ρ_t over time is given by the **continuity equation** (Khalil et al. (2017), Section 4).

Theorem 2 (Continuity equation)**.** *Let M, μ_g, X, ρ_t as defined above. Then the function $\rho \in C^\infty(M \times \mathbb{R})$, $\rho(\cdot, t) := \rho_t$, satisfies the following linear PDE:*

$$X_t(\rho_t) + \rho_t \mathrm{div}(X_t) = -\partial_t \rho. \tag{14}$$

The divergence of a smooth vector field $Y \in \Gamma^\infty(M, TM)$ on a Riemannian manifold is the smooth function $\mathrm{div}(Y) \in C^\infty(M)$ such that

$$\mathrm{div}(Y)\mu_g = \mathcal{L}_Y(\mu_g) = d(Y \lrcorner\, \mu_g), \tag{15}$$

where $\mathcal{L}_Y(\mu_g)$ is the Lie derivative of the the Riemannian volume form with respect to Y. We are interested in computing how the value of ρ_t changes on the flow curves $t \mapsto \phi_X^{t,0}(q_0), q_0 \in M$. Using the continuity equation and the chain rule we have:

$$\frac{d}{dt}\left[\rho_t\left(\phi_X^{t,0}(q_0)\right)\right] = \left[-\mathrm{div}(X_t)\rho_t\right]\left(\phi_X^{t,0}(q_0)\right).$$

If we fix a starting point $q_0 \in M$ we obtain a linear ODE on $\mathbb{R}$. We can then solve for an initial value $\rho_0(q_0)$ of the probability density:

$$\rho_t\left(\phi_X^{t,0}(q_0)\right) = \exp\left(-\int_0^t \mathrm{div}(X_t)\left(\phi_X^{t,0}(q_0)\right) dt\right) \cdot \rho_0(q_0).$$

In many applications we are interested in the log probability density, in which case the expression further simplifies to:

$$\log \rho_t\left(\phi_X^{t,0}(q_0)\right) = \log \rho_0(q_0) - \int_0^t \mathrm{div}(X_t)\left(\phi_X^{t,0}(q_0)\right) dt.$$

5. Related Work

As mentioned in the Introduction, the absence of a general procedure for parameterizing maps between manifolds has been the main obstacle in defining normalizing flows on manifolds. Gemici et al. (2016) try to sidestep this by first mapping points from the manifold M to $\mathbb{R}^n$, applying a normalizing flow in this space and then mapping back to M. However, when the manifold M has a non-trivial topology there exist no continuous and continuously invertible mapping, i.e. a *homeomorphism* between M and $\mathbb{R}^n$, such that this method is bound to introduce numerical instabilities in the computation and singularities in the density. Similarly, Falorsi et al. (2019) create a flexible class of distributions on Lie groups by mapping a complex density from the Lie algebra to the group using the exponential map. While the exponential map is not discontinuous, for some particular groups the resulting density can still present singularities when the initial density in the Lie algebra is not properly constrained. Rezende et al. (2020) define normalizing flows for distributions on hyperspheres and tori. This is done by first showing how to define diffeomorphisms from the circle to itself by imposing special constraints. The method is then generalized to products of circles, and extended to the hypersphere S^n, by mapping it to $S^1 \times [-1, 1]^n$ and imposing additional constraints to ensure that the overall map is a well defined diffeomorphism. Bose et al. (2020) define normalizing flows on hyperbolic spaces by successfully taking into account the different geometry, however the definition of a diffeomorphisms in hyperbolic space is made easier due to the fact that *topologically* the hyperbolic space is homeomorphic to the Euclidean one.

6. Conclusion and future work

Future research will experiment with the presented framework in a wider range of tasks and manifolds. In addition, we will explore how to further improve the scalability of the defined techniques. Possible directions are Monte Carlo approximations of the divergence on manifolds, using numerical integrators adapted to the specific manifold structure and regularization methods based on optimal transport on manifolds (in the spirit of Finlay et al. (2020)).

Acknowledgments

Authors want to thank Nicola De Cao and Tim Davidson for helpful discussions and technical support.

References

Agrachev, A., Barilari, D., and Boscain, U. A comprehensive introduction to sub-riemannian geometry, 2019.

Agrachev, A. A. and Sachkov, Y. *Control theory from the geometric viewpoint*, volume 87. Springer Science & Business Media, 2013.

Bates, J. The embedding dimension of laplacian eigenfunction maps. *Applied and Computational Harmonic Analysis*, 37(3):516–530, 2014.

Bose, A. J., Smofsky, A., Liao, R., Panangaden, P., and Hamilton, W. L. Latent variable modelling with hyperbolic normalizing flows. *arXiv preprint arXiv:2002.06336*, 2020.

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., and Wanderman-Milne, S. JAX: composable transformations of Python+NumPy programs, 2018. URL `http://github.com/google/jax`.

Bullo, F. and Lewis, A. Supplementary chapters for "geometric control of mechanical systems". 06 2004.

Chen, B.-Y. and Verstraelen, L. Laplace transformations of submanifolds, 2013.

Chen, R. T. Q., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. Neural ordinary differential equations. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 31*, pp. 6571–6583. Curran Associates, Inc., 2018.

Da Silva, A. C. *Lectures on symplectic geometry*, volume 3575. Springer, 2001.

Davidson, T. R., Falorsi, L., De Cao, N., Kipf, T., and Tomczak, J. M. Hyperspherical variational auto-encoders. *34th Conference on Uncertainty in Artificial Intelligence (UAI-18)*, 2018.

Falorsi, L., de Haan, P., Davidson, T. R., De Cao, N., Weiler, M., Forré, P., and Cohen, T. S. Explorations in homeomorphic variational auto-encoding. *ICML workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018.

Falorsi, L., de Haan, P., Davidson, T. R., and Forré, P. Reparameterizing distributions on lie groups. *AISTATS*, 2019.

Finlay, C., Jacobsen, J.-H., Nurbekyan, L., and Oberman, A. M. How to train your neural ode: the world of jacobian and kinetic regularization, 2020.

Gemici, M. C., Rezende, D., and Mohamed, S. Normalizing flows on riemannian manifolds. *arXiv preprint arXiv:1611.02304*, 2016.

Grathwohl, W., Chen, R. T. Q., Bettencourt, J., and Duvenaud, D. Scalable reversible generative models with free-form continuous dynamics. In *International Conference on Learning Representations*, 2019.

Haber, E. and Ruthotto, L. Stable architectures for deep neural networks. *Inverse Problems*, 34(1):014004, 2017.

Hairer, E. Solving differential equations on manifolds. 2011.

Howard, R. Analysis on homogeneous spaces, 1994.

Khalil, S. A., Basheer, M. A., and Abdelhaleem, T. A. Types of derivatives: Concepts and applications (ii). *Journal of Mathematics Research*, 9(1), 2017.

Lee, J. M. Smooth manifolds. In *Introduction to Smooth Manifolds*. Springer, 2013.

Nagano, Y., Yamaguchi, S., Fujita, Y., and Koyama, M. A wrapped normal distribution on hyperbolic space for gradient-based learning. In *International Conference on Machine Learning*, pp. 4693–4702, 2019.

Papamakarios, G., Nalisnick, E., Rezende, D. J., Mohamed, S., and Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *stat*, 1050:5, 2019.

Pontryagin, L. S., Mishchenko, E., Boltyanskii, V., and Gamkrelidze, R. *The mathematical theory of optimal processes*. Wiley, 1962.

Pérez Rey, L. A., Menkovski, V., and Portegies, J. W. Diffusion variational autoencoders. *arXiv*, 2019.

Rezende, D. and Mohamed, S. Variational inference with normalizing flows. In *International Conference on Machine Learning*, pp. 1530–1538, 2015.

Rezende, D. J., Papamakarios, G., Racanière, S., Albergo, M. S., Kanwar, G., Shanahan, P. E., and Cranmer, K. Normalizing flows on tori and spheres. *arXiv preprint arXiv:2002.02428*, 2020.

Walschap, G. *Metric Structures in Differential Geometry*. Graduate Texts in Mathematics. Springer, softcover reprint of hardcover 1st ed. 2004 edition, 2004. ISBN 1441919139,9781441919137.

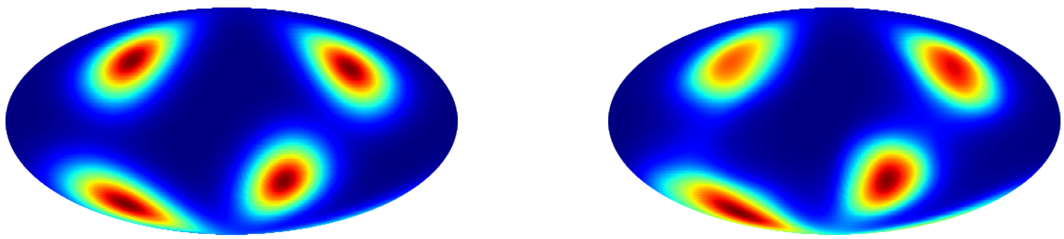

Figure 1. Learned density on S^2

Manifold	Model	KL[nats]	ESS[%]
S^2	MS($N_T = 1$, $K_m = 12$, $K_s = 32$)	.05$\pm$.01	90
	MCNF($H = [10, 10]$)	**.008**$\pm$.001	**98.4**$\pm$.2
S^3	MS($N_T = 1$, $K_m = 32$, $K_s = 64$)	.14	84
	MCNF($H = [10, 10]$)	**.013**$\pm$.001	**97.5**$\pm$.2
$SO(3)$	MCNF($H = [10, 10]$)	**.001**$\pm$.001	**99.7**$\pm$.2
$SU(3)$	MCNF($H = [10, 10]$)	**.006**$\pm$.001	**98.9**$\pm$.2

Table 1. Evaluation of Manifold Continuous Normalizing Flow on Manifold (MCNF) on density matching task. Performance is measured using KL divergence end Effective Sampling Size. H indicates the hidden units in each hidden layer. For S^n the results are compared with recursive Möbius-spline flow (MS) (Rezende et al., 2020). N_T is the number of stacked transformations for each flow; K_m is the number of centres used in Mobius; K_s is the number of segments in the spline flow. Error is computed over 3 replicas of each experiment.

A. Proof of Concept Experiments

As a proof of concept, we show how the proposed Manifold Continuous Normalizing Flow (MCNF) is able to learn complex multi-modal densities on the hyper-sphere and on the matrix Lie groups $SO(3)$ and $SU(3)$. Each model was optimized using Adam for 10000 epochs, learning rate of 10^{-3} and batch size of 256. The parameterized flow was numerically integrated with starting time 0 and final time 1, using the Dormand-Prince ODE integration with adaptive stepsize implemented in Bradbury et al. (2018). All the spaces are considered to be embedded in $\mathbb{R}^m$ and backpropagation was performed using what is observed in Section 3.2.

A.1. Multimodal density matching on spheres

As target densities we used the Mixture of von Mises-Fisher on S^2 and S^3 defined in Rezende et al. (2020). See Rezende et al. (2020) for a detailed description of the task and of the metrics used. The vector fields on S^n are parameterized using the generator defined in Equation 62.

Results are reported in Table 1 while Figure 1 shows the leaned density on S^2. We observe that the proposed model is able to closely match the target densities, with a considerably lower KL divergence than the model by Rezende et al. (2020).

A.2. Multimodal density matching on compact Lie groups

We trained a MCNF on to match multimodal densities on the compact matrix Lie groups $SO(3)$ and $SU(3)$. See Appendix C.2 for definitions and description on how to parameterize vector fields on Lie groups. As target density we used a mixture distribution of the form:

$$\tilde{\rho}(U|k, \{M_i\}_{i=1}^4) \propto \sum_{i=1}^{4} \rho(U|k, M_i) \tag{16}$$

For $SU(3)$ we used:

$$\log \rho(U|k, M) := \text{tr}\left(Real(M^\dagger U)\right) \tag{17}$$

With centers:

$$M_{1,2,3,4} = \begin{bmatrix} \frac{1}{2}i & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & -\frac{1}{2}i & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}}i & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}i \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\frac{\sqrt{3}}{2} & \frac{1}{2}i \\ 0 & \frac{1}{2}i & -\frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{18}$$

For $SO(3)$ we used:

$$\log \rho(U|k, M) := \text{tr}\left(M^\top U\right) \tag{19}$$

With centers:

$$M_{1,2,3,4} = \begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{\sqrt{3}}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{20}$$

In both cases we used concentration $k = 10$. Vector fields are parameterized using genrators formed by left invariant vector fields, which are defined using the Lie algebra bases defined in Equation (34) and Equation (32). Results are reported in Table 1. We observe that the proposed model is able to closely match the target densities.

B. Continuous normalizing flows on semi-Riemannian and nonorientable manifolds

In Section 4 we assumed a orientable manifold with Riemannian metric g to derive the continuous change of volume. This formula uses the divergence, which uniquely depends on the associated Riemannian volume form μ_g. The divergence operator can be generalized assuming a non vanishing volume form or a density [6] on M:

Definition 4. *Let M a smooth manifold and $X \in \Gamma^\infty(M, TM)$ a smooth vector field. Given a non vanishing density $\mu \in \Gamma^\infty(M, \mathcal{D}M)$ we can define the divergence of X with respect to μ as the function* $\text{div}_\mu(X) \in C^\infty(M)$ *such that*

$$\mathcal{L}_X(\mu) = \text{div}_\mu(X) \cdot \mu \tag{21}$$

Similarly, given a non vanishing smooth volume form $\omega \in \Gamma^\infty(M, \Lambda^n T^*M)$ we can define the divergence of X with respect to ω as the function $\text{div}_\omega(X) \in C^\infty(M)$ *such that:*

$$\mathcal{L}_X(\omega) = \text{div}_\omega(X) \cdot \omega \tag{22}$$

This in particular allows to generalize continuous normalizing flows to semi-Riemannian manifolds, where we have non vanishing volume form that arises from the semi-Riemannian metric.

[6] Densities allow to define integration on nonorientable manifolds. Given a volume form ω, the absolute value $|\omega|$ defines a density on M. Contrary to volume forms, any smooth manifold admits a nowhere vanishing density. See Chapter 16 (Lee, 2013) for definitions and further details.

C. Parameterizing vector fields on manifolds

Given a manifold M we are left with the problem of parameterizing a large enough set of vector fields that allows to express a rich class of distributions on the manifold. When we try to parameterize a large set of function we look at neural networks as a natural solution, however they can only parameterize functions $\mathbb{R}^n \to \mathbb{R}^m$, and therefore there is no straightforward way to use them. Finding the best way of parameterizing vector fields on manifolds is an interesting problem with no unique solution, how to tackle it will largely depend on how the manifold is defined and what data structure is used to parameterize it in practice. Nevertheless all the objects and methods discussed in the rest of the paper are defined independently from the specific parameterization method chosen. Therefore, if in the future a better way of parameterizing vector fields will emerge, they will still be applicable.

Notwithstanding the above, in this section we will try to give some guidance on how to approach this problem. In the first part we will show how, using generators, it can be reduced to the much easier task of parameterizing functions on manifolds. We will then give some practical advice in the case where the manifold is described using an embedding in $\mathbb{R}^m$. Throughout this section, given a function $f : M \to \mathbb{R}^m$ we will indicate with $f_i : M \to \mathbb{R}$ its i-th component, such that $f = (f_1, \cdots, f_m)$.

C.1. Local frames and global constraints

We begin by analyzing how we parameterize vector fields in $\mathbb{R}^n$, to investigate to what extent we can generalize this procedure. In the euclidean space vector fields are simply functions $f : \mathbb{R}^n \to \mathbb{R}^n$. In a more geometrical language the function f defines the vector field X in the following way:

$$X = f_1 \partial x^1 + \cdots + f_n \partial x^n. \tag{23}$$

The converse is also true: *for every vector field X there exist a unique continuous function $f : \mathbb{R}^n \to \mathbb{R}^n$ such that Equation* (23) *holds*. On a generic n-dimensional smooth manifold this is only true *locally*. This means that there exists a open cover $\{U_i\}_{i \in \mathcal{I}}$ of M[7], called the **trivialization cover**, such that TM restricted to each U_i is isomorphic to the trivial bundle. This is equivalent to saying that for every set U_i there exist n smooth vector fields $E_1^{(i)}, \cdots, E_n^{(i)} \in \Gamma^\infty(U_i, TU_i))$ such that for every smooth vector field $X \in \Gamma^\infty(M, TM)$ there exists a unique smooth function $f : U_i \to \mathbb{R}^n$ such that:

$$X\big|_{U_i} = f_1 E_1^{(i)} + \cdots + f_n E_n^{(i)}. \tag{24}$$

We then can call $E_1^{(i)}, \cdots, E_n^{(i)}$ a **local frame**. A local frame that is defined on an open domain $U = M$ (this means on the entire manifold) is called a **global frame**. On a manifold there exists plenty of local frames, in fact given a smooth local chart $(U, \{x^i\})$ the fields $\partial_{x^1}, \cdots, \partial_{x_n} \in \Gamma^\infty(U, TU)$ form a local frame called **coordinate frame**. In the special case of $\mathbb{R}^n$ its coordinate frame is a global frame. Unfortunately in general not every manifold has a global frame, the simplest example is the sphere S^2. In the sphere case it is well known that there exists no vector field that is everywhere nonzero, this result goes by the *hairy ball theorem*. It is then clear that no pair of vector fields $E_1, E_2 \in \Gamma^\infty(M, TM)$ can form a global frame, in fact there will always be a point $q \in S^2$ such that:

$$\text{span}\left((E_1)_q, (E_2)_q\right) \leq 1 < 2 = \dim(T_q M). \tag{25}$$

The manifolds for which a global frame $E_1, \cdots, E_n \in \Gamma^\infty(M, TM)$ exists are called **parallelizable manifolds**, for this class we can parameterize all smooth vector fields on M in the same way as we did on $\mathbb{R}^n$. This means choosing a smooth function $f : M \to \mathbb{R}$ and defining a vector field X:

$$X = f_1 E_1 + \cdots + f_n E_n. \tag{26}$$

A manifold is parallelizable iff its tangent bundle is isomorphic to the trivial bundle: $\mathbb{R}^n \times M \cong TM$. A global frame gives an explicit isomorphism:

$$\begin{aligned} &\mathbb{R}^n \times M \to TM, \\ &\quad (z, q) \mapsto z_1 (E_1)_q + \cdots + z_n (E_n)_q . \end{aligned}$$

An important and large class class of parallelizable manifolds is given by Lie Groups, which are smooth manifold which additionally posses a group structure compatible with the manifold structure.

[7]Assuming that the manifold is second countable, there exists $\mathcal{I}$ that is finite and has cardinality $n + 1$, see Lemma 7.1 in (Walschap, 2004).

C.2. Lie groups

A **Lie group** G is a smooth manifold with the additional structure of a group, where the group multiplication and inversion are smooth maps. Lie groups are an important instrument in physics where they are used to model continuous symmetries. Many relevant Lie groups arise as subgroups of the matrix groups $\mathrm{GL}(n, \mathbb{R})$ and $\mathrm{GL}(n, \mathbb{C})$ of real and complex invertible matrices with matrix multiplications as group product.

The **Lie algebra** $\mathfrak{g}$ of a Lie group G is the tangent space of the group at the identity element $\mathfrak{g} := T_e G$. The Lie algebra $\mathfrak{g}$ can be identified with the space of (right) left invariant vector fields on G. In fact any vector $v \in \mathfrak{g}$ defines a left invariant vector field v^L and a right invariant vecotor field v^R in the following way:

$$v_a^L := \mathrm{d}L_a(v), \quad v_a^R := \mathrm{d}R_a(v), \quad \forall a \in G. \tag{27}$$

Where $L_a, R_a : G \to G$ are respectively the left and the right group multiplication. Conversely any left (right) invariant vector field $V \in \Gamma^\infty(G, TG)$ gives a Lie algebra element $V_e \in \mathfrak{g}$. With this identification we can define the Lie bracket in $\mathfrak{g}$ using the Lie bracket between the associated left invariant vector fields:

$$[v, w] := [v^L, w^L], \qquad \forall v, w \in \mathfrak{g}. \tag{28}$$

A fundamental property of Lie groups is that they are parallelizable manifolds. A basis $\{e_1, \cdots, e_n\} \subset T_e G$ defines a **global frame** $\{E_i\}_{i=1}^n$ for TG, where $E_i := e_i^L$, or $E_i := e_i^R$.

Any scalar product $\langle \cdot, \cdot \rangle_{\mathfrak{g}}$ on $\mathfrak{g}$ defines a **left invariant Riemannian metric** g on G:

$$g(v_a, w_a) := \langle \mathrm{d}L_{a^{-1}}(v_a), \mathrm{d}L_{a^{-1}}(w_a) \rangle_{\mathfrak{g}}, \quad \forall a \in G, \ \forall v_a, w_a \in T_a G. \tag{29}$$

This in turn induces a **left-invariant Riemannian volume form** μ_g, which is unique up to a normalizing constant (which depends on the initial scalar product choice). The associated left invariant Borel measure is known as (left) **Haar meausure**. A similar construction can be done to define a **right invariant volume form** on G. A Lie group is **unimodular** if its left and right invariant volume forms coincide. Examples of unimodular groups are *compact* Lie groups and *semisimple* Lie groups. For proofs, additional details and background we refer to Lee (2013) Chapers 7,16 and Falorsi et al. (2019) Appendix D.

Given $v \in \mathfrak{g}$ the exponential map is defined as $\exp(v) := \gamma(1)$ where $\gamma : \mathbb{R} \to G$ is the only 1-parameter subgroup such that $\dot{\gamma}(0) = v$. The exponential map $\exp : \mathfrak{g} \to G$ describes the flow of left and right invariant vector fields:

$$\phi_{v^L}^t(a) = a \exp(vt), \quad \phi_{v^R}^t(a) = \exp(vt) a, \quad \forall a \in G, \forall v \in \mathfrak{g}. \tag{30}$$

Notice that left invariant vector fields act by right translation and vice-versa. Therefore, given any right invariant vector field, $V \in \Gamma^\infty(G, TG)$ its flow $\Phi_V^t : G \to G$ is an **isometry** with respect to the left invariant Riemannian metric g. This means that V has 0 divergence, in fact:

$$\mathrm{div}(V)\mu_g = \mathcal{L}(\mu_g) = \frac{d}{dt}\bigg|_{t=0} \left(\left(\phi_V^t\right)^* \mu_g \right) = \frac{d}{dt}\bigg|_{t=0} \left(\left(L_{\exp(tV_e)}\right)^* \mu_g \right) = \frac{d}{dt}\big|_{t=0} (\mu_g) = 0 \cdot \mu_g, \tag{31}$$

which implies $\mathrm{div}(V) = 0$. We therefore we can obtain a **global frame** $\{E_i\}_{i=1}^n$ **formed by zero divergence vector fields**. We will see in Appendix C.3 that this greatly simplifies the divergence computation. When the Lie group is unimodular we can use left and right invariant vector fields interchangeably.

C.2.1. Special orthonormal matrices SO(n)

The group $SO(n) := \{U \in \mathrm{GL}(n, \mathbb{R}) : U^\top U = I, \ \det(U) = 1\}$ is the matrix group $n \times n$ orthonormal matrices with unit determinant. The lie algebra $\mathfrak{so}(n) := \{A \in \mathrm{M}(n, \mathbb{R}) : A^\top = -A, \ \mathrm{tr}(A) = 0\}$ is given by the traceless skew-symmetric matrices.
When $n = 3$ a basis $\{e_i\}_{i=1}^3$ for $\mathfrak{so}(3)$ is given by:

$$e_{1,2,3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{32}$$

C.2.2. SPECIAL UNITARY MATRICES SU(N)

The group $SU(n) := \{U \in \text{GL}(n, \mathbb{C}) : U^\dagger U = I,\ \det(U) = 1\}$ is the matrix group of complex $n \times n$ unitary matrices with unit determinant. The lie algebra $\mathfrak{su}(n) := \{A \in \text{M}(n, \mathbb{C}) : A^\dagger = -A,\ \text{tr}(A) = 0\}$ is given by the traceless skew-Hermitian matrices.
When $n = 2$ a basis $\{e_i\}_{i=1}^3$ for $\mathfrak{su}(2)$ is given by Pauli matrices:

$$e_{1,2,3} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}. \tag{33}$$

When $n = 3$ a basis $\{e_i\}_{i=1}^8$ for $\mathfrak{su}(3)$ is:

$$e_{1,2,3,4,5,6,7,8} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & i & 0 \end{bmatrix}, \tag{34}$$

$$\begin{bmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} -i & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} i/\sqrt{3} & 0 & 0 \\ 0 & i/\sqrt{3} & 0 \\ 0 & 0 & -2i/\sqrt{3} \end{bmatrix}. \tag{35}$$

C.3. Generators of vector fields

We have seen that for parallelizable manifolds, once we have defined a global frame, we have a bijective correspondence between functions $C^\infty(M, \mathbb{R}^n)$ and smooth vector fields:

$$C^\infty(M, \mathbb{R}^n) \to \Gamma^\infty(M, TM), \tag{36}$$

$$f \mapsto f_1 E_1 + \cdots + f_n E_n. \tag{37}$$

For non parallelizable manifolds, we fail to find a global frame because given any n vector fields $\{E_i\}_{i=1}^n$ there always exist points q where all $\{(E_i)_q\}_{i=1}^n$ fail to span all $T_q M$:

$$\text{span}\left(\{(E_i)_q\}_{i=1}^n\right) \subsetneq T_q M.$$

The idea is then to add vector fields to the set $\{(E_i)_q\}_{i=1}^n$, giving up on the injectivity, until they "generate" all $\Gamma^\infty(M, TM)$. To make this statement more precise we have to use the language of **modules**. In fact in general the space of smooth sections of a vector bundle (E, π, M) forms a module over the ring $C^\infty(M)$ of the smooth functions on M.

Definition 5. *A finite set of vector fields $\{X_i\}_{i=1}^m \subset \Gamma^\infty(M, TM), m \in \mathbb{N}_{>0}$ is a generator of the $C^\infty(M)$-module of the smooth vector fields on M if for every vector field $X \in \Gamma^\infty(M, TM)$ there exist $\{f_i\}_{i=1}^m \subset C^\infty(M)$ such that:*

$$X = f_1 X_1 + \cdots + f_m X_m. \tag{38}$$

If there exist a generator for for $\Gamma^\infty(M, TM)$ we then say that $\Gamma^\infty(M, TM)$ is finitely generated.

Lemma 1. *Let M be a smooth manifold and let $\{X_i\}_{i=1}^m \subset \Gamma^\infty(M, TM), m \in \mathbb{N}_{>0}$ a set of smooth vector fields such that span $\left(\{(X_i)_q\}_{i=1}^n\right) = T_q M,\ \forall q \in M$. Then $\{X_i\}_{i=1}^m$ is a generator for $\Gamma^\infty(M, TM)$.*

Proof. Consider the open sets $U_I := \{q \in M | \{X_i(q)\}_{i \in I}$ are linearly independent$\}$ where $I \subset \{1, \cdots, m\}$ is any subset of indices of cardinality n. let $\mathcal{I} := \{I | I \subset \{1, \cdots, m\},\ \#(I) = n,\ U_I \text{is not empty}\}$. To see that these sets are open, observe that in a local coordinate chart $(U, (x_i))$ we can write $X_{i_k}|_U = \sum_{j=1}^n a^I_{i,j} \partial_{x_j} \forall i \in 1, \cdots, M$. In local coordinates the linear independence of $\{X_i\}_{i \in I}$, is equivalent to $\det(A(x)) \neq 0$. Where if $I = i_1, \cdots, i_n$ $A(x)$ is defined as $A(x)_{jk} = a^I_{i_k, j}$. From the definition of U_I it descends that the family $\{(U_I | I \subset \{1, \cdots, m\},\ \#(I) = n,\ U_I \text{is not empty}\}$ forms a open trivialization cover. Fixed $I \in \mathcal{I}$ there exist smooth functions $\{f^I_i\}_{i \in I \subset C^\infty(U_I)}$ such that

$$X|_{U_i} = \sum_{i \in I} f^I_i E_i|_{U_I} \tag{39}$$

Now let $\{\psi_I\}_{i\in\mathcal{I}}$ be a smooth partition of unity subordinate to $\{U_I\}^n_{I\in\mathcal{I}}$. Defining

$$\grave{f}_i^I := \begin{cases} \psi_I \cdot f_i^I, & \text{on } U_I, \\ 0, & \text{on} M \setminus \text{supp}(\psi_I). \end{cases} \quad \forall I \in \mathcal{I},\ \forall i \in I. \tag{40}$$

We have that:

$$\psi_I \cdot X = \sum_{i\in I} \grave{f}_i^I X_i \tag{41}$$

And therefore

$$X = \sum_{j=1}^{m} \left(\sum_{I\in\mathcal{I}\ s.t.\ j\in I} \grave{f}_i^I \right) X_j \tag{42}$$

□

Theorem 3. *Let M be a (second countable) smooth manifold M. Then the module of smooth vector fields $\Gamma^\infty(M, TM)$ is finitely generated.*

Proof. Since M is second countable we can apply Lemma 7.1 in (Walschap, 2004) and say that there exist an open trivialization cover $\{U_i\}^n_{i=0}$, where n is the dimension of M. We denote with $E_1^{(i)}, \cdots, E_n^{(i)}$ the local frame relative to the domain $U_i \subseteq M$. Now let $\{\psi_i\}^n_{i=0}$ be a smooth partition of unity subordinate to $\{U_i\}^n_{i=0}$. We define the global vector fields on M:

$$\grave{E}_j^{(i)} := \begin{cases} \psi_i \cdot E_j^{(i)}, & \text{on } U_i, \\ 0, & \text{on } M \setminus \text{supp}(\psi_i). \end{cases} \quad \forall i \in \{0, \cdots, n\},\ \forall j \in \{1, \cdots, n\}. \tag{43}$$

Using Lemma 1 we have that $\{\grave{E}_j^{(i)}\}_{\substack{i=0..n \\ j=1..n}}$ is a generator of $\Gamma^\infty(M, TM)$.

□

From this Theorem and the definition of generator we can extract a methodology to parameterize all vector fields on smooth manifolds:

1. choose a suitable set of generators $\{X_i\}^m_{i=1}$,
2. find a way of parameterizing functions $f_i : M \to \mathbb{R}$,
3. model a generic vector field X as a linear combination:

$$X = f_1 X_1 + \cdots + f_m X_m. \tag{44}$$

The above proof also tells us that a simple and general recipe to obtain a generator is to take a collection of local frames and multiply them by a smooth partition of unity. The efficiency of this framework is given by the cardinality of the generator: lower cardinality requires parameterization of less functions. The proof gives us an initial upper bound on the lowest cardinality of the set of generators we can achieve for a generic manifold: $n^2 + n$ where n is the dimension of the manifold. We will see that using the Whitney embedding theorem, and the fact that any smooth manifold admits a Riemannian metric, this number can be further reduced to $2n + 1$.

However the cardinality of the generator is not the only factor to consider when choosing a good generating set. In fact combining equation (44) with the properties of Lie derivative we obtain:

$$\text{div}(X) = \sum_{i=1}^{m} X_i(f_i) + f_i \text{div}(X_i). \tag{45}$$

If we can find a set of generators with known divergence, or for which we can (pre-)compute the divergence, this greatly simplifies the divergence computation.

C.3.1. Time dependent vector fields

When parameterizing time dependent vector fields vector fields we have to model a vector field X_t for all $t \in \mathbb{R}$. Using generators we can easily accomplish this by parametrizing a function $f : \mathbb{R} \times M \to \mathbb{R}^m$ and defining

$$X_t := f_1(t, \cdot)X_1 + \cdots + f_m(t, \cdot)X_m. \tag{46}$$

C.4. Homogeneous spaces

Definition 6. *Let N, M smooth manifolds and $F : M \to N$ a smooth function between them. Given $X \in \Gamma^\infty(M, TM)$ and $Y \in \Gamma^\infty(N, TN)$ smooth vector fields respectively on M and N, we say that they are* **F-related** *if*

$$dF_p(X_p) = Y_{F(p)}, \quad \forall p \in M. \tag{47}$$

A **homogeneous space** is a manifold equipped with a transitive Lie group action.

Definition 7. *A smooth manifold M is homogeneous if a Lie group G acts transitively on M, i.e.:*

There exists a smooth map $G \times M \to M, (a, x) \mapsto a.x$

- $(ab).x = a.(b.x)$,
- $e.x = x, \quad \forall x \in M$,
- *for any $x, y \in M$ there exists an element $a \in G$ such that $a.x = y$.*

We can construct homogeneous spaces taking a Lie group G and $H < G$ a closed subgroup. Then the quotient manifold G/H is a homogeneous space with action $a.(bH) := abH$ and the projection map $\pi : G \to G/H$ is a smooth submersion (Theorem 21.17 Lee (2013)). In particular any Lie group $G = G/\{e\}$ can be considered a homogeneous space. The following theorem tells us that the above construction completely characterizes homogeneous spaces:

Theorem 4 (Theorem 21.18 Lee (2013))**.** *Let G be a Lie group, let M be a homogeneous G-space, and let q be any point of M. The isotropy group $G_q := \{a \in G : a.q = q\}$ is a closed subgroup of G, and the map:*

$$F : G/G_q \to M \tag{48}$$

$$aG_q \mapsto a.q \tag{49}$$

is an equivariant diffeomorphism.

Given $X \in \Gamma^\infty(G, TG)$ a right invariant vector field on G we can define a vector field $\widetilde{X} \in \Gamma^\infty(G/H, TG/H)$ on G/H:

$$\widetilde{X}_{aH} := \frac{d}{dt}\bigg|_{t=0} \exp(tX_e).aH, \quad \forall a \in G \tag{50}$$

Since $\exp(tX_e).aH = (exp(tX_e)a).H$ we have that $\phi^t_{\widetilde{X}} \circ \pi = \pi \circ \phi^t_X$, this is equivalent to say that $\widetilde{X}$ and X are π-related (Proposition 9.6 Lee (2013)). Since $\pi : G \to G/H$ is a smooth submersion and we know that there exist a generator of $\Gamma^\infty(G, TG)$ made by right-invariant vector fields, then there exist a set $\{X_i\}_{i=1}^m \subset \Gamma^\infty(G, TG)$ of right-invariant vector fields such that $\{\widetilde{X}_i\}_{i=1}^m \subset \Gamma^\infty(G/H, TG/H)$. is a generator for $\Gamma^\infty(G/H, TG/H)$. Let g ve an invariant [8] Riemannian metric on G/H associated with the invariant Riemannian volume form μ_g. Given V any right-invariant vector field on G, we have that the flow of $\widetilde{V}$ is an isometry and therefore $\text{div}\left(\widetilde{V}\right) = 0$:

$$\text{div}\left(\widetilde{V}\right)\mu_g = \mathcal{L}(\mu_g) = \frac{d}{dt}\bigg|_{t=0}\left(\left(\phi^t_{\widetilde{V}}\right)^*\mu_g\right) = \frac{d}{dt}\big|_{t=0}(\mu_g) = 0 \cdot \mu_g. \tag{51}$$

We thus have proven the following theorem:

[8] See sections 2.3.2 and 2.3.3 in Howard (1994) for construction of invariant volume forms and Riemannian metrics on homogeneous spaces

Theorem 5. *Let M be a homogeneous G-space. Then there exist $\{X_i\}_{i=1}^m \subset \Gamma^\infty(G, TG)$, $m \in \mathbb{N}_{>0}$ right invariant vector fields on G such that $\{\widetilde{X}_i\}_{i=1}^m \subset \Gamma^\infty(M, TM)$ is a generator for $\Gamma^\infty(M, TM)$. Let g a Riemannian metric invariant by group action, with respect to this metric:*

$$\text{div}\left(\widetilde{X}_i\right) = 0 \quad \forall i = 1 \dots m. \tag{52}$$

Let now $X = \sum_{i=1}^m f_i \widetilde{X}_i$ be an arbitrary vector field on M, with $f \in C^\infty(M, \mathbb{R}^m)$, we have that:

$$\text{div}(X) = \sum_{i=1}^m \widetilde{X}_i(f_i) = \sum_{i=1}^m df_i(\widetilde{X}_i). \tag{53}$$

C.5. Embedded submanifolds of $\mathbb{R}^m$

A general way to work in practice with manifolds is using embedded submanifolds of $\mathbb{R}^m$. An embedding for a manifold M is a continuous injective function $\iota : M \hookrightarrow \mathbb{R}^m$ such that $\iota : M \to \iota(M)$ is a homeomorphism. The embedding is smooth if ι is smooth and M is diffeomorphic to its image. In this case $\iota(M)$ is a smooth submanifold of $\mathbb{R}^m$. For all practical purposes we can directly identify M as a submanifold of $\mathbb{R}^m$, the function $\iota : M \hookrightarrow \mathbb{R}^m$ then simply denotes the inclusion. Through this identification we can then consider the tangent space T_qM, $q \in M$ as a vector subspace of $T_q\mathbb{R}^m$. An embedding is said **proper** if $\iota(M)$ is a closed set in $\mathbb{R}^m$.[9]

Theorem 6 (Whitney Embedding Theorem, 6.15 in (Lee, 2013))**.** *Every smooth n-dimensional manifold admits a proper smooth embedding in $\mathbb{R}^{2n+1}$*

The Whitney embedding theorem tells us that parameterizing manifolds as submanifolds of the Euclidean space gives us a general methodology to work with manifolds. Developing algorithms that assume that the manifold is given as an embedded submanifold of $\mathbb{R}^m$ is therefore of outstanding importance.

For embedded submanifolds parameterizing functions is extremely easy, and can be simply done via restriction: given a smooth function $f : \mathbb{R}^m \to \mathbb{R}$, $f \circ \iota$ then defines a smooth function from M to $\mathbb{R}$.

Unfortunately for vector fields it is not as easy as for functions, in fact in general given a vector field $X \in C(\mathbb{R}^m, T\mathbb{R}^m)$ this does not restrict in general to a vector field on a submanifold $M \subseteq \mathbb{R}^m$, as in general given $q \in M$ we have $X_q \notin T_qM \subseteq T\mathbb{R}^m_q$. In order for X to restrict to a vector field on a submanifold M we need for X to be **tangent to the submanifold** :

$$X_q \in T_qM \subseteq T_q\mathbb{R}^m, \quad \forall q \in M. \tag{54}$$

A tangent vector field then defines a vector field on the submanifold:

Lemma 2. *Let M be a smoothly embedded submanifold of $\mathbb{R}^m$, and let $\iota : M \hookrightarrow \mathbb{R}^m$ denote the inclusion map. If a smooth vector field $Y \in \Gamma^\infty(\mathbb{R}^m, T\mathbb{R}^m)$ is tangent to M there is a unique smooth vector field on M, denoted by $Y|_M$, that is ι-related to Y. Conversely a vector field $\overline{Y} \in \Gamma^\infty(\mathbb{R}^m, T\mathbb{R}^m)$ that is ι-related to Y is tangent to M*

Proof. See proof of Proposition 8.23 in (Lee, 2013) □

More importantly we can parameterize all vector fields on an embedded submanifold using tangent vector fields:

Prop 1. *Let M be a properly embedded submanifold of $\mathbb{R}^m$, and let $\iota : M \hookrightarrow \mathbb{R}^m$ denote the inclusion map. For any smooth vector field $X \in \Gamma^\infty(M, TM)$ there exist a smooth vector field $\overline{X} \in \Gamma^\infty(\mathbb{R}^m, T\mathbb{R}^m)$ tangent to M such that:*

$$\overline{X}|_M = X. \tag{55}$$

We call $\overline{X}$ an extension of X.

[9]Requiring that the embedding is proper excludes embeddings of the form $U \hookrightarrow M$ where U is an open subset of M

Proof. Let $U \subseteq \mathbb{R}^m$ be a tubular neighborhood of M, then by Proposition 6.25 of (Lee, 2013) there exist a smooth map $r : U \to M$ that is both a retraction and a smooth submersion. Then since r is a submersion there exist a vector field $\overline{X} \in \Gamma^\infty(U, TU)$ that is r-related to X [10]. This means $dr_z \overline{X}_z = X_{r(z)} \forall z \in \mathbb{R}^m$. Since r is a retraction

$$d\iota_q \circ dr_q = Id_{T_q\mathbb{R}^m} \;\Rightarrow\; (d\iota_q \circ dr_q)\overline{X}_q = \overline{X}_q \Rightarrow\; d\iota_q X_q = \overline{X}_q \;\forall q \in M, \tag{56}$$

$\overline{X}$ is ι-related to X and therefore tangent to M and such that $\overline{X}|_M = X$. Then $\overline{X}$ can be used to define a tangent vector field on all $\mathbb{R}^m$ using a smooth partition of unity subordinate to the open cover $\{\mathbb{R}^m \setminus M, U\}$. ☐

From the proof of the theorem it's clear that the extension of X is not unique. Our objective is then finding a way to parameterize all vector fields tangent to a submanifold. We first observe that given smooth vector fields $X, Y \in \Gamma^\infty(\mathbb{R}^m, T\mathbb{R}^m)$ tangent to M and smooth functions $f, g \in C^\infty(\mathbb{R}^m)$ then $fX + gY$ is tangent to M. This means that the set of all smooth vector fields tangent to M is a **submodule** of the module of smooth vector fields on $\mathbb{R}^m$. Following the framework outlined in Section C.1 we then need to find l tangent vector fields $\overline{X}_1, \cdots, \overline{X}_l$ such that $\overline{X}_1|_M, \cdots, \overline{X}_l|_M$ generates all $\Gamma^\infty(M, TM)$[11].

C.5.1. Embedded Riemannian Submanifolds

If our embedded submanifold manifold is equipped with a Riemannian metric, the gradient of the embedding gives us a set of generators for the tangent bundle. We first prove the following lemma

Lemma 3. *Let M be a embedded submanifold of N, and let $\iota : M \hookrightarrow N$ denote the inclusion map. Then*

$$\iota^* : \iota^* T^* N \to T^* M \tag{57}$$

$$\beta_{\iota(q)} \mapsto \iota^* \beta_{\iota(q)} : \quad v \mapsto \beta_{\iota(q)}(dv_q) \quad \forall q \in M,\; \forall \beta \in T^*_{\iota(q)}N,\; \forall v_q \in T_q M \tag{58}$$

is a surjective vector bundle homomorphism, where by $\iota^* T^* N$ we denote the pullback bundle $\iota^* T^* N = \{(q, \beta) \in M \times T^* N | \pi(\beta) = \iota(q)\}$

Proof. Fix $q \in M$.Let n be the dimensionality of M and m the dimensionality of N. We need to prove that $\iota^* : T^*_{\iota(q)}N \to T^*_q M$ is surjective. Let $e_1, \cdots, e_n$ a basis for $T_q M$ and $\eta_1, \cdots, \eta_n$ its dual basis. By the linearity of ι^* it's then sufficient to prove that there exists $\beta_1, \cdots, \beta_n \in T^*_{\iota(q)}N$ such that for all $i \in \{1, \cdots, n\}$:

$$\iota^* \beta_i = \eta_i. \tag{59}$$

To see this, consider the set $\{d(\eta_1)_q, \cdots, d(\eta_n)_q\} \subset T_{\iota(q)}N$. Since ι is an embedding, $d\iota$ is injective. Therefore the vectors are linearly independent. We can then complete them to a basis $v_1 := d(\eta_1)_q, \cdots, w_n := d(\eta_n)_q, w_{n+1}, \cdots w_m$ of $T_{\iota(q)}N$. Let $\beta_1, \cdots, \beta_m \in T^*_{\iota(q)}N$ the dual basis. We then have:

$$\iota^* \beta_i(e_j) = \beta_i\left(d\left(\eta_j\right)_q\right) = \beta_i\left(w_j\right) = \delta_{ij} = \eta_i\left(e_j\right) \quad \forall i, j \in \{1, \cdots, n\}. \tag{60}$$

Thus β_i satisfies Equation (59) $\forall i \in \{1, \cdots, n\}$ ☐

Theorem 7. *Let (M, g) be a embedded submanifold of $\mathbb{R}^m$, and let $z : M \hookrightarrow \mathbb{R}^m$ denote the inclusion map. Then $\{\nabla z_i\}_{i=1}^m$ is a set of generators for smooth vector fields $\Gamma^\infty(M, TM)$. Where ∇ denotes the Riemannian gradient with respect to the metric g.*

Proof. Consider the differential forms $\{dz_i\}_{i=1}^m \subset \Gamma^\infty(M, T^*M)$. Using Lemma 3 we have that $\text{span}\left(\{dz_i(q)\}_{i=1}^m\right) = T^*_q M$, which means that at every point they span the cotangent space at the point. Using the musical isomorphism, this implies that the riemannian gradients ∇z_i span the tangent space at every point: $\text{span}\left(\{\nabla z_i(q)\}_{i=1}^m\right) = T_q M$. Using Lemma 1 can conclude that $\{\nabla z_i\}$ is a generator for $\Gamma^\infty(M, TM)$. ☐

[10]See for example Exercise 8-18 of (Lee, 2013)

[11]Since the extension of a vector field is not unique this is different from finding a set of generators for the submodule of vector fields tangent to M

C.5.2. GENERATORS DEFINED BY GRADIENTS OF LAPLACIAN EIGENFUNCTIONS

In general, given a function $f \in C^\infty(M)$ on a Riemannian manifold, its Laplacian is defined as the divergence of its Riemannian gradient:

$$\Delta f := \text{div}(\nabla f). \tag{61}$$

Then the divergence of the fields defined in Theorem 3 is given by the Laplacian of the functions $z_i : M \to \mathbb{R}$.

A **laplacian eigenfunction** f is a smooth function such that

$$\Delta f = \lambda f \tag{62}$$

For some $\lambda \in \mathbb{R}$.

Since any closed Riemannian manifolds has an embedding formed by laplacian eigenfunctions [12], in this case, using Theorem 7 we can build a generator formed by the Riemannian gradient of laplacian eigenfunctions.

An example of particular importance is given by the hypersphere $S^n = \{x \in \mathbb{R}^{n+1} | \, \|x\| = 1\}$. The coordinate functions $z_i(x) = x_i$ form an embedding of laplacian eigenfunctions[13]. This gives us $n+1$ vector fields $\{\overline{\nabla z_i}\}_{i=1}^{n+1} \subset \Gamma^\infty\left(\mathbb{R}^{n+1}, T\mathbb{R}^{n+1}\right)$ tangent to S^n such that their restriction to M forms a generator for $\Gamma^\infty\left(S^n, TS^n\right)$.

$$\overline{\nabla z_i}(x) = e_i - \langle x, e_i \rangle x \quad \forall i \in \{1, \cdots, n+1\} \tag{63}$$
$$\Delta z_i(x) = -n x_i \tag{64}$$

C.5.3. ISOMETRICALLY EMBEDDED SUBMANIFOLDS

If the manifold M is isometrically embedded in $\mathbb{R}^n$. Then ∇z_i is simply given by the orthogonal projection of the constant coordinate field $e_i = \partial_{x_i}$ from $T\mathbb{R}^m$ to TM. In this case the Laplacian of the functions $z_i : M \to \mathbb{R}$ is given by the mean curvature (Chen & Verstraelen (2013), Proposition 2.3):

$$\Delta z = nH = \text{tr}\, \mathrm{I\!I}, \tag{65}$$

where H is the mean curvature and $\mathrm{I\!I}$ is the second fundamental form.

[12] See (Bates, 2014).

[13] Section 1.2 (Bates, 2014).