

# Riemannian continuous models

Vladimirov Eduard, group M05-304a

12 октября 2024 г.

## 1 Introduction

Continuous models like Neural ODEs and normalizing flows have revolutionized machine learning by modeling complex data dynamics over continuous time. However, these models traditionally assume that data resides in Euclidean space, limiting their applicability to datasets with inherent non-Euclidean geometries. Many real-world datasets naturally lie on *Riemannian manifolds* — curved spaces that require specialized mathematical frameworks to capture their intrinsic geometry.

Riemannian geometry extends calculus to curved spaces, providing tools like *tangent spaces* and the *exponential map* to navigate manifolds. Incorporating this geometry into continuous models allows for more accurate and efficient learning from manifold-valued data. This essay explores two significant advancements in this area:

1. **Riemannian Continuous Normalizing Flows:** These models generalize normalizing flows to manifolds by defining flows via manifold ODEs, enabling the modeling of complex probability distributions while respecting geometric constraints.
2. **Neural Manifold ODE:** Extending Neural ODEs to manifolds, these models utilize techniques like the *dynamic chart method* to solve ODEs intrinsically on manifolds.

By leveraging the intrinsic geometry of manifolds, these Riemannian continuous models enhance performance in fields where data resides on curved spaces, such as robotics and medical imaging.

## 2 Background on Riemannian Geometry and Continuous Models

### 2.1 Riemannian Geometry Fundamentals

Let  $\mathcal{M}$  be a smooth manifold of dimension  $n$ . A *smooth manifold* is a topological manifold equipped with an atlas of coordinate charts  $\{(U_\alpha, \varphi_\alpha)\}$ , where each  $U_\alpha$  is an open subset of  $\mathcal{M}$ , and  $\varphi_\alpha : U_\alpha \rightarrow \mathbb{R}^n$  is a homeomorphism onto its image, such that the transition maps  $\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(U_\alpha \cap U_\beta) \rightarrow \varphi_\beta(U_\alpha \cap U_\beta)$  are smooth whenever  $U_\alpha \cap U_\beta \neq \emptyset$ .

At each point  $p \in \mathcal{M}$ , the *tangent space*  $T_p\mathcal{M}$  is a real vector space consisting of the equivalence classes of curves passing through  $p$ , or alternatively, the set of derivations at  $p$ . Elements of  $T_p\mathcal{M}$  can be thought of as the possible directions in which one can tangentially pass through  $p$ .

A *Riemannian metric* on  $\mathcal{M}$  is a smooth assignment of an inner product  $g_p : T_p\mathcal{M} \times T_p\mathcal{M} \rightarrow \mathbb{R}$  for each  $p \in \mathcal{M}$ , such that for any smooth vector fields  $X$  and  $Y$  on  $\mathcal{M}$ , the function  $p \mapsto g_p(X_p, Y_p)$  is smooth. This metric allows for the measurement of angles, lengths, and volumes on the manifold.

Given a piecewise smooth curve  $\gamma : [a, b] \rightarrow \mathcal{M}$ , the *length* of  $\gamma$  is defined by

$$L(\gamma) = \int_a^b \|\dot{\gamma}(t)\|_{g_{\gamma(t)}} dt,$$

where  $\dot{\gamma}(t)$  denotes the derivative of  $\gamma$  at  $t$ , and  $\|\dot{\gamma}(t)\|_{g_{\gamma(t)}} = \sqrt{g_{\gamma(t)}(\dot{\gamma}(t), \dot{\gamma}(t))}$  is the norm induced by the Riemannian metric.

A *geodesic* on  $\mathcal{M}$  is a curve  $\gamma : [a, b] \rightarrow \mathcal{M}$  that locally minimizes the length functional or, equivalently, satisfies the geodesic equation:

$$\frac{D}{dt}\dot{\gamma}(t) = 0,$$

where  $\frac{D}{dt}$  is the covariant derivative along  $\gamma$ . Geodesics generalize the concept of straight lines to curved spaces and represent the paths of shortest distance between points on the manifold.

The *exponential map* at a point  $p \in \mathcal{M}$  is a map

$$\exp_p : T_p\mathcal{M} \rightarrow \mathcal{M},$$

defined by  $\exp_p(v) = \gamma_v(1)$ , where  $\gamma_v : [0, 1] \rightarrow \mathcal{M}$  is the unique geodesic starting at  $p$  with initial velocity  $v$ , i.e.,  $\gamma_v(0) = p$  and  $\dot{\gamma}_v(0) = v$ . The

exponential map provides a diffeomorphism from a neighborhood of the origin in  $T_p\mathcal{M}$  to a neighborhood of  $p$  in  $\mathcal{M}$ .

The exponential map is crucial in transferring linear structures from the tangent space to the manifold. It allows for operations such as adding a tangent vector to a point on the manifold, analogous to vector addition in Euclidean space. This property is essential for defining manifold-valued differential equations and for extending continuous models to Riemannian manifolds.

## 2.2 Continuous Models in Machine Learning

Continuous models in machine learning, such as Neural Ordinary Differential Equations (Neural ODEs) and normalizing flows, have gained prominence due to their ability to model complex, continuous-time dynamics and flexible probability distributions.

**Neural Ordinary Differential Equations (Neural ODEs).** Neural ODEs model the hidden states of a neural network as the solution to an ODE parameterized by a neural network function  $f_\theta$ :

$$\frac{dh(t)}{dt} = f_\theta(h(t), t), \quad h(0) = h_0.$$

This framework allows for adaptive computation and memory efficiency, as the continuous-depth model can be evaluated at arbitrary time points.

**Normalizing Flows.** Normalizing flows transform a simple base distribution  $\pi(z)$  into a complex target distribution  $p(x)$  by applying an invertible, differentiable mapping  $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ :

$$x = f(z), \quad z \sim \pi(z).$$

The change of variables formula computes the density of  $x$ :

$$p(x) = \pi(z) \left| \det \left( \frac{\partial f^{-1}}{\partial x} \right) \right| = \pi(z) \left| \det \left( \frac{\partial f(z)}{\partial z} \right)^{-1} \right|.$$

Taking the logarithm, the log-likelihood of  $x$  is:

$$\log p(x) = \log \pi(z) - \log \left| \det \left( \frac{\partial f(z)}{\partial z} \right) \right|,$$

where  $z = f^{-1}(x)$ .

Computing the determinant of the Jacobian  $\frac{\partial f(z)}{\partial z}$  can be challenging, especially in high dimensions, leading to the development of continuous normalizing flows (CNFs), which use Neural ODEs to define  $f$ .

### 3 Riemannian Continuous Normalizing Flows

Normalizing flows transform a simple base distribution  $\pi(z)$  into a complex target distribution  $p(x)$  using an invertible, differentiable mapping  $f : \mathcal{M} \rightarrow \mathcal{M}$  on a manifold  $\mathcal{M}$ . To generalize normalizing flows to Riemannian manifolds, flows are defined via time-dependent vector fields and the solutions to manifold ordinary differential equations (ODEs).

Let  $\mathcal{M}$  be a  $d$ -dimensional Riemannian manifold with metric tensor  $G(z)$ . Consider a time-dependent vector field  $f_\theta : \mathcal{M} \times [0, T] \rightarrow T\mathcal{M}$ , parameterized by  $\theta$ , where  $T\mathcal{M}$  denotes the tangent bundle of  $\mathcal{M}$ . The flow  $\varphi_t : \mathcal{M} \rightarrow \mathcal{M}$  induced by  $f_\theta$  is obtained by solving the manifold ODE:

$$\frac{dz(t)}{dt} = f_\theta(z(t), t), \quad z(0) = z_0 \in \mathcal{M}, \quad (1)$$

where  $z(t) \in \mathcal{M}$  for all  $t \in [0, T]$ . Under appropriate conditions on  $f_\theta$ , the flow  $\varphi_t$  is a diffeomorphism at each time  $t$ .

The change in log-density along the flow is given by the continuity equation on the manifold. The derivative of the log-density  $p_t(z(t))$  with respect to time is:

$$\frac{d}{dt} \log p_t(z(t)) = -\operatorname{div}(f_\theta)(z(t), t), \quad (2)$$

where  $\operatorname{div}(f_\theta)$  is the Riemannian divergence of the vector field  $f_\theta$ .

Integrating Equation (2) over time from  $t = 0$  to  $t = T$ , the log-density at time  $T$  is:

$$\log p_T(z(T)) = \log p_0(z(0)) - \int_0^T \operatorname{div}(f_\theta)(z(t), t) dt. \quad (3)$$

The Riemannian divergence of the vector field  $f_\theta$  at point  $z \in \mathcal{M}$  is defined as:

$$\operatorname{div}(f_\theta)(z) = \frac{1}{\sqrt{\det G(z)}} \frac{\partial}{\partial z^i} \left( \sqrt{\det G(z)} f_\theta^i(z) \right), \quad (4)$$

where  $\det G(z)$  is the determinant of the metric tensor  $G(z)$ ,  $f_\theta^i(z)$  are the components of  $f_\theta$  in local coordinates, and the Einstein summation convention is used over repeated indices.

Computing the divergence directly requires evaluating the divergence of the vector field in the manifold’s local coordinates, which involves computing derivatives of  $f_\theta$  and the metric tensor. To facilitate computation, the divergence can be expressed in terms of the trace of the covariant derivative:

$$\operatorname{div}(f_\theta)(z) = \operatorname{Tr}(\nabla f_\theta(z)), \quad (5)$$

where  $\nabla f_\theta$  denotes the covariant derivative of  $f_\theta$ .

In practice, to avoid computing the full Jacobian matrix, stochastic trace estimation methods such as the Hutchinson estimator can be employed:

$$\operatorname{div}(f_\theta)(z) = \mathbb{E}_{\varepsilon \sim \mathcal{N}(0, I)} [\varepsilon^\top (\nabla f_\theta(z) \varepsilon)], \quad (6)$$

where  $\varepsilon$  is a random vector sampled from a standard normal distribution, and  $I$  is the identity matrix.

To ensure that the vector field  $f_\theta(z, t)$  lies in the tangent space  $T_z \mathcal{M}$  at each point  $z$ , it can be defined using a projection:

$$f_\theta(z, t) = P_z \tilde{f}_\theta(z, t), \quad (7)$$

where  $\tilde{f}_\theta$  is an unconstrained vector field in the ambient space, and  $P_z$  is the orthogonal projection onto  $T_z \mathcal{M}$ .

Alternatively,  $f_\theta$  can be directly parameterized using operations intrinsic to the manifold. For instance, in embedded manifolds, the vector field can be constructed to automatically satisfy the tangent space constraint.

Solving the manifold ODE (1) requires numerical integration methods that preserve the manifold structure. One approach is to use the exponential map:

$$z(t + \Delta t) = \exp_{z(t)}(\Delta t f_\theta(z(t), t)), \quad (8)$$

where  $\exp_z$  is the exponential map at point  $z$ .

The total log-likelihood of a sample  $x \in \mathcal{M}$  is given by:

$$\log p(x) = \log \pi(z(0)) - \int_0^T \operatorname{div}(f_\theta)(z(t), t) dt, \quad (9)$$

with  $z(T) = x$ .

In summary, Riemannian Continuous Normalizing Flows extend normalizing flows to Riemannian manifolds by defining continuous-time flows via manifold

ODEs. The key components are the vector field  $f_\theta$  constrained to the tangent space, the computation of the Riemannian divergence for the change in log-density, and the use of manifold-aware numerical integration methods to solve the ODE.

## 4 Neural Manifold ODE

Let  $\mathcal{M}$  be a smooth manifold, and consider a time-dependent vector field  $f_\theta : \mathcal{M} \times [0, T] \rightarrow T\mathcal{M}$ , parameterized by  $\theta$ , where  $T\mathcal{M}$  denotes the tangent bundle of  $\mathcal{M}$ .

The *Neural Manifold Ordinary Differential Equation* is defined as:

$$\frac{dz(t)}{dt} = f_\theta(z(t), t), \quad z(0) = z_0 \in \mathcal{M}, \quad (10)$$

where  $z(t) \in \mathcal{M}$  for all  $t \in [0, T]$ , and  $f_\theta(z(t), t) \in T_{z(t)}\mathcal{M}$ .

To solve the ODE (10), we utilize local coordinate charts. Let  $\varphi : U \subset \mathbb{R}^n \rightarrow V \subset \mathcal{M}$  be a coordinate chart such that  $z(t) \in V$ . The inverse chart maps  $z(t)$  to local coordinates:

$$y(t) = \varphi^{-1}(z(t)) \in U. \quad (11)$$

In local coordinates, the ODE (10) becomes:

$$\frac{dy(t)}{dt} = D\varphi_{z(t)}^{-1}(f_\theta(z(t), t)), \quad (12)$$

where  $D\varphi_{z(t)}^{-1} : T_{z(t)}\mathcal{M} \rightarrow \mathbb{R}^n$  is the derivative of  $\varphi^{-1}$  at  $z(t)$ .

By solving (12) in  $\mathbb{R}^n$ , we obtain  $y(t)$ , and map back to the manifold via  $z(t) = \varphi(y(t))$ .

To ensure  $z(t)$  remains within the domain of the chart, we may need to update the chart during integration. This leads to the *dynamic chart method*, where a sequence of charts  $\{\varphi_i\}$  is used to cover the trajectory of  $z(t)$ :

If  $z(t) \notin V_i$ , select a new chart  $\varphi_{i+1} : U_{i+1} \rightarrow V_{i+1}$  such that  $z(t) \in V_{i+1}$ . (13)

For gradient computation with respect to parameters  $\theta$ , we employ the adjoint sensitivity method. Define the adjoint state  $a(t) \in T_{z(t)}^*\mathcal{M}$ , which satisfies:

$$\frac{da(t)}{dt} = -D_{z(t)}f_\theta(z(t), t)^\top a(t), \quad (14)$$

with the terminal condition  $a(T) = \nabla_{z(T)} L$ , where  $L$  is a scalar loss function dependent on  $z(T)$ .

The gradient of the loss with respect to  $\theta$  is given by:

$$\frac{dL}{d\theta} = - \int_0^T a(t)^\top \frac{\partial f_\theta(z(t), t)}{\partial \theta} dt. \quad (15)$$

To ensure that the vector field  $f_\theta(z(t), t)$  lies in the tangent space  $T_{z(t)}\mathcal{M}$  at each point  $z(t)$ , we can define  $f_\theta$  using a projection:

$$f_\theta(z(t), t) = P_{z(t)} \tilde{f}_\theta(z(t), t), \quad (16)$$

where  $\tilde{f}_\theta : \mathcal{M} \times [0, T] \rightarrow \mathbb{R}^d$  is an unconstrained vector field in the ambient space, and  $P_{z(t)} : \mathbb{R}^d \rightarrow T_{z(t)}\mathcal{M}$  is the projection onto the tangent space at  $z(t)$ .

Alternatively,  $f_\theta$  can be parameterized intrinsically using manifold-specific operations to ensure it remains in  $T_{z(t)}\mathcal{M}$  without the need for projection.

When integrating the ODE (10), the numerical solver must account for the manifold structure. One approach is to use the exponential map:

$$z(t + \Delta t) = \exp_{z(t)}(\Delta t f_\theta(z(t), t)), \quad (17)$$

where  $\exp_{z(t)} : T_{z(t)}\mathcal{M} \rightarrow \mathcal{M}$  maps a tangent vector to the manifold along a geodesic starting at  $z(t)$ .

In practice, solving the ODE in local coordinates as in (12) can leverage standard ODE solvers in Euclidean space, while the dynamic chart method ensures the solution remains consistent with the manifold’s topology.

Integrating Neural Manifold ODEs allows for modeling continuous-time dynamics on manifolds, generalizing Neural ODEs to accommodate data with intrinsic geometric structures.

## 5 Conclusion

Riemannian continuous models extend traditional continuous frameworks, such as Neural Ordinary Differential Equations and normalizing flows, to data residing on Riemannian manifolds. By leveraging the intrinsic geometry of manifolds, these models provide a principled approach to modeling manifold-valued data

## Список литературы

- Alimisis, Foivos и др. (2020). «A continuous-time perspective for modeling acceleration in Riemannian optimization». В: *International Conference on Artificial Intelligence and Statistics*. PMLR, с. 1297—1307.
- Huang, Chin-Wei и др. (2022). «Riemannian diffusion models». В: *Advances in Neural Information Processing Systems* 35, с. 2750—2761.
- Jeong, Seungwoo и др. (2023). «Deep efficient continuous manifold learning for time series modeling». В: *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Katsman, Isay и др. (2024). «Riemannian residual neural networks». В: *Advances in Neural Information Processing Systems* 36.
- Lou, Aaron и др. (2020). «Neural manifold ordinary differential equations». В: *Advances in Neural Information Processing Systems* 33, с. 17548—17558.
- Mathieu, Emile и Maximilian Nickel (2020). «Riemannian continuous normalizing flows». В: *Advances in Neural Information Processing Systems* 33, с. 2503—2515.
- Papaioannou, Panagiotis G и др. (2022). «Time-series forecasting using manifold learning, radial basis function interpolation, and geometric harmonics». В: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 32.8.