Unraveling the Single Tangent Space Fallacy: An Analysis and Clarification for Applying Riemannian Geometry in Robot Learning

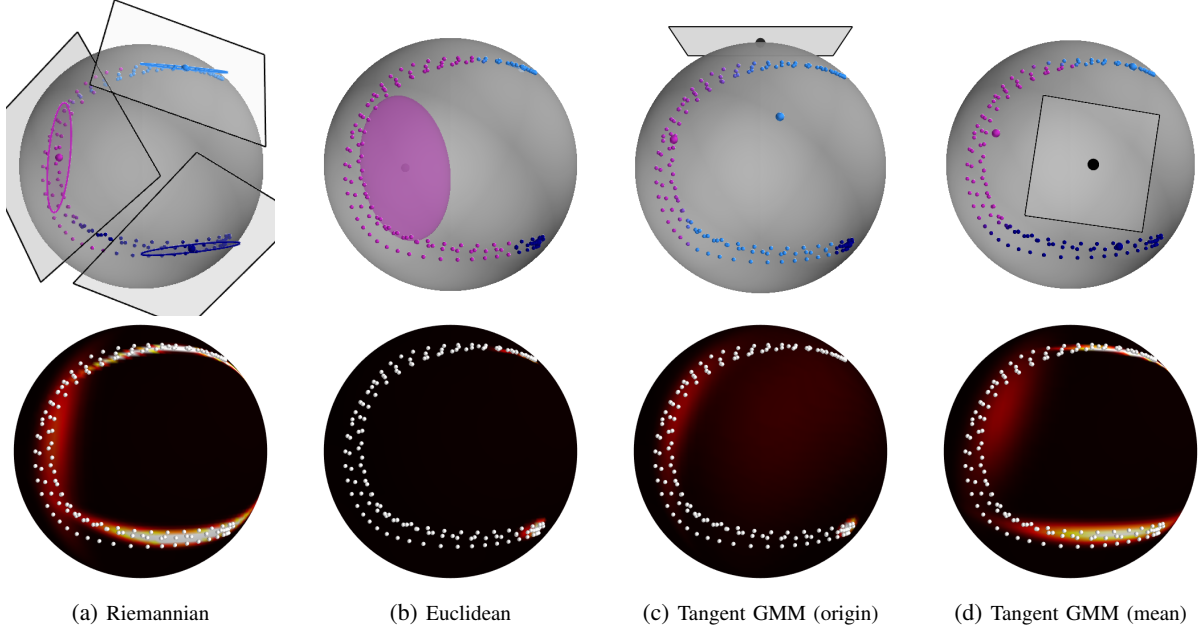Noémie Jaquier*[1], Leonel Rozo*[2], and Tamim Asfour[1]

(a) Riemannian	(b) Euclidean	(c) Tangent GMM (origin)	(d) Tangent GMM (mean)

Fig. 1: Illustration on the sphere $\mathcal{S}^2$ of the importance of geometry in estimating GMM parameters and densities on Riemannian manifolds. *Top row:* Estimated GMM parameters. The colors of the datapoints indicate their (shared) assignments to different GMM components. Two configurations of the single tangent space used by Tangent GMM are displayed in (c)-(d). *Bottom row:* Estimated density on the sphere. Colors ranging from black to yellow indicate low to high probability density $p(\boldsymbol{x})$. Datapoints are depicted as white dots.

Abstract—In the realm of robotics, numerous downstream robotics tasks leverage machine learning methods for processing, modeling, or synthesizing data. Often, this data comprises variables that inherently carry geometric constraints, such as the unit-norm condition of quaternions representing rigid-body orientations or the positive definiteness of stiffness and manipulability ellipsoids. Handling such geometric constraints effectively requires the incorporation of tools from differential geometry into the formulation of machine learning methods. In this context, Riemannian manifolds emerge as a powerful mathematical framework to handle such geometric constraints. Nevertheless, their recent adoption in robot learning has been largely characterized by a mathematically-flawed simplification, hereinafter referred to as the "single tangent space fallacy". This approach involves merely projecting the data of interest onto a single tangent (Euclidean) space, over which an off-the-shelf learning algorithm is applied. This paper provides a theoretical elucidation of various misconceptions surrounding this approach and offers experimental evidence of its shortcomings. Finally, it presents valuable insights to promote best practices when employing Riemannian geometry within robot learning applications.

I. INTRODUCTION

The recent integration of machine learning techniques in robotics, spanning control, perception, and reasoning domains, has substantially enhanced robots' capabilities across diverse environments. Nevertheless, a notable oversight in many of these approaches pertains to the intrinsic non-Euclidean geometry often inherent in robotic data. For instance, three-dimensional rotations can be viewed as elements of the special orthogonal group $\mathrm{SO}(3)$ or the sphere $\mathcal{S}^3$, control gains, inertia, and manipulability ellipsoids lie in the space of symmetric-positive-definite (SPD) matrices $\mathcal{S}^d_{++}$, the joint configuration of a d-degree-of-freedom robot with revolute joints can be seen as a point on a torus $\mathbb{T}^d$, while tree-like data can naturally be embedded in a hyperbolic space $\mathbb{H}^d$. Under these conditions, Riemannian manifolds emerge as powerful mathematical structures used to model such curved spaces, allowing for a more accurate understanding of complex phenomena in various robot learning applications. To perform computations on these manifolds, tangent spaces play a crucial role, as they provide local linear approximations of the manifold's geometry, enabling calculations of derivatives and vector operations. However, the conventional approach of using a single tangent space for the entire manifold presents significant limitations that can

lead to erroneous results and misinterpretations.

As proposed in several works [1]–[4], the single tangent space approximation is carried out as follows: First, the Riemannian data or variable of interest is projected to a single tangent space; later, an off-the-shelf Euclidean learning algorithm models or processes the projected data; and finally the results are projected back on the Riemannian manifold. This simplification, though common, is fundamentally flawed and overlooks essential geometric properties of Riemannian manifolds. Although few works have recently discussed the drawbacks of employing naive Euclidean approximations in different Riemannian settings [5]–[7], a more theoretical and experimental analysis is still lacking in the current literature. This paper is aimed at filling this gap by delving into the fallacies of employing a single tangent space for computations on Riemannian manifolds, and by providing experimental evidence of the inability of such an approach to capture the manifold's intrinsic curvature and its global structure, which can severely affect robot learning applications.

With the goal of rising awareness about the correct use of Riemannian methods in robot learning applications, **the main contributions of this paper are**: *(1)* we present a simple yet elucidating example to introduce the concept and effects of the single tangent space fallacy; *(2)* we point out and explain five common misconceptions of the single tangent space approach by building on several concepts of Riemannian geometry; and *(3)* we experimentally illustrate the shortcomings of using a single tangent space in two common robot learning settings, namely, data density estimation and first-order dynamical systems learning. Our experimental findings show that proper Riemannian formulations of a robot learning problem clearly outperform Euclidean and single-tangent space approximations.

II. BACKGROUND

A. Riemannian Manifolds

A smooth manifold $\mathcal{M}$ can be intuitively conceptualized as a set of points that locally, but not globally, resemble the Euclidean space $\mathbb{R}^d$ [8], [9]. An abstract definition of a manifold encompasses the topological, differential, and geometric structure via the so-called *charts*: maps between parts of $\mathcal{M}$ to $\mathbb{R}^d$. The collection of these charts — seen as local parameterizations — is called *atlas*. A chart on a smooth manifold $\mathcal{M}$ is a diffeomorphic mapping $\varphi : \mathcal{U} \rightarrow \tilde{\mathcal{U}}$ from an open set $\mathcal{U} \subset \mathcal{M}$ to an open set $\tilde{\mathcal{U}} \subseteq \mathbb{R}^d$ (see Fig. 2-*left*). The transition map between two intersecting sets $\mathcal{U}_1$ and $\mathcal{U}_2$, given by $\varphi_1 \circ \varphi_2^{-1}$ or $\varphi_2 \circ \varphi_1^{-1} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is also a diffeomorphism and guarantees consistency where the two charts overlap. The smooth structure of $\mathcal{M}$ makes it possible to take derivatives of curves on the manifold, leading to tangent vectors in $\mathbb{R}^d$. The set of tangent vectors of all curves at $\boldsymbol{x} \in \mathcal{M}$ forms a d-dimensional affine subspace of $\mathbb{R}^d$, known as the *tangent space* $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ of $\mathcal{M}$ at $\boldsymbol{x}$. The collection of all such tangent spaces is called the *tangent bundle* $\mathcal{TM}$. Finally, note that the local coordinates on $\mathcal{U}$ induced by the chart φ define a basis for $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ (see Fig. 2).
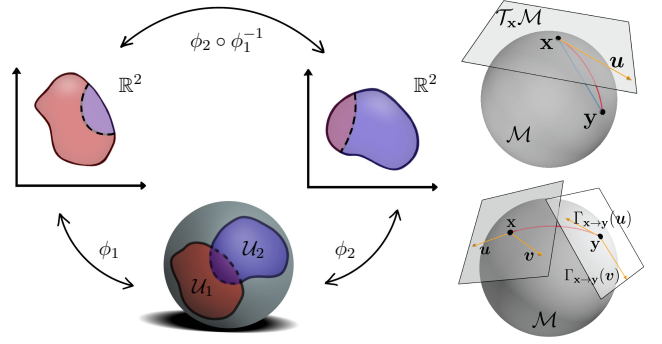

Fig. 2: Transition function between two charts on the sphere *(left)*, and basic Riemannian operations *(right)* on the sphere $\mathcal{S}^2$. The vector $\boldsymbol{u}$ lies on the tangent space of $\boldsymbol{x}$ such that $\boldsymbol{u} = \text{Log}_{\boldsymbol{x}}(\mathbf{y})$. $\Gamma_{\mathbf{x} \rightarrow \mathbf{y}}(\boldsymbol{u})$ and $\Gamma_{\mathbf{x} \rightarrow \mathbf{y}}(\boldsymbol{v})$ are parallel-transported vectors.

A *Riemannian manifold* $(\mathcal{M}, g)$ is a smooth manifold endowed with a Riemannian metric g, that is a family of inner products $g_{\boldsymbol{x}} : \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \times \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \rightarrow \mathbb{R}$ associated to each point $\boldsymbol{x} \in \mathcal{M}$ [9]. To operate with Riemannian manifolds, it is common practice to exploit their Euclidean tangent spaces and resort to mappings back and forth between $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ and $\mathcal{M}$. Specifically, the exponential map $\text{Exp}_{\boldsymbol{x}}(\boldsymbol{u}) : \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \rightarrow \mathcal{M}$ maps a point $\boldsymbol{u} \in \mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ to a point $\boldsymbol{y}$ on the manifold, so that it lies on the geodesic starting at $\boldsymbol{x}$ in the direction $\boldsymbol{u}$ and such that the geodesic distance $d_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{y}) = d_{\mathbb{R}, g_{\boldsymbol{x}}}(\boldsymbol{x}, \boldsymbol{u})$ (see Fig. 2-*right*). The inverse operation is the logarithmic map $\text{Log}_{\boldsymbol{x}}(\boldsymbol{y}) : \mathcal{M} \rightarrow \mathcal{T}_{\boldsymbol{x}}\mathcal{M}$. Finally, the parallel transport $\Gamma_{\boldsymbol{x} \rightarrow \boldsymbol{y}}(\boldsymbol{u}) : \mathcal{T}_{\boldsymbol{x}}\mathcal{M} \rightarrow \mathcal{T}_{\boldsymbol{y}}\mathcal{M}$ describes how elements of $\mathcal{M}$ can be transported along curves on $\mathcal{M}$ while maintaining their intrinsic geometric properties. This allows us to operate elements lying on different tangent spaces (see Fig. 2-*right*).

B. Riemannian Gaussian Distributions

A Riemannian Gaussian distribution (RGD) $\mathcal{N}_{\mathcal{M}}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$, depends on two parameters, the mean $\boldsymbol{\mu} \in \mathcal{M}$ and covariance $\boldsymbol{\Sigma} \in \mathcal{T}_{\boldsymbol{\mu}}\mathcal{M}$. Its probability density function, generalising that of a Gaussian distribution on the Euclidean space $\mathbb{R}^d$, can be locally approximated by[1],

$$\mathcal{N}_{\mathcal{M}}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^d|\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}\text{Log}_{\boldsymbol{\mu}}(\boldsymbol{x})\boldsymbol{\Sigma}^{-1}\text{Log}_{\boldsymbol{\mu}}(\boldsymbol{x})^{\top}}. \quad (1)$$

This RGD corresponds to an approximated maximum-entropy distribution for Riemannian manifolds, as introduced in [10]. The parameters of the RGD can be estimated via maximum likelihood estimation (MLE). The MLE of $\boldsymbol{\mu}$ corresponds to the Fréchet mean $\boldsymbol{\mu} \in \mathcal{M}$ [11] given by,

$$\boldsymbol{\mu} = \underset{\boldsymbol{\mu} \in \mathcal{M}}{\text{argmin}} \frac{1}{N} \sum_{n=1}^{N} d_{\mathcal{M}}(\boldsymbol{\mu}, \boldsymbol{x}_n)^2. \quad (2)$$

The covariance $\boldsymbol{\Sigma}$ is estimated in $\mathcal{T}_{\boldsymbol{\mu}}\mathcal{M}$ as,

$$\boldsymbol{\Sigma} = \frac{1}{N-1} \sum_{n=1}^{N} \text{Log}_{\boldsymbol{\mu}}(\boldsymbol{x}_n) \text{Log}_{\boldsymbol{\mu}}(\boldsymbol{x}_n)^{\top}. \quad (3)$$

---

[1]Under the assumption of having low-variance data.

III. LEARNING RIEMANNIAN DATA DISTRIBUTIONS: A MOTIVATIONAL EXAMPLE

We begin our analysis by studying a simple density estimation problem on the sphere $\mathcal{S}^2$. Modeling complex, multimodal probability distributions on smooth manifolds is at the heart of various robotics tasks ranging from state estimation [12] to motion planning [13] and learning from demonstrations [6], [14]. We aim at estimating the underlying density of C-shaped data lying on the sphere $\mathcal{S}^2$ with a simple, yet popular probabilistic model, i.e., a Gaussian mixture model (GMM) with $K = 3$ components. As our data is not Euclidean, several GMM variants can be considered for estimating their density. All are initialized identically.

A. Euclidean GMM

As $\mathcal{S}^2 \subset \mathbb{R}^3$, a first straightforward approach is to apply a traditional Euclidean GMM acting on the embedding space $\mathbb{R}^3$. The corresponding density is of the classical form,

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \qquad (4)$$

where the mean $\boldsymbol{\mu}_k$, covariance $\boldsymbol{\Sigma}_k$, and prior π_k of each Gaussian component are estimated via MLE. Fig. 1b displays the estimated model and resulting density on $\mathcal{S}^2$. We observe that the GMM components are located inside the sphere, as they are not constrained to lie on its surface. By acting on $\mathbb{R}^3$, the model fully disregards the intrinsic data geometry. As a result, the estimated density is mostly concentrated inside the sphere and fails to represent the observed data.

B. Euclidean GMM on a Single Tangent Space

Next, we apply a second naive approach, referred to as Tangent GMM, which consists in projecting all datapoints in a single tangent space using the logarithmic map, fitting a Euclidean GMM in this tangent space, and projecting the resulting probability density back onto the manifold using the exponential map. This approach might seem appealing as it uses an off-the-shelf Euclidean GMM, and the resulting density lies on the sphere surface. However, it displays various shortcomings, as shown next. Fig. 1c shows the Tangent GMM estimated via the tangent space at the origin. Despite the post-projection to $\mathcal{S}^2$, the model poorly fits the data: The mean of the light-blue component is located far away from the data and the density inside the C is non-zero, although this region is out of the data support. This is due to the severe distortions induced by the projection onto the Euclidean tangent space As shown in Fig. 3-*left*, the C-shape horizontally elongates and its upper part is reversely bent.

Importantly, such distortions depend on the choice of tangent space. Fig. 1d shows the Tangent GMM applied at the Fréchet mean of the data, i.e., via the tangent space resulting in minimal distortions (see Fig. 3-*right*). However, the model still fails at entirely capturing the underlying data distribution: It estimates a close-to-zero density at the bottom-left part of the C, and a high density around the mean of the pink component, located out of the data support.
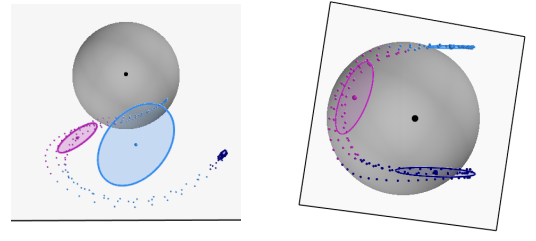

Fig. 3: Datapoints and GMM of Figs. 1c, 1d in the single tangent spaces at the origin with a top view *(left)* and Fréchet mean *(right)*.

C. Riemannian GMM

Last, we examine a Riemannian GMM [15], whose probability density is based on multiple RGDs (1), so that,

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}_{\mathcal{M}}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \qquad (5)$$

In other words, each mixture component is centered at a point $\boldsymbol{\mu}_k \in \mathcal{M}$, while its directional dispersion is characterized by a covariance $\boldsymbol{\Sigma}_k \in \mathcal{T}_{\boldsymbol{\mu}_k}\mathcal{M}$. The GMM parameters are estimated via MLE with a Riemannian Expectation-Maximization algorithm [15]. Fig. 1a depicts the resulting Riemannian GMM. The learned model is consistent with the data geometry, and the estimated density matches the underlying data distribution, achieving the highest likelihood among the considered GMMs (see C column in Table I). By simultaneously leveraging multiple tangent spaces, the Riemannian GMM limits the distortions of individual tangent spaces, while guaranteeing that the estimated probability density exclusively lies on the manifold. As explained next, considering multiple tangent spaces is key to the design of high-performance geometric learning models.

IV. THE SINGLE TANGENT SPACE FALLACY

The example of Section III illustrated some of the limitations of learning on a single tangent space. The flaws of this approach were evident even on a relatively simple learning setting on a low-dimensional sphere, i.e., the simplest Riemannian manifold to operate on after $\mathbb{R}^d$. Here, we delve into the mathematical aspects that explain why most single-tangent-space-based approaches are fundamentally flawed, and highlight the common misunderstandings leading to their usage. Section V then illustrates their flaws experimentally.

A. A Deep Dive into Mathematics of Tangent Spaces

To understand the single tangent space fallacy, we first formally review the notions of tangent vectors and spaces. Here, we define tangent vectors via an equivalence class of curves as in [16, Chap. 8], [17, Chap. 3]. We consider the set of all smooth curves $\boldsymbol{\gamma} : I \rightarrow \mathcal{M}$, passing through the point $\boldsymbol{x} \in \mathcal{M}$ such that $\boldsymbol{\gamma}(0) = \boldsymbol{x}$, with I an interval in $\mathbb{R}$. Given a chart $(\mathcal{U}, \varphi)$ at $\boldsymbol{x}$, two curves $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ are equivalent at $\boldsymbol{x}$ if and only if their derivatives equal at $\boldsymbol{x}$, i.e.,

$$\boldsymbol{\gamma}_1 \sim \boldsymbol{\gamma}_2 \iff (\varphi \circ \boldsymbol{\gamma}_1)'(0) = (\varphi \circ \boldsymbol{\gamma}_2)'(0). \qquad (6)$$

The equivalence class of such curves, denoted $[\boldsymbol{\gamma}] = \boldsymbol{\gamma}'(0)$, defines a *tangent vector*. The *tangent space* $\mathcal{T}_{\boldsymbol{x}}\mathcal{M}$ is then the set $\{[\boldsymbol{\gamma}]\}$ of all tangent vectors at $\boldsymbol{x}$. The foregoing definitions
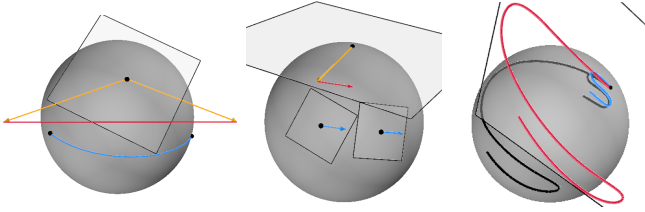
Fig. 4: Illustration of the fallacies 2, 3, and 5 (in red). *Left:* the blue and red curves depict the geodesic and Euclidean distances, respectively. *Middle:* the blue arrows represent two velocities w.r.t two different tangent spaces, while the red arrow depicts a linear velocity on the single tangent space. *Right:* the black curves display Riemannian trajectories, while the blue and red curves show the corresponding distorted projections on a single tangent space.

lead to crucial insights regarding the single tangent space fallacy. First, tangent vectors are *attached* to a specific point x. Second, a *linear structure* is naturally induced on $\mathcal{T}_x\mathcal{M}$ via the bijective map $d\varphi_x : \mathcal{T}_x\mathcal{M} \to \mathbb{R}^d : [\gamma] \mapsto (\varphi \circ \gamma)'(0)$, as it satisfies, independently of the choice of chart,

$$a[\gamma_1] + b[\gamma_2] = d\varphi_x^{-1}\left(a\, d\varphi_x([\gamma_1]) + b\, d\varphi_x([\gamma_2])\right). \quad (7)$$

In other words, two tangent vectors can be added into a third, and a real multiple of a tangent vector remains a tangent vector. Third, each tangent space $\mathcal{T}_x\mathcal{M}$ is a *local linear approximation* of $\mathcal{M}$ at x. As such, tangent spaces do not globally capture the intrinsic manifold geometry.

B. Fallacies of Single Tangent Space Approaches

Single-tangent-space-based approaches might appear appealing as *(1)* the linear structure of tangent spaces complies with traditional Euclidean methods, and *(2)* the results obtained on a tangent space can be mapped onto the manifold via the exponential maps. However, such approaches are fundamentally flawed. Next, we explain five core misconceptions, which relate to the single tangent space fallacy.

As mentioned earlier, exponential and logarithmic maps are key operations to leverage the linear structure of tangent spaces. However, two key aspects of these mappings are often overlooked when operating on a single tangent space.

Fallacy 1. *The exponential map is a local diffeomorphism.*

Namely, it is a diffeomorphism between a neighborhood $\mathcal{B}_r(\mathbf{0}) \subseteq \mathcal{T}_x\mathcal{M}$ around $\mathbf{0} \in \mathcal{T}_x\mathcal{M}$ and the manifold $\mathcal{M}$. The neighborhood $\mathcal{B}_r(\mathbf{0})$ is characterized by the injectivity radius r, so that it neither contains nor goes beyond the cut locus of $\mathcal{M}$. In other words, the injectivity radius fundamentally quantifies the extent to which geodesics behave like straight lines in $\mathbb{R}^d$ within the largest possible neighborhood $\mathcal{B}_r$. For instance, the sphere exponential map is diffeomorphic only for tangent vectors u satisfying $\|u\|_x < d_{\mathcal{S}^d}(x, -x) = r$. Off-the-shelf Euclidean learning approaches applied on a single tangent space disregard the injectivity radius. Thus, they are prone to generate different tangent vectors associated to the same point on the manifold, leading to inconsistencies.

Fallacy 2. *In general, the exponential map is not an isometric mapping and consequently does not preserve distances.*

As local approximations, tangent spaces intrinsically distort distances as, e.g., in Figs. 3, 5d. The following inequality holds for manifolds with non-zero curvature $\forall\, y_1, y_2 \neq x$,

$$d_{\mathcal{M}}(y_1, y_2) \neq \|\mathrm{Log}_x(y_2) - \mathrm{Log}_x(y_1)\|_x, \quad (8)$$

so that only distances with respect to the basepoint x are preserved (see Fig. 4-*left*). In particular, distances are under/over-estimated on tangent spaces of manifolds with constant positive/negative curvature. Therefore, applying distance-based learning algorithms on data projected in a single tangent space unavoidably leads to distorted models as they overlook the intrinsic geometry of the manifold. Distortions are more pronounced for data lying far away from the basepoint and exacerbated for complex Riemannian manifolds, e.g., non-constant curvature and high-dimensional manifolds. For instance, orientation kernelized movement primitives based on a single tangent space [4] learn distorted models that can only be applied to data closely grouped on $\mathcal{S}^3$. This limitation is further worsened on symmetric positive-definite matrix (SPD) manifolds [3]. Leveraging several tangent spaces significantly reduces distortions and increases the performance of learning algorithms, e.g., for clustering [18], and classification [19] tasks, among others.

Fallacy 3. *Velocities on Riemannian manifolds are attached to specific positions and are not independent entities.*

This implies that Riemannian vector fields, e.g., first-order dynamical systems (DS), assign a tangent vector in $\mathcal{T}_x\mathcal{M}$ to each point x on the manifold. Such vectors are attached to x and associated with the physical notion of velocity. In contrast, vectors $z = \mathrm{Log}_x(y_2) - \mathrm{Log}_x(y_1)$ resulting from tangent space projections with $y_1, y_2 \neq x$ are physically-meaningless and can neither be associated with nor be integrated as velocities at y_1 or y_2 (see Fig. 4-*middle*). Hence, a DS learned on a single tangent space [20] unavoidably leads to distortions and to physically-incoherent models. Riemannian second-order DS [21] are intrinsically formulated via tangent vectors and covariant accelerations [22], and cannot be learned in a single tangent space.

Fallacy 4. *In general, data on Riemannian manifolds and Lie groups are not treated equivalently.*

Several Riemannian manifolds can be endowed with a Lie group structure. This allows the formulation of learning algorithms that leverage their Lie algebra, i.e., the tangent space at the group identity. Exponential and logarithmic maps on Lie groups are mappings between the Lie algebra and the group. These are often defined in combination with composition operations [23], which intuitively play the role of transportation operators w.r.t the group identity (a.k.a. the adjoint operator). This differs from the Riemannian exponential and logarithmic maps. Therefore, learning on the Lie algebra is not equivalent to learning on a single tangent space of a Riemannian manifold. Learning algorithms must be formulated adequately for each type of manifold, see, e.g., the differences between learning stable DS on Lie groups [24] and on Riemannian manifolds [7].

d	C	Mixture of RGDs			Mixture of WGDs			Mixture of vMFs		
	2	3	7	10	3	7	10	3	7	10
Euclidean GMM	552	-132 ± 49	-299 ± 69	-425 ± 70	-74 ± 151	-210 ± 249	-300 ± 250	195 ± 68	240 ± 92	241 ± 122
Tangent GMM	2117	814 ± 38	514 ± 66	232 ± 82	809 ± 103	470 ± 299	284 ± 254	1010 ± 44	1010 ± 77	949 ± 122
Riemannian GMM	**2176**	**853 ± 40**	**642 ± 66**	**432 ± 92**	**850 ± 98**	**543 ± 291**	**318 ± 334**	**1052 ± 40**	**1080 ± 104**	**1068 ± 156**

TABLE I: Average test log-likelihood over 100 target densities estimated via GMMs on sphere manifolds $\mathcal{S}^d$.

d	Mixture of RGDs		Mixture of inv. Wisharts	
	2	3	2	3
Eucl. GMM	-552 ± 143	-1228 ± 218	-381 ± 133	-860 ± 218
Tan. GMM	-409 ± 55	-909 ± 112	-207 ± 89	-355 ± 158
Riem. GMM	**-380 ± 48**	**-792 ± 136**	**-140 ± 83**	**-134 ± 133**

TABLE II: Average test log-likelihood over 100 target densities estimated via GMMs on SPD manifolds $\mathcal{S}^d_{++}$.

Dataset	DTWD		Success rate	
	S	W	S	W
Proj. Eucl. DS	3.2 ± 3.1	4.4 ± 6.5	0.97 ± 0.02	0.87 ± 0.15
Tangent DS	6.4 ± 0.3	9.2 ± 3.0	0.96 ± 0.07	0.67 ± 0.04
Riem. DS	**0.91 ± 1.2**	**1.1 ± 3.9**	**1.0 ± 00**	**0.99 ± 0.02**

TABLE III: Average DTWD w.r.t the demonstrations and success rate over 1000 runs for Proj. Eucl. DS, Tangent DS, and Riem. DS.

Fallacy 5. *Experiments on constant curvature and low-dimensional Riemannian manifolds may be misleading.*

Inaccuracies and distortions induced by using a single tangent space are less apparent under certain conditions. For instance, Sommer et al. [25] showed that the variance of certain symmetric datasets on spheres are equivalently captured via Tangent and Exact Principal Geodesic Analysis (PGA) [1], while significant differences arise for similar datasets on hyperbolic spaces. Moreover, the reconstruction error of Tangent PGA drops significantly faster than Exact PGA for increasing data variance [26]. Such behaviors lead to inaccurate and overly-optimistic interpretations of results obtained via single tangent space-based approaches on small data ranges (see Fig. 4-*right*), as in [3], [20], and on low-dimensional simple manifolds, e.g., spheres as in [4].

C. Cases Where a Single Tangent Space Is Appropriate

Generalization of linear methods are often intrinsically formulated via a single tangent space. These include geodesic regression [27], [28] and geodesic basis function models of movement primitives [29]. Such learning models are based on a single geodesic, which is completely defined by a basepoint $x \in \mathcal{M}$ and an initial velocity $v \in \mathcal{T}_x\mathcal{M}$, i.e., via a single tangent space $\mathcal{T}_x\mathcal{M}$. Consequently, learning naturally occurs in $\mathcal{T}_x\mathcal{M}$ only. Notice that an appropriate basepoint x must still be selected to achieve optimal performance.

V. EXPERIMENTS

We here illustrate the single tangent space fallacy via two different robot learning applications.

A. Learning Densities

We return to the density estimation problem of Section III and analyse the performance of Euclidean GMM, Tangent GMM at the Fréchet mean, and Riemannian GMM to estimate several target densities on various Riemannian manifolds. We train the models on 100 samples from the target densities and compute their likelihood on 100 additional test points. In each case, all GMMs are initialized identically.

First, we estimate densities on spheres $\mathcal{S}^d$ with $d = \{3, 7, 10\}$. We consider target densities in the form of mixtures of 3 RGDs, 3 Wrapped Gaussian distributions (WGDs) [30], [31], or 3 isotropic von Mises-Fisher distributions (vMFs) [30]. The means of the 3 mixture components are defined as $\boldsymbol{\mu}_1 = \frac{1}{d+1}\mathbf{1}_{d+1}$, $\boldsymbol{\mu}_2 = \frac{1}{d+1}\left(\begin{smallmatrix}1\\-\mathbf{1}_d\end{smallmatrix}\right)$, and $\boldsymbol{\mu}_3 = \frac{1}{d+1}\left(\begin{smallmatrix}-1\\\mathbf{1}_d\end{smallmatrix}\right)$, with $\mathbf{1}_d$ denoting a size d vector of ones.

The covariances $\boldsymbol{\Sigma}_k$ of the RGDs and WGDs are randomly-generated SPD matrices, and the concentration parameters of the vMFs are randomly generated from a uniform distribution on $[20, 70]$. As show in Table I, the Riemannian GMMs consistently outperform the Euclidean and Tangent GMMs, thus matching best the target densities. Although Tangent GMMs largely outperform Euclidean GMMs, their performance remains below that of the Riemannian models. This performance gap increases for complex target densities, e.g., mixtures of RGDs and WGDs, compared to simpler densities defined by isotropic components (e.g., vMFs), which are easier to model on a single tangent space. Note that the performance gap also grows with the manifold dimensionality.

Next, we study manifolds with non-constant curvature, and estimate densities on SPD manifolds $\mathcal{S}^d_{++}$ of dimensions $d = \{2, 3\}$. We consider target densities in the form of mixtures of 5 RGDs, or 5 inverse-Wishart distributions (iWDs) [32]. The mean and covariance of the RGDs are randomly generated as SPD matrices of appropriate dimensions with eigenvalues in $[0.1, 2]$ and $[0.1, 0.5]$, respectively. The scale matrix and degree of freedom of the iWDs are generated as the means of the RGD, and from a uniform distribution on $[d + 1, d + 3]$, respectively. The Riemannian GMM consistently achieves the highest likelihood and, similarly as for sphere manifolds, the performance gap between Riemannian GMMs and the other approaches grows with the manifold dimensionality.

B. Learning Dynamical Systems with Normalizing Flows

We shift our attention to the problem of learning stable first-order DS, whose state evolves on a Riemannian manifold. Specifically, we consider DS of the form $\dot{x} = f(x)$, with $x \in \mathcal{M}$ and $\dot{x} \in \mathcal{T}_x\mathcal{M}$. Such systems are conventionally learned from demonstrations, by using example trajectories of the desired motion to train a parametrized function $f_{\boldsymbol{\theta}}$ under certain stability guarantees. Most prior works assume the state x to evolve in an Euclidean space [33]–[38]. However, there has been a growing interest in extending these approaches to non-Euclidean settings [7], [20], [24].

We evaluate the effects of considering a single tangent space to learn first-order DS on Riemannian manifolds. Given a set of demonstrations, our learning problem boils down to learn $f_{\boldsymbol{\theta}}$, which can be represented using a GMM as in [33], [35], [36], or normalizing flows [39] as in [7], [20], [24], [37]. We focus on the latter as it offers a general and extensible framework to model highly-complex
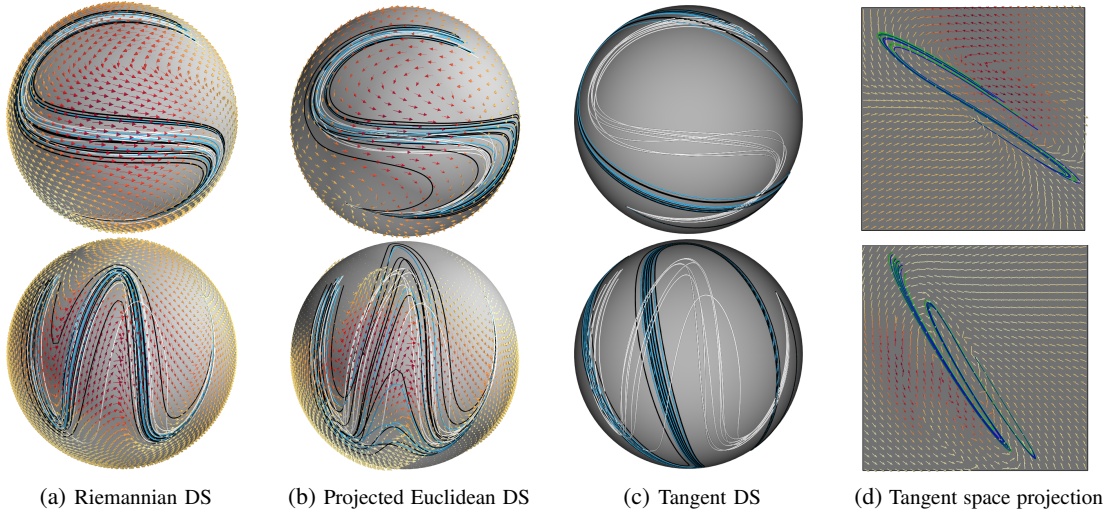
(a) Riemannian DS   (b) Projected Euclidean DS   (c) Tangent DS   (d) Tangent space projection

Fig. 5: Learning first-order DS on the LASA datasets S *(top)* and W *(bottom)* projected on $\mathcal{S}^2$. The demonstrations are displayed as white curves and the learned vector field is depicted by arrows (color-coded based on the magnitude). Blue trajectories are reproductions starting at the same initial points as the demonstrations, while black trajectories start from randomly-sampled points in their neighborhood. The vector field of Tangent DS is not displayed in (c) as it is learned on the single tangent space in (d) and does not produce valid velocities.

vector fields. We tackle the learning problem via: *(1)* a full Riemannian framework [7], *(2)* a projection-based Euclidean approach, projecting solutions in $\mathbb{R}^d$ to $\mathcal{M}$, and *(3)* a single-tangent space perspective as proposed in [20]. We evaluate all approaches on the LASA dataset [40] projected on $\mathcal{S}^2$ [7], so that $\boldsymbol{x} \in \mathcal{S}^2$, $\dot{\boldsymbol{x}} \in \mathcal{T}_{\boldsymbol{x}}\mathcal{S}^2$. A dataset $\{\{\boldsymbol{x}_{m,t}, \dot{\boldsymbol{x}}_{m,t}\}_{t=1}^{T_m}\}_{m=1}^{M}$ of M demonstrations is used to learn $f_{\boldsymbol{\theta}}$, which is represented by a Riemannian continuous normalizing flow (CNF) [41]. The CNF is trained on a different spaces: *(1)* $\mathcal{S}^2$ for the Riemannian DS, *(2)* $\mathbb{R}^3$ for the projected Euclidean DS, and *(3)* $\mathcal{T}_{\boldsymbol{g}}\mathcal{S}^2$ for the Tangent DS with $\boldsymbol{g}$ being the DS attractor.

Fig. 5 shows the demonstrations for the datasets S and W, the learned vector fields, and the reproduced trajectories for all considered approaches. As expected, the Riemannian framework provides reproductions that closely match the demonstrations. The projection-based approach retrieves trajectories that roughly match the demonstrations pattern, although some reproductions significantly deviate from the data support. The vector field learned on a single-tangent space leads to very distorted reproductions, even for starting points coinciding with those of the demonstrations. Despite the learned vector field on $\mathcal{T}_{\boldsymbol{g}}\mathcal{S}^2$ capturing the demonstrations pattern, it is evident that the demonstrations are severely distorted in $\mathcal{T}_{\boldsymbol{g}}\mathcal{S}^2$. Such distortions are then reflected on $\mathcal{S}^2$ when projecting the reconstructions using the exponential map. Note that the trajectories of both datasets cover a large part of the semisphere of $\mathcal{S}^2$, thus being hard to approximate via a single tangent space. Indeed, as the starting points of the demonstrations are far away from the goal, they display a high distortion which is exacerbated by the exponential and logarithmic maps needed to reproduce the motion.

We quantitatively assess the performance of the considered approaches via *(1)* the dynamic time warping distance (DTWD) as a measure of trajectory reproduction accuracy, and *(2)* the stability of the learned DS as the percentage of trajectories converging to the goal out of 1000 uniformly-sampled initial points on $\mathcal{S}^2$. Table III reports the DTWD

and success rate metrics, where the Riemannian formulation consistently outperforms its Euclidean counterparts. Notice that a low success rate means that the model is unable to guarantee stable reproductions, which is imperative when controlling real robots with this kind of approaches.

VI. TAKE HOME MESSAGES

This paper emphasized the critical importance of avoiding the fallacy of using a single tangent space for operating on Riemannian manifolds in robot learning problems. We experimentally showed how the performance of machine learning methods can be severely affected by such an approximation, even in low-dimensional manifolds of constant curvature and noticeably more in complex, high-dimensional manifolds. Good practices for operating with Riemannian manifolds in robot learning include: *(1)* designing coordinate-invariant algorithms that naturally emerge from the geometric constructions defined via Riemannian metrics; *(2)* leveraging a bundle of tangent spaces and parallel transport operators to capture the manifold's intricate geometry and global structure; *(3)* formulating learning models that leverage the core principles of Riemannian theory. For example, Ijspeert et al. [42], [43] formulated dynamic movement primitives based on second-order DS and nonlinear oscillators. Naturally, a sound Riemannian approach should build on nonlinear second-order DS on Riemannian manifolds [44], where concepts such as the covariant derivative are crucial for a proper treatment of the problem. Similarly, reinforcement learning (RL) algorithms on Riemannian manifolds should be formulated on a case-by-case basis by adapting their core principles. For instance, Riemannian Bayesian optimization requires Riemannian kernels and optimization techniques [45], [46]. This is fundamentally opposed to using a single tangent space in which any off-the-shelf RL method is applied as in [47]. To conclude, our main message is to embrace mathematically-sound techniques so that researchers and practitioners unlock the full potential of Riemannian manifolds in robot learning.

References

[1] P. T. Fletcher, C. Lu, S. M. Pizer, and S. Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE Trans. on Medical Imaging*, vol. 23, no. 8, pp. 995–1005, 2004. [Online]. Available: https://doi.org/10.1109/tmi.2004.831793

[2] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 10, pp. 1713–1727, 2008. [Online]. Available: https://doi.org/10.1109/TPAMI.2008.75

[3] F. J. Abu-Dakka, Y. Huang, J. Silvério, and V. Kyrki, "A probabilistic framework for learning geometry-based robot manipulation skills," *Robotics and Autonomous Systems*, vol. 141, p. 103761, 2021. [Online]. Available: https://doi.org/10.1016/j.robot.2021.103761

[4] Y. Huang, F. J. Abu-Dakka, J. Silvério, and D. G. Caldwell, "Toward orientation learning and adaptation in Cartesian space," *IEEE Trans. on Robotics*, vol. 37, no. 1, pp. 82–98, 2021. [Online]. Available: https://doi.org/10.1109/TRO.2020.3010633

[5] E. Mathieu and M. Nickel, "Riemannian continuous normalizing flows," in *Neural Information Processing Systems (NeurIPS)*, 2020, pp. 2503–2515. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/1aa3d9c6ce672447e1e5d0f1b5207e85-Paper.pdf

[6] N. Jaquier, L. Rozo, D. G. Caldwell, and S. Calinon, "Geometry-aware manipulability learning, tracking and transfer," *Intl. Journal of Robotics Research*, vol. 20, no. 3, pp. 624–650, 2021. [Online]. Available: https://doi.org/10.1177/0278364920946815

[7] J. Zhang, H. B. Mohammadi, and L. Rozo, "Learning Riemannian stable dynamical systems via diffeomorphisms," in *Conference on Robot Learning (CoRL)*, ser. Proceedings of Machine Learning Research, vol. 205, 2023, pp. 1211–1221.

[8] M. do Carmo, *Riemannian Geometry*. Birkhäuser Basel, 1992.

[9] J. M. Lee, *Introduction to Riemannian Manifolds*. Springer, 2018.

[10] X. Pennec, "Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision*, vol. 25, pp. 127–154, 2006. [Online]. Available: https://doi.org/10.1007/s10851-006-6228-4

[11] M. Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Annales de l'Institut Henry Poincaré*, vol. 10, no. 4, pp. 215–310, 1948.

[12] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "On-manifold preintegration for real-time visual–inertial odometry," *IEEE Trans. on Robotics*, vol. 33, no. 1, pp. 1–21, 2017. [Online]. Available: https://doi.org/10.1109/tro.2016.2597321

[13] W. Park, Y. Liu, Y. Zhou, M. Moses, and G. S. Chirikjian, "Kinematic state estimation and motion planning for stochastic nonholonomic systems using the exponential map," *Robotica*, vol. 26, no. 4, p. 419–434, 2008. [Online]. Available: https://doi.org/10.1017/S0263574708004475

[14] M. Zeestraten, "Programming by demonstration on Riemannian manifolds," PhD thesis, University of Genova, Italy, 2018. [Online]. Available: https://iris.unige.it/handle/11567/930621#.YCgbRuoo85k

[15] E. Simo-Serra, C. Torras, and F. Moreno-Noguer, "Geodesic finite mixture models," in *Proceedings of the British Machine Vision Conference (BMVC)*, 2014. [Online]. Available: https://doi.org/10.5244/c.28.91

[16] N. Boumal, *An introduction to optimization on smooth manifolds*. Cambridge University Press, 2023. [Online]. Available: https://doi.org/10.1017/9781009166164

[17] J. M. Lee, *Introduction to smooth manifolds*. Springer, 2013.

[18] E. Simo-Serra, C. Torras, and F. Moreno-Noguer, "3d human pose tracking priors using geodesic mixture models," *Intl. Journal on Computer Vision*, vol. 122, no. 2, pp. 388–408, 2017. [Online]. Available: https://doi.org/10.1007/s11263-016-0941-2

[19] A. Sanin, C. Sanderson, M. T. Harandi, and B. C. Lovell, "K-tangent spaces on Riemannian manifolds for improved pedestrian detection," in *IEEE International Conference on Image Processing*, 2012, pp. 473–476. [Online]. Available: https://doi.org/10.1109/ICIP.2012.6466899

[20] W. Wang, M. Saveriano, and F. J. Abu-Dakka, "Learning deep robotic skills on Riemannian manifolds," *IEEE Access*, vol. 10, pp. 114143–114152, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3217800

[21] S. Fiori, I. Cervigni, M. Ippoliti, and C. Menotta, "Synthetic nonlinear second-order oscillators on Riemannian manifolds and their numerical simulation," *Discrete & Continuous Dynamical Systems - B*, vol. 27, no. 3, p. 1227, 2022. [Online]. Available: https://doi.org/10.3934/dcdsb.2021088

[22] J. A. Rojas-Quintero, F. Dubois, and H. C. R. de Ávila, "Riemannian formulation of Pontryagin's maximum principle for the optimal control of robotic manipulators," *Mathematics*, vol. 10, no. 7, pp. 1–22, 2022.

[23] J. Solà, J. Deray, and D. Atchuthan, "A micro Lie theory for state estimation in robotics," *arXiv preprint 1812.01537*, 2018.

[24] J. Urain, D. Tateo, and J. Peters, "Learning stable vector fields on Lie groups," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 12569–12576, 2022. [Online]. Available: https://doi.org/10.1109/LRA.2022.3219019

[25] S. Sommer, F. Lauze, and M. Nielsen, "Optimization over geodesics for exact principal geodesic analysis," *Advances in Computational Mathematics*, vol. 40, no. 2, pp. 283–313, 2014. [Online]. Available: https://doi.org/10.1007/s10444-013-9308-1

[26] X. Fan, C. Yang, and B. C. Vemuri, "Nested hyperbolic spaces for dimensionality reduction and hyperbolic nn design," in *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 356–365. [Online]. Available: 10.1109/CVPR52688.2022.00045

[27] P. T. Fletcher, "Geodesic regression and the theory of least squares on Riemannian manifolds," *Intl. Journal on Computer Vision*, vol. 105, pp. 171–185, 2013. [Online]. Available: https://doi.org/10.1007/s11263-012-0591-y

[28] H. J. Kim, N. Adluru, M. D. Collins, M. K. Chung, B. B. Bendin, S. C. Johnson, R. J. Davidson, and V. Singh, "Multivariate general linear models (MGLM) on Riemannian manifolds with applications to statistical analysis of diffusion weighted images," in *Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 2705–2712. [Online]. Available: https://doi.org/10.1109/CVPR.2014.352

[29] L. Rozo and V. Dave, "Orientation probabilistic movement primitives on Riemannian manifolds," in *Conference on Robot Learning (CoRL)*, ser. Proceedings of Machine Learning Research, vol. 164, 2022, pp. 373–383.

[30] K. V. Mardia and P. E. Jupp, *Distributions on Spheres*. John Wiley and Sons, Ltd, 1999, ch. 9, pp. 159–192. [Online]. Available: https://doi.org/10.1002/9780470316979.ch9

[31] F. Galaz-Garcia, M. Papamichalis, K. Turnbull, S. Lunagomez, and E. Airoldi, "Wrapped distributions on homogeneous Riemannian manifolds," *arXiv preprint 2204.09790*, 2022. [Online]. Available: https://arxiv.org/abs/2204.09790

[32] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian Data Analysis*, 3rd ed. Chapman and Hall/CRC, 2013. [Online]. Available: https://doi.org/10.1201/b16018

[33] S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with Gaussian mixture models," *IEEE Trans. on Robotics*, vol. 27, no. 5, pp. 943–957, 2011. [Online]. Available: https://doi.org/10.1109/TRO.2011.2159412

[34] K. Neumann and J. J. Steil, "Learning robot motions with stable dynamical systems under diffeomorphic transformations," *Robotics and Autonomous Systems*, vol. 70, pp. 1–15, 2015. [Online]. Available: https://doi.org/10.1016/j.robot.2015.04.006

[35] C. Blocher, M. Saveriano, and D. Lee, "Learning stable dynamical systems using contraction theory," in *International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, 2017, pp. 124–129. [Online]. Available: https://doi.org/10.1109/URAI.2017.7992901

[36] N. Figueroa and A. Billard, "A physically-consistent bayesian non-parametric mixture model for dynamical system learning," in *Conference on Robot Learning (CoRL)*, 2018, pp. 927–946. [Online]. Available: https://proceedings.mlr.press/v87/figueroa18a.html

[37] M. A. Rana, A. Li, D. Fox, B. Boots, F. Ramos, and N. Ratliff, "Euclideanizing flows: Diffeomorphic reduction for learning stable dynamical systems," in *Conference on Learning for Dynamics and Control (L4DC)*, 2020, pp. 630–639. [Online]. Available: https://proceedings.mlr.press/v120/rana20a.html

[38] Y. Zhang, L. Cheng, H. Li, and R. Cao, "Learning accurate and stable point-to-point motions: A dynamic system approach," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 1510–1517, 2022. [Online]. Available: https://doi.org/10.1109/LRA.2022.3140677

[39] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference," *Journal of Machine Learning Research*, vol. 22, no. 57, pp. 1–64, 2021. [Online]. Available: http://jmlr.org/papers/v22/19-1028.html

[40] A. Lemme, Y. Meirovitch, M. Khansari-Zadeh, T. Flash, A. Billard, and J. J. Steil, "Open-source benchmarking for learned reaching

motion generation in robotics," *Paladyn, Journal of Behavioral Robotics*, vol. 6, no. 1, 2015. [Online]. Available: https://doi.org/10.1515/pjbr-2015-0002

[41] A. Lou, D. Lim, I. Katsman, L. Huang, Q. Jiang, S. N. Lim, and C. M. De Sa, "Neural manifold ordinary differential equations," in *Neural Information Processing Systems (NeurIPS)*, 2020, pp. 17 548–17 558. [Online]. Available: https://proceedings.neurips.cc/paper/2020/file/cbf8710b43df3f2c1553e649403426df-Paper.pdf

[42] A. J. Ijspeert, J. Nakanishi, and S. Schaal, "Movement imitation with nonlinear dynamical systems in humanoid robots," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2002, pp. 1398–1403. [Online]. Available: https://doi.org/10.1109/ROBOT.2002.1014739

[43] ——, "Learning rhythmic movements by demonstration using nonlinear oscillators," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2002, pp. 958–963. [Online]. Available: https://doi.org/10.1109/IRDS.2002.1041514

[44] S. Fiori and A. Bonci, "Nonlinear second-order dynamical systems on Riemannian manifolds: Damped oscillators," in *Intl. Conf. on Modeling, Simulation and Visualization Methods (MSV)*, 2015, pp. 16–22.

[45] N. Jaquier, L. Rozo, S. Calinon, and M. Bürger, "Bayesian optimization meets Riemannian manifolds in robot learning," in *Conference on Robot Learning (CoRL)*, ser. Proceedings of Machine Learning Research, vol. 100, 2020, pp. 233–246. [Online]. Available: https://proceedings.mlr.press/v100/jaquier20a.html

[46] N. Jaquier, V. Borovitskiy, A. Smolensky, A. Terenin, T. Asfour, and L. Rozo, "Geometry-aware Bayesian optimization in robotics using Riemannian Matérn kernels," in *Conference on Robot Learning (CoRL)*, ser. Proceedings of Machine Learning Research, vol. 164, 2021. [Online]. Available: https://proceedings.mlr.press/v164/jaquier22a.html

[47] N. Alhousani, M. Saveriano, I. Sevinc, T. Abdulkuddus, H. Kose, and F. J. Abu-Dakka, "Geometric reinforcement learning for robotic manipulation," *IEEE Access*, 2023.