

# NEURAL CONTRACTIVE DYNAMICAL SYSTEMS

Hadi Beik-Mohammadi<sup>1,2</sup>, Søren Hauberg<sup>3</sup>, Georgios Arvanitidis<sup>3</sup>, Nadia Figueroa<sup>4</sup>,  
Gerhard Neumann<sup>2</sup>, and Leonel Rozo<sup>1</sup>

<sup>1</sup> Bosch Center for Artificial Intelligence (BCAI), <sup>2</sup> Karlsruhe Institute of Technology (KIT),

<sup>3</sup> Technical University of Denmark (DTU), <sup>4</sup> University of Pennsylvania (UPenn).

Emails: [hadi.beik-mohammadi, leonel.rozo]@de.bosch.com, [sohau, gear]@dtu.dk,  
nadiafig@seas.upenn.edu, gerhard.neumann@kit.edu

## ABSTRACT

Stability guarantees are crucial when ensuring that a fully autonomous robot does not take undesirable or potentially harmful actions. Unfortunately, global stability guarantees are hard to provide in dynamical systems learned from data, especially when the learned dynamics are governed by neural networks. We propose a novel methodology to learn *neural contractive dynamical systems*, where our neural architecture ensures contraction, and hence, global stability. To efficiently scale the method to high-dimensional dynamical systems, we develop a variant of the variational autoencoder that learns dynamics in a low-dimensional latent representation space while retaining contractive stability after decoding. We further extend our approach to learning contractive systems on the Lie group of rotations to account for full-pose end-effector dynamic motions. The result is the first highly flexible learning architecture that provides contractive stability guarantees with capability to perform obstacle avoidance. Empirically, we demonstrate that our approach encodes the desired dynamics more accurately than the current state-of-the-art, which provides less strong stability guarantees.

## 1 STABILITY IN ROBOT LEARNING

Deploying a fully autonomous robot requires guarantees of stability to ensure that the robot does not perform unwanted, potentially dangerous, actions. Consider the robot in Fig. 1, which is trained from demonstrations (black dots) to execute a sinusoidal motion (orange). If the robot is pushed away (red perturbation) from its planned trajectory, then it should still be guaranteed to reproduce a motion similar to the demonstration pattern (blue). Manually designed robot movements could achieve such stability guarantees, but even skilled engineers struggle to hand-code highly dynamic motions (Billard et al., 2016). In contrast, learning robot dynamics from demonstrations has shown to be an efficient and intuitive approach for encoding highly dynamic motions into a robot’s repertoire (Schaal et al., 2003). Unfortunately, these learning-based approaches often struggle to ensure stability as they rely on the machine learning model to extrapolate in a controlled manner. In particular, it has turned out to be surprisingly difficult to control the extrapolating behavior of neural network models (Xu et al., 2021), which hampers stability guarantees.

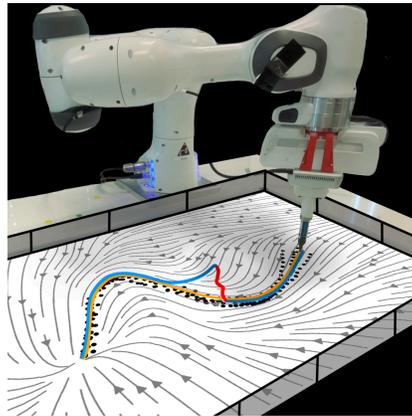


Figure 1: Robot motion executed via a neural contractive dynamical system (NCDS).

**Stability is commonly ensured** through *asymptotic* or *contraction* guarantees. Asymptotic stability ensures that all motions converge to a fixed *point* (known as the system’s *attractor*) (Khansari-Zadeh & Billard, 2011; 2014; Neumann & Steil, 2015; Zhang et al., 2022b;a). This is suitable when the only requirement is that the robot eventually reaches a certain configuration, e.g. its end-effector is at a specific goal position. Many tasks, however, require the robot to dynamically follow desired trajectories, e.g. in flexible manufacturing, human-robot interaction, or in entertainment settings.

In these cases, asymptotic stability is insufficient. A stronger notion of stability is provided by *contraction theory* (Lohmiller & Slotine, 1998), which ensures that all the path integrals regardless of their initial state incrementally converge over time. Unfortunately, the mathematical requirements of a contractive system are difficult to ensure in popular neural network architectures.

**Existing contractive learning methods** focus on low-capacity models fitted under contraction constraints. Ravichandar et al. (2017) and Ravichandar & Dani (2019) both use Gaussian mixture models to encode the demonstrations and model the dynamical system with Gaussian mixture regression (GMR), which is fitted under contraction constraints. Blocher et al. (2017) also use GMR to model the system dynamics but instead include a stabilizing control input that guarantees a local contractive behavior. Likewise, Sindhwani et al. (2018) model the dynamical system with a well-crafted kernel, under which optimization remains convex w.r.t. contraction constraints. Note that the above works impose the contraction constraints during optimization, which complicates model training.

Developing high-capacity (deep) models is difficult as off-the-shelf neural network architectures provide no stability guarantees. Instead, current approaches split the task into two steps: first, learn a non-contractive dynamical system, and then estimate a Riemannian metric (Sec. 2.1), parametrized by a neural network, under which the system becomes contractive (Tsukamoto et al., 2021a; Dawson et al., 2023). The approach, thus, requires training two separate networks, which is impractical and comes with additional challenges. For example, Sun et al. (2020) learns the metric by regularizing towards a contractive system, such that stability can *only* be ensured near training data. Alternatively, Tsukamoto & Chung (2021) develop a convex optimization problem for learning the metric from sampled *optimal* metrics. Unfortunately, accessing such optimal metrics is a difficult problem in itself. Additionally, Kozachkov et al. (2022) introduce a network of small and simple recurrent neural networks with built-in contractive behavior as an alternative to a single, larger network.

**In this paper**, we propose the *neural contractive dynamical system (NCDS)* which is the first neural network architecture that is guaranteed to be contractive for all parameter values. This implies that NCDS can easily be incorporated into existing pipelines to provide stability guarantees and then trained using standard unconstrained optimization. As demonstration data may be high dimensional, we secondly propose an injective variant of the variational autoencoder that allows learning low-dimensional latent dynamics, while ensuring contraction guarantees for the *decoded* dynamical system. Thirdly, as most complex robotic tasks involve full-pose end-effector movements (including *orientation* dynamics), we extend our approach to cover the Lie group  $SO(3)$ . Finally, we show that NCDS can trivially be modified to avoid (dynamic) obstacles without sacrificing stability guarantees.

## 2 LEARNING CONTRACTIVE VECTOR FIELDS

Our ambition is a flexible neural architecture, which is guaranteed to always be contractive. We, however, first introduce contraction theory as this is a prerequisite for our design.

### 2.1 BACKGROUND: CONTRACTIVE DYNAMICAL SYSTEMS

To begin with, assume an autonomous dynamical system  $\dot{\mathbf{x}}_t = f(\mathbf{x}_t)$ , where  $\mathbf{x}_t \in \mathbb{R}^D$  is the state variable,  $f : \mathbb{R}^D \rightarrow \mathbb{R}^D$  is a  $C^1$  function and  $\dot{\mathbf{x}}_t = \frac{d\mathbf{x}}{dt}$  denotes temporal differentiation. As illustrated in Fig. 1, contraction stability ensures that all solution trajectories of a nonlinear system  $f$  incrementally converge regardless of initial conditions  $\mathbf{x}_0, \dot{\mathbf{x}}_0$ , and temporary perturbations (Lohmiller & Slotine, 1998). The system stability can, thus, be analyzed differentially, i.e. we can ask if two nearby trajectories converge to one another. Specifically, contraction theory defines a measure of distance between neighboring trajectories, known as the *contraction metric*, in which the distance decreases exponentially over time (Jouffroy & I. Fossen, 2010; Tsukamoto et al., 2021b).

Formally, an autonomous dynamical system yields the differential relation  $\delta\dot{\mathbf{x}} = \mathbf{J}(\mathbf{x})\delta\mathbf{x}$ , where  $\mathbf{J}(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}}$  is the system Jacobian and  $\delta\mathbf{x}$  is a virtual displacement (i.e., an infinitesimal spatial displacement between the nearby trajectories at a fixed time). Note that we have dropped the time index  $t$  to limit notational clutter. The rate of change of the corresponding infinitesimal squared distance  $\delta\mathbf{x}^\top \delta\mathbf{x}$  is then,

$$\frac{d}{dt}(\delta\mathbf{x}^\top \delta\mathbf{x}) = 2\delta\mathbf{x}^\top \delta\dot{\mathbf{x}} = 2\delta\mathbf{x}^\top \mathbf{J}(\mathbf{x})\delta\mathbf{x}. \quad (1)$$

It follows that if the symmetric part of the Jacobian  $\mathbf{J}(\mathbf{x})$  is negative definite, then the infinitesimal squared distance between neighboring trajectories decreases over time. This is formalized as:

**Definition 1 (Contraction stability (Lohmiller & Slotine, 1998))** *An autonomous dynamical system  $\dot{\mathbf{x}} = f(\mathbf{x})$  exhibits a contractive behavior if its Jacobian  $\mathbf{J}(\mathbf{x}) = \partial f / \partial \mathbf{x}$  is negative definite, or equivalently if its symmetric part is negative definite. This means that there exists a constant  $\tau > 0$  such that  $\delta \mathbf{x}^\top \delta \mathbf{x}$  converges to zero exponentially at rate  $2\tau$ . This can be summarized as,*

$$\exists \tau > 0 \quad \text{s.t.} \quad \forall \mathbf{x}, \quad \frac{1}{2}(\mathbf{J}(\mathbf{x}) + \mathbf{J}(\mathbf{x})^\top) \prec -\tau \mathbf{I} \prec 0. \quad (2)$$

The above analysis can be generalized to account for a more general notion of distance of the form  $\delta \mathbf{x}^\top \mathbf{M}(\mathbf{x}) \delta \mathbf{x}$ , where  $\mathbf{M}(\mathbf{x})$  is a positive-definite matrix known as the (Riemannian) contraction metric (Tsukamoto et al., 2021b; Dawson et al., 2023). In our work, we learn a dynamical system  $f$  so that it is inherently contractive since its Jacobian  $\mathbf{J}(\mathbf{x})$  fulfills the condition given in Eq. 2. Consequently, it is not necessary to specifically learn a contraction metric as it corresponds to an identity matrix. For our purposes, we, thus, define contraction guarantees following Eq. 2.

## 2.2 NEURAL CONTRACTIVE DYNAMICAL SYSTEMS (NCDS)

From Definition 1, we seek a flexible neural network architecture, such that the symmetric part of its Jacobian is negative definite. We consider the dynamical system  $\dot{\mathbf{x}} = f(\mathbf{x})$ , where  $\mathbf{x} \in \mathbb{R}^D$  denotes the system’s state and  $f : \mathbb{R}^D \rightarrow \mathbb{R}^D$  is parametrized by a neural network. It is not obvious how to impose a negative definiteness constraint on a network’s Jacobian without compromising its expressiveness. To achieve this, we first design a neural network  $\hat{\mathbf{J}}_f$  representing the Jacobian of our final network. This produces matrix-valued negative definite outputs. The final neural network will then be formed by integrating the Jacobian network. Specifically, we define the Jacobian as,

$$\hat{\mathbf{J}}_f(\mathbf{x}) = -(\mathbf{J}_\theta(\mathbf{x})^\top \mathbf{J}_\theta(\mathbf{x}) + \epsilon \mathbb{I}_D), \quad (3)$$

where  $\mathbf{J}_\theta : \mathbb{R}^D \rightarrow \mathbb{R}^{D \times D}$  is a neural network parameterized by  $\theta$ ,  $\epsilon \in \mathbb{R}^+$  is a small positive constant, and  $\mathbb{I}_D$  is an identity matrix of size  $D$ . Intuitively,  $\mathbf{J}_\theta$  can be interpreted as the (approximate) square root of  $\hat{\mathbf{J}}_f$ . Clearly,  $\hat{\mathbf{J}}_f$  is negative definite as all eigenvalues are bounded from above by  $-\epsilon$ .

Next, we take inspiration from Lorraine & Hossain (2019) and integrate  $\hat{\mathbf{J}}_f$  to produce a function  $f$ , which is implicitly parametrized by  $\theta$ , and has Jacobian  $\hat{\mathbf{J}}_f$ . The fundamental theorem of calculus for line integrals tells us that we can construct such a function by a line integral of the form

$$\dot{\mathbf{x}} = f(\mathbf{x}) = \dot{\mathbf{x}}_0 + \int_0^1 \hat{\mathbf{J}}_f(c(\mathbf{x}, t, \mathbf{x}_0)) \dot{c}(\mathbf{x}, t, \mathbf{x}_0) dt, \quad (4)$$

$$c(\mathbf{x}, t, \mathbf{x}_0) = (1-t)\mathbf{x}_0 + t\mathbf{x} \quad \text{and} \quad \dot{c}(t, \mathbf{x}_0, \mathbf{x}) = \mathbf{x} - \mathbf{x}_0, \quad (5)$$

where  $\mathbf{x}_0$  and  $\dot{\mathbf{x}}_0 = f(\mathbf{x}_0)$  represent the initial conditions of the state variable and its first-order time derivative, respectively. The input point  $\mathbf{x}_0$  can be chosen arbitrarily (e.g. as the mean of the training data or it can be learned), while the corresponding function value  $\dot{\mathbf{x}}_0$  has to be estimated alongside the parameters  $\theta$ . Therefore, given a set of demonstrations denoted as  $\mathcal{D} = \{\mathbf{x}_i, \dot{\mathbf{x}}_i\}$ , our objective is to learn a set of parameters  $\theta$  along with the initial conditions  $\mathbf{x}_0$  and  $\dot{\mathbf{x}}_0$ , such that the integration in Eq. 4 enables accurate reconstruction of the velocities  $\dot{\mathbf{x}}_i$  given the state  $\mathbf{x}_i$ .

The integral in Eq. 4 is similar to neural ordinary differential equations (Chen et al., 2018b), with the subtle difference that it is a second-order equation as the outcome pertains to the *velocity* at state  $\mathbf{x}$ . We can, thus, view this system as a second-order neural ordinary equation (Norcliffe et al., 2021) and solve it using off-the-shelf numerical integrators. The resulting function  $f$  will have a negative definite Jacobian for any choice of  $\theta$  and is consequently contractive by construction. In other words, we can control the extrapolation behavior of the neural network that parametrizes our dynamical system  $f$  via the negative-definiteness of  $\hat{\mathbf{J}}_f(\mathbf{x})$ . We call this the *neural contractive dynamical system (NCDS)*, and emphasize that this approach has two key benefits: (1) we can use any smooth neural network  $\mathbf{J}_\theta$  as our base model, and (2) training can then be realized using ordinary *unconstrained* optimization, unlike previous approaches. Figure 2 shows an example NCDS vector field, which is clearly highly flexible even if it guarantees contractive stability.

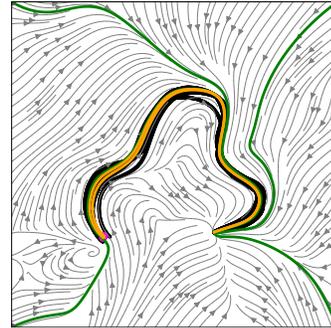


Figure 2: The learned vector field (grey) and demonstrations (black). Yellow and green trajectories show path integrals starting from demonstration starting points and random points, respectively.

### 3 LEARNING LATENT CONTRACTIVE DYNAMICS

Learning highly nonlinear contractive dynamical systems in high-dimensional spaces is difficult. These systems may exhibit complex trajectories with intricate interdependencies among the system variables, making it challenging to capture the underlying dynamics. In the experiments, we show that NCDS works very well for low-dimensional problems, but when only limited data is available, the approach becomes brittle in higher dimensions. A common approach in such cases is to first reduce the data dimensionality, and work as before in the resulting low-dimensional latent space (Chen et al., 2018a; Hung et al., 2022; Beik-Mohammadi et al., 2021). The main difficulty is that even if the latent dynamics are contractive, the associated high-dimensional dynamics need not be. This we solve next.

#### 3.1 BACKGROUND: DEEP GENERATIVE MODELS

**Variational autoencoders (VAEs).** The main goal of *deep generative models* is to approximate the true underlying probability density  $p(\mathbf{x})$  given a finite set of training data in an ambient space  $\mathcal{X}$ , by considering a lower-dimensional latent space  $\mathcal{Z}$ . In particular, the *variational autoencoder (VAE)* (Kingma & Welling, 2014; Rezende et al., 2014) is a latent variable model, often specified through a prior  $p(\mathbf{z}) = \mathcal{N}(\mathbf{z} | \mathbf{0}, \mathbb{I}_d)$  where  $\mathbf{z} \in \mathcal{Z}$ , and a likelihood  $p_\phi(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x} | \mu_\phi(\mathbf{z}), \mathbb{I}_D \sigma_\phi^2(\mathbf{z}))$  with  $\mathbf{x} \in \mathcal{X}$ . Typically, the mean and the variance of the likelihood are parametrized using deep neural networks  $\mu_\phi : \mathcal{Z} \rightarrow \mathcal{X}$  and  $\sigma_\phi^2 : \mathcal{Z} \rightarrow \mathbb{R}_+^D$  with parameters  $\phi$ , and  $\mathbb{I}_D$  and  $\mathbb{I}_d$  being identity matrices of size  $D$  and  $d$ , respectively. These neural networks are trained by maximizing the evidence lower bound (ELBO) (Kingma & Welling, 2014). The latent variable  $\mathbf{z}$  is approximated using a variational encoder  $\mu_\xi(\mathbf{x})$  and is often interpreted as the low-dimensional representation of an observation  $\mathbf{x}$ . In our work, we use a VAE to provide low-dimensional representations of individual points along observed trajectories in order to learn a latent contractive dynamical system.

**Injective flows.** A limitation of VAEs is that their marginal likelihood is intractable and we have to rely on a bound. When  $\dim(\mathcal{X}) = \dim(\mathcal{Z})$ , we can apply the change-of-variables theorem to evaluate the marginal likelihood exactly, giving rise to *normalizing flows* (Tabak & Turner, 2013). This requires the decoder to be diffeomorphic, i.e. a smooth invertible function with a smooth inverse. In order to extend this to the case where  $\dim(\mathcal{X}) > \dim(\mathcal{Z})$ , Brehmer & Cranmer (2020) proposed an *injective flow*, which implements a zero-padding operation (see Sec.3.2 and Equation 6) on the latent variables alongside a diffeomorphic decoder, such that the resulting function is injective.

#### 3.2 LATENT NEURAL CONTRACTIVE DYNAMICAL SYSTEMS

As briefly discussed above, we want to reduce data dimensionality with a VAE, but we further require that any latent contractive dynamical system remains contractive after it has been decoded into the data space. To do so, we can leverage the fact that contraction is invariant under coordinate changes (Manchester & Slotine, 2017; Kozachkov et al., 2023). This means that the transformation between the latent and data spaces may be generally achieved through a diffeomorphic mapping.

**Theorem 1 (Contraction invariance under diffeomorphisms (Manchester & Slotine, 2017))**

*Given a contractive dynamical system  $\dot{\mathbf{x}} = f(\mathbf{x})$  and a diffeomorphism  $\psi$  applied on the state  $\mathbf{x} \in \mathbb{R}^D$ , the transformed system preserves contraction under the change of coordinates  $\mathbf{y} = \psi(\mathbf{x})$ . Equivalently, contraction is also guaranteed under a differential coordinate change  $\delta_{\mathbf{y}} = \frac{\partial \psi}{\partial \mathbf{x}} \delta_{\mathbf{x}}$ .*

Following Theorem 1, we learn a VAE with an injective decoder  $\mu : \mathcal{Z} \rightarrow \mathcal{X}$ . Letting  $\mathcal{M} = \mu(\mathcal{Z})$  denote the image of  $\mu$ , then  $\mu$  is a diffeomorphism between  $\mathcal{Z}$  and  $\mathcal{M}$ , such that Theorem 1 applies. Geometrically,  $\mu$  spans a  $d$ -dimensional submanifold of  $\mathcal{X}$  on which the dynamical system operates.

Here we leverage the zero-padding architecture from Brehmer & Cranmer (2020) for the decoder. Formally, an injective flow  $\mu : \mathcal{Z} \rightarrow \mathcal{X}$  learns an injective mapping between a low-dimensional latent space  $\mathcal{Z}$  and a higher-dimensional data space  $\mathcal{X}$ . Injectivity of the flow ensures that there are no singular points or self-intersections in the flow, which may compromise the stability of the system dynamics in the data space. The injective decoder  $\mu$  is composed of a zero-padding operation on the latent variables followed by a series of  $K$  invertible transformations  $g_k$ . This means that,

$$\mu = g_K \circ \dots \circ g_1 \circ \text{Pad}, \tag{6}$$

where  $\text{Pad}(\mathbf{z}) = [z_1 \cdots z_d 0 \cdots 0]^\top$  represents a  $D$ -dimensional vector  $\mathbf{z}$  with additional  $D - d$  zeros. We emphasize that this decoder is an injective mapping between  $\mathcal{Z}$  and  $\mu(\mathcal{Z}) \subset \mathcal{X}$ , such that a decoded contractive dynamical system remains contractive.

Specifically, we propose to learn a latent data representation using a VAE, where the decoder mean  $\mu_\xi$  follows the architecture in Eq. 6. Empirically, we have found that training stabilizes when the variational encoder takes the form  $q_\xi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z} | \mu_\xi^{-1}(\mathbf{x}), \mathbb{I}_d \sigma_\xi^2(\mathbf{x}))$ , where  $\mu_\xi^{-1}$  is the approximate inverse of  $\mu_\xi$  given by,

$$\mu_\xi^{-1} = \text{Unpad} \circ g_1^{-1} \circ \cdots \circ g_K^{-1}, \quad (7)$$

where  $\text{Unpad} : \mathbb{R}^D \rightarrow \mathbb{R}^d$  removes the last  $D - d$  dimensions of its input as an approximation to the inverse of the zero-padding operation. We emphasize that an exact inverse is not required to evaluate a lower bound of the model evidence.

It is important to note that the state  $\mathbf{x}$  solely encodes the positional information of the system, disregarding the velocity  $\dot{\mathbf{x}}$ . In order to decode the latent velocity  $\dot{\mathbf{z}}$  into the data space velocity  $\dot{\mathbf{x}}$ , we exploit the Jacobian matrix associated with the decoder mean function  $\mu_\xi$ , computed as  $\mathbf{J}_{\mu_\xi}(\mathbf{z}) = \partial \mu_\xi / \partial \mathbf{z}$ . This enables the decoding process formulated as below,

$$\dot{\mathbf{x}} = \mathbf{J}_{\mu_\xi}(\mathbf{z}) \dot{\mathbf{z}}. \quad (8)$$

The above tools let us learn a contractive dynamical system on the latent space  $\mathcal{Z}$ , where the contraction is guaranteed by employing the NCDS architecture (Sec. 2.2). Then, the latent velocities<sup>1</sup>  $\dot{\mathbf{z}}$  given by such a contractive dynamical system can be mapped to the data space  $\mathcal{X}$  using Eq. 8. Assuming the initial robot configuration  $\mathbf{x}_0$  is in  $\mathcal{M}$  (i.e.,  $\mathbf{x}_0 = f(\mathbf{z}_0)$ ), the subsequent motion follows a contractive system along the manifold. If the initial configuration  $\mathbf{x}_0$  is not in  $\mathcal{M}$ , the encoder is used to approximate a projection onto  $\mathcal{M}$ , producing  $\mathbf{z}_0 = \mu^{-1}(\mathbf{x}_0)$ . Note that the movement from  $\mathbf{x}_0$  to  $f(\mathbf{z}_0)$  need not be contractive, but this is a finite-time motion, after which the system is contractive.

### 3.3 LEARNING POSITION AND ORIENTATION DYNAMICS

So far, we have focused on Euclidean robot states, but in practice, the end-effector motion also involves rotations, which do not have an Euclidean structure. We first review the group structure of rotation matrices and then extend NCDS to handle non-Euclidean data using Theorem 1.

**Orientation parameterization.** Three-dimensional spatial orientations can be represented in several ways, including Euler angles, unit quaternions, and rotation matrices (Shuster, 1993). We focus on the latter. The set of rotation matrices forms a Lie group, known as the *special orthogonal group*  $\mathcal{SO}(3) = \{\mathbf{R} \in \mathbb{R}^{3 \times 3} \mid \mathbf{R}^\top \mathbf{R} = \mathbb{I}, \det(\mathbf{R}) = 1\}$ . Every Lie group is associated with its Lie algebra, which represents the tangent space at its origin (Fig. 3). This Euclidean tangent space allows us to operate with elements of the group via their projections on the Lie algebra (Solà et al., 2018). In the context of  $\mathcal{SO}(3)$ , its Lie algebra  $\mathfrak{so}(3)$  is the set of all  $3 \times 3$  skew-symmetric matrices  $[\mathbf{r}]_\times$ . This skew-symmetric matrix exhibits three degrees of freedom, which can be reparameterized as a 3-dimensional vector  $\mathbf{r} = [r_x, r_y, r_z] \in \mathbb{R}^3$ .

We can map back and forth between the Lie group  $\mathcal{SO}(3)$  and its associated Lie algebra  $\mathfrak{so}(3)$  using the *logarithmic* and *exponential maps*, denoted  $\text{Log} : \mathcal{SO}(3) \rightarrow \mathfrak{so}(3)$  and  $\text{Exp} : \mathfrak{so}(3) \rightarrow \mathcal{SO}(3)$ , respectively. We provide explicit formulae for these in Appendix A.2. Due to wrapping (e.g.  $360^\circ$  rotation corresponds to  $0^\circ$ ), the exponential map is surjective. This implies that the inverse, i.e.  $\text{Log}$ , is multivalued, which complicates matters. However, for vectors  $\mathbf{r} \in \mathfrak{so}(3)$ , both  $\text{Exp}$  and  $\text{Log}$  are diffeomorphic if  $\|\mathbf{r}\| < \pi$  (Falorsi et al., 2019; Urain et al., 2022). This  $\pi$ -ball  $\mathcal{B}_\pi$  is known as the *first cover* of the Lie algebra and corresponds to the part where no wrapping occurs.

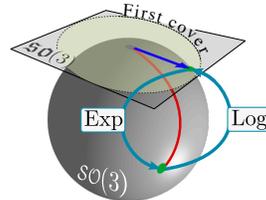


Figure 3: Aspects of the Lie group  $\mathcal{SO}(3)$ .

**NCDS on Lie groups.** Consider the situation where the system state is a rotation, i.e.  $\mathbf{x} \in \mathcal{SO}(3)$ , which we seek to model with a latent NCDS. From a generative point of view, we first construct a decoder  $\mu : \mathcal{Z} \rightarrow \mathfrak{so}(3)$  with outputs in the Lie algebra. We can then apply the exponential map to generate a rotation matrix  $\mathbf{R} \in \mathcal{SO}(3)$ , such that the complete decoder becomes  $\text{Exp} \circ \mu$ .

<sup>1</sup>For training, the latent velocities are simply estimated by a numerical differentiation w.r.t the latent state  $\mathbf{z}$ .

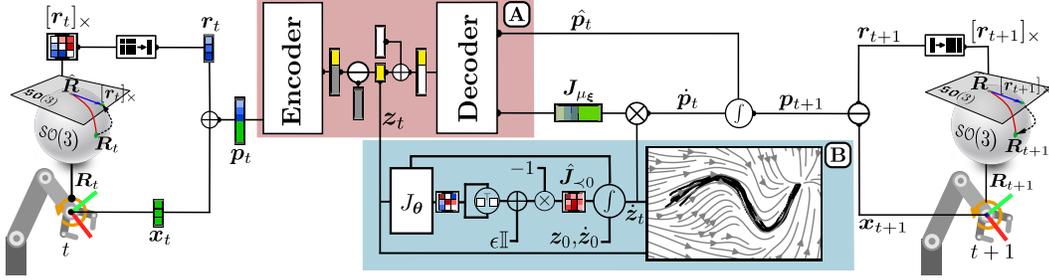


Figure 4: Architecture overview: a single iteration of NCDS simultaneously generating position and orientation dynamics. **(A) VAE (pink box):** The encoder processes the concatenated position-orientation data  $p_t$ , yielding a resulting vector that is subsequently divided into two components: the latent code  $z$  (yellow squares) and the surplus (gray squares). The Unpad function in Equation 7 removes the unused segment. The unpadded latent code  $z$  is fed to the contraction module and simultaneously padded with zeros (white squares) before being passed to the injective decoder. **(B) Contraction (blue box):** The Jacobian network output, given the latent codes, is reshaped into a square matrix and transformed into a negative definite matrix using Equation 2. The numerical integral solver then computes the latent velocity  $\dot{z}$ . Later, using Eq. 8,  $\dot{z}$  is mapped to the input-space velocity via the decoder’s Jacobian  $J_{\mu_\xi}$ .

Unfortunately, even if  $\mu$  is injective, we cannot ensure that  $\text{Exp} \circ \mu$  is also injective (since  $\text{Exp}$  is surjective), which then breaks the stability guarantees of NCDS. Here we leverage the result that  $\text{Exp}$  is a diffeomorphism as long as we restrict ourselves to the first cover of  $\mathfrak{so}(3)$ . Specifically, if we choose a decoder architecture such that  $\mu : \mathcal{Z} \rightarrow \mathcal{B}_\pi$  is injective and have outputs on the first cover, then  $\text{Exp} \circ \mu$  is injective, and stability is ensured (see Appendix A.5). Note that during training, observed rotation matrices can be mapped directly into the first cover of  $\mathfrak{so}(3)$  using the logarithmic map. When the system state consists of both rotations and positions, we simply decode to higher dimensional variables and apply exponential and logarithmic maps on the appropriate dimensions. Figure 4 depicts this general form of NCDS.

### 3.4 OBSTACLE AVOIDANCE VIA MATRIX MODULATION

In real-world scenarios, obstacle avoidance is critical to achieving safe autonomous robots. Thus, the learned dynamical system should effectively handle previously unseen obstacles, without interfering with the overall contracting behavior of the system. Fortunately, NCDS can easily be adapted to perform contraction-preserving obstacle avoidance using the dynamic modulation matrix  $G$  from Huber et al. (2022). This approach locally reshapes the learned vector field in the proximity of obstacles, while preserving contraction guarantees. To avoid obstacles effectively, it is imperative to know both the position and geometry of the obstacle. This makes obstacle avoidance on the VAE latent space difficult as we need to map both obstacle location and geometry to the latent space  $\mathcal{Z}$ .

Instead, we directly apply the modulation matrix to the data space  $\mathcal{X}$  of the decoded dynamical system. Specifically, Huber et al. (2022) shows how to construct a modulation matrix  $G$  from the object location and geometry, such that the vector field  $\dot{x} = G(x)J_{\mu_\xi}(z)\dot{z}$  is both contractive and steers around the obstacle. With this minor modification, NCDS can be adapted to avoid obstacles. We provide the details in Appendix A.3. Note that this is tailored for obstacle avoidance at the end-effector level. In order to extend this to multiple-limb obstacle avoidance, the modulation matrix must be refined to incorporate the robot body, and NCDS must be trained using joint space trajectories.

## 4 EXPERIMENTAL RESULTS

To evaluate the efficiency of NCDS, we consider several synthetic and real tasks. Comparatively, we show that NCDS is the only method to scale gracefully to higher dimensional problems, due to the latent structure. We further demonstrate the ability to build dynamic systems on the Lie group of rotations and to avoid obstacles while ensuring stability. Neither of the baseline methods have such capabilities. Experimental details are in Appendix A.4, while ablation studies of NCDS modeling choices are in Appendix A.6.3. Furthermore, the videos are available at: <https://sites.google.com/view/neuralcontraction/home>.

**Datasets.** There are currently no established benchmarks for contraction-stable robot motion learning, so we focus on two datasets. First, we test our approach on the LASA dataset (Lemme et al., 2015), often used for benchmarking asymptotic stability. This consists of 26 different two-

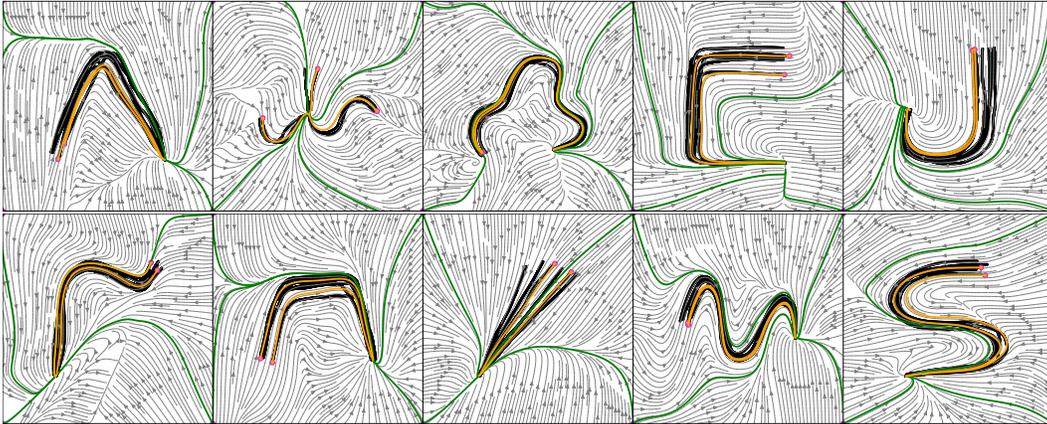


Figure 5: Visualization of the LASA-2D dataset: Gray contours represent the learned vector field, black trajectories depict demonstrations, and orange/green trajectories illustrate path integrals starting from the initial points of the demonstrations and plot corners. The magenta circles indicate the initial points of the path integrals.

dimensional hand-written trajectories, which the robot is tasked to follow. To evaluate our method, we have selected 10 different trajectories from this dataset. To ensure consistency and comparability, we preprocess all trajectories to all stop at the same target state. Additionally, we omit the initial few points of each demonstration, to ensure that the only state exhibiting zero velocity is the target state. To further complicate this easy task, we task the robot to follow 2, or 4 stacked trajectories, resulting in 4 (LASA-4D), or 8-dimensional (LASA-8D) data, respectively. We further consider a dataset consisting of 5 trajectories from a 7-DoF Franka-Emika Panda robot using a Cartesian impedance controller.

**Metrics.** We use dynamic time warping distance (DTWD) as the established quantitative measure of reproduction accuracy w.r.t. a demonstrated trajectory, assuming equal initial conditions (Ravichandar et al., 2017; Sindhwani et al., 2018). This is defined as,

$$\text{DTWD}(\tau_x, \tau_{x'}) = \sum_{j \in l(\tau_{x'})} \min_{i \in l(\tau_x)} \left( d(\tau_{x_i}, \tau_{x'_j}) \right) + \sum_{i \in l(\tau_x)} \min_{j \in l(\tau_{x'})} \left( d(\tau_{x_i}, \tau_{x'_j}) \right),$$

where  $\tau_x$  and  $\tau_{x'}$  are two trajectories (e.g. a path integral and a demonstration trajectory),  $d$  is a distance function (e.g. Euclidean distance), and  $l(\tau)$  is the length of trajectory  $\tau$ .

**Baseline methods.** Our work is the first contractive neural network architecture, so we cannot compare NCDS directly to methods with identical goals. Instead, we compare to existing methods that provide asymptotic stability guarantees. In particular, *Euclideanizing flow* (Rana et al., 2020), *Imitation flow* (Urain et al., 2020) and *SEDS* (Khansari-Zadeh & Billard, 2011).

#### 4.1 COMPARATIVE RESULTS

We first evaluate NCDS on two-dimensional trajectories for ease of visualization. Here data is sufficiently low-dimensional that we do not consider a latent structure. Figure 5 shows the learned vector fields (gray contours) for 10 diverse trajectory shapes chosen according to their difficulty from the LASA dataset, covering a wide range of demonstration patterns (black curves) and dynamics. We observe that NCDS effectively captures and replicates the underlying dynamics. We compute additional path integrals starting from outside the data support (green curves) to assess the generalization capability of NCDS beyond the observed demonstrations. Remarkably, NCDS successfully generates plausible trajectories even in regions not covered by the original demonstration data. Moreover, the yellow path integrals, starting from the initial points of the demonstrations, show that our approach can also reproduce the demonstrated trajectories accurately. This is evidence that the contractive construction is a viable approach to controlling the extrapolation properties of the neural network.

Table 1 shows average DTWD distances among five generated path integrals and five demonstration trajectories for both NCDS and baseline methods. For the two-dimensional problem, both Euclideanizing flow and Imitation flow outperform NCDS and SEDS, though all methods perform quite well. To investigate stability, Fig. 6 displays the average distance over time among five path integrals starting from random nearby initial points for different methods. We observe that only for NCDS does this distance decrease monotonically, which indicates that it is the only contractive method.

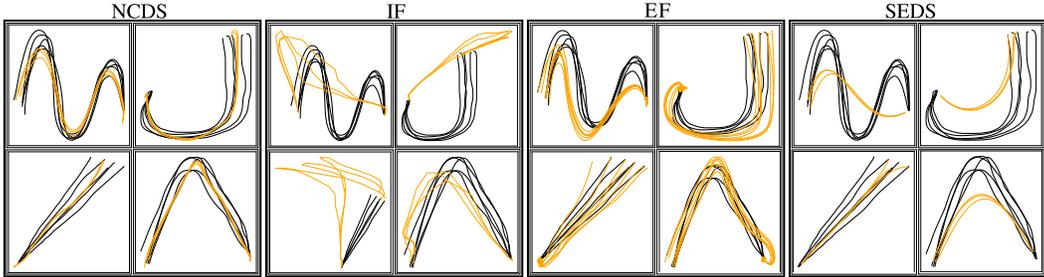


Figure 7: Path integrals generated using different methods on LASA-8D. Black curves are training data, while yellow are the learned path integrals. To construct the 8D dataset, we have concatenated 4 different 2D datasets.

To test how these approaches scale to higher-dimensional settings, we consider the LASA-4D and LASA-8D datasets, where we train NCDS with a two-dimensional latent representation. From Table 1, it is evident that the baseline methods quickly deteriorate as the data dimension increases, and only NCDS gracefully scales to higher dimensional data. This is also evident from Fig. 7 which shows the training data and reconstructions of different LASA-8D dimensions with different methods. These results clearly show the value of having a low-dimensional latent structure (more details in Appendix A.6.5).

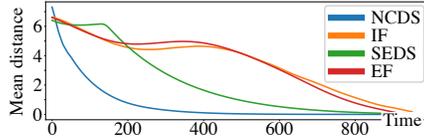


Figure 6: Average distance between random nearby trajectories over time for LASA-2D. Only NCDS monotonically decreases, i.e. it is the only contractive method.

To further compare our method, we consider a robotic setting where we evaluate all the methods on a real dataset of joint-space trajectories collected on a 7-DoF robot. The last column of Table 1 shows that only NCDS scales gracefully to high-dimensional joint-space data, and outperforms the baseline methods. The supplementary material includes a video showcasing the simulation environment in which each robot executes a drawing task, depicting V-shape trajectories on the table surface.

#### 4.2 OBSTACLE AVOIDANCE

Unlike baseline methods, our NCDS framework also provides dynamic collision-free motions. To assess the obstacle avoidance capabilities of our approach, we conducted a couple of experiments on the LASA dataset. The results of this experiment are reported in the third panel of Figure 8. As observed, the presence of an obstacle that completely blocks the demonstrations is represented by a red circle. Remarkably, NCDS successfully reproduced safe trajectories by effectively avoiding the obstacle and ultimately reaching the target. Moreover, the last panel in Figure 8 shows the obstacle avoidance in action as the robot’s end-effector navigates around the orange cylinder. These results experimentally show how the dynamic modulation matrix approach leads to obstacle-free trajectories while still guaranteeing contraction.

#### 4.3 ROBOT EXPERIMENT

To demonstrate the potential application of NCDS to robotics, we conducted several experiments on a 7-DoF Franka-Emika robotic manipulator, where an operator kinesthetically teaches the robot drawing tasks. The learned dynamics are executed using a Cartesian impedance controller. The details of the implementation and the specific network architecture are provided in Appendix A.4 and A.6.4.

	LASA-2D	LASA-4D	LASA-8D	7 DoF robot
Euclideanizing flow	<b>0.72 ± 0.12</b>	3.23 ± 0.34	10.22 ± 0.40	5.11 ± 0.30
Imitation flow	0.80 ± 0.24	<b>0.79 ± 0.22</b>	4.69 ± 0.52	2.63 ± 0.37
SEDS	1.60 ± 0.44	3.08 ± 0.20	4.85 ± 1.64	2.69 ± 0.18
NCDS	1.37 ± 0.40	0.98 ± 0.15	<b>2.28 ± 0.24</b>	<b>1.18 ± 0.16</b>

Table 1: Average dynamic time warping distances (DTWD) between different approaches. NCDS exhibits comparable performance in low dimensions, while clearly being superior in high dimensions.

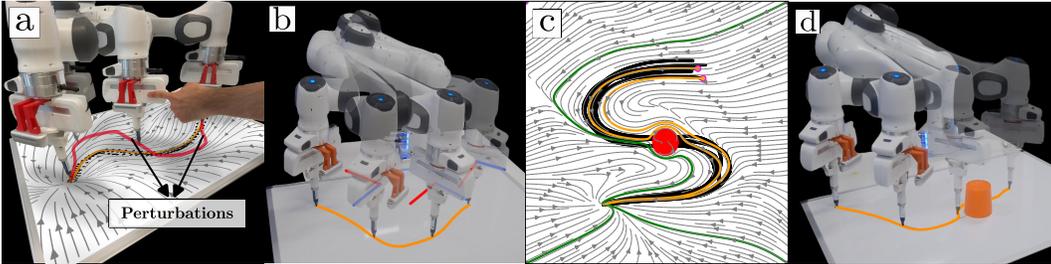


Figure 8: Robot and obstacle avoidance. Left two panels show robot experiments with orange and red paths illustrating unperturbed and perturbed motion. The first panel shows a learned vector field in  $\mathbb{R}^3$  with a constant orientation. In the second panel, the vector field expands to  $\mathbb{R}^3 \times \mathcal{SO}(3)$ , aligning the end-effector’s orientation with the motion direction. Superimposed robot images depict frames of executed motion. The right two panels display how an obstacle locally reshapes the learned vector field using the modulation matrix. In the third panel, gray contours represent the learned vector field, black trajectories depict the demonstrations, and orange/green trajectories are path integrals starting from both initial points of the demonstrations and plot corners. Magenta and red circles indicate the initial points of the path integrals and the obstacle, correspondingly. The last panel shows how the robot successfully avoids the obstacle (orange cylinder).

Firstly, we demonstrated 7 sinusoidal trajectories while keeping the end-effector orientation constant, which means that the robot end-effector dynamics only evolved in  $\mathbb{R}^3$ . As Fig. 8-a shows, the robot was able to successfully reproduce the demonstrated dynamics by following the learned vector field encoded by NCDS. Importantly, we also tested the extrapolation capabilities of our approach by introducing physical perturbations to the robot end-effector, under which the robot satisfactorily adapted and completed the task (see the video attached to the supplementary material).

The next experiment was aimed at testing our extension to handle orientation dynamics by learning dynamics evolving in  $\mathbb{R}^3 \times \mathcal{SO}(3)$ , i.e. full-pose end-effector movements. We collected several V-shape trajectories on a table surface in which the robot end-effector always faces the direction of the movement, as shown in Fig. 8-b. To evaluate the performance of this setup on a more complex dataset, we generated a synthetic LASA dataset in  $\mathbb{R}^3 \times \mathcal{SO}(3)$ . The results show that NCDS adeptly learns a contractive dynamical system from this dataset (see Appendix A.6.1). Another set of experiment was aimed at validating our approach on joint-space dynamic motions in a simulation environment (Corke & Haviland, 2021). These results show that among the baseline methods, NCDS outperforms others in generating motion within the joint space. (see Appendix A.6.6). Furthermore, we have further extended the joint space experiments to learning human motion skills. The findings indicate that NCDS successfully learns a contractive dynamical system from complex demonstrations in high-dimensional spaces. (see Appendix A.6.2).

## 5 DISCUSSION

In this paper, we proposed the first mechanism for designing neural networks that are contractive. This allows us to build a flexible class of models, called *neural contractive dynamical systems* (NCDS), that can learn dynamical systems from demonstrations while retaining contractive stability guarantees. Since contraction is built into the neural network architecture, the resulting model can be fitted directly to demonstration data using standard unconstrained optimization. To cope with the complexities of high-dimensional dynamical systems, we further developed a variant of VAEs that ensures latent contractive dynamical systems decode to contractive systems. Empirically, we found that the latent representation tends to behave simpler than the original demonstrations, which further increases the robustness of latent NCDS. We have further extended NCDS to the Lie group consisting of rotation matrices in order to model real-world robotic motions. Finally, we showed that our approach can be adapted to avoid obstacles while remaining contractive using the modulation techniques from Huber et al. (2022).

**Limitations.** Empirically, we find that NCDS is sensitive to the choice of numerical integration scheme when solving Eq. 4. Specifically, it is important to use integration schemes with adaptive step sizing for the method to be reliable. In all experiments, we used the standard DOPRI method. One concern with adaptive step sizing is that computation time is not fixed, which may negatively influence real-time implementations. On a related note, the need for numerical integration does increase the running time associated with our neural architecture compared to standard feedforward networks (see Appendix A.6.3). We, however, consider that a small price for stability guarantees.

## ACKNOWLEDGMENTS

This work was supported by a research grant (42062) from VILLUM FONDEN. This project received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement 757360). The work was partly funded by the Novo Nordisk Foundation through the Center for Basic Machine Learning Research in Life Science (NNF20OC0062606).

## REFERENCES

- H. Beik-Mohammadi, S. Hauberg, G. Arvanitidis, G. Neumann, and L. Rozo. Learning Riemannian manifolds for geodesic motion skills. In *Robotics: Science and Systems (R:SS)*, 2021. URL <https://roboticsconference.org/2021/program/papers/082/index.html>.
- A. Billard, S. Calinon, and R. Dillmann. Learning from humans. In B. Siciliano and O. Khatib (eds.), *Handbook of Robotics*, chapter 74, pp. 1995–2014. Springer, Secaucus, NJ, USA, 2016. URL [https://doi.org/10.1007/978-3-319-32552-1\\_74](https://doi.org/10.1007/978-3-319-32552-1_74). 2nd Edition.
- Caroline Blocher, Matteo Saveriano, and Dongheui Lee. Learning stable dynamical systems using contraction theory. In *IEEE International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pp. 124–129, 2017. URL <https://ieeexplore.ieee.org/document/7992901>.
- Johann Brehmer and Kyle Cranmer. Flows for simultaneous manifold learning and density estimation. In *Neural Information Processing Systems (NeurIPS)*, pp. 442–453, 2020. URL [https://proceedings.neurips.cc/paper\\_files/paper/2020/file/051928341be67dcb03f0e04104d9047-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2020/file/051928341be67dcb03f0e04104d9047-Paper.pdf).
- Nutan Chen, Alexej Klushyn, Alexandros Paraschos, Djalel Benbouzid, and Patrick Van der Smagt. Active learning based on data uncertainty and model sensitivity. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 1547–1554, 2018a. URL <http://doi.org/10.1109/IROS.2018.8593552>.
- Ricky T. Q. Chen. torchdiffeq, 2018. URL <https://github.com/rtqichen/torchdiffeq>.
- Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Neural Information Processing Systems (NeurIPS)*, 2018b. URL <https://proceedings.neurips.cc/paper/2018/file/69386f6bb1dfed68692a24c8686939b9-Paper.pdf>.
- Peter Corke and Jesse Haviland. Not your grandmother’s toolbox – the robotics toolbox reinvented for Python. In *IEEE International Conference on Robotics and Automation (ICRA)*, pp. 11357–11363, 2021. URL <https://10.1109/ICRA48506.2021.9561366>.
- Charles Dawson, Sicun Gao, and Chuchu Fan. Safe control with learned certificates: A survey of neural Lyapunov, barrier, and contraction methods for robotics and control. *IEEE Transactions on Robotics (T-RO)*, 39(3):1749–1767, 2023. URL <https://doi.org/10.1109/TRO.2022.3232542>.
- Luca Falorsi, Pim de Haan, Tim R. Davidson, and Patrick Forré. Reparameterizing distributions on Lie groups. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 3244–3253, 2019. URL <https://proceedings.mlr.press/v89/falorsi19a.html>.
- Lukas Huber, Aude Billard, and Jean-Jacques Slotine. Avoidance of convex and concave obstacles with convergence ensured through contraction. *IEEE Robotics and Automation Letters (RA-L)*, 4(2):1462–1469, 2019. doi: <http://doi.org/10.1109/LRA.2019.2893676>.
- Lukas Huber, Jean-Jacques Slotine, and Aude Billard. Avoiding dense and dynamic obstacles in enclosed spaces: Application to moving in crowds. *IEEE Transactions on Robotics*, 38(5): 3113–3132, 2022. URL <https://ieeexplore.ieee.org/document/9765824>.

- Chia-Man Hung, Shaohong Zhong, Walter Goodwin, Oiwi Parker Jones, Martin Engelcke, Ioannis Havoutis, and Ingmar Posner. Reaching through latent space: From joint statistics to path planning in manipulation. *IEEE Robotics and Automation Letters (RA-L)*, 7(2):5334–5341, 2022. URL <http://doi.org/10.1109/LRA.2022.3152697>.
- Jerome Jouffroy and Thor I. Fossen. A tutorial on incremental stability analysis using contraction theory. *Modeling, Identification and Control*, 2010. ISSN 0332-7353. URL <https://doi.org/10.4173/mic.2010.3.2>.
- S. Mohammad Khansari-Zadeh and Aude Billard. Learning stable nonlinear dynamical systems with Gaussian mixture models. *IEEE Transactions on Robotics*, 27(5):943–957, 2011. doi: 10.1109/TRO.2011.2159412.
- S. Mohammad Khansari-Zadeh and Aude Billard. Learning control Lyapunov function to ensure stability of dynamical system-based robot reaching motions. *Robotics and Autonomous Systems (RAS)*, 62(6):752–765, 2014. doi: <https://doi.org/10.1016/j.robot.2014.03.001>.
- Diederik P. Kingma and Max Welling. Auto-encoding variational Bayes. In *International Conference on Learning Representations (ICLR)*, 2014. URL <https://openreview.net/forum?id=33X9fd2-9FyZd>.
- Leo Kozachkov, Michaela M Ennis, and Jean-Jacques Slotine. RNNs of RNNs: Recursive construction of stable assemblies of recurrent neural networks. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL <https://openreview.net/forum?id=2dgB38geVEU>.
- Leo Kozachkov, Patrick Wensing, and Jean-Jacques Slotine. Generalization as dynamical robustness—the role of riemannian contraction in supervised learning. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856. URL <https://openreview.net/forum?id=Sb6p5mcefw>.
- 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*, 6(1), 2015. URL <http://doi.org/10.1515/pjbr-2015-0002>.
- Winfried Lohmiller and Jean-Jacques E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. ISSN 0005-1098. URL [https://doi.org/10.1016/S0005-1098\(98\)00019-3](https://doi.org/10.1016/S0005-1098(98)00019-3).
- Jonathan Lorraine and Safwan Hossain. JacNet: Learning functions with structured Jacobians. *INNF Workshop at the International Conference on Machine Learning (ICML)*, 2019. URL [https://invertibleworkshop.github.io/INNF\\_2019/accepted\\_papers/pdfs/INNF\\_2019\\_paper\\_10.pdf](https://invertibleworkshop.github.io/INNF_2019/accepted_papers/pdfs/INNF_2019_paper_10.pdf).
- Ian R. Manchester and Jean-Jacques E. Slotine. Control contraction metrics: Convex and intrinsic criteria for nonlinear feedback design. *IEEE Transactions on Automatic Control*, 62(6):3046–3053, 2017. URL <https://doi.org/10.1109/TAC.2017.2668380>.
- Christian Mandery, Ömer Terlemez, Martin Do, Nikolaus Vahrenkamp, and Tamim Asfour. Unifying representations and large-scale whole-body motion databases for studying human motion. *IEEE Transactions on Robotics*, 32(4):796–809, 2016. URL <https://ieeexplore.ieee.org/document/7506114>.
- Klaus Neumann and Jochen J. Steil. Learning robot motions with stable dynamical systems under diffeomorphic transformations. *Robotics and Autonomous Systems (RAS)*, 70:1–15, 2015. URL <https://doi.org/10.1016/j.robot.2015.04.006>.
- Alexander Luke Ian Norcliffe, Cristian Bodnar, Ben Day, Nikola Simidjievski, and Pietro Lio. On second order behaviour in augmented neural ODEs: A short summary. In *Workshop on the Symbiosis of Deep Learning and Differential Equations at NeurIPS*, 2021. URL <https://openreview.net/forum?id=XpmaGtI04ki>.

- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019. URL <https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>.
- Michael Poli, Stefano Massaroli, Atsushi Yamashita, Hajime Asama, Jinkyoo Park, and Stefano Ermon. Torchdyn: Implicit models and neural numerical methods in pytorch. URL <https://github.com/DiffEqML/torchdyn/>.
- Muhammad Asif Rana, Anqi Li, Dieter Fox, Byron Boots, Fabio Ramos, and Nathan Ratliff. Euclideanizing flows: Diffeomorphic reduction for learning stable dynamical systems. In *Conference on Learning for Dynamics and Control (LADC)*, pp. 630–639, 2020. URL <https://proceedings.mlr.press/v120/rana20a.html>.
- Harish Chaandar Ravichandar and Ashwin Dani. Learning position and orientation dynamics from demonstrations via contraction analysis. *Autonomous Robots*, 43(4):897–912, 2019. ISSN 0929-5593. URL <https://doi.org/10.1007/s10514-018-9758-x>.
- Harish Chaandar Ravichandar, Iman Salehi, and Ashwin Dani. Learning partially contracting dynamical systems from demonstrations. In *Conference on Robot Learning (CoRL)*, pp. 369–378, 2017. URL <https://proceedings.mlr.press/v78/ravichandar17a.html>.
- Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning (ICML)*, pp. 1278–1286, 2014.
- Stefan Schaal, Auke Ijspeert, and Aude Billard. Computational approaches to motor learning by imitation. *Phil. Trans. R. Soc. Lond. B*, 358:537–547, 2003. URL <http://doi.org/10.1098/rstb.2002.1258>.
- Malcolm David Shuster. A survey of attitude representation. *Journal of The Astronautical Sciences*, 41:439–517, 1993.
- Vikas Sindhwani, Stephen Tu, and Mohi Khansari. Learning contracting vector fields for stable imitation learning. *arXiv*, 1804.04878, 2018. URL <https://arxiv.org/abs/1804.04878>.
- Joan Solà, Jérémie Deray, and Dinesh Atchuthan. A micro lie theory for state estimation in robotics. *CoRR*, abs/1812.01537, 2018. URL <http://arxiv.org/abs/1812.01537>.
- Dawei Sun, Susmit Jha, and Chuchu Fan. Learning certified control using contraction metric. In *Conference on Robot Learning (CoRL)*, 2020. URL <http://arxiv.org/abs/2011.12569>.
- Esteban G Tabak and Cristina V Turner. A family of nonparametric density estimation algorithms. *Communications on Pure and Applied Mathematics*, 66(2):145–164, 2013.
- Hiroyasu Tsukamoto and Soon Jo Chung. Neural contraction metrics for robust estimation and control: A convex optimization approach. *IEEE Control Systems Letters*, 5:211–216, 2021. ISSN 24751456. URL <https://doi.org/10.1109/LCSYS.2020.3001646>.
- Hiroyasu Tsukamoto, Soon-Jo Chung, Jean-Jacques Slotine, and Chuchu Fan. A theoretical overview of neural contraction metrics for learning-based control with guaranteed stability. In *IEEE Conference on Decision and Control (CDC)*, pp. 2949–2954, 2021a. URL <https://ieeexplore.ieee.org/document/9682859>.
- Hiroyasu Tsukamoto, Soon-Jo Chung, and Jean-Jacques E. Slotine. Contraction theory for nonlinear stability analysis and learning-based control: A tutorial overview. *Annual Reviews in Control*, 52:135–169, 2021b. ISSN 1367-5788. URL <https://doi.org/10.1016/j.arcontrol.2021.10.001>.

- J. Urain, M. Ginesi, D. Tateo, and J. Peters. Imitationflow: Learning deep stable stochastic dynamic systems by normalizing flows. In *International Conference on Intelligent Robots and Systems (IROS)*, pp. 5231–5237. IEEE, 2020. URL <http://ras.papercept.net/images/temp/IROS/files/1340.pdf>.
- Julen Urain, Davide Tateo, and Jan Peters. Learning stable vector fields on lie groups. *IEEE Robotics and Automation Letters*, 7(4):12569–12576, 2022. doi: 10.1109/LRA.2022.3219019. URL <https://ieeexplore.ieee.org/document/9935105>.
- Keyulu Xu, Mozhi Zhang, Jingling Li, Simon Shaolei Du, Ken-Ichi Kawarabayashi, and Stefanie Jegelka. How neural networks extrapolate: From feedforward to graph neural networks. In *International Conference on Learning Representations (ICLR)*, 2021. URL <https://openreview.net/forum?id=UH-cmocLJC>.
- Jiechao Zhang, Hadi Beik-Mohammadi, and Leonel Rozo. Learning Riemannian stable dynamical systems via diffeomorphisms. In *Conference on Robot Learning (CoRL)*, 2022a. URL <https://openreview.net/pdf?id=o8dLx8OVcNk>.
- Yu Zhang, Long Cheng, Houcheng Li, and Ran Cao. Learning accurate and stable point-to-point motions: A dynamic system approach. *IEEE Robotics and Automation Letters (RA-L)*, 7(2): 1510–1517, 2022b. URL <https://doi.org/10.1109/LRA.2022.3140677>.

## A APPENDIX

### A.1 NEGATIVE DEFINITENESS OF A SYMMETRIC MATRIX

Let  $\mathbf{A}$  be a real  $n \times n$  symmetric matrix, and let  $\mathbf{A}_S = \frac{1}{2}(\mathbf{A} + \mathbf{A}^\top)$  denote its symmetric part. Suppose that  $\mathbf{A}_S$  is negative definite, which means that for all nonzero vectors  $\mathbf{v} \in \mathbb{R}^n$ , we have  $\mathbf{v}^\top \mathbf{A}_S \mathbf{v} < 0$ . Now let  $\mathbf{w}$  be an eigenvector of  $\mathbf{A}$  with corresponding eigenvalue  $\lambda$ , so that  $\mathbf{A}\mathbf{w} = \lambda\mathbf{w}$ . Then we have  $\mathbf{w}^\top \mathbf{A}\mathbf{w} = \mathbf{w}^\top \lambda\mathbf{w} = \lambda\mathbf{w}^\top \mathbf{w} = \lambda|\mathbf{w}|^2$ . Taking the transpose of the equation  $\mathbf{A}\mathbf{w} = \lambda\mathbf{w}$ , we have  $\mathbf{w}^\top \mathbf{A}^\top = \lambda\mathbf{w}^\top$ . Multiplying on the left by  $\mathbf{w}$  and using the fact that  $\mathbf{A}$  is symmetric, we obtain  $\mathbf{w}^\top \mathbf{A}\mathbf{w} = \lambda\mathbf{w}^\top \mathbf{w}$ . Therefore, we have  $\mathbf{w}^\top \mathbf{A}\mathbf{w} = \mathbf{w}^\top \mathbf{A}_S \mathbf{w} + \mathbf{w}^\top (\mathbf{A} - \mathbf{A}_S) \mathbf{w} \leq \mathbf{w}^\top \mathbf{A}_S \mathbf{w} < 0$ . Since  $\mathbf{w}^\top \mathbf{A}\mathbf{w} < 0$  for any nonzero eigenvector  $\mathbf{w}$  of  $\mathbf{A}$ , we conclude that  $\mathbf{A}$  is negative definite. Thus, the negative definiteness of the symmetric part  $\mathbf{A}_S$  implies the negative definiteness of  $\mathbf{A}$ .

### A.2 LIE OPERATORS OF $\mathcal{SO}(3)$

To map back and forth between the Lie group  $\mathcal{SO}(3)$  and its associated Lie algebra  $\mathfrak{so}(3)$ , we employ the logarithmic  $\text{Log} : \mathcal{SO}(3) \rightarrow \mathfrak{so}(3)$ , and exponential  $\text{Exp} : \mathfrak{so}(3) \rightarrow \mathcal{SO}(3)$  maps, which are defined as follows,

$$\text{Exp}([\mathbf{r}]_\times) = \mathbb{I} + \frac{\sin(\zeta)}{\zeta} [\mathbf{r}]_\times + \frac{1 - \cos(\zeta)}{\zeta^2} [\mathbf{r}]_\times^2, \quad (9)$$

$$\text{Log}(\mathbf{R}) = \zeta \frac{\mathbf{R} - \mathbf{R}^\top}{2 \sin(\zeta)}, \quad (10)$$

where  $\zeta = \arccos\left(\frac{\text{Tr}(\mathbf{R}) - 1}{2}\right)$ . Moreover,  $\mathbf{R}$  represents the rotation matrix, and  $[\mathbf{r}]_\times$  denotes the skew-symmetric matrix associated with the coefficient vector  $\mathbf{r}$ .

### A.3 OBSTACLE AVOIDANCE VIA MATRIX MODULATION

We here detail the used matrix modulation technique, due to Huber et al. (2022), which we use for obstacle avoidance. This approach locally reshapes the learned vector field in the proximity of obstacles, while preserving contraction guarantees. To avoid obstacles effectively, it is imperative to know both the position and geometry of the obstacle. This makes obstacle avoidance on the VAE latent space particularly difficult as we need to map the obstacle location and geometry to the latent space  $\mathcal{Z}$ . Hence, we directly apply the modulation matrix to the data space  $\mathcal{X}$  of the decoded dynamical system.

Formally, given the modulation matrix  $\mathbf{G}$ , we can reshape the data space vector field to dynamically avoid an obstacle as follows,

$$\dot{\mathbf{x}} = \mathbf{G}(\mathbf{x}) \mathbf{J}_{\mu_\xi}(\mathbf{z}) \dot{\mathbf{z}}, \quad \text{with} \quad \mathbf{G}(\mathbf{x}) = \mathbf{E}(\mathbf{x}) \mathbf{D}(\mathbf{x}) \mathbf{E}(\mathbf{x})^{-1}, \quad (11)$$

where  $\mathbf{E}(\mathbf{x})$  and  $\mathbf{D}(\mathbf{x})$  are the basis and diagonal eigenvalue matrices computed as,

$$\mathbf{E}(\mathbf{x}) = [r(\mathbf{x}) \mathbf{e}_1(\mathbf{x}) \dots \mathbf{e}_{d-1}(\mathbf{x})], \quad \text{and} \quad \mathbf{D}(\mathbf{x}) = \text{diag}(\lambda^r(\mathbf{x}) \lambda^e(\mathbf{x}), \dots, \lambda^e(\mathbf{x})), \quad (12)$$

where  $r(\mathbf{x}) = \frac{\mathbf{x} - \mathbf{x}_r}{\|\mathbf{x} - \mathbf{x}_r\|}$  is a reference direction computed w.r.t. a reference point  $\mathbf{x}_r$  on the obstacle, and the tangent vectors  $\mathbf{e}_i$  form an orthonormal basis to the gradient of the distance function  $\Gamma(\mathbf{x})$  (see (Huber et al., 2022) for its full derivation). Moreover, the components of the matrix  $\mathbf{D}$  are defined as  $\lambda^r(\mathbf{x}) = 1 - \left(\frac{1}{\Gamma(\mathbf{x})}\right)^{\frac{1}{\rho}}$ ,  $\lambda^e(\mathbf{x}) = 1 + \left(\frac{1}{\Gamma(\mathbf{x})}\right)^{\frac{1}{\rho}}$ , where  $\rho \in \mathbb{R}_+$  is a reactivity factor. Note that the matrix  $\mathbf{D}$  modulates the dynamics along the directions of the basis defined by the set of vectors  $r(\mathbf{x})$  and  $\mathbf{e}(\mathbf{x})$ . As stated in (Huber et al., 2022), the function  $\Gamma(\cdot)$  monotonically increases w.r.t the distance from the obstacle's reference point  $\mathbf{x}_r$ , and it is, at least, a  $C^1$  function. Importantly, the modulated dynamical system  $\dot{\mathbf{x}} = \mathbf{G}(\mathbf{x}) \mathbf{J}_{\mu_\xi}(\mathbf{z}) \dot{\mathbf{z}}$  still guarantees contractive stability, which can be proved by following the same proof provided in (Huber et al., 2019, App. B.5).

### A.4 IMPLEMENTATION DETAILS

Both the Variational Autoencoder (VAE) and the Jacobian network  $\mathbf{J}_\theta$  were implemented using the PyTorch framework (Paszke et al., 2019). For the VAE, we employed an injective generator based

**Algorithm 1:** Neural Contractive Dynamical Systems (NCDS): Training in task space**Data:** Demonstrations:  $\tau_n = \{(\mathbf{x}_t, \mathbf{R}_t)\}_n, n \in [1, N], t \in [1, T_n]$ **Result:** Learned contractive dynamical system $\mathbf{r}_{n,t} = \text{Log}(\mathbf{R}_{n,t}); //$  Obtaining skew-symmetric coefficients via the Lie algebra $\mathbf{p}_{n,t} = [\mathbf{x}_{n,t}, \mathbf{r}_{n,t}]; //$  Create new state vector $\text{argmin}_{\xi} \mathcal{L}_{\text{ELBO}}(\xi^*; \mathbf{p}_{n,t}); //$  Train the VAE $\mathbf{z}_{n,t} = \mu_{\xi}^{-1}(\mathbf{p}_{n,t}); //$  Encode all poses using the trained VAE $\dot{\mathbf{z}}_{n,t} = \frac{\mathbf{z}_{n,t+1} - \mathbf{z}_{n,t}}{\Delta t}; //$  Estimate latent velocities via finite differences $\text{argmin}_{\theta} \mathcal{L}_{\text{Jac}}(\theta^*; \mathbf{p}_{n,t}); //$  Train the Jacobian network**Algorithm 2:** Neural Contractive Dynamical Systems (NCDS): Robot Control Scheme**Data:** Current state of the robot end-effector at time  $t$ :  $[\mathbf{x}_t, \mathbf{R}_t]$ **Result:** Velocity of the end-effector at current time step  $\dot{\mathbf{x}}_t$  $\mathbf{r}_t = \text{Log}(\mathbf{R}_t); //$  Obtaining skew-symmetric coefficients $\mathbf{p}_t = [\mathbf{x}_t, \mathbf{r}_t]; //$  Create new state vector $\mathbf{z}_t = \mu_{\xi}^{-1}(\mathbf{p}_t); //$  Compute the latent state $\hat{\mathbf{z}}_t = f(\mathbf{z}_t); //$  Compute the latent velocity $\mathbf{J}_{\mu_{\xi}}(\mathbf{z}_t) = \frac{\partial \mu_{\xi}}{\partial \mathbf{z}_t}; //$  Compute the Jacobian of the decoder $\dot{\mathbf{x}}_t = \mathbf{J}_{\mu_{\xi}}(\mathbf{z}_t) \hat{\mathbf{z}}_t; //$  Compute input space velocity

on M-flows (Brehmer & Cranmer, 2020), specifically using rational-quadratic neural spline flows. These flows were structured with three coupling layers. Within each coupling transform, half of the input values underwent elementwise transformation using a monotonic rational-quadratic spline. The parameters of these splines were determined through a residual network comprising two residual blocks, with each block consisting of a single hidden layer containing 30 nodes. Throughout the network, we employed Tanh activations and did not incorporate batch normalization or dropout techniques. The rational-quadratic splines were constructed with ten bins, evenly distributed over the range of  $(-10, 10)$ . The Jacobian network was implemented using a neural network architecture consisting of two hidden layers, with each layer containing 500 nodes. The output of this network was reshaped to form a square matrix. To facilitate the integration process, we used an off-the-shelf numerical integrator called 'odeint' from the `torchdiffeq` Python package (Chen, 2018). This package provides efficient implementations of various numerical integration methods. In our experiments, we specifically utilized the Runge-Kutta and `dopri5` integration methods, which are well-known and widely used for solving ordinary differential equations.

**Loss functions:** The VAE is trained using the standard evidence lower bound (ELBO):

$$\mathcal{L}_{\text{ELBO}} = \mathbb{E}_{q_{\xi}(\mathbf{z}|\mathbf{x})} [\log(p_{\phi}(\mathbf{x}|\mathbf{z}))] - \text{KL}(q_{\xi}(\mathbf{z}|\mathbf{x})||p(\mathbf{z})), \quad (13)$$

where KL denotes the Kullback-Leibler divergence. Thereafter, the Jacobian network is trained according to the loss function  $\mathcal{L}_{\text{Jac}}$ , which measures the error between the observed and approximated next state,

$$\mathcal{L}_{\text{Jac}} = \|\mathbf{z}_{t+1} - (\mathbf{z}_t + \hat{\mathbf{z}}_t)\|^2, \quad (14)$$

where  $\mathbf{z}_t$ ,  $\mathbf{z}_{t+1}$ , and  $\hat{\mathbf{z}}_t$  represent the current and next observed latent states, along with the calculated latent velocity.

**Algorithms:** In this section, we elaborate on the training and robot control schemes for NCDS. The steps for training the Variational Autoencoder (VAE) and Jacobian network are outlined in Algorithm 1. Furthermore, the steps for employing NCDS to control a robot are detailed in Algorithm 2. As the algorithms show, the training of the latent NCDS is not fully end-to-end. In the latter case, we first train the VAE (end-to-end), and then train the latent NCDS using the encoded data.

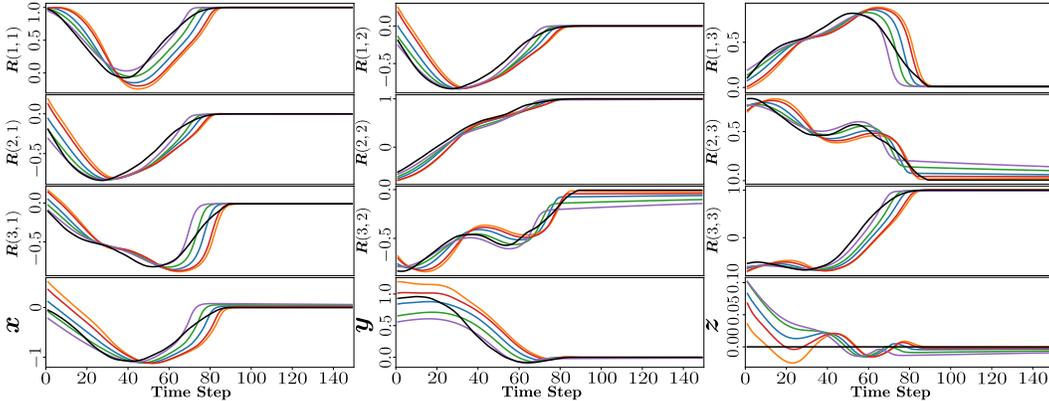


Figure 11: Time series of demonstrations and path integrals in  $\mathbb{R}^3 \times \mathcal{SO}(3)$

### A.5 AN INJECTIVE DECODER ARCHITECTURE OVER THE $\pi$ -BALL $\mathcal{B}_\pi$

To ensure that a decoder neural network has outputs over the  $\pi$ -ball (Sec. 3.3), we introduce a simple layer. Let  $f : \mathbb{R}^d \rightarrow \mathbb{R}^D$  be an injective neural network, where injectivity is only considered over the image of  $f$ . If we add a TanH-layer, then the output of the resulting network is the  $[-1, 1]^D$  box, i.e.  $\tanh(f(z)) : \mathbb{R}^d \rightarrow [-1, 1]^D$ . We can further introduce the function,

$$b(x) = \begin{cases} \frac{\|x\|_\infty}{\|x\|_2} x & x \neq \mathbf{0} \\ x & x = \mathbf{0} \end{cases} \quad (15)$$

This function smoothly (and invertible) deforms the  $[-1, 1]^D$  box into the unit ball (Fig. 9). Then, the function,

$$\hat{f}(z) = \pi b(\tanh(f(z))), \quad (16)$$

is injective and has the signature  $\hat{f} : \mathbb{R}^d \rightarrow \mathcal{B}_\pi(D)$ .

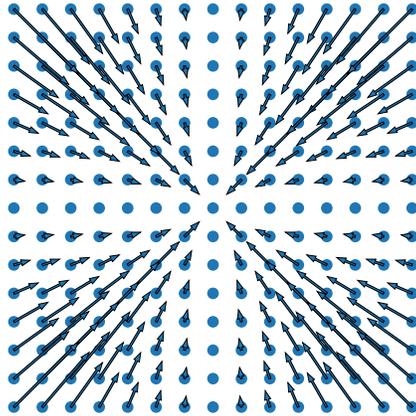


Figure 9: An illustration of the function  $b(x)$  in two dimensions.

### A.6 ADDITIONAL EXPERIMENTS

#### A.6.1 LEARNING $\mathbb{R}^3 \times \mathcal{SO}(3)$

The orientation trajectories of the robot experiments tend to show relatively simple dynamics. Consequently, the learning of these trajectories may not fully showcase the true potential of the approach. As a result, we conducted additional experiments to demonstrate the capacity of NCDS to encode full-pose (i.e., position-orientation) dynamic motions, which present a more challenging and comprehensive evaluation scenario. To construct a challenging dataset on  $\mathcal{SO}(3)$ , we projected the LASA datasets onto a 3-sphere, thereby generating synthetic quaternion data. Then we transformed this dataset to rotation matrices on  $\mathcal{SO}(3)$ . Later, we applied the Log map to project these points onto the Lie algebra. To construct the position data in  $\mathbb{R}^3$ , we simply project the 2D LASA dataset to  $\mathbb{R}^3$  by concatenating a zero to the vector. Consequently, we concatenate this vector with the vector on the Lie algebra representing the orientation. This will produce a 6 dimensional state vector in  $\mathbb{R}^6$ . To increase the level of complexity, we employed two distinct LASA datasets for the position and orientation dimensions. Figure 10, depicts the latent dynamical system, the black dots represent the demonstration points, yellow path integrals start at the initial point of the demonstration, and green path integrals start in

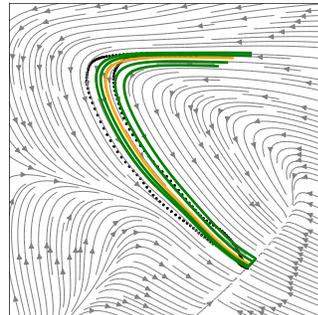


Figure 10: The background depicts the contours of the latent vector field, green and yellow curves depict the path integrals, and black dots represent demonstration in latent space.

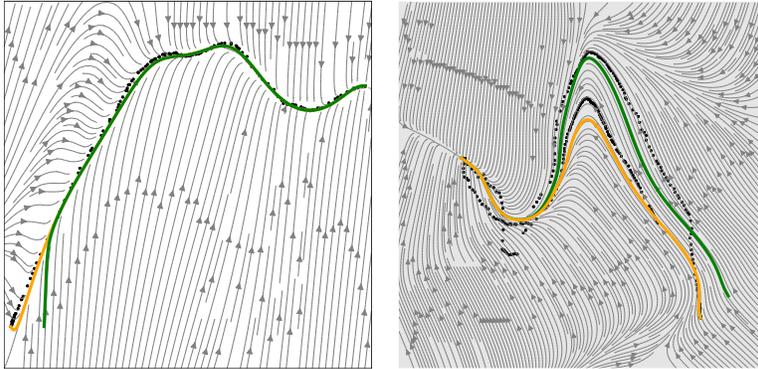


Figure 12: *Left*: Latent path integrals with NCDS trained in joint space  $\mathbb{R}^{44}$ . *Right*: Latent path integrals with NCDS trained on full human motion space ( $\mathbb{R}^{44} \times \mathbb{R}^3 \times \mathcal{SO}(3)$ ). The background depicts the contours of the latent vector field, green and yellow curves depict the path integrals, and black dots represent demonstration in latent space.

random points around the demonstrations. Figure 11 shows a time series of input space trajectory in  $\mathbb{R}^3$  and  $\mathcal{SO}(3)$ . Moreover, the black curves show the demonstration data and other trajectories show the path integrals in each dimension. The results show that NCDS is able to learn complex non-linear contractive dynamical systems in  $\mathbb{R}^3 \times \mathcal{SO}(3)$ .

#### A.6.2 LEARNING HUMAN MOTIONS

To assess the model’s performance in a more complex environment, we conducted a recent experiment using the KIT Whole-Body Human Motion Database (Mandery et al., 2016). This dataset captures subject-specific motions, which are standardized based on the subject’s height and weight using a reference kinematics and dynamics model known as the master motor map (MMM).

**Motion in Joint space  $\mathbb{R}^{44}$ :** Focusing on the Tennis forehand motion, we selected this specific skill from the dataset, where each point in the motion trajectories is defined in a 44-dimensional joint space. Figure 12–*left* displays the latent contractive vector field generated by NCDS. The background depicts the contours of the latent vector field, and green and yellow curves represent latent path integrals when the initial point is respectively distant from and coincident with the initial point of the demonstration. Figure 13 illustrates how the human dummy replicates each of the generated motions. The upper row of this plot depicts the progressive evolution of the demonstrated motion from left to right. Meanwhile, the middle and bottom rows illustrate the motions generated by NCDS when the initial point is respectively distant from and coincident with the initial point of the demonstration.

**Motion in on full human motion space  $\mathbb{R}^{44} \times \mathbb{R}^3 \times \mathcal{SO}(3)$ :** The previous experiment has primarily concentrated on reconstructing motions within the joint space alone, which may be insufficient for some particular human motions. The data recorded from the human also includes the base (hip) link pose of the human dummy, which is essential for a complete representation of the motion. To address this, we introduced the Kick motion skill into our experiments, where each point of the motion trajectory is defined in a 44-dimensional joint space, along with the position and orientation data for the base link (hip link). This approach extends the representation of the motion to a comprehensive 50-dimensional input space.

Figure 12–*right* shows the latent path integrals generated by NCDS. Figure 14 shows the human dummy mimicking each generated motion. The upper row of this plot depicts the progressive evolution of the demonstration motion from left to right. Meanwhile, the middle and bottom rows illustrate the motions generated by NCDS when the initial point is respectively distant from and coincident with the initial point of the demonstration. This result demonstrates the inherent ability of Latent NCDS to compress a 50-dimensional nonlinear motion into a 2-dimensional space.

#### A.6.3 QUANTITATIVE RESULTS

**Regularization:** It is imperative to clarify that the regularization term in Equation 2, known as epsilon  $\epsilon$ , plays a crucial role in defining the upper bound for the eigenvalues of the Jacobian  $\hat{J}_f(x)$ ,

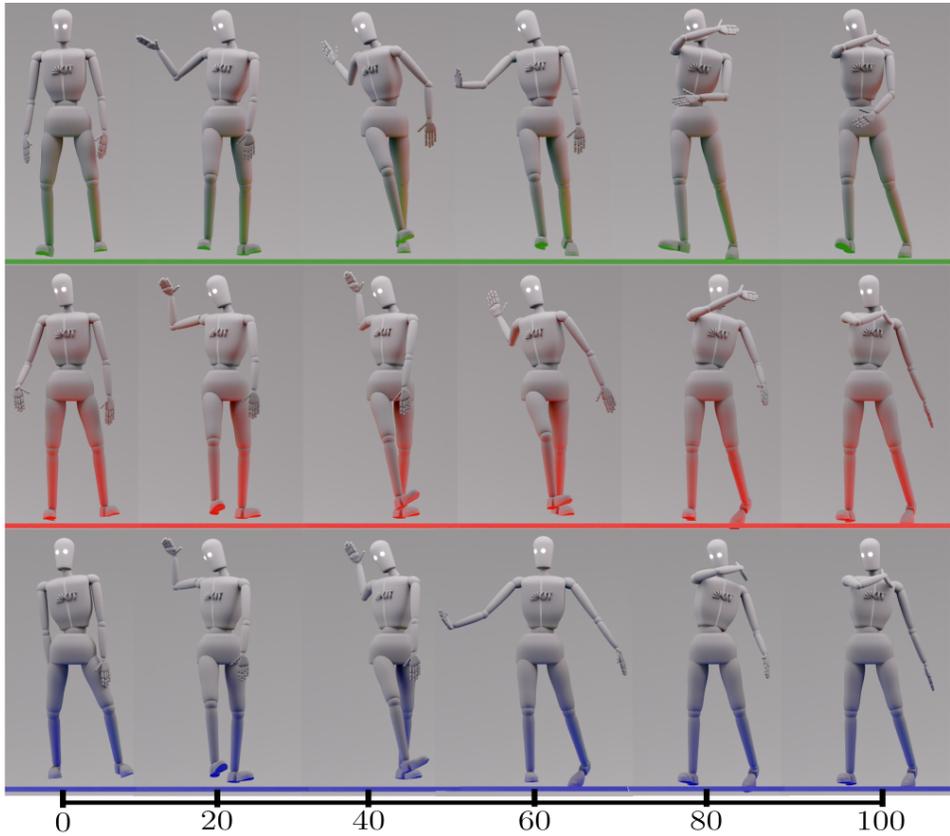


Figure 13: Generated Human Motion with NCDS trained in joint space  $\mathbb{R}^{44}$ : The upper row depicts the progressive evolution of the demonstration motion from left to right. Meanwhile, the middle and bottom rows illustrate the motions generated by NCDS when the initial point is respectively distant from and coincident with the initial point of the demonstration.

which inherently governs the contraction rate of the system. Figure 15 illustrates the effect of  $\epsilon$  on the accuracy of the system. Adjusting this value, particularly through increments, has the potential to negatively affect the network’s expressiveness, therefore compromising the reconstruction process. The observed negative effect can be attributed to the fact that this regularization is applied to the entire Jacobian matrix and not exclusively to the elements responsible for preserving the negative definiteness of the Jacobian. As a result, it might inadvertently penalize essential components of the Jacobian, leading to adverse consequences on the model’s expressivity. A more targeted approach that focuses solely on the relevant elements could potentially mitigate these undesired effects and enhance the regularization’s overall effectiveness.

**Unconstrained dynamical system:** To illustrate the influence of having a negative definite Jacobian in NCDS, we contrast with a system identical to NCDS except that we remove the definiteness constraint on the Jacobian network. Figure 16 shows the path integrals generated by both contractive (Figure 16-a) and unconstrained (Figure 16-b) dynamical systems. As anticipated, the dynamics generated by the unconstrained system lack stable behavior. However, the vector field aligns with the data trends in the data support regions. Furthermore, we performed similar experiments with two different baseline approaches: an MLP and a Neural Ordinary Differential Equation (NeuralODE) network, to highlight the effect of the contraction constraint on the behavior of the dynamical system. The MLP baseline uses a neural network with 2 hidden layers each with 100 neurons with tanh activation function. The NeuralODE baseline is implemented based on (Poli et al.), with 2 hidden layers each with 100 neurons. The model is then integrated into a NeuralODE framework. The NeuralODE is configured with the ‘adjoint’ sensitivity method and utilizes the ‘dopri5’ solver for both the forward and adjoint passes. It is important to mention that these baselines directly reproduce the velocity according to  $\dot{x} = f(x)$ . These models were trained with the same cost function as our NCDS’s Jacobian network loss defined in Equation 14. Both networks are trained for 1000 epochs

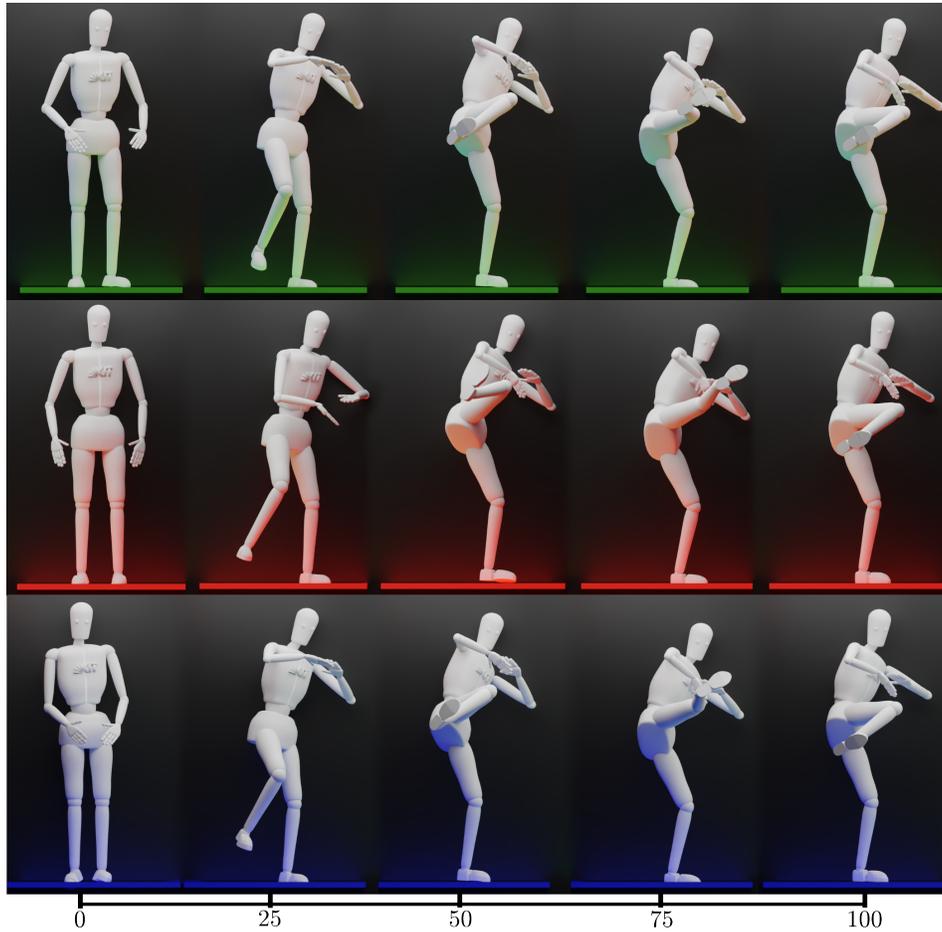


Figure 14: Generated Human Motion with NCDS trained on full human motion space ( $\mathbb{R}^{44} \times \mathbb{R}^3 \times SO(3)$ ): The upper row depicts the progressive evolution of the demonstration motion from left to right. Meanwhile, the middle and bottom rows illustrate the motions generated by NCDS when the initial point is respectively distant from and coincident with the initial point of the demonstration.

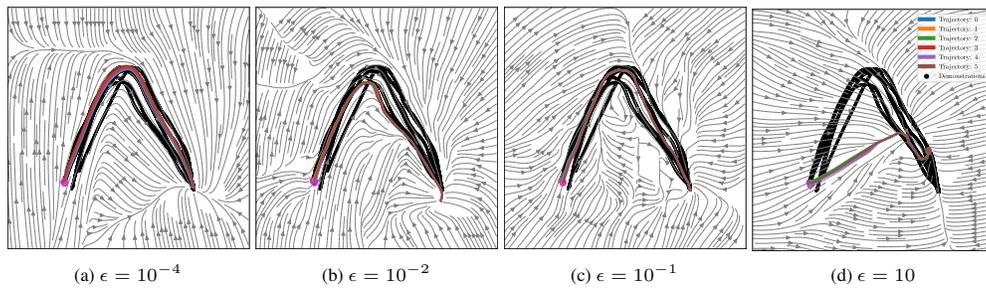


Figure 15: Path integrals generated by NCDS trained using different regularization term  $\epsilon$ .

with ADAM optimizer. Figure 16-c and 16-d shows that both models effectively model the observed dynamics (vector field); however, as anticipated, contraction stability is not achieved.

**Integration:** Furthermore, we tested the integrator in a toy example constructed using a Sine wave under different conditions. Figure 17 shows the integrator under different conditions. Figure 17-a and -b displays cases when the absolute and relative tolerance parameters of the integrator need to be at least as  $10^{-3}$ . Figure 17-c shows that the number of time steps to solve the integration does not make any difference. Furthermore, Figure 17-d displays different integration methods and their data reconstruction accuracy.

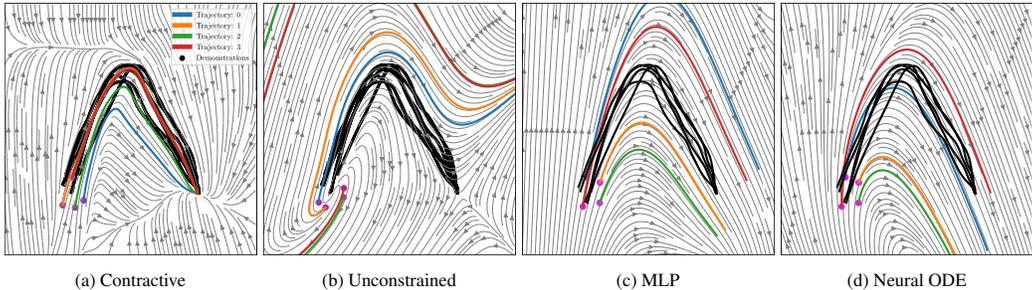


Figure 16: Path integrals generated under the Neural Contractive Dynamical Systems (NCDS) setting, along with baseline comparisons using Multilayer Perceptron (MLP) and Neural Ordinary Differential Equation (NeuralODE) models.

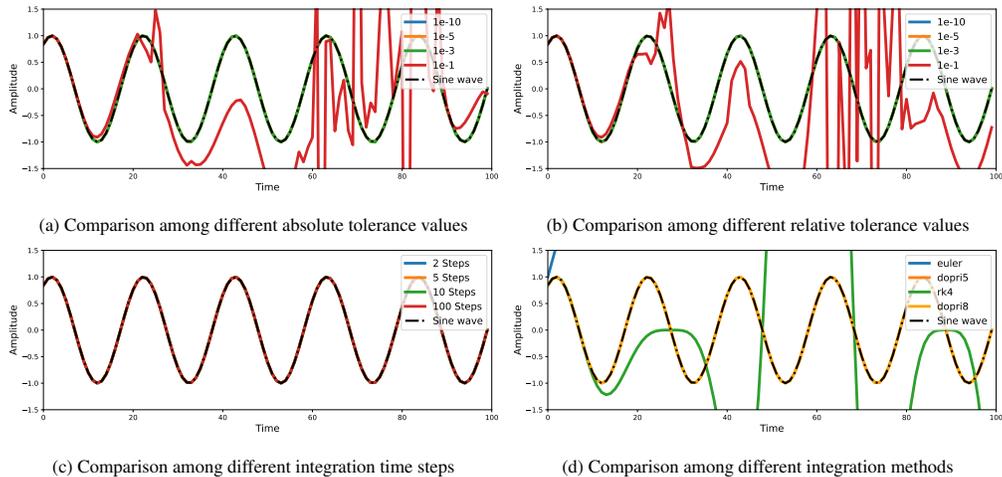


Figure 17: Evaluation of the integrator

**Activation function:** The choice of activation function certainly affects the generalization in the dynamical system (where there is no data). To show this phenomena, we ablated few common activation functions. For this experimental setup, we employed a feedforward neural network comprising two hidden layers, each consisting of 100 units. As shown in Figure 18, both the *Tanh* and *Softplus* activation functions display superior performance, coupled with satisfactory generalization capabilities. This indicates that the contour of the vector field aligns with the overall demonstration behavior outside of the data support. Conversely, opting for the *Sigmoid* activation function might yield commendable generalization beyond the confines of the data support. However, it is essential to acknowledge that this choice could potentially encounter challenges when it comes to reaching and stopping at the target.

**Dimensionality and execution time:** Table 2 compares 5 different experimental setups. First, Table 2-a reports the average execution time under the setup where the dimensionality reduction (i.e., no VAE) is discarded and the dynamical system is learned directly in the input space. As the results show, increasing the dimensionality of the input space from 2 to 8 entails a remarkable 40-fold increase in execution time for a single integration step. This empirical relationship reaffirms the inherent computational intensity tied to Jacobian-based computations. To address this issue, our approach leverages the dimensionality reduction provided by a custom Variational Autoencoder (VAE) that takes advantage of an injective generator (as decoder) and its inverse (as encoder). As showcased in Table 2-b, the VAE distinctly mitigates the computational burden. Comparing Table 2-a and Table 2-b, the considerable advantages of integrating the contractive dynamical system with the VAE into the full pipeline become evident as it leads to a significant halving of the execution time in the 8D scenario. Additionally, the findings indicate that increasing the dimensionality from 8 to 44 results in a sixfold increment in execution time. Moreover, the presented results indicate that our

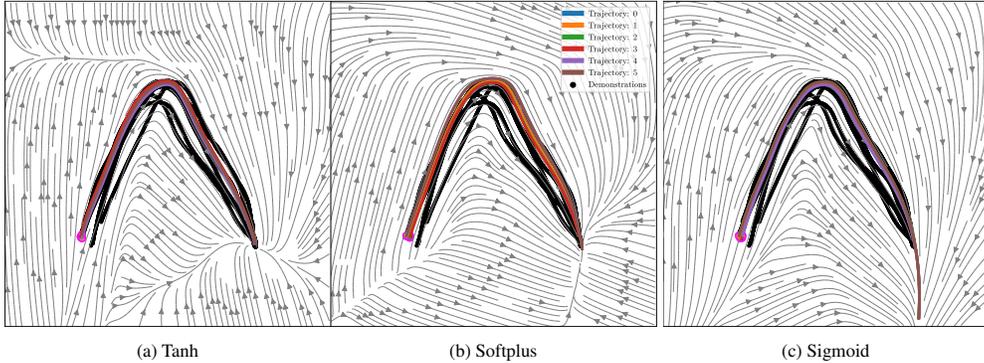


Figure 18: Path integrals generated by NCDS trained using different activation functions.

Input. Dim	2D	8D
time (ms)	1.23	40.9

Learning the vector field directly in the input space.

Input. Dim	3D	8D	44D
time (ms)	10.6	20.2	121.0

Learning the vector field in the latent space (full pipeline).

Table 2: Average execution time (in milliseconds) of a single integration step

current system may not achieve real-time operation. However, it is noteworthy that the video footage from our real-world robot experiments convincingly demonstrates the system’s rapid query response time, sufficiently efficient to control the robot arm, devoid of any operational concerns.

#### A.6.4 COMPARISONS

Table 3 summarizes the comparisons presented in the main papers in the contraction literature. Here we note a lack of consensus regarding the method or baseline for comparison in the context of learning contractive dynamical systems. A significant number of approaches have overlooked the contraction properties of their baseline, opting instead to base their comparisons solely on asymptotic stability guarantees. Following this trend, we have conducted a thorough and extensive comparative analysis involving Stable Estimator of Dynamical Systems (SEDS), Euclideanizing Flows (EF), and Imitation Flows (IF). We emphasize that neither of these baselines provides contraction guarantees but rather asymptotic stability guarantees. Therefore, the comparison is provided solely to evaluate the general stability and accuracy of the reconstructed motion.

To perform comparison between these methods and ours, we have provided two groups of datasets. The first group comprises multiple synthetic datasets, derived from the LASA dataset. The second category combines joint space trajectories of a 7-DOF Franka-Emika Panda robot arm.

#### Experimental setups:

- **Imitation Flow:** For Imitation Flow we have used a imitation flow network with the depth of 5. Each level within this depth was constructed utilizing CouplingLayer, RandomPermutation, and LULinear techniques, as recommended by the authors for managing high-dimensional data. The training process includes 1000 epochs, with a empirically chosen learning rate of  $10^{-3}$ . These hyper parameters were found experimentally and the best results have been selected to be compared against.
- **Euclideanizing Flow:** Here, we have used coupling network with random Fourier features parameterization, with 10 coupling layers each with 200 hidden units, and length scale of 0.45 as the length scale for random Fourier features. The training process includes 1000 epochs, with a empirically chosen learning rate of  $10^{-4}$ . These hyper parameters were found experimentally and the best results have been selected to be compared against.
- **SEDS:** Here, we use 5 Gaussian functions. The training process includes 1000 epochs with MSE objective.

### A.6.5 SYNTHETIC LASA

First, we evaluate the performance and accuracy of the methods on a 2D LASA dataset as the baseline for the experiments. As depicted in Fig. 19, all methods demonstrate strong performance, effectively capturing the data trends and successfully reaching the target points.

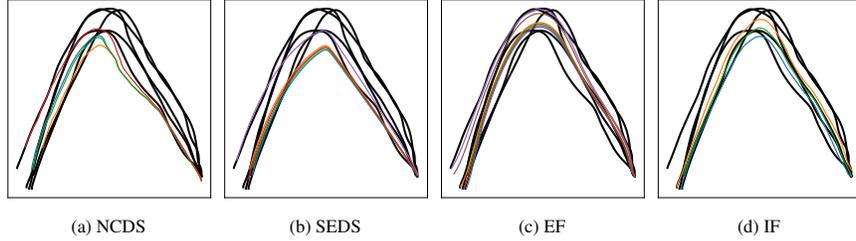


Figure 19: Path integrals generated using different approaches for the two-dimensional LASA problem. All methods provide similar reconstructions.

Examining the average time warping distance in Fig. 20, it becomes evident that EF and IF have better performance in reproducing the system dynamics.

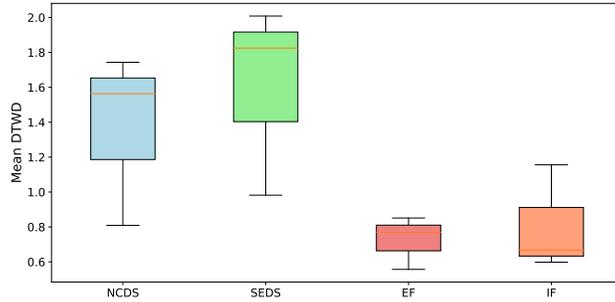


Figure 20: Average dynamic time warping distance between path integrals and the demonstrations for the two-dimensional LASA problem.

We extend our evaluation to a 4D LASA synthetic dataset created by merging two distinct datasets. In Figure 21, we observe that while NCDS and IF maintain a strong level of performance, SEDS and EF exhibit comparatively inferior results.

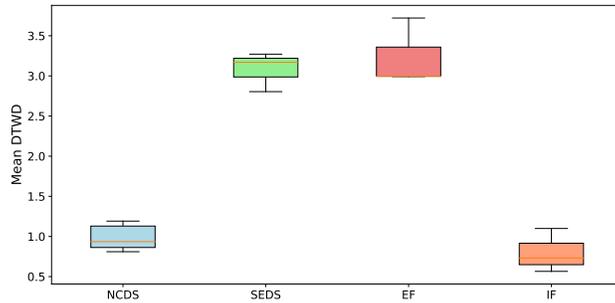


Figure 21: Average dynamic time warping distance between path integrals and the demonstrations for the four-dimensional LASA problem.

Lastly, we assess the methods using an 8D LASA synthetic dataset, constructed by concatenating 4 different datasets,

$$\mathcal{D} = [x_{\text{Angle}}, y_{\text{Angle}}, x_{\text{Line}}, y_{\text{Line}}, x_{\text{Sine}}, y_{\text{Sine}}, x_{\text{JShape}}, y_{\text{JShape}}] \in \mathbb{R}^{T \times 8}. \quad (17)$$

As illustrated in Fig. 22, NCDS has the best performance in higher-dimensional spaces.

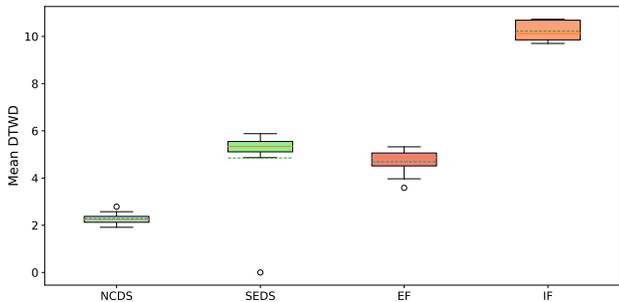


Figure 22: Average dynamic time warping distance between 5 trajectories constructed with different methods and demonstration trajectories

It is worth noting that in the 2D scenario, NCDS was directly trained in the input space, whereas in the 4D and 8D experiments, the dynamics were learned in the latent space

To make sure that we have a fair comparison we have scaled up all the methods. 1. Imitation flow: the depth of the network was scaled up by factor of 2 (from 5 to 10), We evaluated 20 different architecture with higher and lower depth and used the one with the best performance. 2. Euclideanizing Flow: the number of coupling blocks in the BijectionNet in the original code released by the Authors was increased by the factor of 2 from (10 to 20). We evaluated 20 different architecture with higher number of blocks and nodes and used the one with the best performance. 3. SEDS: we increased the number of Gaussian functions from 5 to 20. We evaluated 10 different architecture with different number of Gaussian functions used the one with the best performance. 4. NCDS: The Jacobian network did not need more parameters since the latent space has stayed 2D. Therefore, we just added the VAE to the architecture when moving from 2D to 8D.

#### A.6.6 JOINT SPACE DYNAMICS

Since using the pose information including Quaternion is not feasible due to the fact that none of the methods that we compared against support non-Euclidean spaces, we have chosen to use joint space trajectories of a 7-DOF robot. Therefore, demonstration data that is used is 7-dimensional and it represents “V” shape character on a table. Fig. 23 shows the comparison on the average dynamic time warping distance (DTWD) among different methods when learning 7D joint space trajectories.

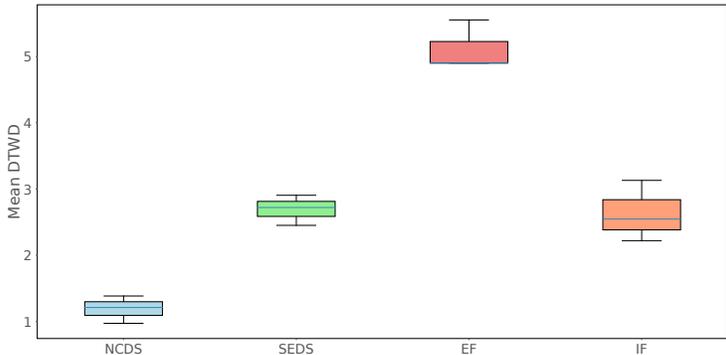


Figure 23: Average dynamic time warping distance between path integrals and the demonstrations for the 7-dimensional Joint space problem.

As shown, NCDS provides the highest reconstruction accuracy, therefore showing how our method better scales to higher-dimensional settings. The video of this experiment is int the supplementary material. Moreover, the video shows the recording process of the demonstrations and also 4 different videos from simulated robots performing the path integrals in the joint space using the four benchmarked methods.

<b>Paper title</b>	<b>Abbr.</b>	<b>Compared against</b>	<b>Year</b>
Safe Control with Learned Certificates: A Survey of Neural Lyapunov, Barrier, and Contraction methods			2022
Learning Contraction Policies from Offline Data		MPC (ILQR), RL (CQL)	2022
Neural Contraction Metrics for Robust Estimation and Control: A Convex Optimization Approach	NCM	CV-STEM, LQR	2021
Learning Stabilizable Nonlinear Dynamics with Contraction-Based Regularization	CCM-R		2021
Learning Certified Control Using Contraction Metric	C3M	SoS (Sum-of-Squares programming), MPC, RL (PPO)	2020
Learning position and orientation dynamics from demonstrations via contraction analysis	CDSP	SEDS, CLF-DM (Control Lyapunov Function-based Dynamic Movements), Tau-SEDS, NIVF (neurally imprinted vector fields)	2019
Learning Stable Dynamical Systems using Contraction Theory	C-GMR	SEDS	2017
Learning Contracting Vector Fields For Stable Imitation Learning	CVF	DMP, SEDS and CLF-DM	2017
Learning Partially Contracting Dynamical Systems from Demonstrations	CDSP (only position)	DMP, CLF-DM	2017

Table 3: The present state-of-the-art literature pertaining to the learning of contractive dynamical systems.