Latent SDEs on Homogeneous Spaces

Sebastian Zeng Florian Graf Roland Kwitt

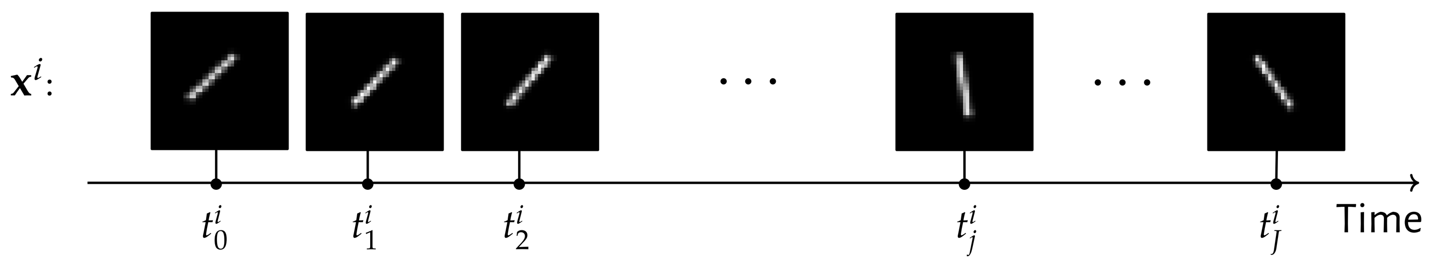
University of Salzburg, Austria

NEURAL INFORMATION PROCESSING SYSTEMS

Problem setting

We seek to learn from sequential data, i.e., from a set of N multivariate time series

$\mathbf{x}^1, \ldots, \mathbf{x}^N$; e.g.:



We assume:

(1) $\mathbf{x}^i$ to be a **partially observed** continuous path from a **stochastic process**

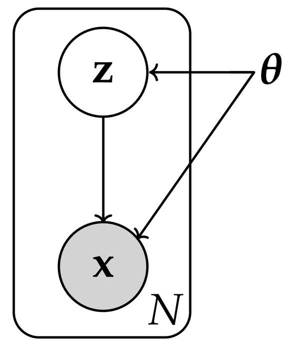
$$X : \Omega \times [0, T] \to \mathbb{R}^d, \text{ and}$$

(2) that this process is governed by some **latent stochastic process** Z (with paths $\mathbf{z}^i$).

We seek to learn X!

Fitting a process to data

We follow a **Variational Bayes** approach with the following **directed graphical model**:

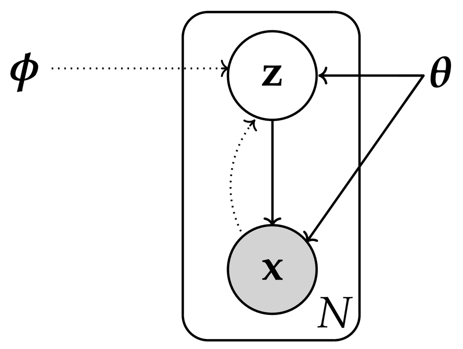


1. A path $\mathbf{z}^i$ is drawn from a (latent) parametric **prior path distribution** $p_{\theta^*}(\mathbf{z})$.
2. An (observed) path $\mathbf{x}^i$ is drawn from the **conditional path distribution** $p_{\theta^*}(\mathbf{x}|\mathbf{z}^i)$.

Well-explored in the **vector-valued** setting (e.g., $\mathbf{x}^i \in \mathbb{R}^d$).
❓ Less well-explored in the **path-valued** setting (e.g., $\mathbf{x}^i \in \mathcal{C}([0,T], \mathbb{R}^d)$) — **Ours**!

Approximate variational inference

Learning directed probabilistic models in case of intractable posterior path distributions $p_{\theta}(\mathbf{z}|\mathbf{x}^i)$,



entails the choice of …

- … a tractable **prior path distribution** $p_{\theta}(\mathbf{z})$ over latent paths $\mathbf{z}$,
- … a tractable **approximate posterior path distribution** $q_{\phi}(\mathbf{z}|\mathbf{x}^i)$,

and the maximization of the **evidence lower bound (ELBO)** w.r.t. θ, ϕ:

$$\log p_{\theta}(\mathbf{x}^i) \ \geq \ -D_{\mathrm{KL}}\left(q_{\phi}(\mathbf{z}|\mathbf{x}^i) \| p_{\theta}(\mathbf{z})\right) + \mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x}^i)}\left[\log p_{\theta}(\mathbf{x}^i|\mathbf{z})\right].$$

Model **evidence**	KL divergence between **approximate posterior** and the **prior** path distribution	Expected log-likelihood of **observed path** given the **latent path**

We consider **path distributions** of stochastic processes that are solutions to **SDEs**.

Choice of a latent space

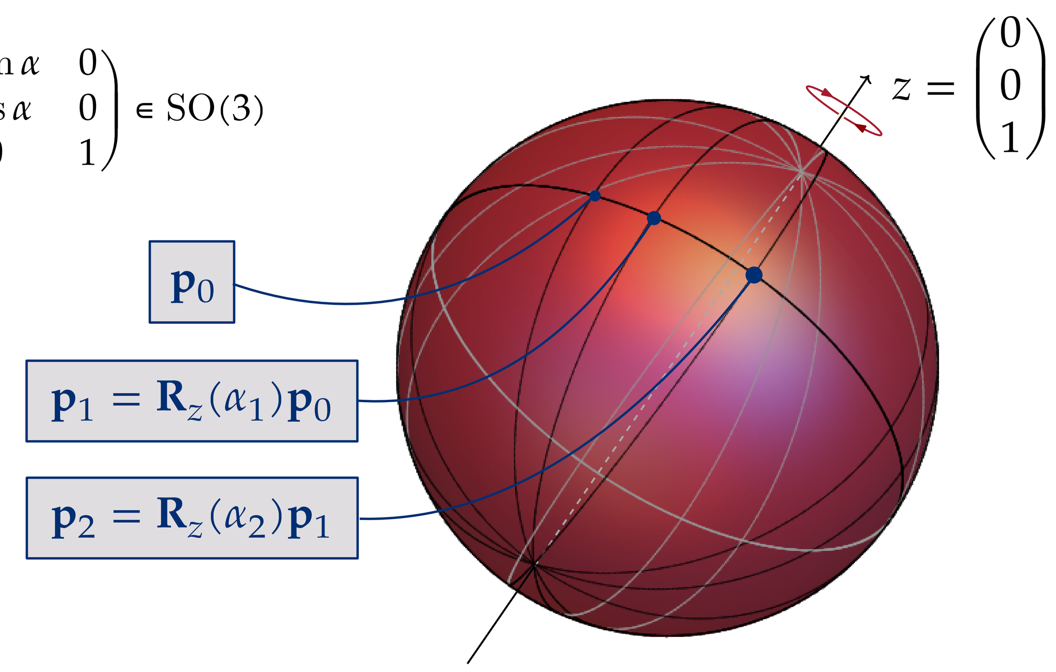
Some **favorable properties** of latent spaces for carrying distributions connected to random dynamics are …

- … to offer flexibility for modelling **non-linear structures** (i.e., manifolds),
- … to include the focus on **geometric features** (e.g., symmetry), and
- … to **respect the latter under discretization** (e.g., for sampling).

Considering these aspects, choosing SDEs that evolve on a **homogeneous space** as the consequence of some **(matrix) Lie group action** appears to be a reasonable choice.

Example: $\mathbb{S}^2$ with (quadratic) matrix Lie group $\mathrm{SO}(3)$.

- $$\mathbf{R}_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \in \mathrm{SO}(3)$$



Dynamics in homogeneous spaces

Example continued: $\mathbb{S}^2$ with (quatratic) matrix Lie group $\mathrm{SO}(3)$.

$$\mathrm{SO}(n) = \{\mathbf{A} \in \mathrm{Mat}(n) : \mathbf{A}^\top\mathbf{A} = \mathbf{I}_n, \det(\mathbf{A}) = +1\}$$

$\mathrm{Mat}(n)$ — space of real $n \times n$ matrices

Lie group

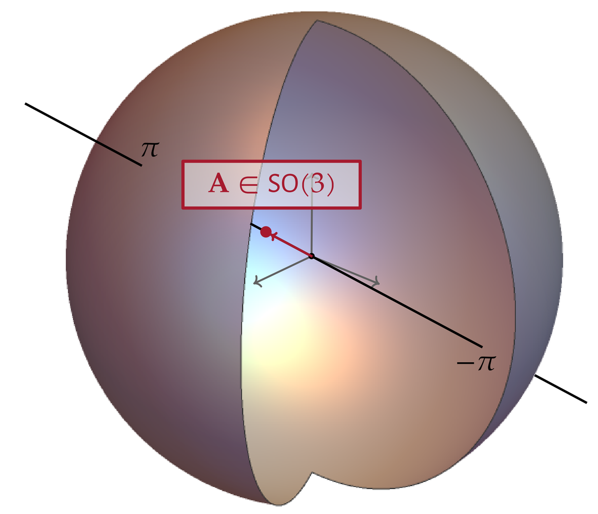
$$\mathfrak{so}(n) = \{\mathbf{A} \in \mathrm{Mat}(n) : \mathbf{A} + \mathbf{A}^\top = \mathbf{0}_n\}$$

Lie algebra
(Vector space together with Lie bracket $[\cdot,\cdot]$)

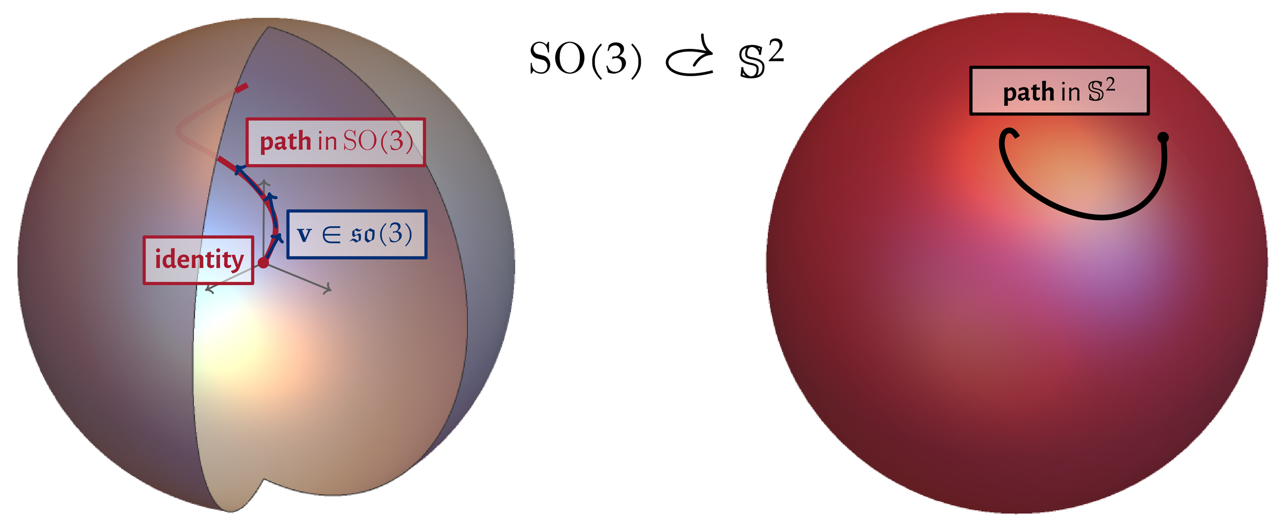
Sketch for $\mathrm{SO}(3) \cong \mathbb{B}^3_{r=\pi} / _{(x \sim -x \text{ if } x \in \partial\mathbb{B}^3_{r=\pi})}$

(3-Ball of radius $r = \pi$ with identified antipodal points on the boundary $\partial\mathbb{B}^3_{r=\pi}$)

- radial component $\simeq$ roation angle
- angular component $\simeq$ roation axis



Generating a path on the sphere:



SDEs in (quadratic) matrix Lie groups

Leveraging the Lie algebra $\mathfrak{g}$, we can define (Itô) SDEs in a (quadratic) matrix **Lie group** $\mathcal{G}$ of the form:

$$\mathrm{d}G_t = \Big(\underbrace{\mathbf{V}_0(t)\mathrm{d}t}_{\text{Drift}} + \underbrace{\sum_{i=1}^m \mathrm{d}w_t^i \mathbf{V}_i}_{\text{Diffusion}}\Big) G_t\,, \quad G_0 = \mathbf{I}_n\,; \qquad \mathbf{V}_0(t) = \mathbf{K}(t) + \frac{1}{2}\sum_{i=1}^m \mathbf{V}_i^2\,.$$

$\mathbf{K}(t), \mathbf{V}_1, \ldots, \mathbf{V}_m \in \mathfrak{g}$.

This induces an SDE for $Z = G \cdot Z_0$ in the **homogeneous space** :

$$\mathrm{d}Z_t = \Big(\mathbf{V}_0(t)\mathrm{d}t + \sum_{i=1}^m \mathrm{d}w_t^i \mathbf{V}_i\Big) Z_t\,, \quad Z_0 \sim \mathcal{P}\,.$$

$\mathcal{P}$ … distribution over the initial state.

To implement a drift parameterization depending on $\mathbf{x}$, we realize

$$\mathbf{K}^{\phi}(\mathbf{x})(t) : [0, T] \to \mathfrak{g}$$

via Chebyshev polynomials with **learnable coefficients**.

The **prior** $p_{\theta}(\mathbf{z})$ and **approximate posterior** $q_{\theta}(\mathbf{z}|\mathbf{x})$ are determined by an SDE of this form!

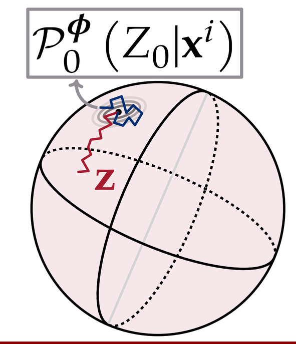
The **overall objective** (for our example of latent paths on $\mathbb{S}^{n-1}$):

$$\begin{aligned} \mathcal{L}(\phi, \theta; \mathbf{x}^i) = \ & D_{\mathrm{KL}}\left(\mathcal{P}_0^{\phi}(Z_0|\mathbf{x}^i) \,\|\, \mathcal{U}_{\mathbb{S}^{n-1}}\right) \\ & + \frac{1}{2}\int_0^T \int_{\mathbb{S}^{n-1}} q_{Z_t}(\mathbf{z}) \|\mathbf{K}^{\phi}(\mathbf{x}^i)(t)\mathbf{z}\|^2 \mathrm{d}\mathbf{z}\mathrm{d}t \\ & + \mathbb{E}_{\mathbf{z} \sim q_{\phi}(\mathbf{z}|\mathbf{x}^i)}\left[\log p_{\theta}(\mathbf{x}^i|\mathbf{z})\right] \end{aligned}$$

KL-div. to **uniform distribution** on $\mathbb{S}^{n-1}$

KL-div. between **approximate posterior** and a **driftless prior**

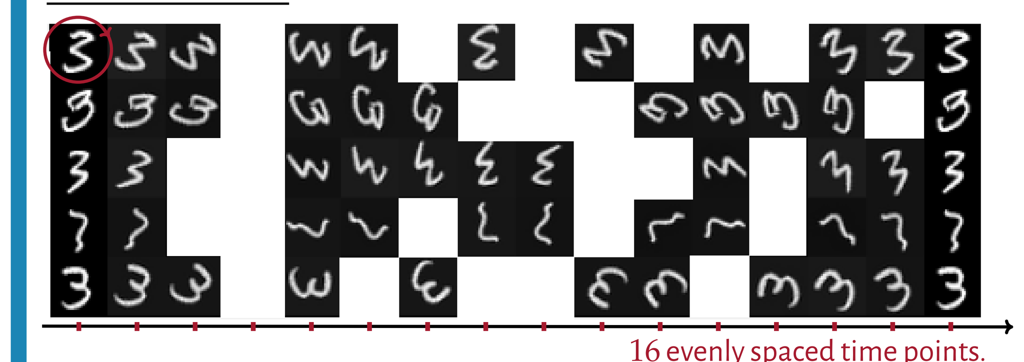
Expected **log-likelihood** of observed path given the latent path



- **Sampling** from the approximate posterior:
  We use a one-step **geometric** Euler-Maryuama SDE solver, that is particularly easy to implement! [Marjanovic & Solo 2015; Muniz et al., 2022]

Some results

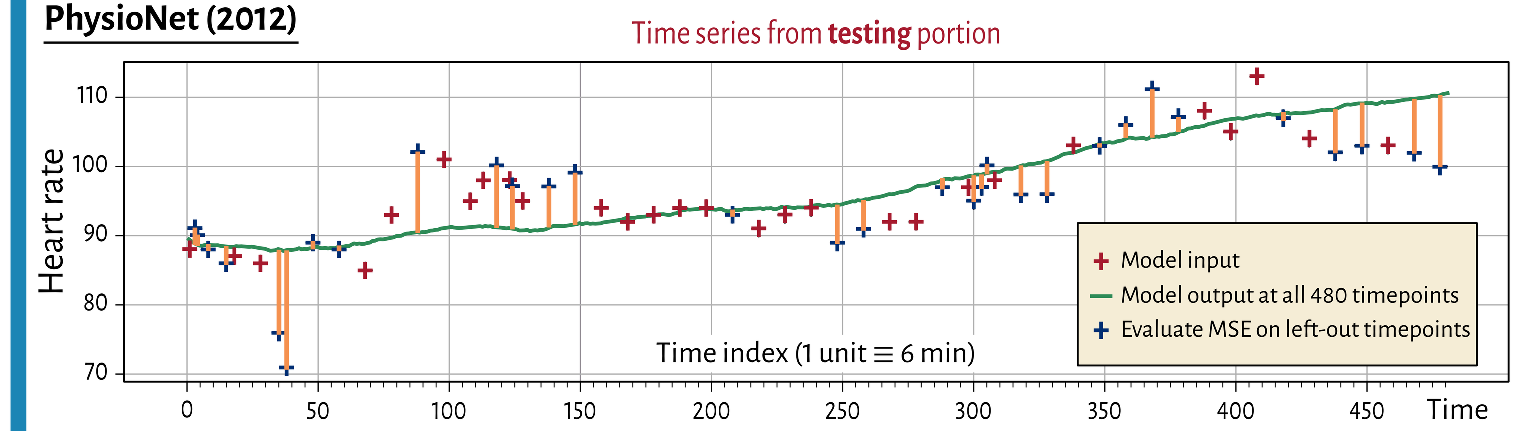
Rotating MNIST



† indicates results from [Yildiz et al., 2019].

	↓ **MSE** $(\times 10^{-3})$
†GPPVAE-dis	30.9 ± 0.02
†GPPVAE-joint	28.8 ± 0.05
†ODE^2VAE	19.4 ± 0.06
†ODE^2VAE-KL	18.8 ± 0.31
CNN-ODE	14.5 ± 0.73
Ours	$\mathbf{11.8 \pm 0.25}$

PhysioNet (2012)



	↓ **MSE** $(\times 10^{-3})$	
CRU	5.11 ± 0.40	[Schirmer et al., 2022]
f-CRU	5.24 ± 0.49	[Schirmer et al., 2022]
mTAND-Full	3.61 ± 0.08	[Shukla & Marlin, 2021]
mTAND-ODE	3.38 ± 0.03	[Shukla & Marlin, 2021] (*with added ODE*)
Ours	$\mathbf{3.11 \pm 0.02}$	

SCAN ME