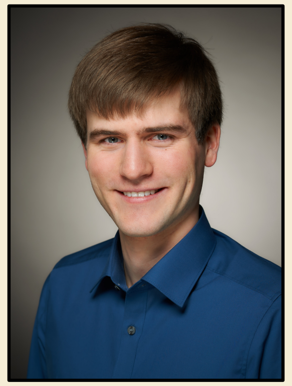
Latent SDEs on Homogeneous Spaces


Sebastian Zeng, Florian Graf & Roland Kwitt

University of Salzburg, Austria

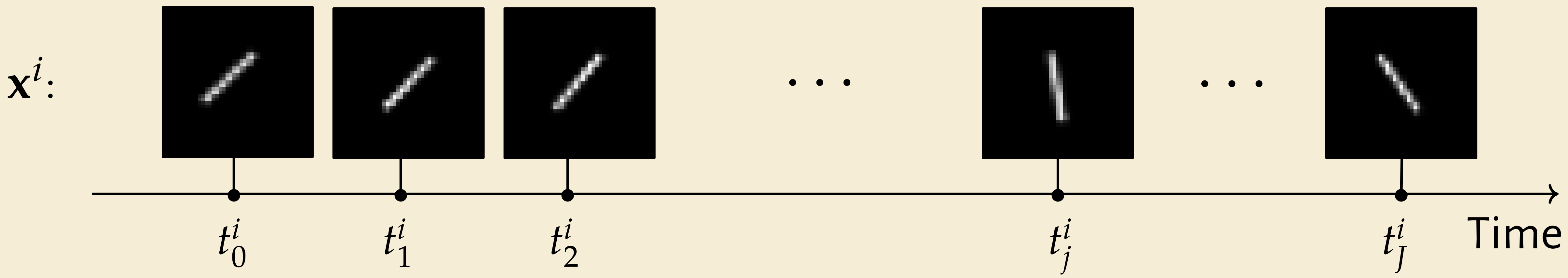
Problem setting

We seek to learn from sequential data, i.e., from a set of N multivariate time series $\mathbf{x}^1, \ldots, \mathbf{x}^N$.

Problem setting

We seek to learn from sequential data, i.e., from a set of N multivariate time series $\mathbf{x}^1, \ldots, \mathbf{x}^N$.

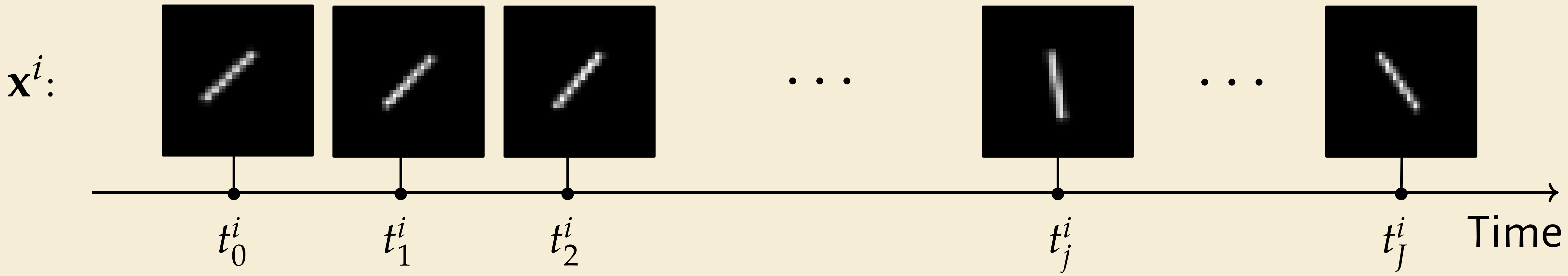
Example data sample:



Problem setting

We seek to learn from sequential data, i.e., from a set of N multivariate time series $\mathbf{x}^1, \dots, \mathbf{x}^N$.

Example data sample:



We assume

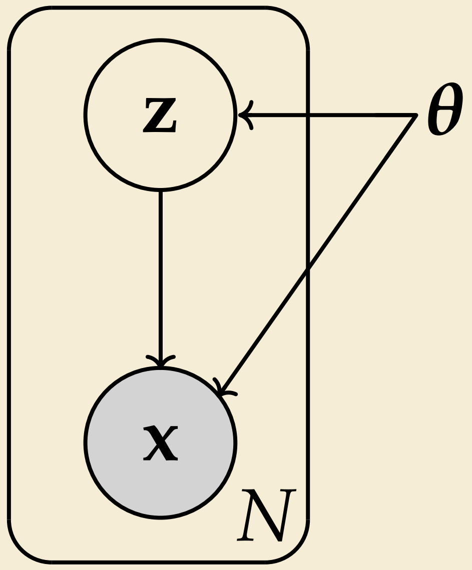
(1) $\mathbf{x}^i$ to be a **partially observed** continuous path from a **stochastic process** $X : \Omega \times [0, T] \to \mathbb{R}^d$, and

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

$\boxed{\text{We seek to learn } X\text{!}}$

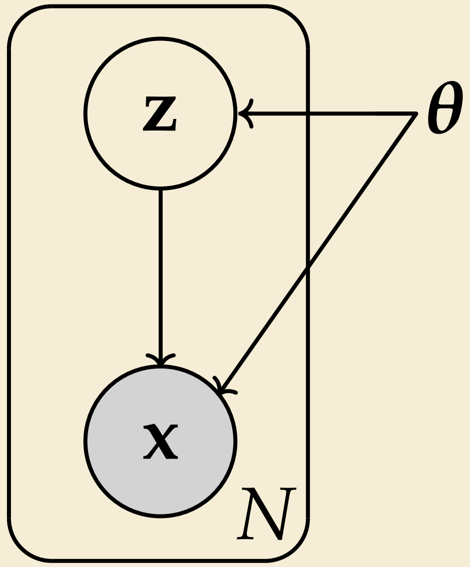
Fitting a process to data

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



Fitting a process to data

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

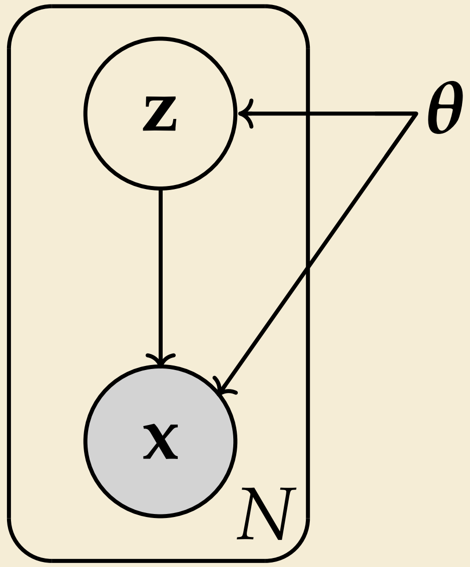


1 A path $\mathbf{z}^i$ is drawn from a (latent) parametric **prior path distribution** $p_{\boldsymbol{\theta}^*}(\mathbf{z})$.

$\boldsymbol{\theta}^*$ - true (unknown) parameter

Fitting a process to data

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



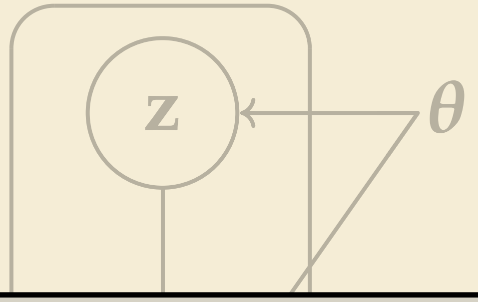
$\boxed{1}$ A path $\mathbf{z}^i$ is drawn from a (latent) parametric **prior path distribution** $p_{\boldsymbol{\theta}^*}(\mathbf{z})$.

$\boxed{2}$ An (observed) path $\mathbf{x}^i$ is drawn from the **conditional path distribution** $p_{\boldsymbol{\theta}^*}(\mathbf{x}|\mathbf{z}^i)$.

$\boldsymbol{\theta}^*$ - true (unknown) parameter

Fitting a process to data

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



💡 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**!

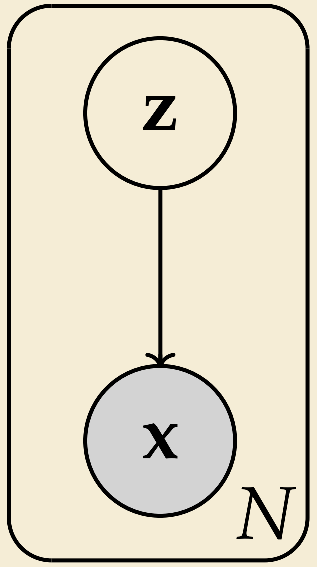
1 A path $\mathbf{z}^i$ is drawn from a (latent) parametric **prior path distribution** $p_{\theta^*}(\mathbf{z})$.

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

θ^* - true (unknown) parameter

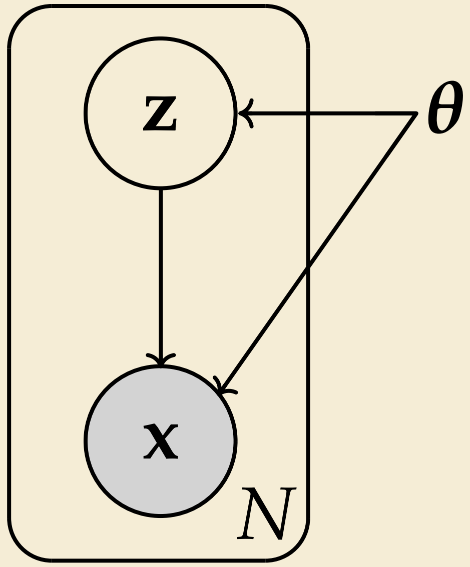
Approximate variational inference

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



Approximate variational inference

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

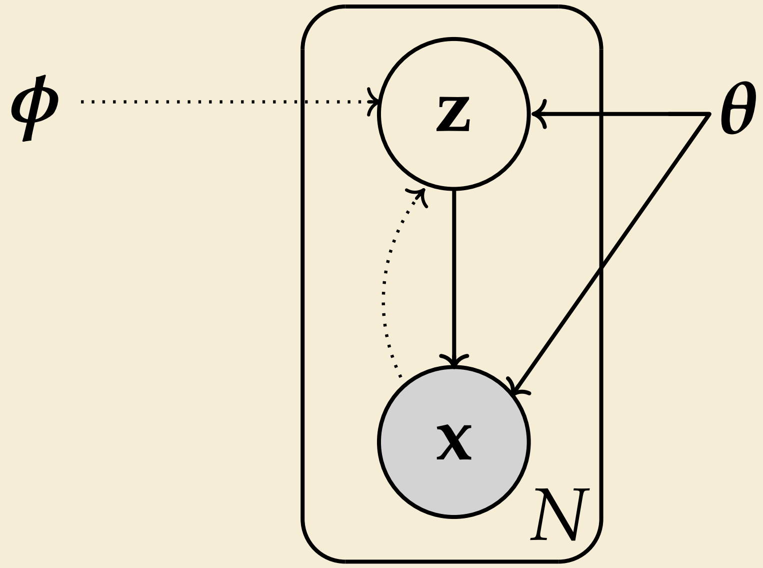


entails the choice of ...

- ... a tractable **prior path distribution** $p_{\boldsymbol{\theta}}(\mathbf{z})$ over latent paths $\mathbf{z}$,

Approximate variational inference

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

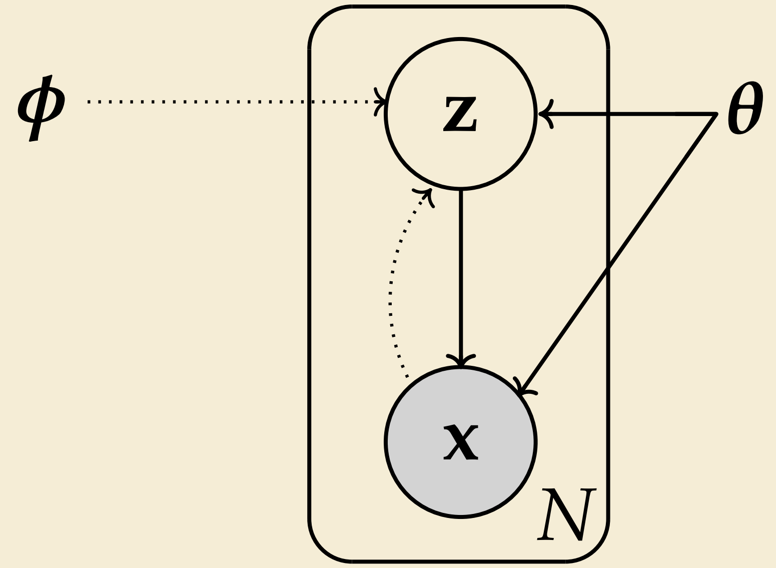


entails the choice of ...

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

Approximate variational inference

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



entails the choice of ...

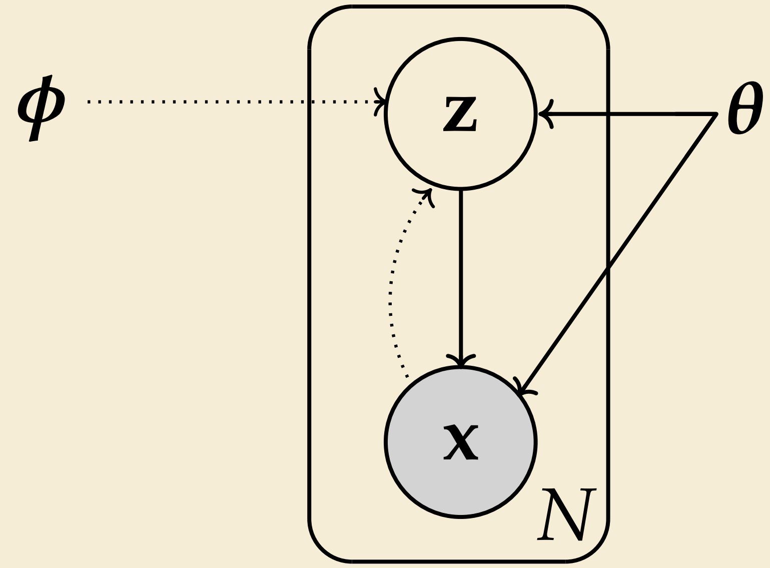
- ... a tractable **prior path distribution** $p_{\boldsymbol{\theta}}(\mathbf{z})$ over latent paths $\mathbf{z}$,
- ... a tractable **approximate posterior path distribution** $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}^i)$,

and the maximization of the **evidence lower bound (ELBO)** w.r.t. $\boldsymbol{\theta}, \boldsymbol{\phi}$:

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

Approximate variational inference

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



entails the choice of ...

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

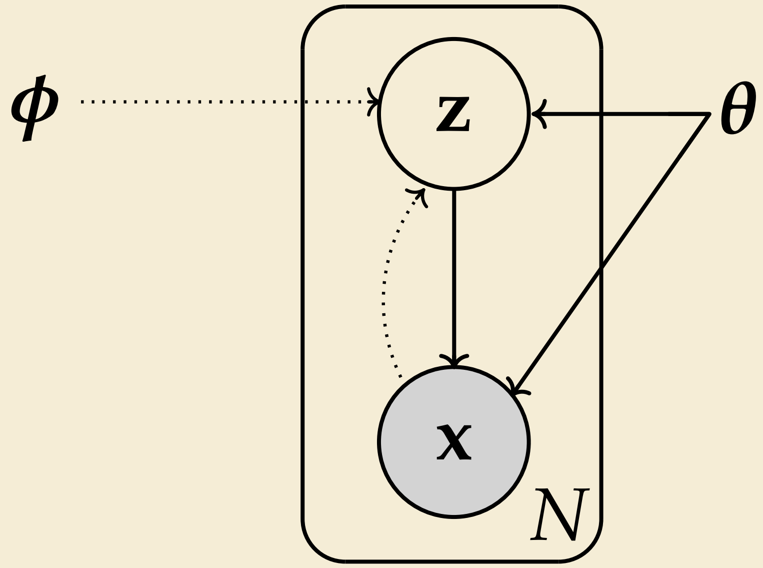
and the maximization of the **evidence lower bound (ELBO)** w.r.t. $\boldsymbol{\theta}, \boldsymbol{\phi}$:

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

Model **evidence**

Approximate variational inference

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



entails the choice of ...

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

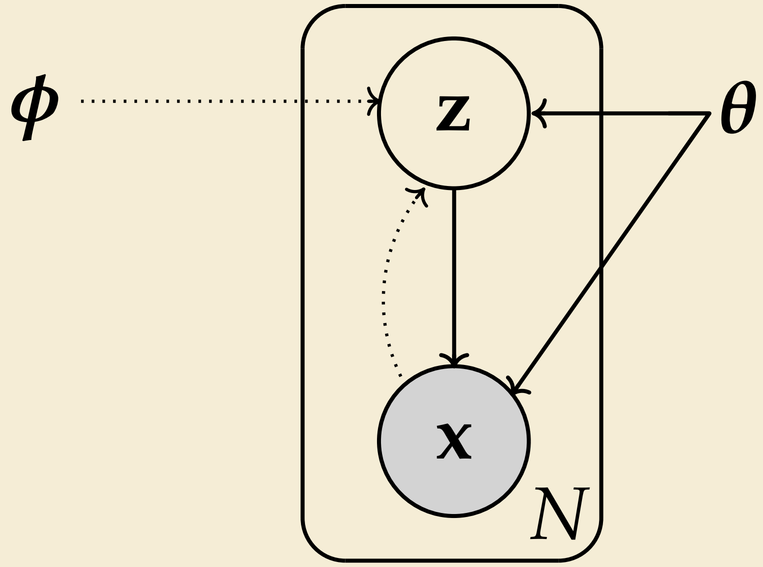
and the maximization of the **evidence lower bound (ELBO)** w.r.t. $\boldsymbol{\theta}, \boldsymbol{\phi}$:

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

KL divergence between **approximate posterior** and the **prior** path distribution

Approximate variational inference

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



entails the choice of ...

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

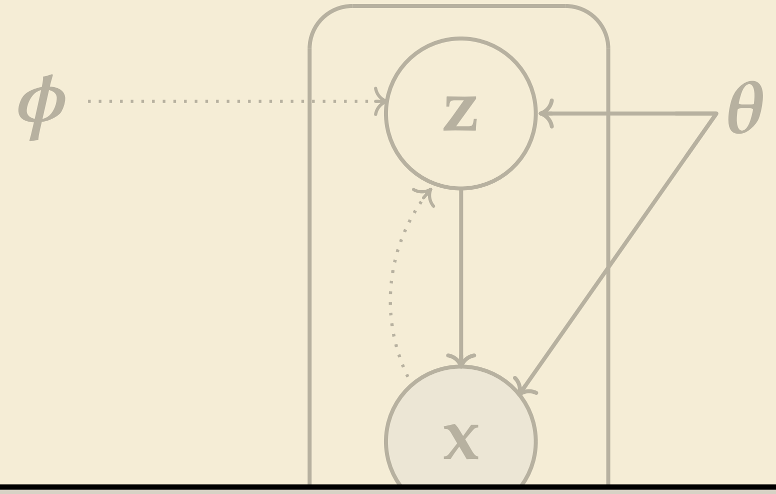
and the maximization of the **evidence lower bound (ELBO)** w.r.t. $\boldsymbol{\theta}, \boldsymbol{\phi}$:

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

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

Approximate variational inference

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



In this work, we consider **path distributions** of stochastic processes that are solutions to **SDEs**.

entails the choice of ...

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

and the maximization of the **evidence lower bound (ELBO)** w.r.t. $\boldsymbol{\theta}, \boldsymbol{\phi}$:

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

Choice of latent space

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

Choice of 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).

Choice of latent space

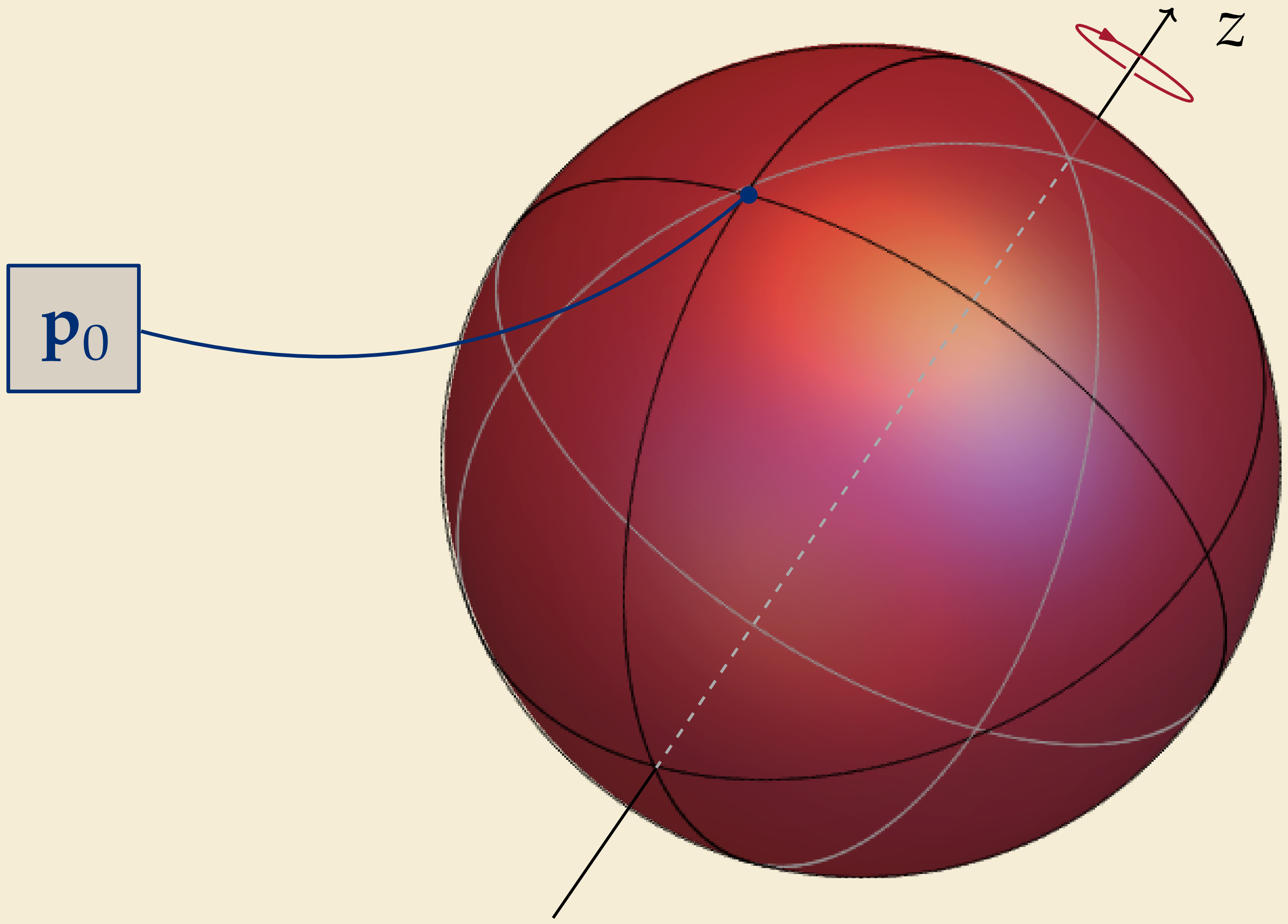
Some **favorable properties** of latent spaces for carrying distributions connected to random dynamics are …

- … 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.

Choice of latent space

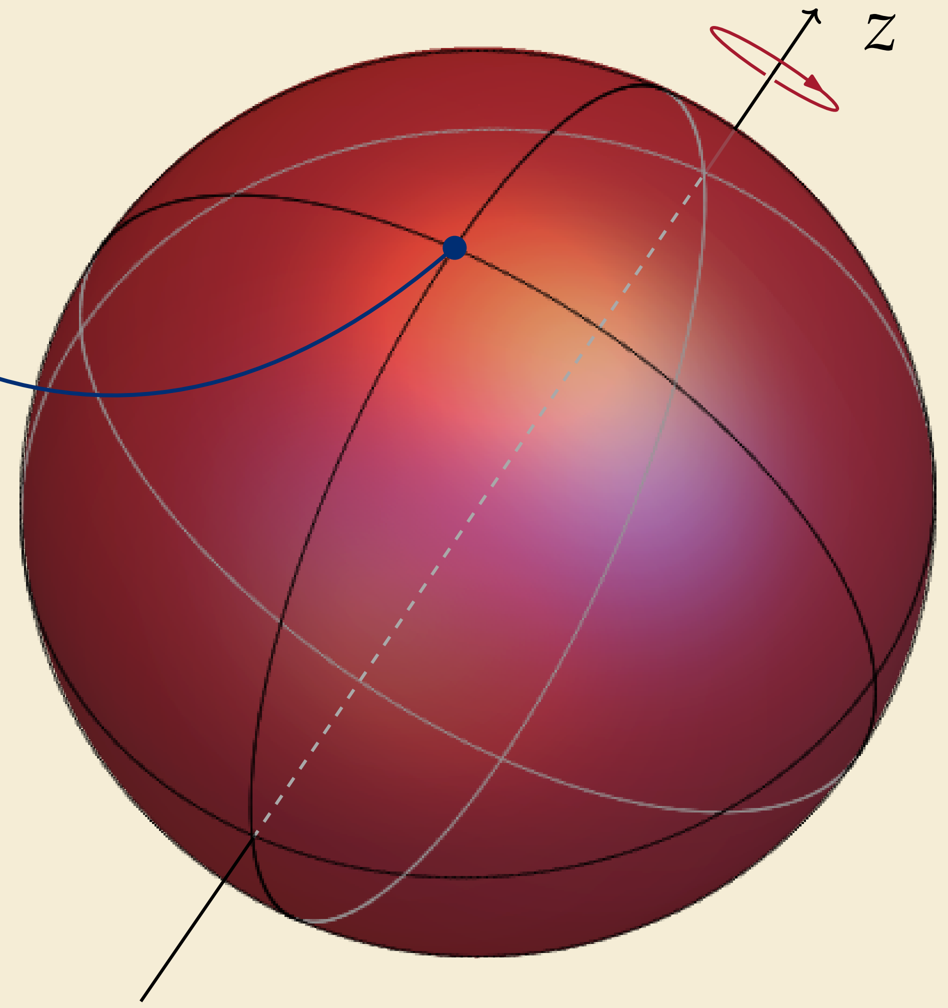
Example: $\mathbb{S}^2$ with (quadratic) matrix Lie group $\mathrm{SO}(3)$.



Choice of latent space

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

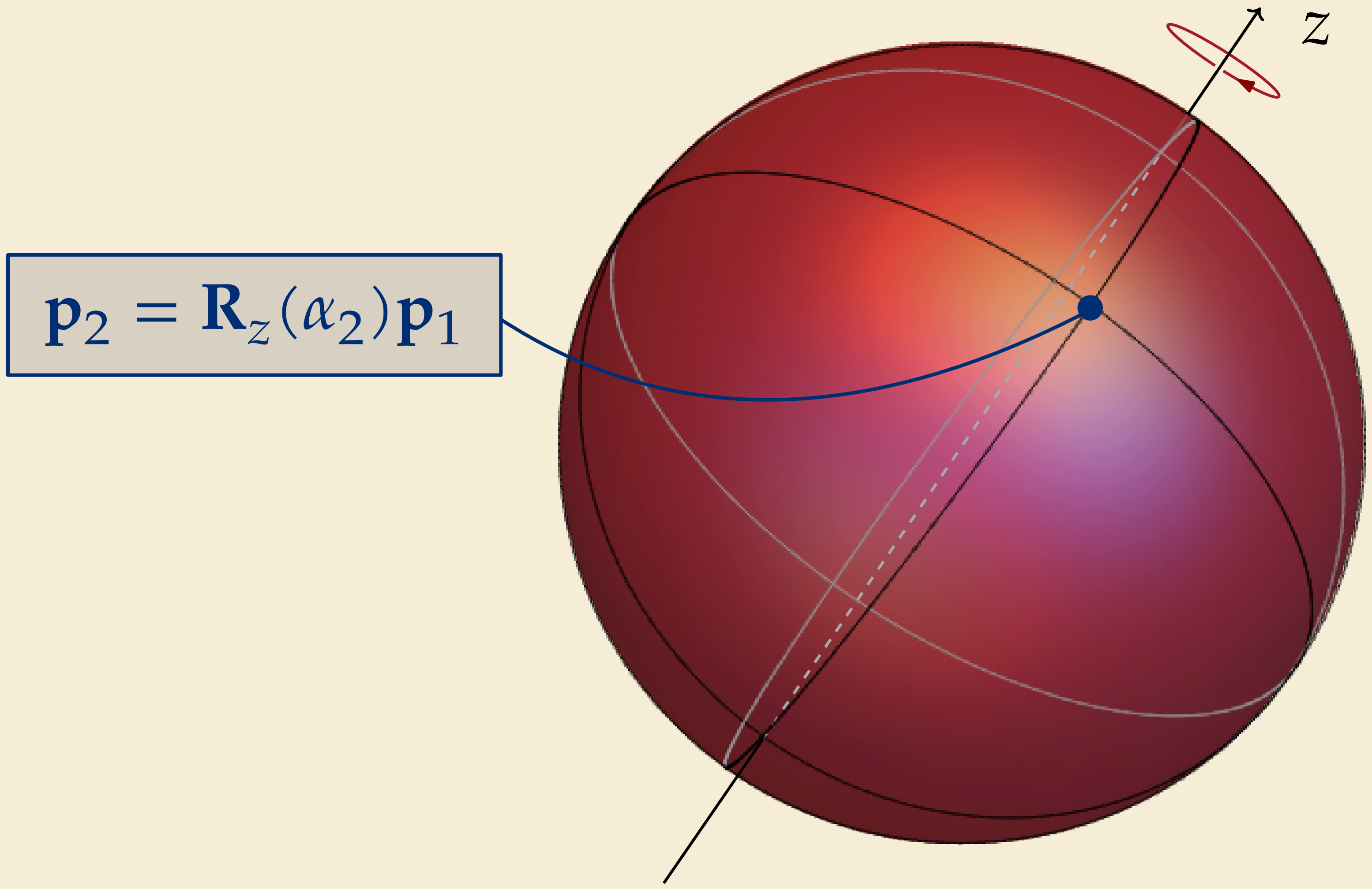
$$\mathbf{p}_1 = \mathbf{R}_z(\alpha_1)\mathbf{p}_0$$



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

Choice of latent space

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

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

Lie group

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

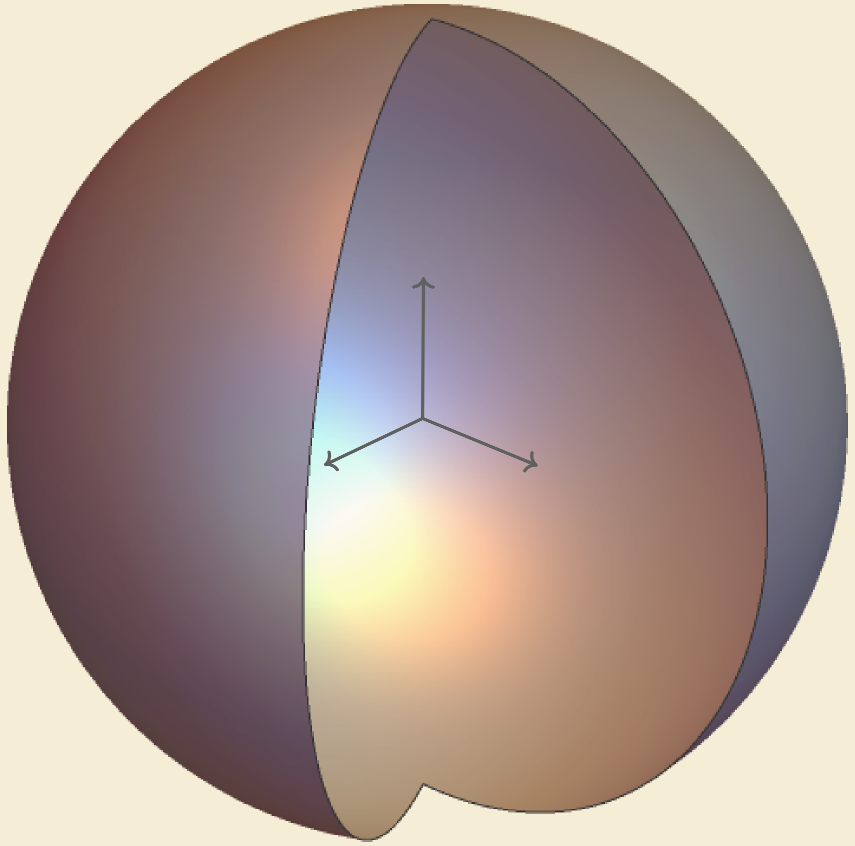
Dynamics in homogeneous spaces

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

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

Lie group

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi}/ \sim$:



$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \vee x = y$

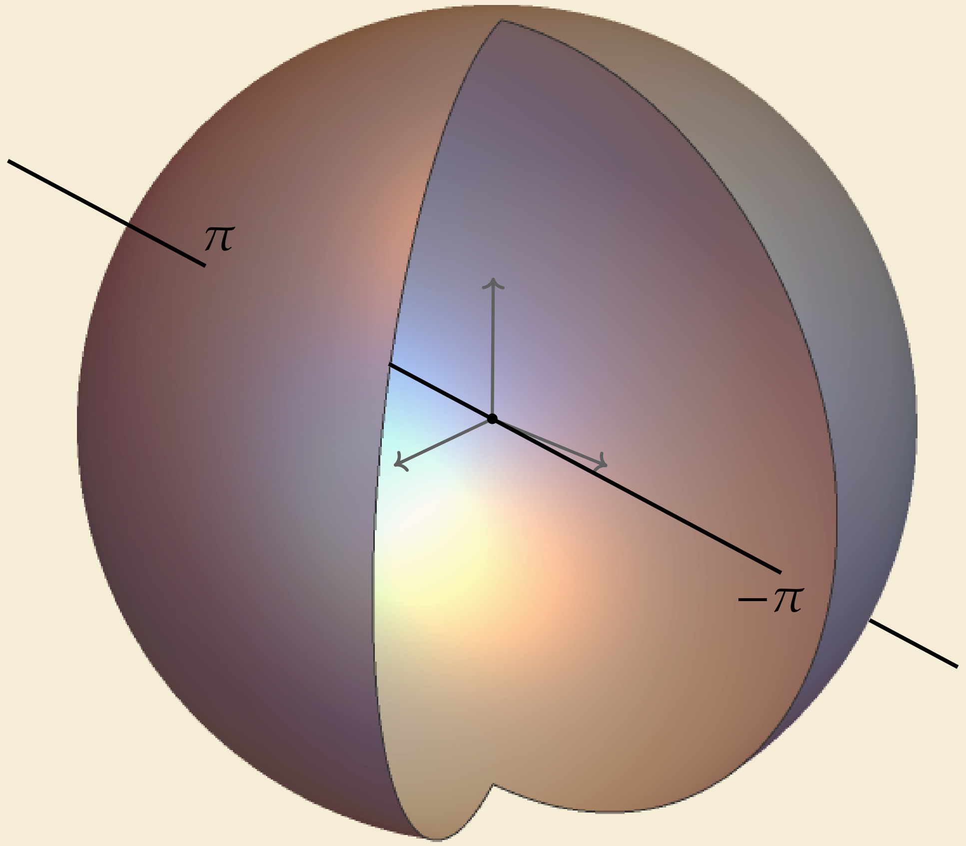
Dynamics in homogeneous spaces

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

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

Lie group

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi} / \sim$:



$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \lor x = y$

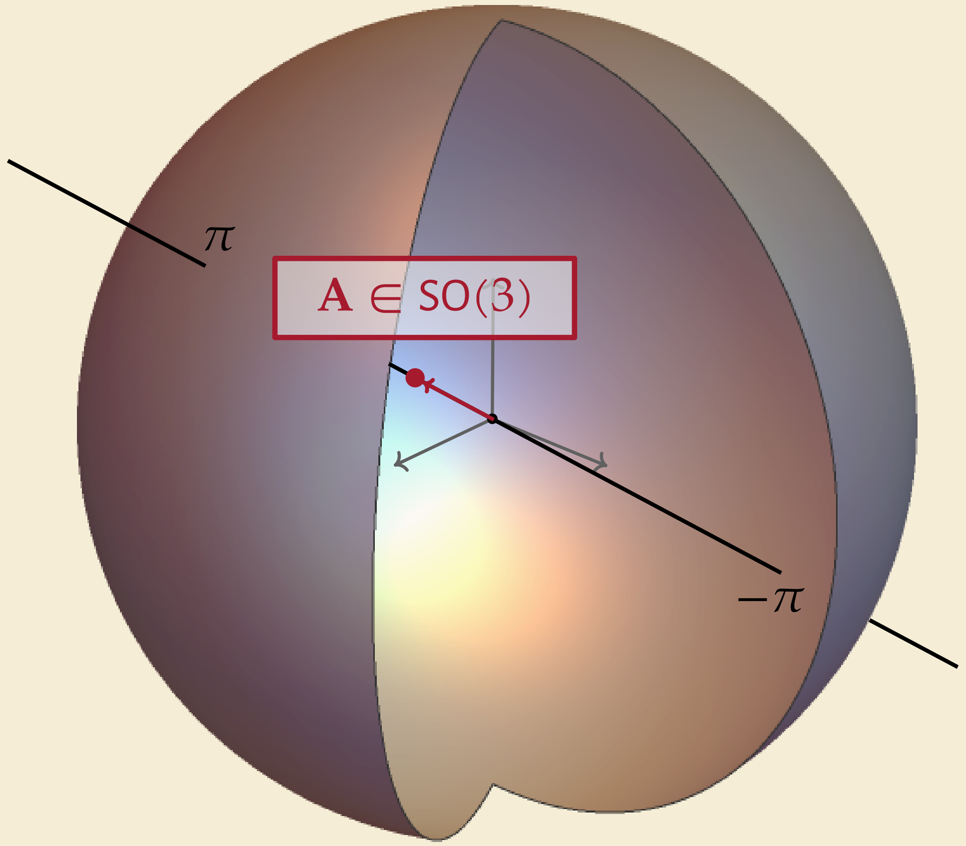
Dynamics in homogeneous spaces

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

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

Lie group

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi} / \sim$:



$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \vee x = y$

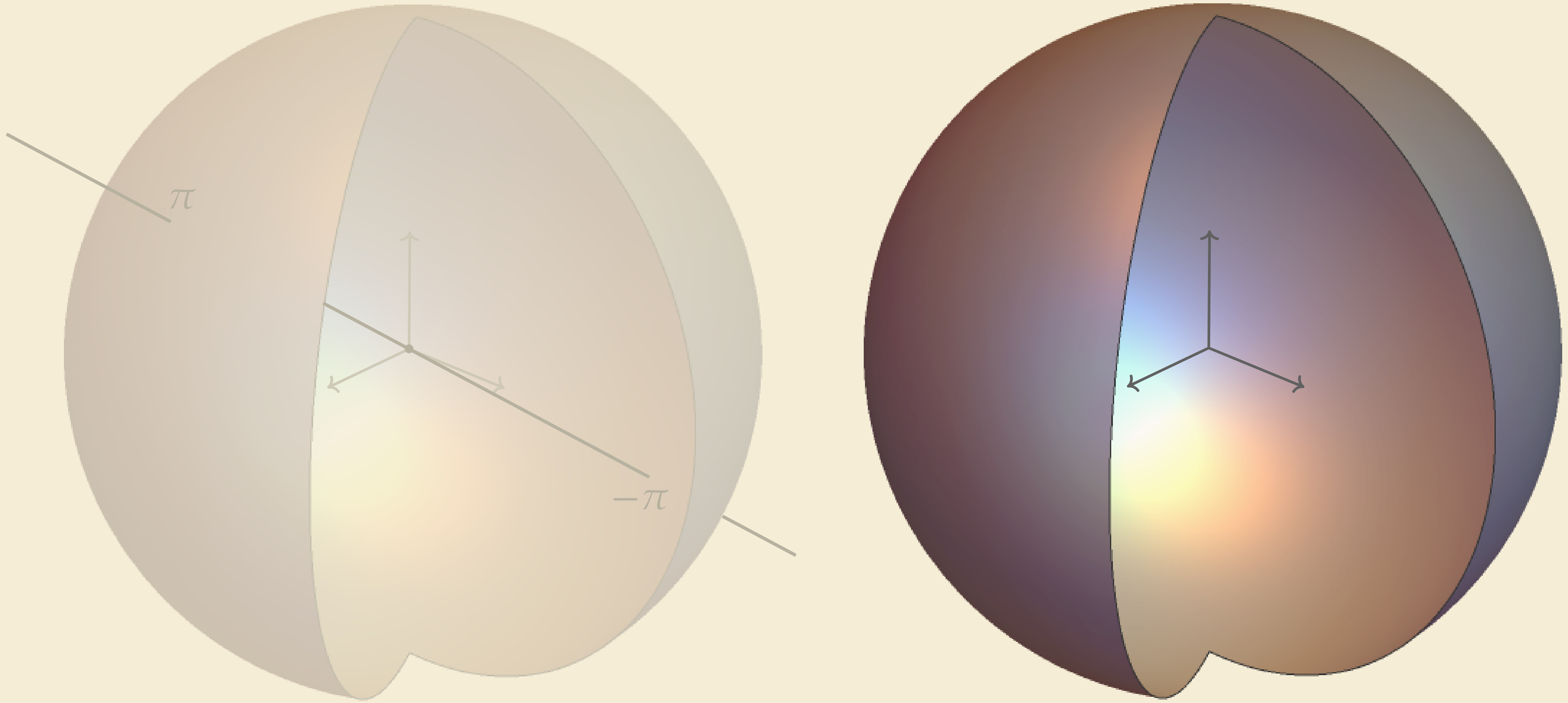
Dynamics in homogeneous spaces

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

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

Lie group

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi}/\sim$:



$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \vee x = y$

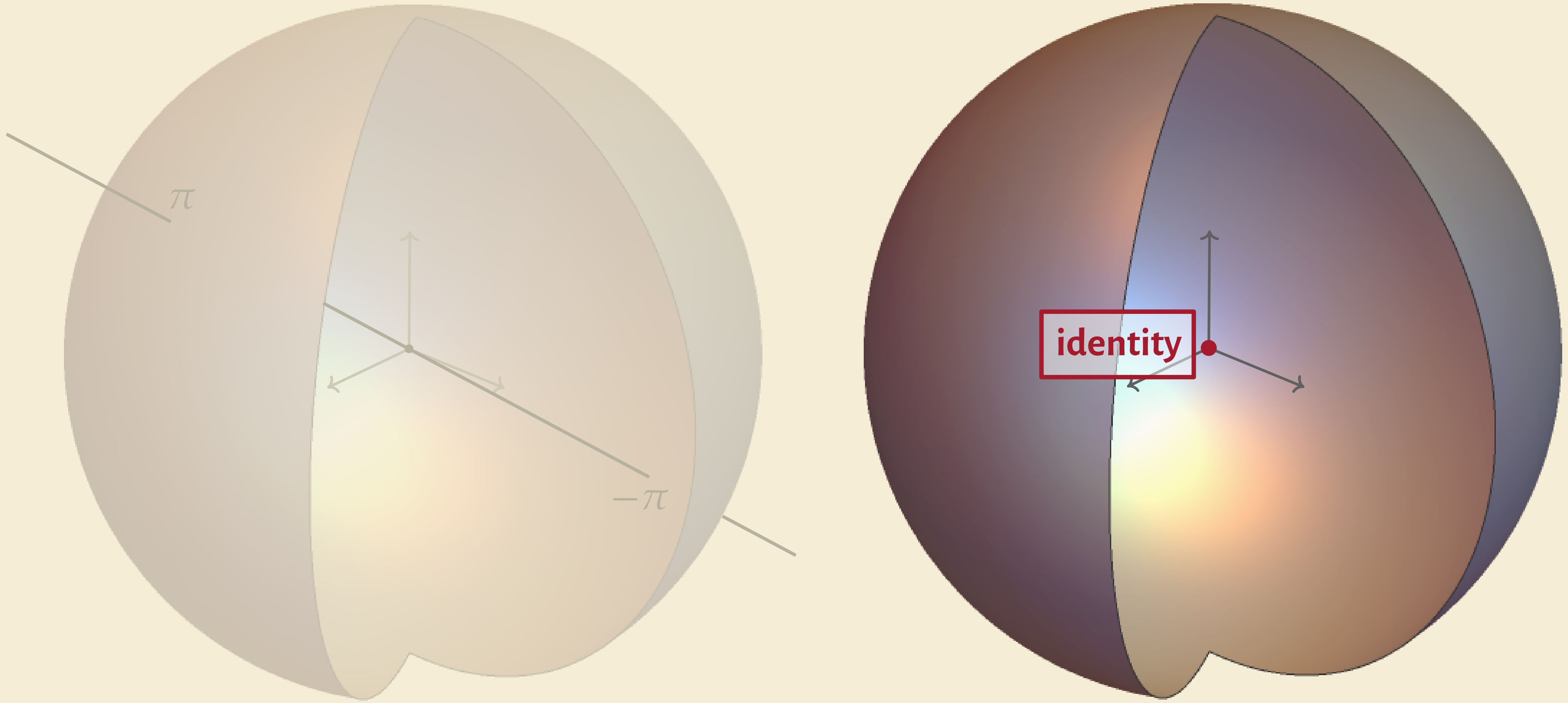
Dynamics in homogeneous spaces

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

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

Lie group

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi}/\sim$:



$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \vee x = y$

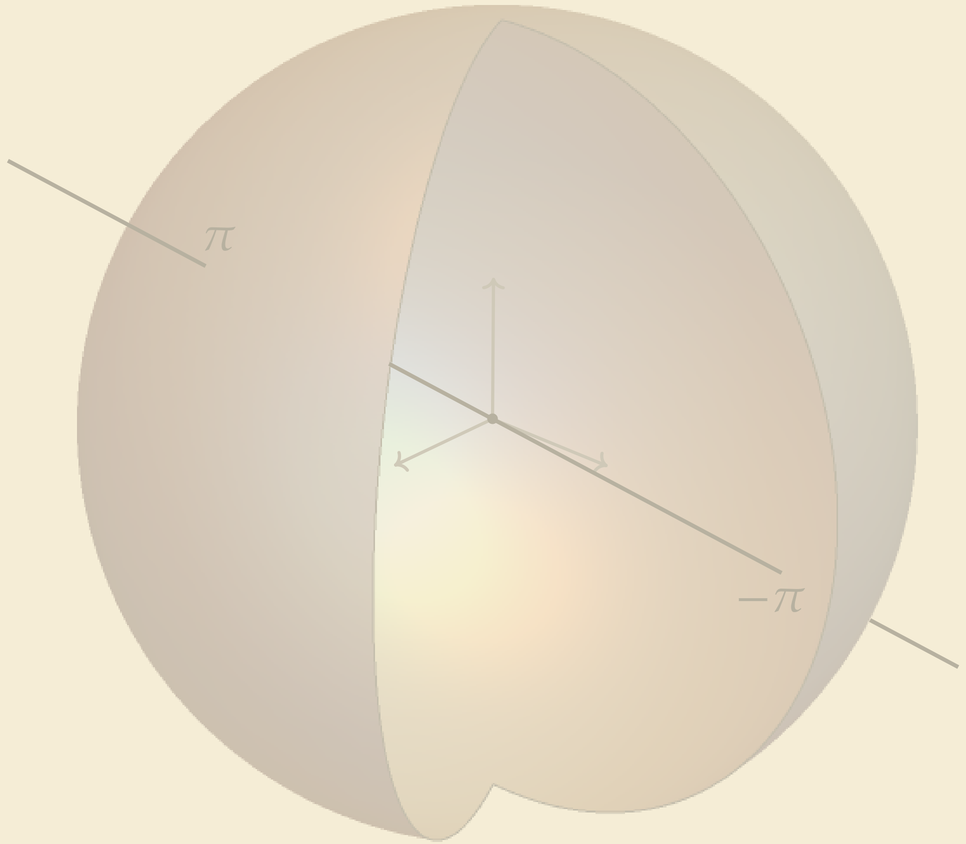
Dynamics in homogeneous spaces

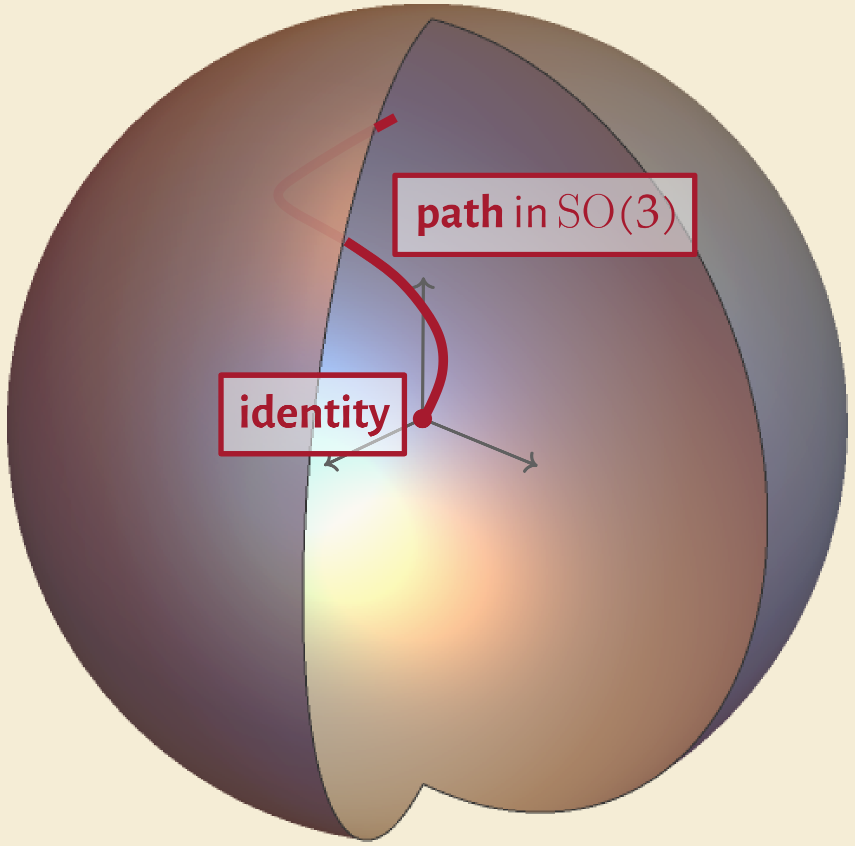
Example continues: $\mathbb{S}^2$ with (quatratic) matrix Lie group $\mathrm{SO}(3)$.

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

Lie group

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi}/\sim$:





$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \vee x = y$

Dynamics in homogeneous spaces

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

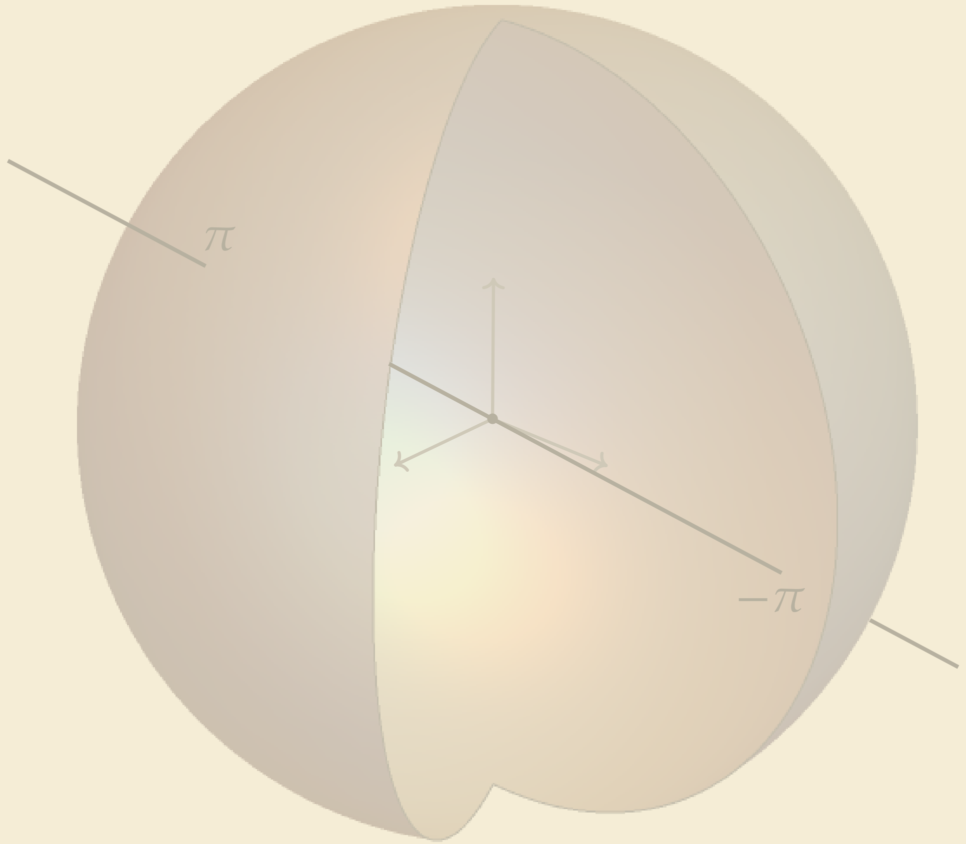
$$\mathrm{SO}(n) = \{\mathbf{A} \in \mathrm{Mat}(n) : \mathbf{A}^\top \mathbf{A} = \mathbf{I}_n, \det(\mathbf{A}) = +1\}$$

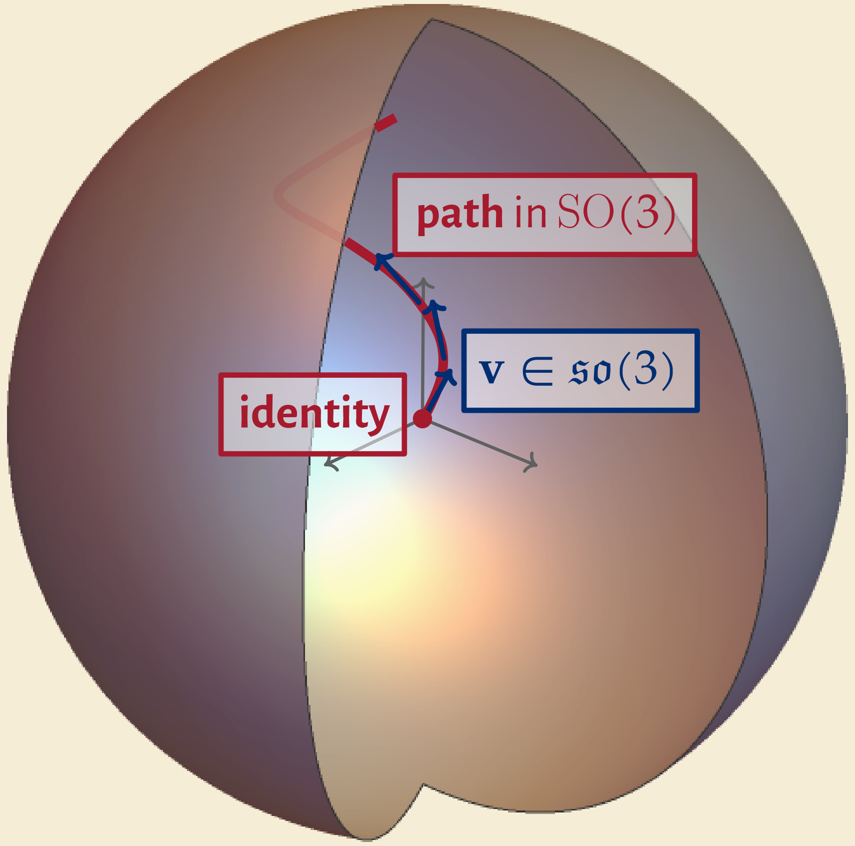
Lie group

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

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

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi}/\sim$:





$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \vee x = y$

Dynamics in homogeneous spaces

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

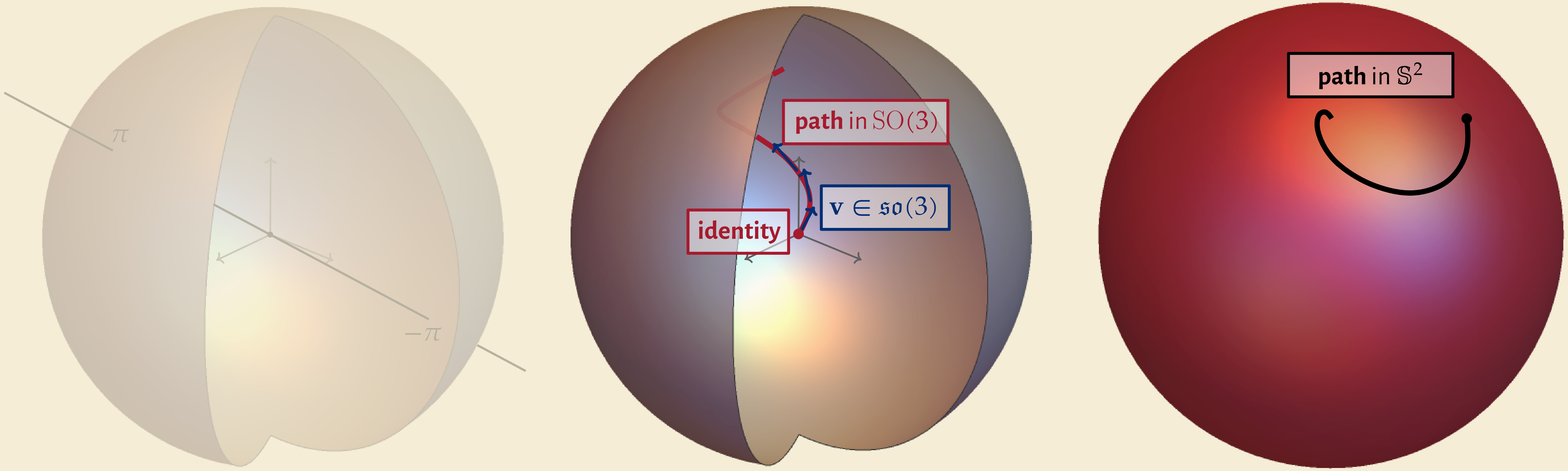
$$\boxed{\mathrm{SO}(n) = \{\mathbf{A} \in \mathrm{Mat}(n) : \mathbf{A}^\top \mathbf{A} = \mathbf{I}_n, \det(\mathbf{A}) = +1\}}$$

Lie group

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

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

Sketch for $\mathrm{SO}(3) \cong \mathrm{Int}\mathbb{B}^3_{r=\pi} \cup \partial\mathbb{B}^3_{r=\pi} / \sim$:



$\mathrm{Mat}(n)$ – space of real $n \times n$ matrices; $\mathbb{B}^3_{r=\pi}$ – 3-Ball of radius π; $x \sim y :\Leftrightarrow x = -y \vee x = y$

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

$$\boxed{\mathrm{d}G_t = \left(\mathbf{V}_0(t)\mathrm{d}t + \sum_{i=1}^{m} \mathrm{d}w_t^i \mathbf{V}_i\right) G_t\,, \quad G_0 = \mathbf{I}_n\,.}$$

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

$$\boxed{\mathbf{V}_0(t) = \mathbf{K}(t) + \frac{1}{2}\sum_{i=1}^{m} \mathbf{V}_i^2}$$

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

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

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

$$\mathbf{V}_0(t) = \mathbf{K}(t) + \frac{1}{2}\sum_{i=1}^{m} \mathbf{V}_i^2$$

$$\mathrm{d}G_t = \left(\mathbf{V}_0(t)\mathrm{d}t + \sum_{i=1}^{m} \mathrm{d}w_t^i \mathbf{V}_i \right) G_t\,, \quad G_0 = \mathbf{I}_n\,.$$

Diffusion

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

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

$$\mathbf{V}_0(t) = \mathbf{K}(t) + \frac{1}{2}\sum_{i=1}^{m} \mathbf{V}_i^2$$

$$\mathrm{d}G_t = \left(\mathbf{V}_0(t)\mathrm{d}t + \sum_{i=1}^{m} \mathrm{d}w_t^i \mathbf{V}_i\right) G_t\,, \quad G_0 = \mathbf{I}_n\,.$$

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

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

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

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

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

$$\mathbf{V}_0(t) = \mathbf{K}(t) + \frac{1}{2}\sum_{i=1}^{m}\mathbf{V}_i^2$$

$$\mathrm{d}G_t = \left(\mathbf{V}_0(t)\mathrm{d}t + \sum_{i=1}^{m}\mathrm{d}w_t^i\mathbf{V}_i\right)G_t\,, \quad G_0 = \mathbf{I}_n\,.$$

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

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

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

$\mathcal{P}$ ... 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**.

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

$$\mathbf{V}_0(t) = \mathbf{K}(t) + \frac{1}{2}\sum_{i=1}^{m} \mathbf{V}_i^2$$

$$\mathrm{d}G_t = \left(\mathbf{V}_0(t)\mathrm{d}t + \sum_{i=1}^{m} \mathrm{d}w_t^i \mathbf{V}_i\right) G_t\,, \quad G_0 = \mathbf{I}_n\,.$$

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

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

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

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

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

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**.

Learning objective

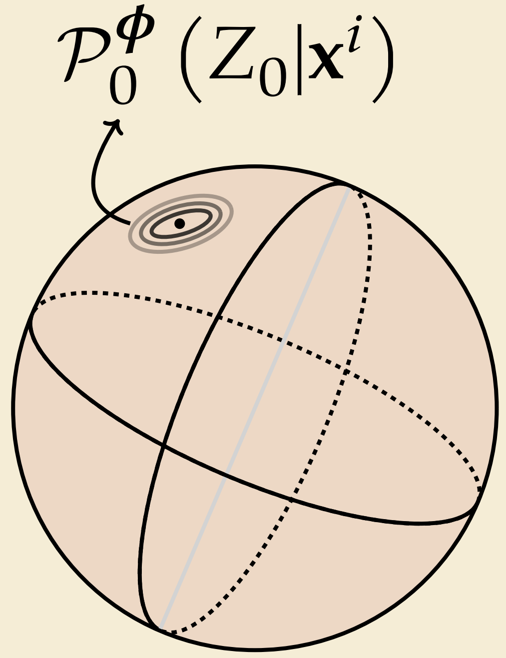
The **overall objective** (for our example of latent paths on $\mathbb{S}^{n-1}$):

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

Learning objective

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

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

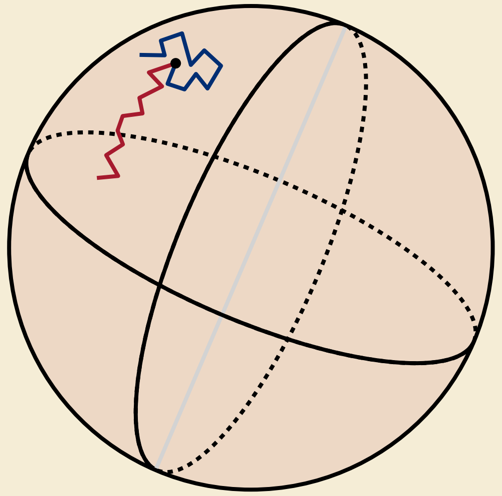


KL-div. to **uniform distribution** on $\mathbb{S}^{n-1}$

Learning objective

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

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



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

(essentially penalizes large rotations)

Learning objective

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

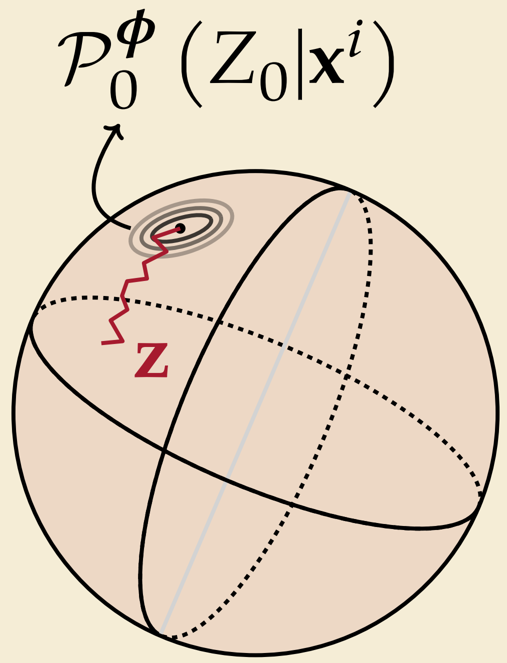
$$\begin{aligned}\mathcal{L}(\boldsymbol{\phi}, \boldsymbol{\theta}; \mathbf{x}^i) = D_{\mathrm{KL}}\left(\mathcal{P}_0^{\boldsymbol{\phi}}\left(Z_0|\mathbf{x}^{\mathbf{i}}\right) \| \mathcal{U}_{\mathbb{S}^{n-1}}\right) \\ + \frac{1}{2}\int_0^T \int_{\mathbb{S}^{n-1}} q_{Z_t}(\mathbf{z}) \|\mathbf{K}^{\boldsymbol{\phi}}(\mathbf{x}^i)(t)\mathbf{z}\|^2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

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

Learning objective

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

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



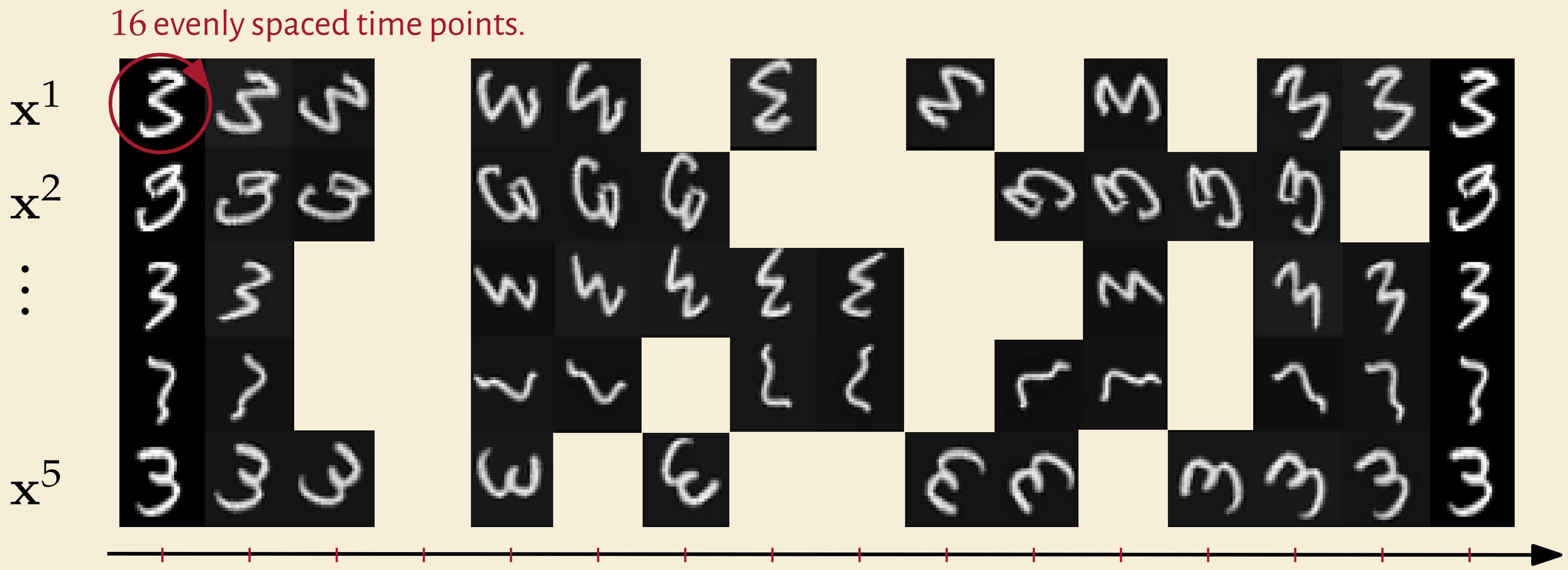
- **Sampling** from the approximate posterior:

  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

Five training samples of handwritten **rotating 3's** from `Rotating MNIST`

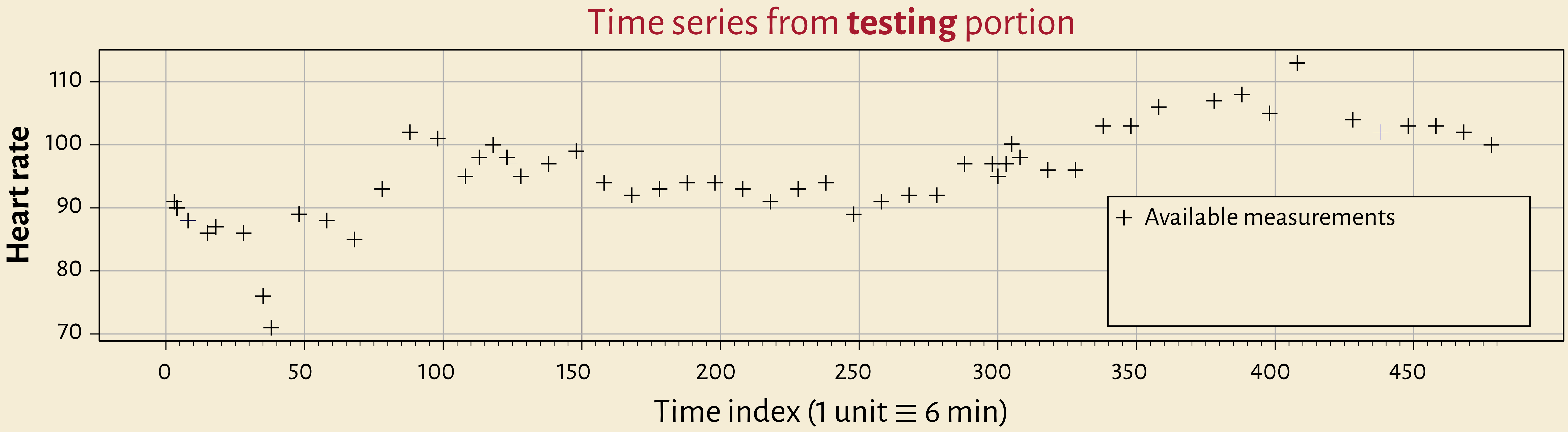


	$\downarrow$**MSE** $(\times 10^{-3})$
†GPPVAE-dis	30.9 ± 0.02
†GPPVAE-joint	28.8 ± 0.05
†ODE^{2}VAE	19.4 ± 0.06
†ODE^{2}VAE-KL	18.8 ± 0.31
CNN-ODE	14.5 ± 0.73
Ours	$\mathbf{11.8 \pm 0.25}$

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

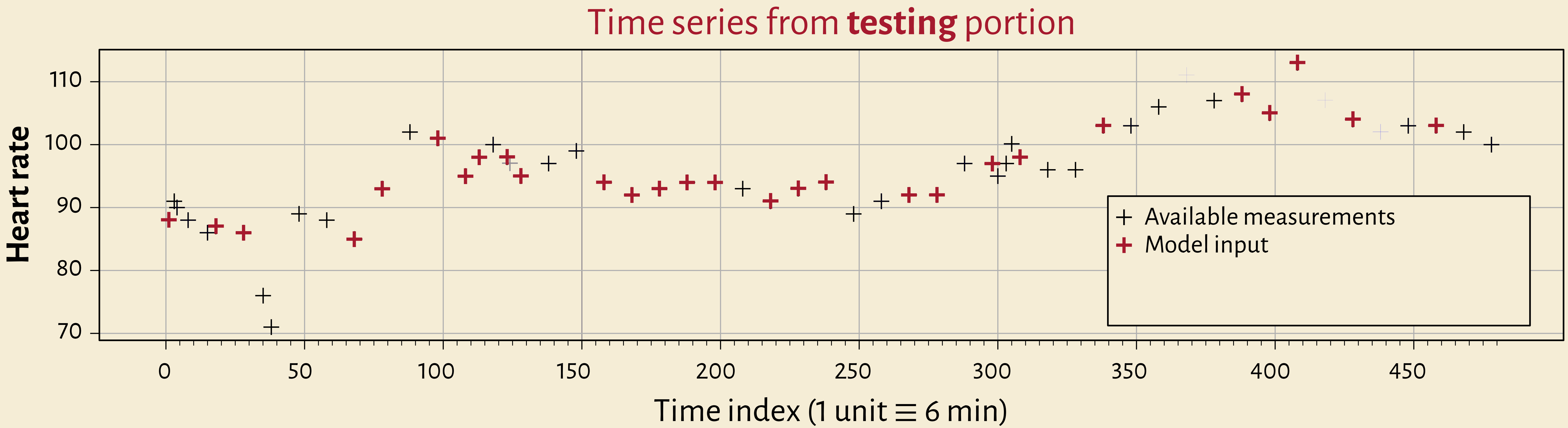
Some results

`PhysioNet (2012)` **interpolation task** (see [Shukla & Marlin, 2021]):



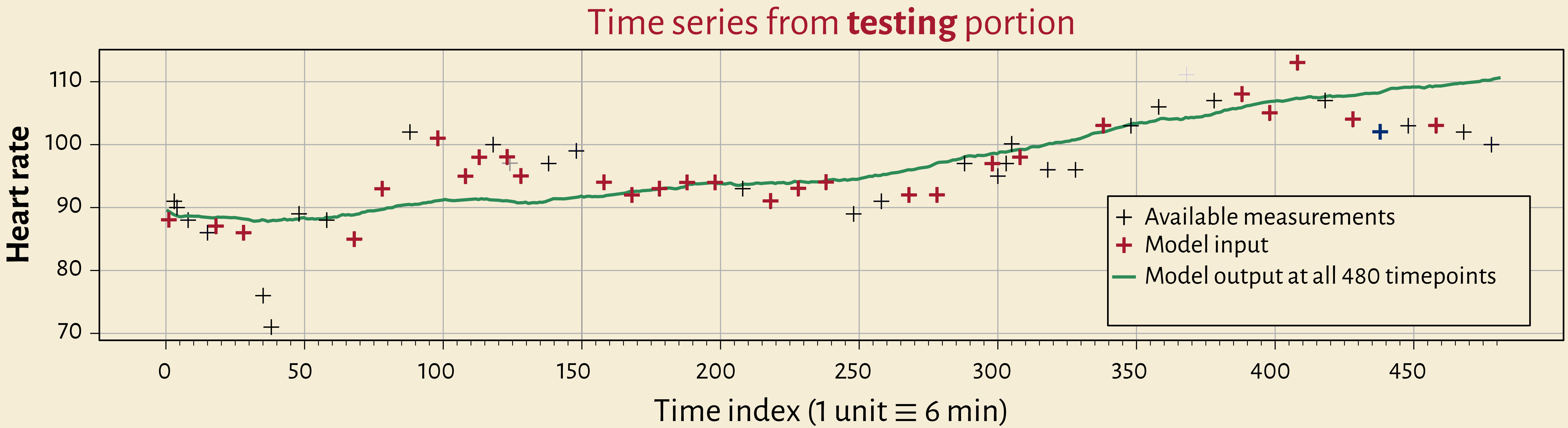
Some results

`PhysioNet (2012)` **interpolation task** (see [Shukla & Marlin, 2021]):



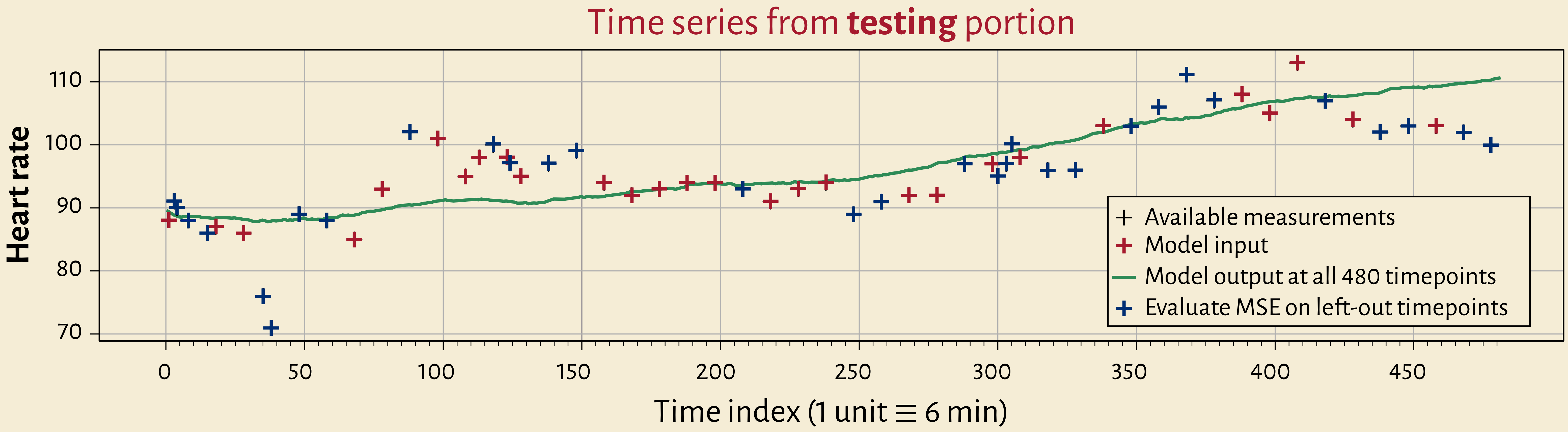
Some results

`PhysioNet (2012)` **interpolation task** (see [Shukla & Marlin, 2021]):



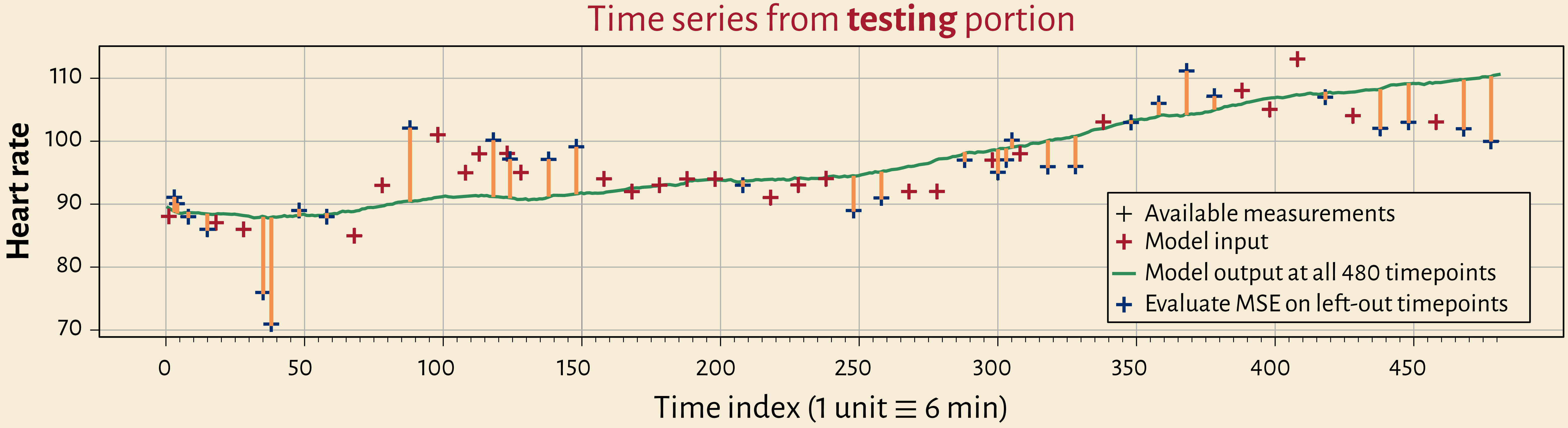
Some results

`PhysioNet (2012)` **interpolation task** (see [Shukla & Marlin, 2021]):



Some results

`PhysioNet (2012)` **interpolation task** (see [Shukla & Marlin, 2021]):



	↓**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	$\mathbf{3.11 \pm 0.02}$	

Thanks for your attention!

Come see us at our **poster # 1400**
Wed 13 Dec 5 p.m. CST @ Great Hall & Hall B1+B2

Full source code available at
`https://github.com/plus-rkwitt/LatentSDEonHS`

References

[G. Marjanovic and V. Solo]
"Numerical Methods for Stochastic Differential Equations in Matrix Lie Groups Made Simple".
In: *IEEE Transactions on Automatic Control* 63.12 (2018), pp. 4035–4050.

[M. Muniz, M. Ehrhardt, M. Günther, and R. Winkler]
"Higher strong order methods for linear Itô SDEs on matrix Lie groups".
In: *BIT Numerical Mathematics* 62.4 (2022), pp.1095–1119.

[Ç. Yildiz, M. Heinonen, and H. Lahdesmäki]
"ODE^{2}VAE: Deep generative second order ODEs with Bayesian neural networks".
In: *NeurIPS*. 2019.

[M. Schirmer, M. Eltayeb, S. Lessmann, and M. Rudolph]
"Modeling Irregular Time Series with Continuous Recurrent Units".
In: *ICML*. 2022.

[S. N. Shukla and B. M. Marlin]
"Multi-Time Attention Networks for Irregularly Sampled Time Series".
In: *ICLR*. 2021.