Latent GP-ODEs with Informative Priors

Ilze Amanda Auzina [1] Çağatay Yıldız [2] Efstratios Gavves [1]

[1]University of Amsterdam [2]University of Tübingen

Probablistic Generative Model: VAE-GP-ODE

For a given high dimensional input sequence $\mathbf{x}_{0:N}$ our probabilistic generative model can be described as:

$\mathbf{s}_0 \sim p(\mathbf{s}_0)$ position encoder

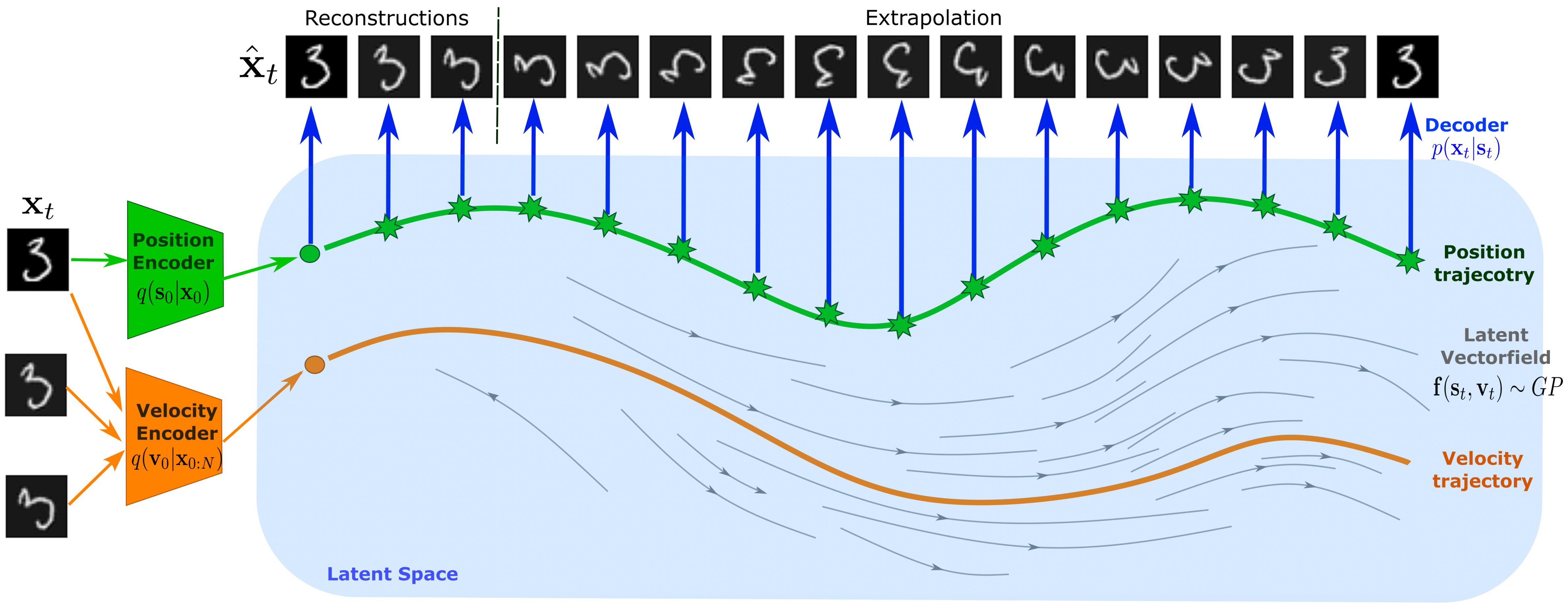
$\mathbf{v}_0 \sim p(\mathbf{v}_0)$ velocity encoder

$\mathbf{s}_t = \mathbf{s}_0 + \int_0^t \mathbf{v}_\tau d\tau$ position trajectory

$\mathbf{v}_t = \mathbf{v}_0 + \int_0^t \mathbf{f}(\mathbf{s}_\tau, \mathbf{v}_\tau) d\tau$ velocity trajectory

$\mathbf{x}_i \sim p(\mathbf{x}_i|\mathbf{s}_i), i \in [0, N]$ decoder

where the vector field $\mathbf{f}$ is modelled via Gaussian Process (GP-ODE).



Earlier Work and Contribution

Background: We consider multivariate autonomous ODEs of the form

$$\dot{\mathbf{x}}_t := \frac{d\mathbf{x}_t}{dt} = \mathbf{f}(\mathbf{x}_t) \qquad \mathbf{f}: \mathbb{R}^D \to \mathbb{R}^D$$

where the vector field $\mathbf{f}$ is learned via Gaussian Process prior

$$\mathbf{f}(\mathbf{x}) \sim GP(\mathbf{0}, K(\mathbf{x}, \mathbf{x}')), \qquad K(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{D \times D}$$

Task: Learning dynamics from high-dimensional data sequences

Existing GP-ODE models:

- Limited to low-dimensional data setting [2]
- No use of prior knowledge of the dynamical system [4]
- Decoupled training of the latent space embedding and the dynamics [3]

We propose: A ***probabilistic dynamical*** model that extends previous work by

- Learning dynamics from high-dimensional data
- End-to-end training via Variational Inference (VI)
- Informative GP priors

details

High-dimensional data for GPs

We combine a VAE with a GP-ODE. The resulting dynamics are learned in a latent space and our model supports both 1st and 2nd order differential euqations.

End-to-end training

We train our model end-to-end because decoupled training leads to embeddings that are unconstrained by the dynamics of the observed process, leading to poor generalization (see Pretrained VAE).

This is achieved by optimising the following ELBO:

$$\begin{aligned} L = \; & \mathbb{E}_{q(\mathbf{f},\mathbf{U},\mathbf{z}_0|X)}\left[\log p(X|\mathbf{z}_0, \mathbf{f})\right] && \text{(VAE reconstruction term)} \\ & - KL\left[q_{\text{enc}}(\mathbf{z}_0|\mathbf{x}_{0:N})||p(\mathbf{z}_0)\right] && \text{(VAE regularization term)} \\ & - KL\left[q(\mathbf{U})||p(\mathbf{U})\right] && \text{(GP inducing KL)} \end{aligned}$$

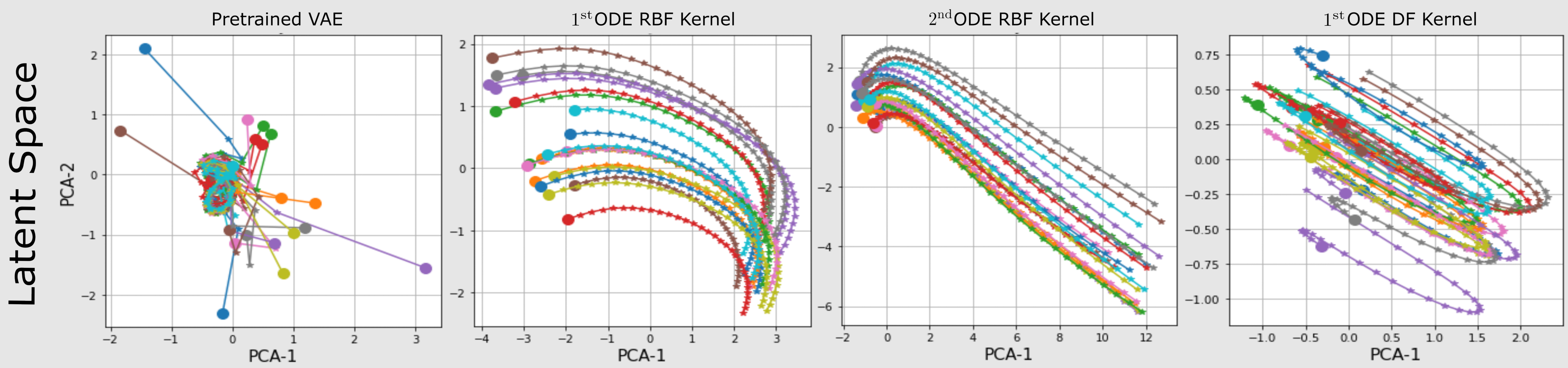
Informative Priors:

In order to bridge the gap between the unobserved true and the inferred dynamics, we set the kernel of the GP-ODE model to have **divergence-free** (DF) properties.

Benefit → reduced search space of the *true* model.

Results: Extrapolation and Forecasting

Dataset: Sequences of rotating handwritten 3s with random initial starting angle. The model is conditioned on either only the initial frame $\mathbf{x}_1$ (1st ODE) or initial 5 frames $\mathbf{x}_{1:5}$ (2nd ODE).



References

[1] Mauricio A Alvarez, Lorenzo Rosasco, Neil D Lawrence, et al. Kernels for vector-valued functions: A review. *Foundations and Trends® in Machine Learning*, 4(3):195–266, 2012.

[2] Pashupati Hegde, Çağatay Yıldız, Harri Lähdesmäki, Samuel Kaski, and Markus Heinonen. Bayesian inference of odes with gaussian processes. *arXiv preprint arXiv:2106.10905*, 2021.

[3] Arno Solin, Ella Tamir, and Prakhar Verma. Scalable inference in sdes by direct matching of the fokker–planck–kolmogorov equation. *Advances in Neural Information Processing Systems*, 34:417–429, 2021.

[4] Cagatay Yildiz, Markus Heinonen, and Harri Lahdesmaki. Ode2vae: Deep generative second order odes with bayesian neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

ilze.amanda.auzina@gmail.com