A Variational Approximation for Analyzing the Dynamics of Panel data

JURIJS NAZAROVS[1], RUDRASIS CHAKRABORTY[2], SONGWONG TASNEEYAPANT[1], SATHYA N. RAVI[3], VIKAS SINGH[1]

[1] University of Wisconsin - Madison, USA [2] Amazon Lab 126, USA [3] University of Illinois at Chicago, USA

37th Conference on Uncertainty in Artificial Intelligence
July 27-30, 2021
Online
uai2021

OVERVIEW

- We propose a probabilistic framework (ME-NODE), incorporating (fixed + random) mixed effect in ODE, to appropriately model the variability of panel data.
- Our model can be considered as a smooth approximation of SDE, but the underlying neural network can be trained efficiently using ODE solvers.
- For training the model, we derive the new Evidence Lower Bound loss for our ME-NODE model.
- We show applications in different dynamical/temporal settings including longitudinal brain image analysis.

MOTIVATION

- Panel data involves longitudinal measurements of the same set of participants, which requires not only modeling the temporal dynamics, but also account for *variability within and across individuals*.
- Two ways to model dynamical systems:

Types	Speed	Libraries	Uncertainty
SDE	✗	✓	✓
ODE	✓	✓	✗

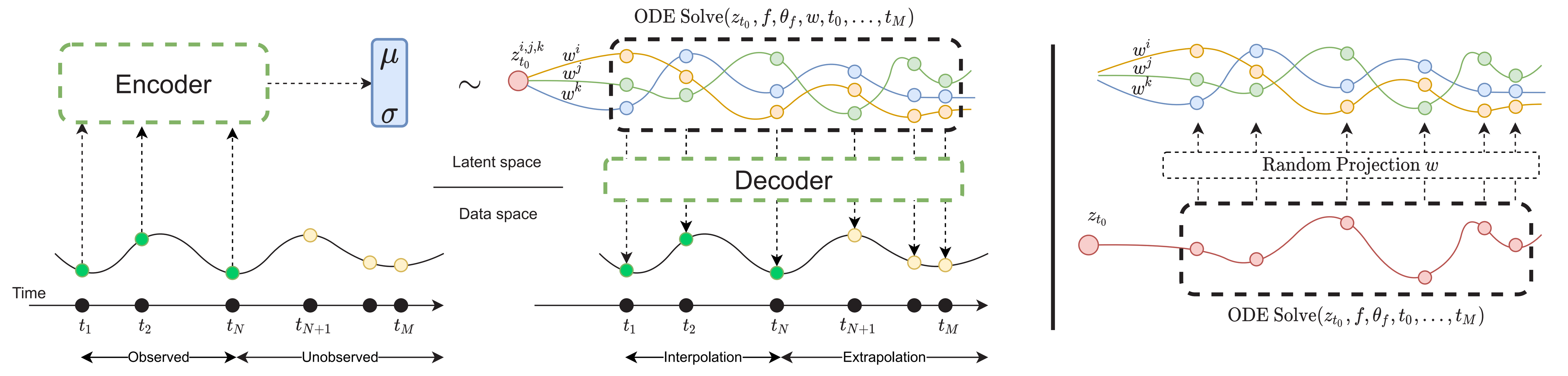
- To capture some of the characteristics of panel data, literature incorporates white noise type functions in differential equation models, leading to various forms of stochastic differential equation (SDE):

$$z_t = f_\mu(z,t)dt + L_\Sigma(z,t) \circ d\beta(t). \qquad (1)$$

- Recent work shows interfacing deep methods with SDEs but SDE solvers are far more involved than ODEs.
- Is there a way to utilize ODE solvers, while preserving stochastic nature of Equation (1)?

In special cases, yes!

METHODS AND MAIN RESULTS



- Smooth Approximations of Stratonovich SDE

$$z_t = f_\mu(z,t)dt + L_\Sigma(z,t) \circ d\beta(t)$$

$$\Big\downarrow \text{Wong and Zakai Thm.}$$

$$\dot{z}_t = f(z,t) + L(z,t)\textstyle\sum_{n=1}^N b_n\varphi_n(t)$$

$$\Big\downarrow$$

$$\dot{z}_t = f(z,t) + g(z,t)b$$

$$\Big\downarrow {\scriptstyle f(z,t)=\Gamma(z,t)\beta,\; g(z,t)=\Gamma(z,t)\sqrt{\Sigma_b}}$$

$$\dot{z}_t = \Gamma(z,t)\mathbf{w},$$

$$\text{where } \mathbf{w} = \beta + b\sqrt{\Sigma_b} \sim \mathcal{N}(\beta, \sqrt{\Sigma_b}).$$

- Benefits of smooth approximation: **(a)** models uncertainty like SDE, but **(b)** need a ODE solver and the associated computational benefits.
- Final model

$$\begin{bmatrix} z_0^i \sim \mathcal{N}(\mu,\sigma), \text{ where } \mu,\sigma = E(\boldsymbol{x}^i) \\ \mathbf{w}^i = \boldsymbol{\beta} + b^i \sim \mathcal{N}(\boldsymbol{\beta}, \Sigma_b) \\ \dot{z}_t^i = \Gamma\left(z_t^i\right)\mathbf{w}^i \\ \mathbf{x}_t^i = D(z_t^i) + \boldsymbol{\epsilon}_t \end{bmatrix}$$

- Loss: given the nature of constructed trajectories, ME-NODE results in a special parameterization of approximate posterior (and prior):

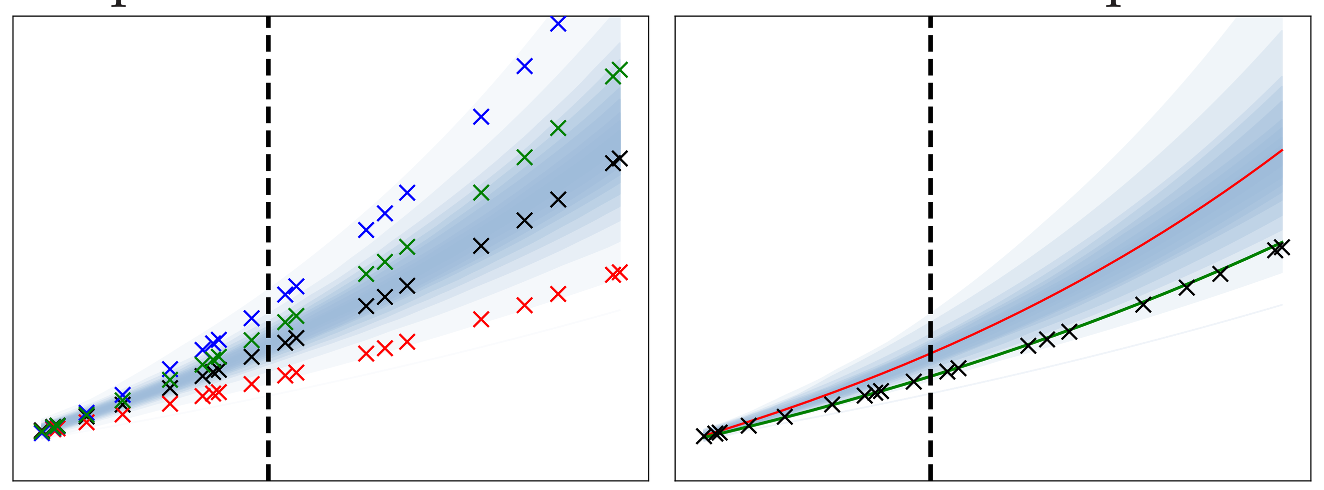
$$q(\boldsymbol{z},\mathbf{w}) = 1_{\boldsymbol{z}_{-0}^{\text{obs}}}\{\boldsymbol{z}_{-0}|z_0,\mathbf{w}\}q\left(z_0\right)q\left(\mathbf{w}\right)$$

The final loss takes the form:

$$\frac{1}{|S|}\sum_{s\in S}\left(\log p\left(x|\boldsymbol{z}^s,\mathbf{w}^s\right) - \log\frac{q(z_0^s)q(\mathbf{w}^s)}{p(z_0^s)p(\mathbf{w}^s)}\right),$$

where $S = \{\forall s \in S : 1_{\boldsymbol{z}_{-0}^{\text{obs}}}\{\boldsymbol{z}_{-0}|z_0^s,\mathbf{w}^s\} = 1\}$.

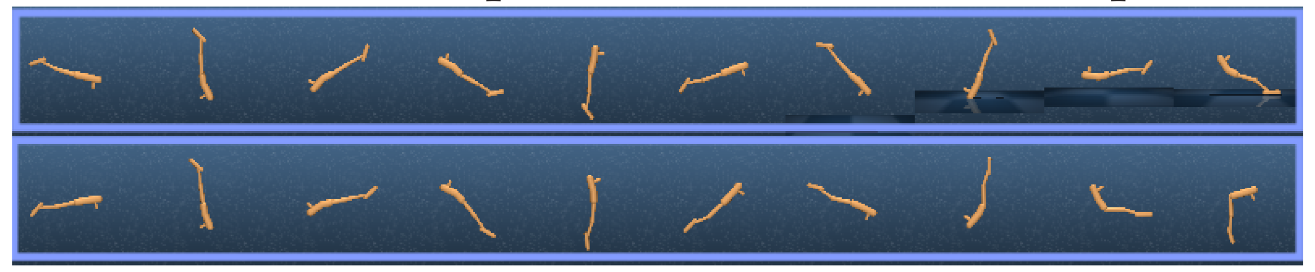
- Efficient sampling: approximating $1_{\boldsymbol{z}_{-0}^{\text{obs}}}\{\boldsymbol{z}_{-0}|z_0,\mathbf{w}\}$. ABC approximation of $1_y\{z\}$ as $1_{A_{\epsilon,y}}\{z\}$
- Personalized prediction: utilize observed data to select personal mixed effect $\Longrightarrow$ better extrapolation.



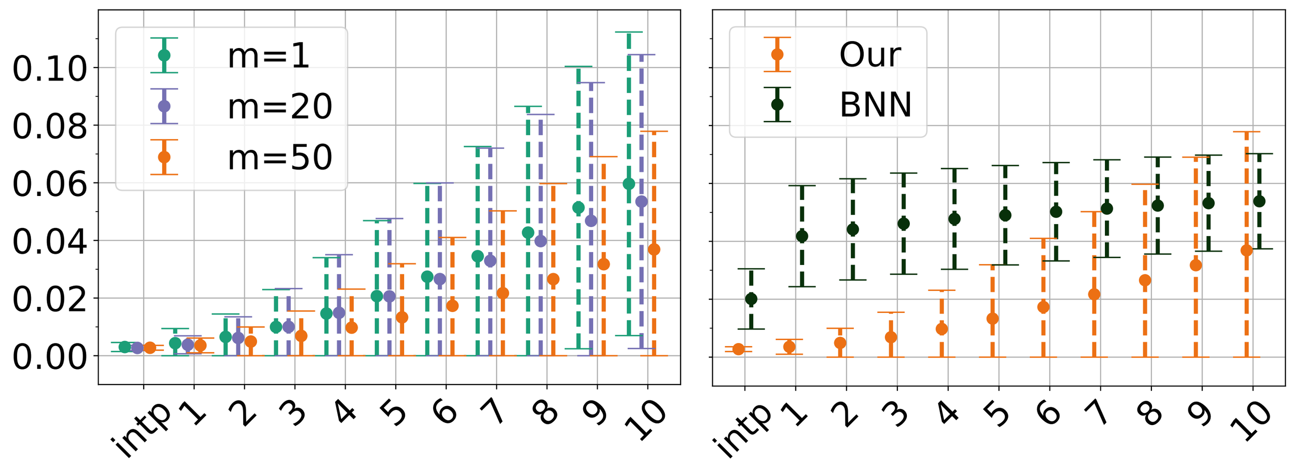
Research supported in part by by NIH grants RF1 AG059312 and RF1 AG062336. SNR was supported by UIC start-up funds.

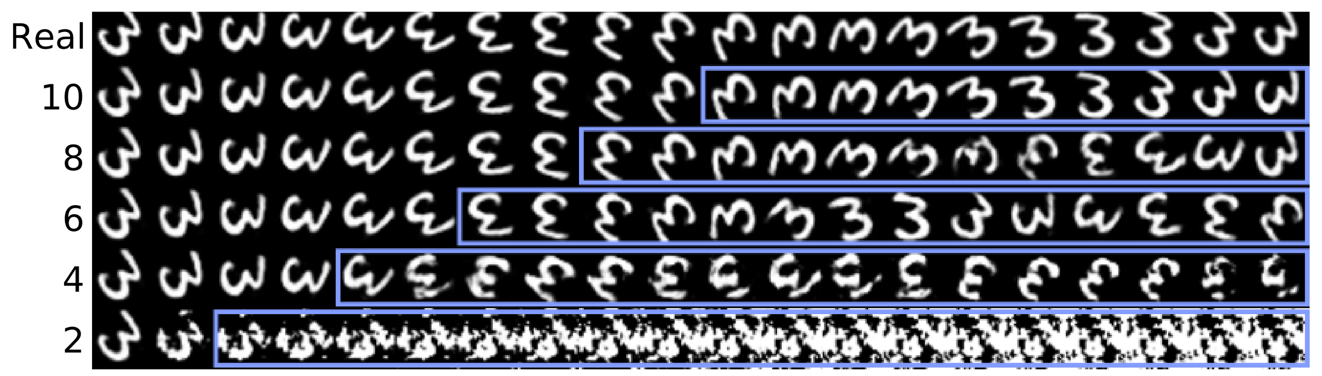
DETAILED RESULTS

- **Mujoco**: Model learns correlation structure between samples, so improvements in both interpolation and extrapolation



- **Rotation MNIST**: The role of dimensionality/size of w on results is sensible. For dataset with 8 possible angles. As dimension m grows we see improvements in MSE (left). The MSE remains stable for extrapolation steps (right).





- **Brain imaging**: Experiments on two different types of data (image derived summaries and whole brain images) suggest viability of ME-NODE for modeling longitudinal 3D imaging data.