Variational Integrator Networks for Physically Structured Embeddings

Author: Igor Alentev

Robotics Track

i.alentev@innopolis.university

INNOPOLIS UNIVERSITY

Table of Contents

INNOPOLIS UNIVERSITY

What is *Neural Differential Equation* anyway?

Neural Differential Equation

A *neural differential equation* is a differential equation using a neural network to parameterise the vector field. The canonical example is a *neural ordinary differential equation*.

$$y(0) = y_0 \tag{1}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t}(t) = f_\theta(t, y(t)) \tag{2}$$

Where θ is some vector of learnt parameters. Usually, f_θ is a feedforward network.

[1] Chen et al. 2019.

[2] Kidger 2022.

Residual Network (ResNet)

$$y_{j+1} = y_j + f_\theta(j, y_j) \tag{3}$$

Where $f_\theta(j, \cdot)$ is j-th residual block. With θ as vector of parameters from all layers.

If we try the discretization of neural ODE, it might start looking familiar.

$$\frac{y(t_{j+1}) - y(t_j)}{\Delta t} \approx \frac{\mathrm{d}y}{\mathrm{d}t}(t) = f_\theta(t_j, y(t_j)) \tag{4}$$

If we absorb the discretization step into the f_θ, we can derive:

$$y(t_{j+1}) = y(t_j) + f_\theta(t, y_j) \tag{5}$$

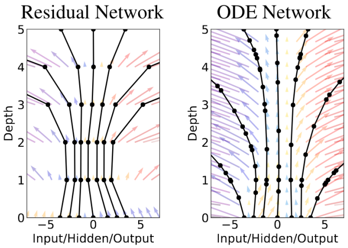


Figure 1: *Left:* A Residual network defines a discrete sequence of finite transformations. *Right:* A ODE network defines a vector field, which continuously transforms the state. *Both:* Circles represent evaluation locations.

Why to bother with constant discrete number of hidden layers?
Continuous-layer architecture allows:

1. Precision+Accuracy tuning
2. Constant memory
3. Fast backprop
4. Continuous evaluation

[3] Yin et al. 2019.

1. *Neural ODEs* are the continuous limit of residual networks.
2. **GRU** and **LSTM** updates rules suspiciously similar to discretised differential equations.
3. **StyleGAN2** is simply discretised *SDE*
4. Invertible NN coupling layers are reversible DE solvers

Many of the DL architectures resemble DEs.

Neural Network $\Leftrightarrow$ Differential Equation

Variational Integrator Networks[4]

A Bridge

VIN is the bridge between the viewpoint of representing deep residual networks as discretisation of differential equations and the viewpoint of geometric embeddings.

Geometric Embeddings

Geometric Embeddings[a b c] is the way of embedding data into it's natural geometry, preserving relational information.

[a] Chamberlain, Clough, and Deisenroth 2017.
[b] Davidson et al. 2022.
[c] Xiong et al. 2023.

[4] Saemundsson et al. 2020.

Principle of Least Action

$$S(\mathbf{q}) = \int_0^T L(\mathbf{q}(t), \dot{\mathbf{q}}(t))\mathrm{d}t \tag{6}$$

Euler-Lagrange Equations

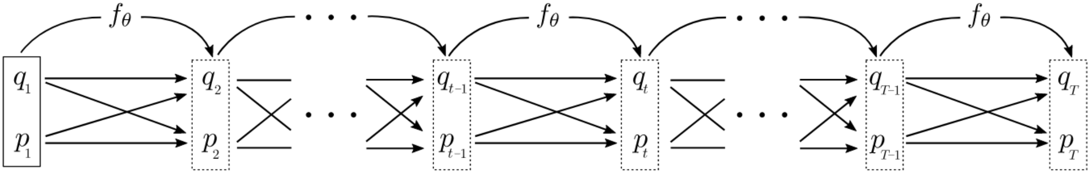
$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\mathbf{q}}} - \frac{\partial L}{\partial \mathbf{q}} = 0 \tag{7}$$

Discrete Euler-Lagrange in Momentum Form

$$\mathbf{p}_k = -D_1 L_d(\mathbf{q}_k, \mathbf{q}_{k+1}, \Delta t) \tag{8}$$

$$\mathbf{p}_{k+1} = D_2 L_d(\mathbf{q}_k, \mathbf{q}_{k+1}, \Delta t) \tag{9}$$

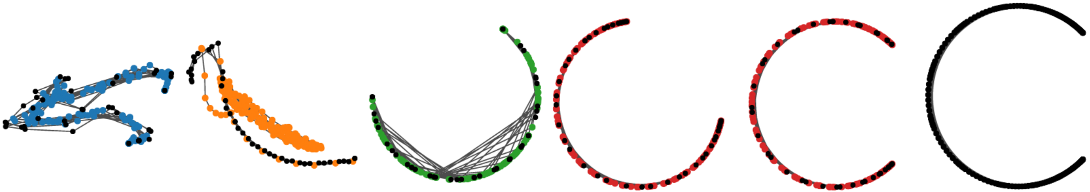
[5] West 2004.



Overall structure of the VIN resembles the residual network we seen before. We consider it to be a multilayered (discretised) dynamics of the system. $(\mathbf{q}, \mathbf{p})$ is the hidden space, which is 2-form of position-momentum phase-space and f_θ we have seen earlier.

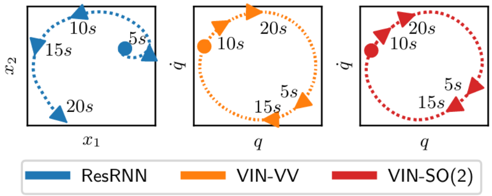
1. Physical properties conservation automatically enforced by underlying structure
2. Interpretability is increased.
   Embedding evolves in the phase-space, which has notions of kinetic and potential energy.
3. Complex phenomena modelling is allowed due to the black-box nature of potential.

Case study: *Ideal pendulum*. Noisy observations. Image data.



(a) VAE (b) DVAE (c) LG-VAE (d) VIN-$SO(2)$ (e) VIN-$SO(2)$ (fixed) (f) Ground Truth

What if we will consider VIN in terms of VAE?



Let's infer and evolve the embeddings of ResRNN and VIN. The main question is the preservation of evolution patterns.

Conclusion & Questions

References I

- Chamberlain, Benjamin Paul, James Clough, and Marc Peter Deisenroth (May 2017). *Neural Embeddings of Graphs in Hyperbolic Space*. arXiv: 1705.10359 [cs, stat].
- Chen, Ricky T. Q. et al. (Dec. 2019). *Neural Ordinary Differential Equations*. arXiv: 1806.07366 [cs, stat].
- Davidson, Tim R. et al. (Sept. 2022). *Hyperspherical Variational Auto-Encoders*. arXiv: 1804.00891 [cs, stat].
- Kidger, Patrick (Feb. 2022). *On Neural Differential Equations*. arXiv: 2202.02435 [cs, math, stat].
- Saemundsson, Steindor et al. (Mar. 2020). *Variational Integrator Networks for Physically Structured Embeddings*. arXiv: 1910.09349 [cs, stat].
- West, Matthew (May 2004). "Variational Integrators". PhD thesis. Pasadena, California: California Institute of Technology.

Xiong, Bo et al. (Apr. 2023). *Geometric Relational Embeddings: A Survey*. arXiv: 2304.11949 [cs].

Yin, Minghao et al. (Apr. 2019). *On the Mathematical Understanding of ResNet with Feynman Path Integral*. arXiv: 1904.07568 [hep-th, stat].