Supplementary Material for Latent Diffeomorphic Dynamic Mode Decomposition

Willem Diepeveen[a], Jon Schwenk[b], Andrea Bertozzi[a]

[a]Department of Mathematics, UCLA, Los Angeles, CA 90095, United States
[b]Los Alamos National Laboratory, Los Alamos, NM 87545, United States

Keywords: Interpretable Machine Learning, Time Series, Non-linear Systems, Data Reduction

1. Data set details

1.1. Synthetic data

The synthetic data in Section 5 of the main article is of the form

$$(\mathbf{x}^i, \mathbf{p}^i) = (\Phi^{-1} \circ \mathcal{K} \circ \Phi)(\mathbf{x}^{i-1}, \mathbf{p}^{i-1}), \quad \mathbf{v}^i = g(\mathbf{p}^i).$$

with

$$\Phi(\mathbf{x}, \mathbf{p}) := (\varphi_1(\mathbf{x}), f(\mathbf{x}) + \varphi_2(\mathbf{p})),$$

and

$$\mathcal{K} = \begin{bmatrix} \mathbf{K}_1 & \\ & \mathbf{K}_2 \end{bmatrix}.$$

The diffeomorphisms are given by

$$\varphi_1(\mathbf{x}) := (2(\mathbf{x}_1 - \sin(\mathbf{x}_2)), \frac{1}{4}\mathbf{x}_2),$$

and

$$\varphi_2(\mathbf{x}) := (2(\mathbf{x}_1 - \mathbf{x}_2^2 - 3), 3\mathbf{x}_2),$$

the coupling is given by

$$f(\mathbf{x}) := (\mathbf{x}_1^2 + \mathbf{x}_2^2, \mathbf{x}_1 - \mathbf{x}_2),$$

and the matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ are parametrized as

$$e^{-\mu_j \Delta t} \begin{bmatrix} \cos(\omega_j \Delta t) & -\sin(\omega_j \Delta t) \\ \sin(\omega_j \Delta t) & \cos(\omega_j \Delta t) \end{bmatrix},$$

with $\mu_1 = \mu_2 = 0$, $\omega_1 = \frac{\pi}{100}$, $\omega_2 = \frac{\pi}{100\sqrt{10}}$ and $\Delta t = 1$.

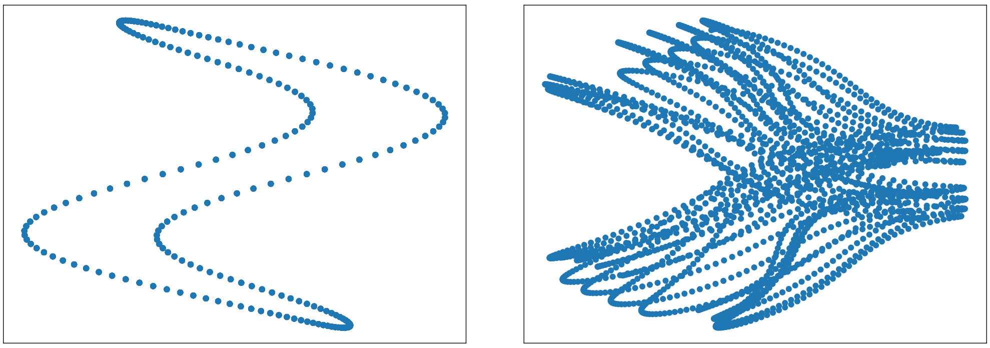
The system is initialized as

$$(\mathbf{x}^0, \mathbf{p}^0) := \Phi^{-1}([0, 1]^\top, [1, 1]^\top).$$

Finally,

$$g(\mathbf{p}) := \text{softplus}(-\mathbf{p}_1 - \frac{3}{4}\mathbf{p}_2 + 1\frac{1}{2}).$$

As the $\mathbf{x}$ and $\mathbf{p}$-variables are 2-dimensional, we can visualize them separately (see Figure 1).



(a) Clean **x**-variable data

(b) Clean **p**-variable data

Figure 1: A visualization of the synthetic data: whereas the **x**-variable follows a periodic orbit, the latent **p**-variable data are much more complicated due to the interplay with the memory of the system.

1.2. Real data

Streamflow observations (**v** variable) were provided by the United States Geological Survey (USGS). The example here is from Muddy Creek near Emery, Utah (USGS id 09330500). These data are available at https://waterdata.usgs.gov/monitoring-location/09330500/#dataTypeId=continuous-00065-0&period=P7D. Streamflow data are provided as daily averages in cubic feet per second. These data were normalized by the drainage area of the watershed of the gage, or 271.9 km^2 in this case, and converted to millimeters per day to arrive at **v**.

The **x**-variables were sampled over the watershed of the Muddy Creek gage by spatial averaging (or summing in the case of total precipitation). These variables include:

Variable Name	Variable Meaning
dewpoint_temperature_2m__mean__era5l_daily	Daily mean dewpoint temperature at 2 meters
potential_evaporation__sum__era5l_daily	Daily sum of potential evaporation
snow_depth_water_equivalent__mean__era5l_daily	Daily mean snow depth water equivalent
surface_net_solar_radiation__mean__era5l_daily	Daily mean surface net solar radiation
surface_net_thermal_radiation__mean__era5l_daily	Daily mean surface net thermal radiation
surface_pressure__mean__era5l_daily	Daily mean surface atmospheric pressure
temperature_2m__mean__era5l_daily	Daily mean air temperature at 2 meters
total_precipitation__sum__era5l_daily	Daily sum of total precipitation
u_component_of_wind_10m__mean__era5l_daily	Daily mean east-west wind component at 10 meters
v_component_of_wind_10m__mean__era5l_daily	Daily mean north-south wind component at 10 meters
volumetric_soil_water_layer_1__mean__era5l_daily	Daily mean volumetric soil water content, Layer 1 (0–7 cm)
volumetric_soil_water_layer_2__mean__era5l_daily	Daily mean volumetric soil water content, Layer 2 (7–28 cm)
volumetric_soil_water_layer_3__mean__era5l_daily	Daily mean volumetric soil water content, Layer 3 (28–100 cm)
volumetric_soil_water_layer_4__mean__era5l_daily	Daily mean volumetric soil water content, Layer 4 (100–289 cm)

2. Training details

Common parameters.

- **Batch size**: 256
- **Optimizer**: Adam with `betas` = (0.9, 0.99) and learning rate 10^{-3}.
- **Model Architecture**:
  - **Diffeomorphisms**: Additive coupling layer [1], which adds the mapping of the sum of two adjacent parity-0 inputs through a learnable order-2 polynomials to corresponding parity-1 entries (i.e., the parity-1 entry in between the two parity-0 entries).
  - **Couplings**: Multi layer perceptron network with ℓ_f hidden layers with dimension m_f (different per data set) and learnable polynomial activation functions of order 2 (unique polynomial per neuron).
  - **Regression Mappings**: Multi layer perceptron with one hidden layer with dimension m_g and softplus activation.

Data set-specific parameters. The remaining parameters are summarized below:

Data set	ℓ_f	m_f	m_g	**Epochs**
Synthetic	2	2	4	1000
Real	1	40	4	200

3. Additional numerical results

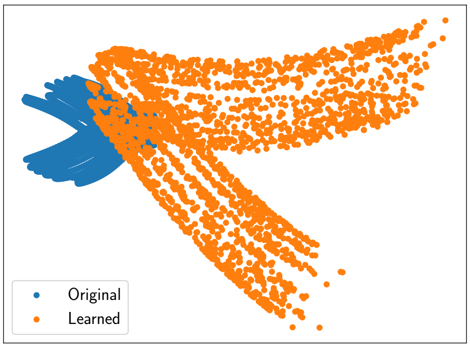
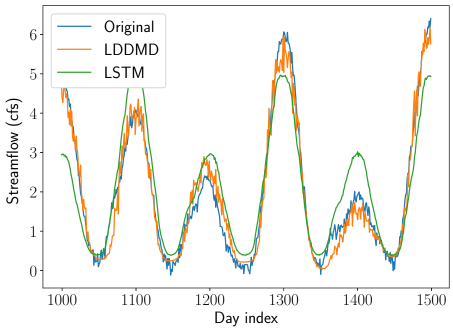
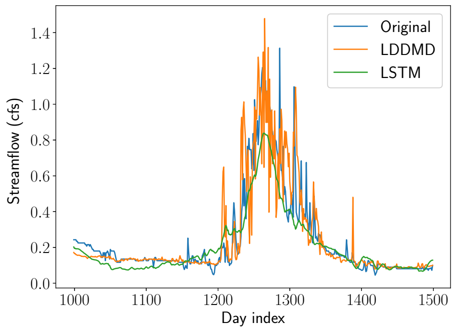


Figure 2: The ground truth and learned latent (**p**-variable) dynamics look very similar, but differ roughly by a rigid body transformation and rescaling. This indicates that the latent space does carry information for the synthetic data, i.e., LDDMD has an interpretable latent space.



(a) Syntethic data results



(b) Real-world data results

Figure 3: When examining a specific interval in the plots in Figure 1, it is evident that the LSTM predictions are considerably less noisy than those of LDDMD. This difference is likely due to the inherent averaging in LSTMs, where predictions at each time step incorporate information from many preceding time points, effectively smoothing out noise. In contrast, LDDMD lacks a comparable denoising mechanism.

References

[1] L. Dinh, D. Krueger, Y. Bengio, Nice: Non-linear independent components estimation, arXiv preprint arXiv:1410.8516 (2014).