SOLVER-FREE NEURAL ORDINARY DIFFERENTIAL EQUATIONS FOR FORECASTING LONG HORIZON TIME SERIES

Szymon Haponiuk

About me

- DL Algorithms team at NVIDIA
- Master’s student at University of Warsaw
- Music
    - Playing, composing, producing
    - DL applications, AI assisted workflows

Outline

- Forecasting, long horizon, why?
- Quick LTSF landscape analysis, inc. NeuralODE/LatentODE
- Curriculum Learning for long horizon time series
- Unified Long-Horizon Time-Series Benchmark
- Solver-free latent ODE

in this talk: trajectory = series (loosely speaking)

Forecasting

- We will focus on forecasting without static/dynamic covariates
- Onput: sequence of history states, sequence of history timestamps, sequence of horizon timestamps
- Output: sequence of horizon states
- Usually history is a long sequence and horizon is short
  - eg. history of 192 points, horizon of 24 points
- LTSF: long-term time-series forecasting
  - eg. history of 500 points, horizon of 500 points

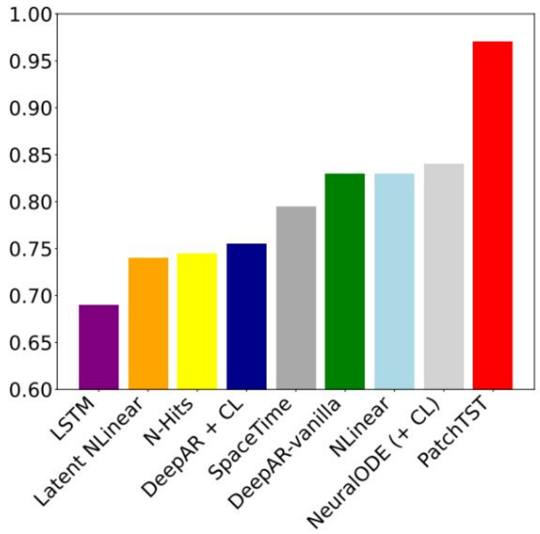
Why LTSF is hard?

- Long range dependencies
- Computational complexity
  - transformer models have quadratic-time complexity
  - RNN-based models deal with vanishing/exploding gradients
- Compounding errors
- It may simply be impossible to predict that far into the future with such data…

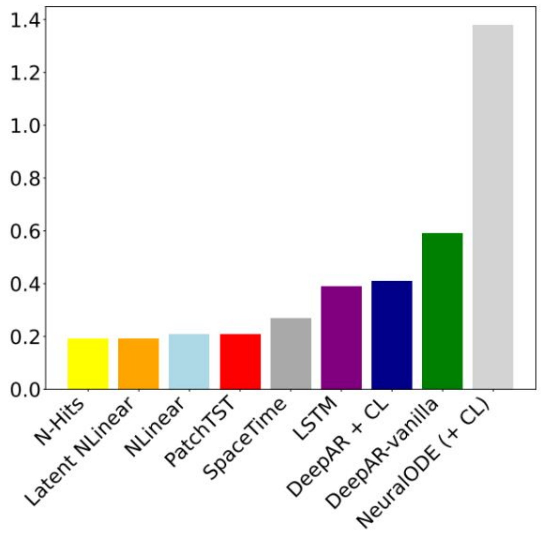
LTSF Landscape analysis

- Baselines
- Statistical methods
- Tree-based methods
- Classical deep learning
- Transformer variants
- State-space models
- N-Beats/N-Hits
- LTSF Linear
- LatentODE

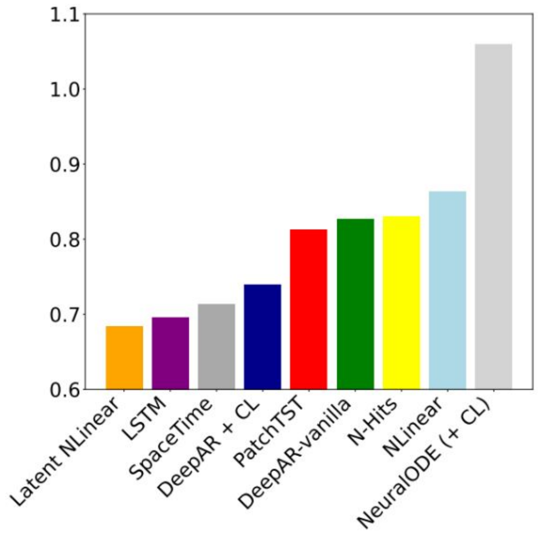
LTSF Landscape analysis



(a) MSE averaged over chaotic and MuJoCo datasets

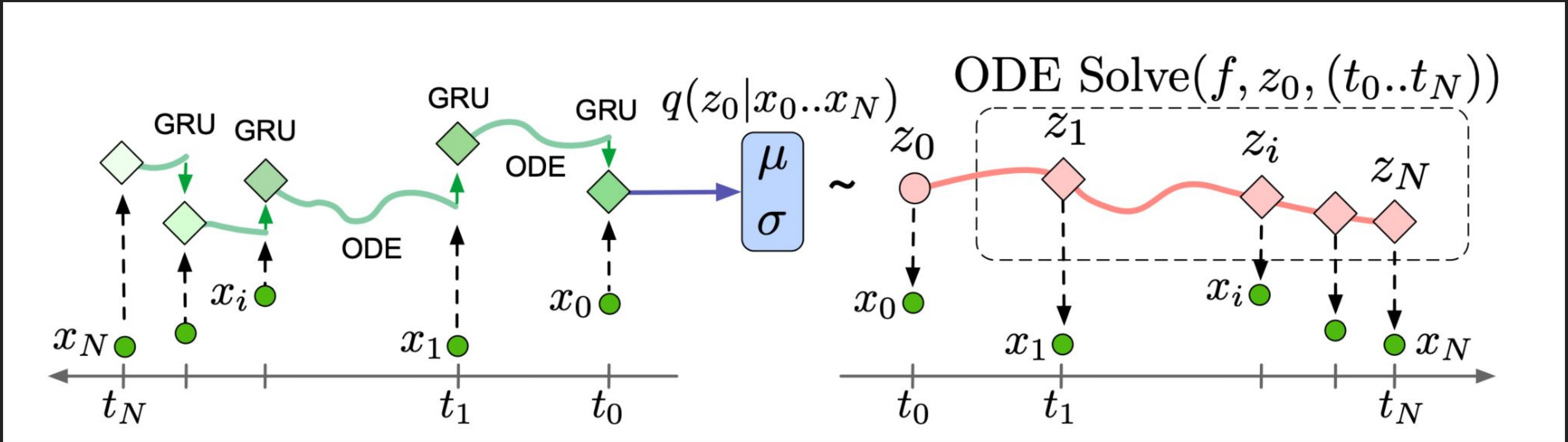


(b) MSE averaged over univariate real-life datasets



(c) MSE averaged over the Weather dataset results

LatentODE



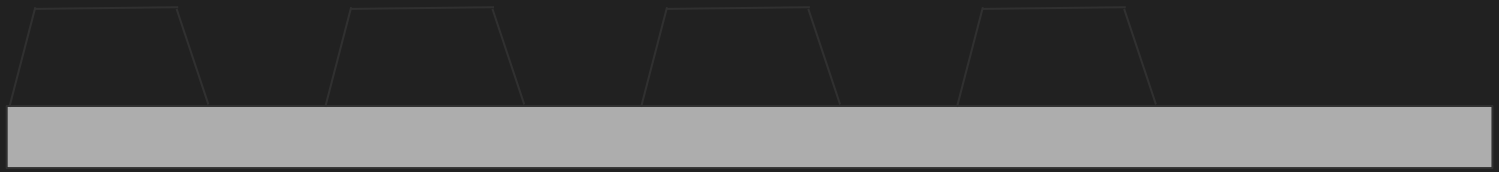
source: https://arxiv.org/abs/1907.03907

Curriculum Learning

- Boosts training convergence speed of models for LTSF
- Applicable to models with variable output length (eg. DeepAR, LatentODE)
- Three distinct phases
  - Short length pretraining
  - Increasing length training
  - Full length training

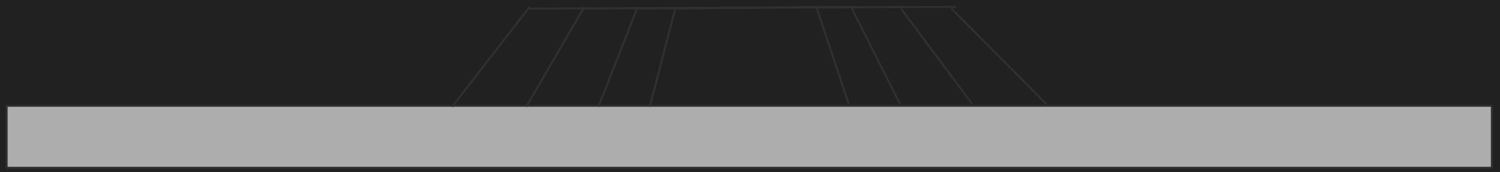
Short length pretraining

- Sampling short length subtrajectories from each trajectory in the dataset
- Exposing the model to various histories, not only the beginning of the trajectory
- Fixed number of epochs
- Model trained to forecast short horizon data usually converges much faster



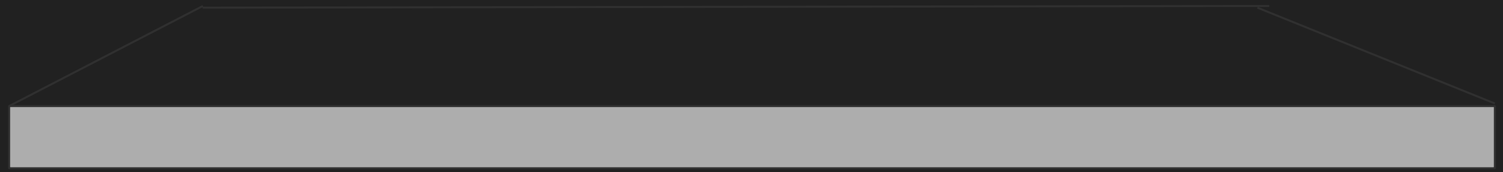
Increasing length training

- Gradually increasing horizon length each epoch
- Similar to the Scheduled Sampling in https://arxiv.org/abs/1506.03099
- Connects first stage to the last stage



Full length training

- Standard way of training
- By the time the training reaches this stage, the model could be already quite far in the convergence
- Model has seen a larger set of series histories, which may lead to better generalization



Ablation on DeepAR

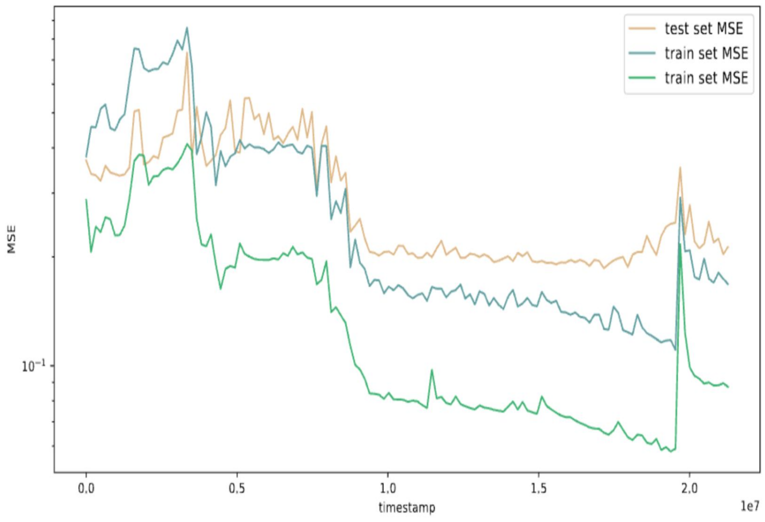


Figure 3.3: Simultaneous plots of training evaluation, test evaluation and the current training loss for DeepAR vanilla, lookback=720

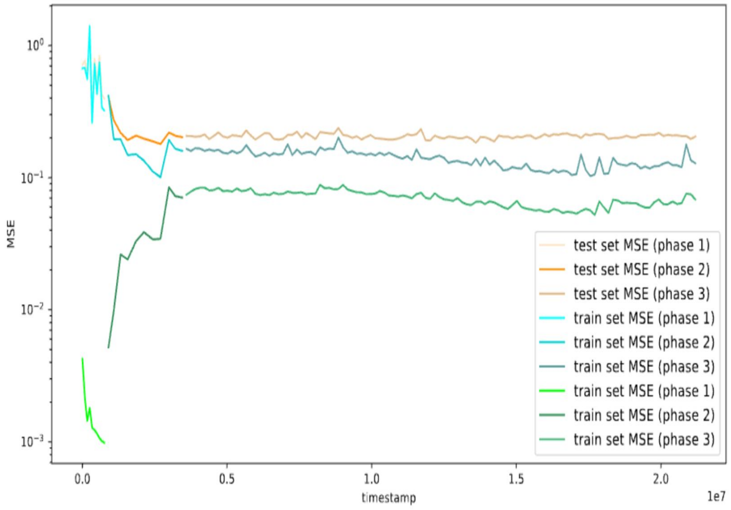


Figure 3.6: Simultaneous plots of training evaluation, test evaluation and the current training loss for DeepAR + CL, lookback=720, plots are divided into 3 curriculum learning phases

Unified Long-Horizon Time Series Benchmark

- 5 categories of time series
  - Real-life, univariate
  - Real-life, multivariate
  - Synthetic, MuJoCo
  - Synthetic, chaotic
  - Synthetic, PDE
- 17 datasets, 100+ GB
- Comparing “SOTA” and classical deep learning models
  - New models tend to be fine-tuned to univariate real life datasets
  - Classical deep learning models perform very well on various categories
  - Introduces Latent NLinear model

Solver-free latent ODE

- Benefits of LatentODE
  - Trajectories can be extrapolated into the future and the past, infinitely
  - Evaluable at arbitrary timestep
- Shortcomings of LatentODE
  - Slow training speed (use of sequential solver)
  - Slow inference speed (not that important in forecasting, though)
- A naively simple solution that retains the benefits and deals with the shortcomings can be constructed

homogeneous linear ODE with constant coefficients

$$\frac{dx}{dt} = Ax$$

$$x(t) = x(t_0)e^{A(t-t_0)}$$

- We have used matrix exponentiation implemented in PyTorch, which is a differentiable operation and has a low memory footprint

Architecture - SFMODE

- SFMODE - Solver-free multi-linear latent ODE
- A nonlinear encoder as in LatentODE (we use LSTM) outputs M states

  $z_1(t_0), \ldots, z_M(t_0)$

- For N timestamps in each state is transformed in a just described manner to

  $z_1(t_1), \ldots, z_M(t_1), \ldots, z_1(t_N), \ldots, z_M(t_N)$

  using M different ODE learnable matrices $A_1, \ldots, A_M$

- using a single nonlinear decoder Dec the final output is of the form

  $(Dec(z_1(t_1)) + \cdots + Dec(z_M(t_1)), \ldots, Dec(z_1(t_N)) + \cdots + Dec(z_M(t_N s)))$

Technical remarks

- ODE matrices may be constrained
  - eg. skew symmetric matrix with diagonal helps in stabilizing the training
- Using many smaller ODE matrices helps to mitigate the cubic time complexity of matrix exponentiation wrt. latent size

Results - chaotic

DATASET	L	SFMODE MSE	SFMODE MAE	LSTM MSE	LSTM MAE	N-HITS MSE	N-HITS MAE	LAT. NLIN. MSE	LAT. NLIN. MAE	DEEPAR CL MSE	DEEPAR CL MAE	SPACETIME MSE	SPACETIME MAE	NODE MSE	NODE MAE	NLINEAR MSE	NLINEAR MAE
L.-V.	96	0.80±0.00	0.61±0.00	***0.80***	***0.61***	0.83	0.63	0.81	0.61	0.81	0.62	0.83	0.63	0.90	0.68	0.89	0.68
	500	0.78±0.00	0.59±0.00	***0.78***	0.59	0.80	0.61	0.79	0.59	0.79	***0.59***	0.81	0.63	0.87	0.66	0.84	0.64
	1000	0.64±0.00	0.50±0.00	***0.63***	***0.49***	0.71	0.55	0.70	0.53	0.64	0.50	0.87	0.67	0.78	0.60	0.87	0.67
M.-G.	96	0.64±0.01	0.57±0.01	0.67	0.59	***0.64***	***0.55***	0.68	0.58	0.80	0.69	0.74	0.64	0.96	0.79	0.82	0.70
	500	0.67±0.00	0.60±0.00	***0.66***	***0.58***	0.74	0.63	0.80	0.67	0.70	0.62	0.81	0.71	0.88	0.76	0.90	0.76
	1000	0.56±0.02	0.52±0.01	***0.49***	***0.46***	0.73	0.64	0.78	0.66	0.96	0.60	0.99	0.82	0.86	0.75	0.92	0.77
LORENZ	96	0.51±0.01	0.48±0.00	0.56	0.51	***0.48***	***0.43***	0.54	0.49	0.61	0.55	0.63	0.57	0.76	0.67	0.69	0.60
	500	0.62±0.01	0.57±0.01	0.60	0.54	***0.58***	***0.52***	0.61	0.53	0.67	0.60	0.76	0.68	0.84	0.74	0.84	0.73
	1000	0.54±0.00	0.51±0.00	***0.47***	***0.43***	0.67	0.59	0.71	0.62	0.83	0.63	0.97	0.82	0.80	0.70	0.88	0.75
AVG.		0.64	0.55	***0.63***	***0.53***	0.69	0.57	0.71	0.59	0.76	0.60	0.82	0.69	0.85	0.71	0.85	0.70

Results - MuJoCo

Dataset	L	SFMODE		deepAR cl		LSTM		SpaceTime		Lat. NLin.		N-Hits		NLinear		deepAR v.	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
cheetah(S)	96	0.80±0.00	0.72±0.00	***0.79***	***0.71***	0.80	0.72	0.80	0.72	0.80	0.72	0.82	0.73	0.81	0.73	0.90	0.79
	250	0.77±0.00	0.70±0.00	***0.76***	***0.69***	0.77	0.70	0.78	0.71	0.78	0.71	0.82	0.73	0.80	0.72	0.95	0.82
	500	0.69±0.00	0.65±0.00	***0.68***	***0.64***	0.68	0.65	0.70	0.66	0.70	0.66	0.77	0.69	0.73	0.68	0.94	0.82
hopper(S)	96	0.72±0.00	0.48±0.00	0.72	***0.48***	***0.72***	0.48	0.73	0.49	0.73	0.48	0.74	0.49	0.75	0.51	0.72	0.48
	250	0.75±0.00	0.49±0.00	0.75	0.48	***0.74***	***0.48***	0.77	0.50	0.77	0.50	0.79	0.52	0.81	0.53	0.75	0.48
	500	0.63±0.00	0.45±0.00	***0.63***	0.44	0.63	***0.44***	0.67	0.47	0.68	0.48	0.69	0.50	0.73	0.52	0.65	0.45
walker(S)	96	0.86±0.00	0.65±0.00	0.86	0.65	***0.86***	***0.64***	0.87	0.65	0.87	0.65	0.88	0.65	0.88	0.66	0.87	0.65
	250	0.85±0.00	0.62±0.00	***0.85***	0.62	0.85	***0.62***	0.87	0.64	0.89	0.65	0.91	0.67	0.94	0.70	0.85	0.62
	500	0.69±0.00	0.51±0.00	***0.68***	0.50	0.69	0.50	0.76	0.57	0.75	0.56	0.80	0.59	0.83	0.63	0.69	***0.50***
avg.		0.75	0.59	***0.75***	***0.58***	0.75	0.58	0.77	0.60	0.77	0.60	0.80	0.62	0.81	0.63	0.81	0.62

Results - PDE

DATASET	L	SFMODE MSE	SFMODE MAE	PATCHT MSE	PATCHT MAE	SPACETIME MSE	SPACETIME MAE	DEEPAR CL MSE	DEEPAR CL MAE	LSTM MSE	LSTM MAE	LAT. NLINEAR MSE	LAT. NLINEAR MAE
K.-S.	96	0.95±0.00	0.80±0.00	1.05	0.85	0.97	0.81	***0.92***	***0.78***	0.96	0.81	0.99	0.82
	250	0.94±0.00	0.80±0.00	1.06	0.85	0.97	0.82	***0.90***	***0.77***	0.97	0.81	1.00	0.83
	500	0.94±0.01	0.80±0.01	1.04	0.84	0.97	0.82	***0.86***	***0.74***	0.94	0.79	0.99	0.82
C.-H.	96	0.82±0.01	0.76±0.01	***0.46***	***0.52***	0.57	0.63	1.01	0.89	0.74	0.71	0.83	0.78
	250	0.92±0.12	0.82±0.07	***0.36***	***0.45***	0.49	0.58	0.59	0.64	0.73	0.71	0.87	0.79
	500	0.91±0.24	0.82±0.15	***0.27***	***0.39***	0.79	0.74	0.50	0.57	0.67	0.66	0.89	0.80
AVG.		0.91	0.80	***0.71***	***0.65***	0.79	0.73	0.80	0.73	0.84	0.75	0.93	0.81

Results - univariate real life

DATA.	L	SFMODE		N-HITS		LAT. NLIN.		NLIN.		PATCHT		SPACET.		LSTM		DEEPAR CL	
		MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE	MAE
ETT	96	0.22±0.00	0.33±0.00	0.21	0.33	0.22	0.33	0.23	0.35	0.22	0.34	***0.21***	***0.33***	0.24	0.35	0.23	0.34
	336	0.18±0.01	0.31±0.01	0.17	***0.30***	0.18	0.30	0.19	0.32	***0.17***	0.31	0.17	0.31	0.21	0.33	0.18	0.32
	720	***0.15±0.01***	0.28±0.01	0.15	0.29	0.15	***0.28***	0.16	0.29	0.15	0.29	0.17	0.31	0.17	0.30	0.18	0.31
M4	96	0.22±0.00	0.24±0.00	0.21	0.22	0.22	0.22	0.25	0.25	0.21	***0.21***	***0.20***	0.22	0.22	0.23	0.23	0.24
	168	0.14±0.00	0.17±0.00	***0.12***	0.16	0.14	0.16	0.14	0.17	0.13	***0.15***	0.12	0.16	0.14	0.16	0.13	0.17
ETT (L)	1000	0.20±0.03	0.34±0.03	0.17	0.30	***0.16***	***0.29***	0.16	0.30	0.16	0.30	1.02	0.92	0.19	0.32	0.19	0.32
AVG.		0.18	0.28	***0.17***	0.27	0.18	***0.27***	0.19	0.28	0.17	0.27	0.32	0.37	0.20	0.28	0.19	0.28

Results - multivariate real life

Data.	L	SFMODE MSE	SFMODE MAE	Lat. NLin. MSE	Lat. NLin. MAE	LSTM MSE	LSTM MAE	SpaceT MSE	SpaceT MAE	DeepAR cl MSE	DeepAR cl MAE	DeepAR v. MSE	DeepAR v. MAE	PatchT MSE	PatchT MAE	N-Hits MSE	N-Hits MAE
P.-B.	144	***0.65±0.02***	0.36±0.01	0.68	0.41	0.67	***0.36***	0.69	0.39	0.71	0.38	0.73	0.38	N/A	N/A	N/A	N/A
Weat.	96	***0.71±0.01***	0.43±0.00	0.71	***0.43***	0.72	0.43	0.73	0.45	0.75	0.45	0.88	0.54	0.91	0.46	0.82	0.47
	250	0.69±0.01	0.42±0.00	***0.69***	0.42	0.69	***0.42***	0.71	0.43	0.75	0.45	0.87	0.54	0.81	0.43	0.85	0.46
	500	***0.65±0.00***	***0.39±0.00***	0.66	0.41	0.67	0.41	0.69	0.43	0.72	0.43	0.73	0.44	0.72	0.40	0.83	0.45
Avg.		***0.67***	***0.40***	0.68	0.42	0.69	0.41	0.71	0.42	0.73	0.43	0.80	0.47	0.81	0.43	0.83	0.46

Further directions

- Explore VAE for generating trajectories
- Use multiple different matrix constraints

Thank you

Questions?