Efficient Parametric SVD of Koopman Operator for Stochastic Dynamical Systems

Minchan Jeong*, **Jongha (Jon) Ryu***, Se-Young Yun, Gregory W. Wornell

KAIST AI Graduate School of AI

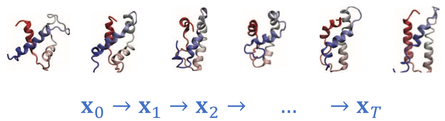
virtual poster

arXiv

NEURAL INFORMATION PROCESSING SYSTEMS 2025

Motivation & Problem Setting

Various scientific systems (e.g., molecular dynamics, climate data) can be viewed as stochastic dynamical systems.



$\mathbf{x}_0 \to \mathbf{x}_1 \to \mathbf{x}_2 \to \quad \dots \quad \to \mathbf{x}_T$

Q. Given noisy trajectories, can we analyze the dynamics?

Goal: extract hidden **slow & dominant modes**, estimate **characteristic timescales** (e.g., folding rates), ...

Challenge 1. Nonlinearity ($\mathbf{x}_t \to \mathbf{x}_{t+1}$ is nonlinear)

Solution. "Koopman analysis": analyze the **conditional expectation operator** $(\mathcal{K}g)(\mathbf{x}) \stackrel{\text{def}}{=} \mathbb{E}_{p(\mathbf{x}'|\mathbf{x})}[g(\mathbf{x}')]$ (a.k.a. **Koopman operator**)

Note: the Koopman operator is now **LINEAR**!
→ we can borrow spectral tools from operator theory

Challenge 2. Dimensionality ($\mathbf{x_t}$ lives in high-dim. space)

Early methods (e.g., DMD, EDMD) rely on pre-defined, fixed measurement functions to capture the operator (cf. PCA, kernel PCA)

(+) closed-form solution / easy analysis; (−) NOT scalable!!!

Solution. Use neural networks to capture the dynamics with flexibility!

But... how should we train them?

Existing Deep-Learning Based Methods

VAMPnet and **DPNet** were proposed to train neural nets to capture top singular subspaces of $\mathcal{K}$ from data.

VAMPnet: $\mathcal{L}_{\text{vamp-}r}(\mathbf{f},\mathbf{g}) \triangleq -\left\| \left(\mathsf{M}_{\rho_0}[\mathbf{f}]\right)^{-\frac{1}{2}} \mathsf{T}[\mathbf{f},\mathbf{g}] \left(\mathsf{M}_{\rho_1}[\mathbf{g}]\right)^{-\frac{1}{2}} \right\|_2^r$

DPNet: $\mathcal{L}_{\text{dp}}^{(\gamma)}(\mathbf{f},\mathbf{g}) \triangleq -\left\| \left(\mathsf{M}_{\rho_0}[\mathbf{f}]\right)^{-\frac{1}{2}} \mathsf{T}[\mathbf{f},\mathbf{g}] \left(\mathsf{M}_{\rho_1}[\mathbf{g}]\right)^{-\frac{1}{2}} \right\|_{\text{F}}^2$

$\mathsf{M}_{\rho_0}[\mathbf{f}] \triangleq \mathbb{E}_{\rho_0(\mathbf{x})}[\mathbf{f}(\mathbf{x})\mathbf{f}(\mathbf{x})^\top]$
$\mathsf{M}_{\rho_1}[\mathbf{g}] \triangleq \mathbb{E}_{\rho_1(\mathbf{x}')}[\mathbf{g}(\mathbf{x}')\mathbf{g}(\mathbf{x}')^\top]$
$\mathsf{T}[\mathbf{f},\mathbf{g}] \triangleq \mathbb{E}_{\rho_0(\mathbf{x})p(\mathbf{x}'\,|\,\mathbf{x})}[\mathbf{f}(\mathbf{x})\mathbf{g}(\mathbf{x}')^\top]$

These **second moment matrices** are core quantities. Can be estimated by minibatch samples.

These objectives suffer from numerical instability and biased gradient (bad for large scale optimization).

Our Proposal

Directly minimize the **low-rank approximation (LoRA) error:**

$$\left\| \mathcal{K} - \sum_{i=1}^k f_i \otimes g_i \right\|_{\text{HS}}^2$$

$$\mathcal{L}_{\text{lora}}(\mathbf{f},\mathbf{g}) \triangleq -2\,\text{tr}(\mathsf{T}[\mathbf{f},\mathbf{g}]) + \text{tr}(\mathsf{M}_{\rho_0}[\mathbf{f}]\mathsf{M}_{\rho_1}[\mathbf{g}])$$

- **No need for regularization** for orthogonality or matrix inverse.
- Allows natural **unbiased gradient estimate**.
- **Can also learn ordered singular functions** with the **nesting technique**. (experiments show empirical advantage over non-nested version)

Summary

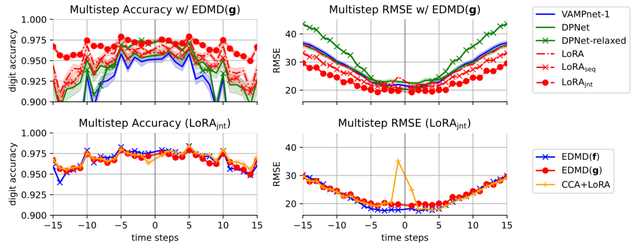
- **Challenge**: Existing deep Koopman methods suffer from numerical instability and biased gradients.
- **Solution**: We propose to use **LoRA-based optimization**, which allows unbiased gradient estimates and requires no regularization
- **Impact**: Scalable and stable training of dominant modes in high-dimensional systems (e.g., Chignolin molecular simulations), accurately recovering eigenfunctions.

Experimental Results

Superior in prediction & eigenfunction recovery

Exp 1. Ordered MNIST 0→1→2→3→4→0

Prediction Accuracy: Evaluated by an oracle classifier
RMSE: L2 distance to class centroids 0 1 2 3 4



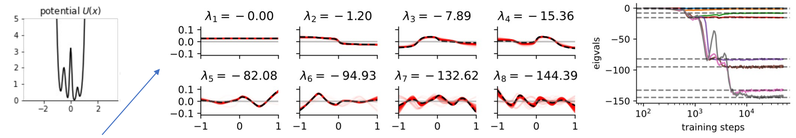
LoRA variants consistently outperforms existing baselines.

Exp 2. 1D Langevin dynamics

A continuous-time dynamics described by a stochastic diff eq. $dX_t = -\frac{U'(x_t)}{\gamma}dt + \sqrt{\frac{2}{\beta\gamma}}dW_t$

Koopman operator = Infinitesimal generator (differential operator) $(\mathcal{L}f)(x) = -\frac{U'(x_t)}{\gamma}f'(x) + \frac{1}{\beta\gamma}f''(x)$

Black dashed: true / **Red**: learned (earlier steps are dimmed)



Dominant eigenfunctions capture transitions between metastable states.

While existing methods fail, LoRA_{seq} accurately recovers ordered eigenfunctions without any regularizations.

Robust & scalable, achieving better orthogonality

Exp 3. Chignolin molecular dynamics

(10 amino-acids, 77 non-H atoms)

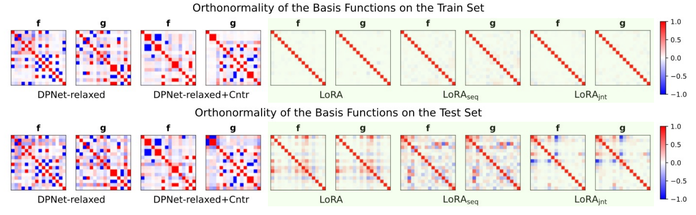
Operator fitting scores on the test dataset

k: # modes, ***H***: feature dimension, ***B***: Batch size

(k, H, B)	LoRA_{seq}	LoRA_{jnt}	LoRA	DPNet-r	DPNet-rC
(64, 128, 384)	31.61	29.89	29.51	-9.76	5.20
(64, 64, 96)	25.83	23.82	25.51	4.75	5.93
(64, 64, 384)	26.66	27.25	24.11	-1.97	4.72
(32, 128, 384)	17.68	15.66	16.26	1.75	3.08
(32, 64, 96)	17.19	15.64	17.36	4.03	6.44
(32, 64, 384)	16.69	17.02	15.79	4.36	4.91
(16, 64, 384)	9.73	8.79	8.80	3.05	4.19
(16, 64, 96)	10.22	8.80	8.52	2.74	4.40

LoRA variants exhibit consistent stability, while baselines fail to converge or result in negative scores.

Orthonormality of basis functions on train/test set



LoRA variants implicitly learns orthonormal basis (diagonal Gram matrices) without any regularization.