Koopman-Assisted Reinforcement Learning

Preston Rozwood [*,1,2] Edward Mehrez [*,1,2] Ludger Paehler [3] Wen Sun [1] Steven L. Brunton [4]

[*]These Authors Contributed Equally [1]Cornell University [2]Subharmonic Technologies Inc. [3]Technical University of Munich

[4]University of Washington

Abstract

- Koopman operator methods **address limitations and interpretability** in the discovery of solutions to the Bellman and Hamilton-Jacobi-Bellman equations
- Introduce a **new class of RL algorithms** by connecting the Koopman operator with Markov Decision Processes (MDP)
- Demonstrate flexibility of the framework with applications to deterministic or stochastic systems, and discrete or continuous-time dynamics, achieving **SOTA results on four controlled dynamical systems**

Koopman Analysis and MDPs

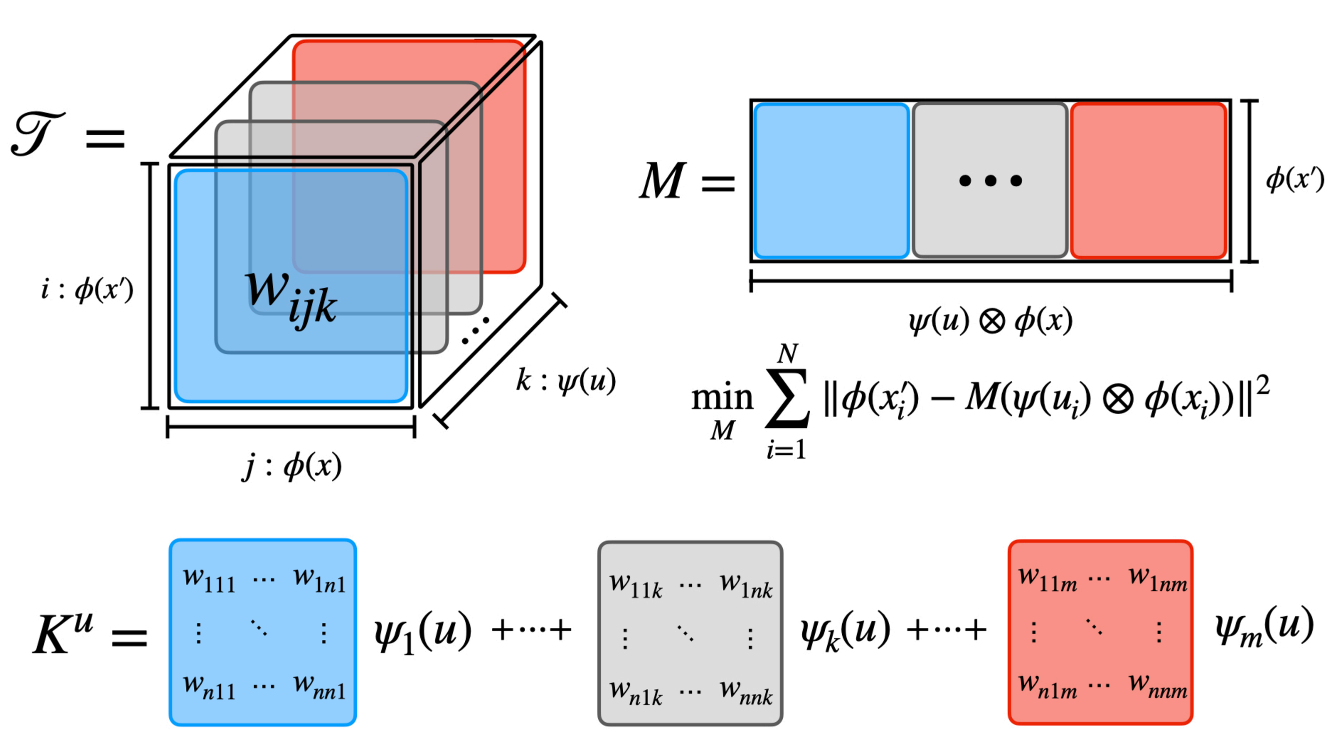
- Presume an infinite-horizon MDP Bellman Equation with a cost function $c = x^\top Qx + u^\top Ru$

$$V^\pi(x) = \max_{\pi_t} \mathbb{E}_{u\sim\pi_t(\cdot|x)} \left[-c(x,u) + \gamma \mathbb{E}_{x'\sim p(\cdot|x,u)}[V(x')]\right]$$

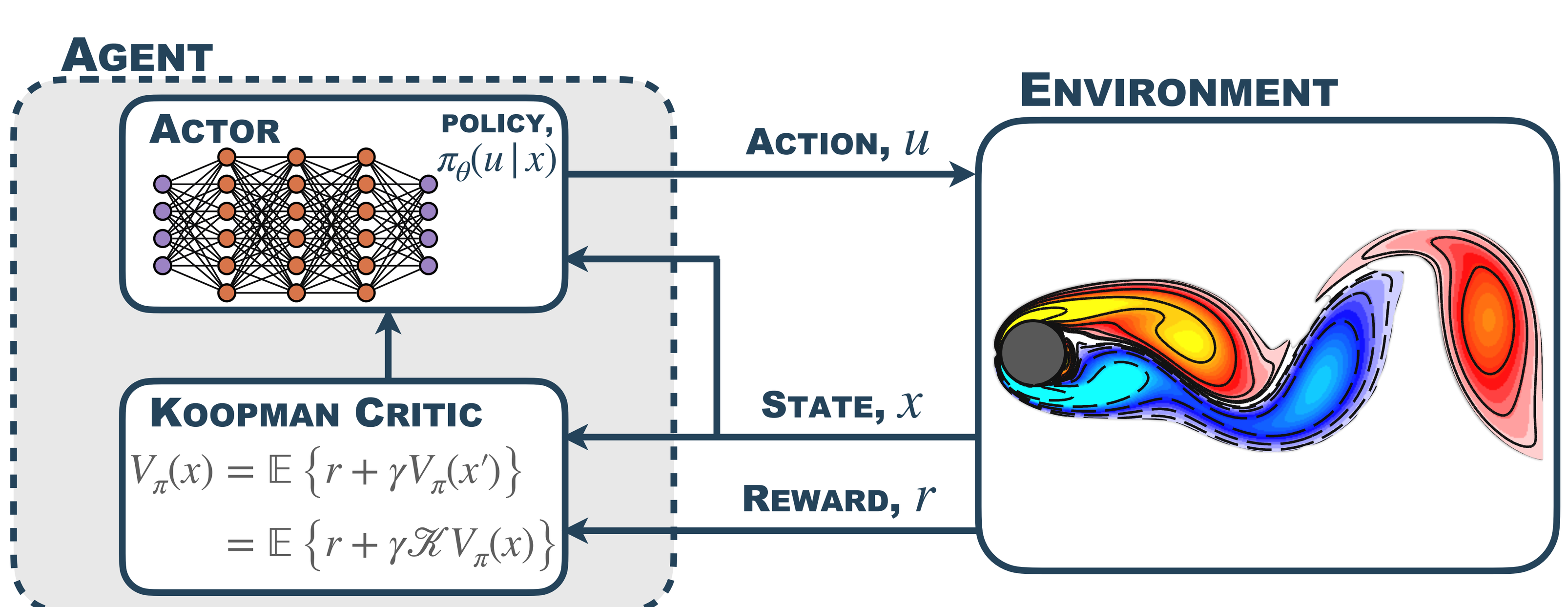
- Re-express MDP in terms of the Koopman operator

$$V^\pi(x) = \mathbb{E}_{u\sim\pi(\cdot|x)} \left\{r(x,u) + \gamma \mathbb{E}_{x'\sim p(\cdot|x,u)}\left[V(x')\right]\right\}$$
$$\implies V^\pi(x) = \mathbb{E}_{u\sim\pi(\cdot|x)} \left\{r(x,u) + \gamma \mathcal{K}^u V(x)\right\}$$

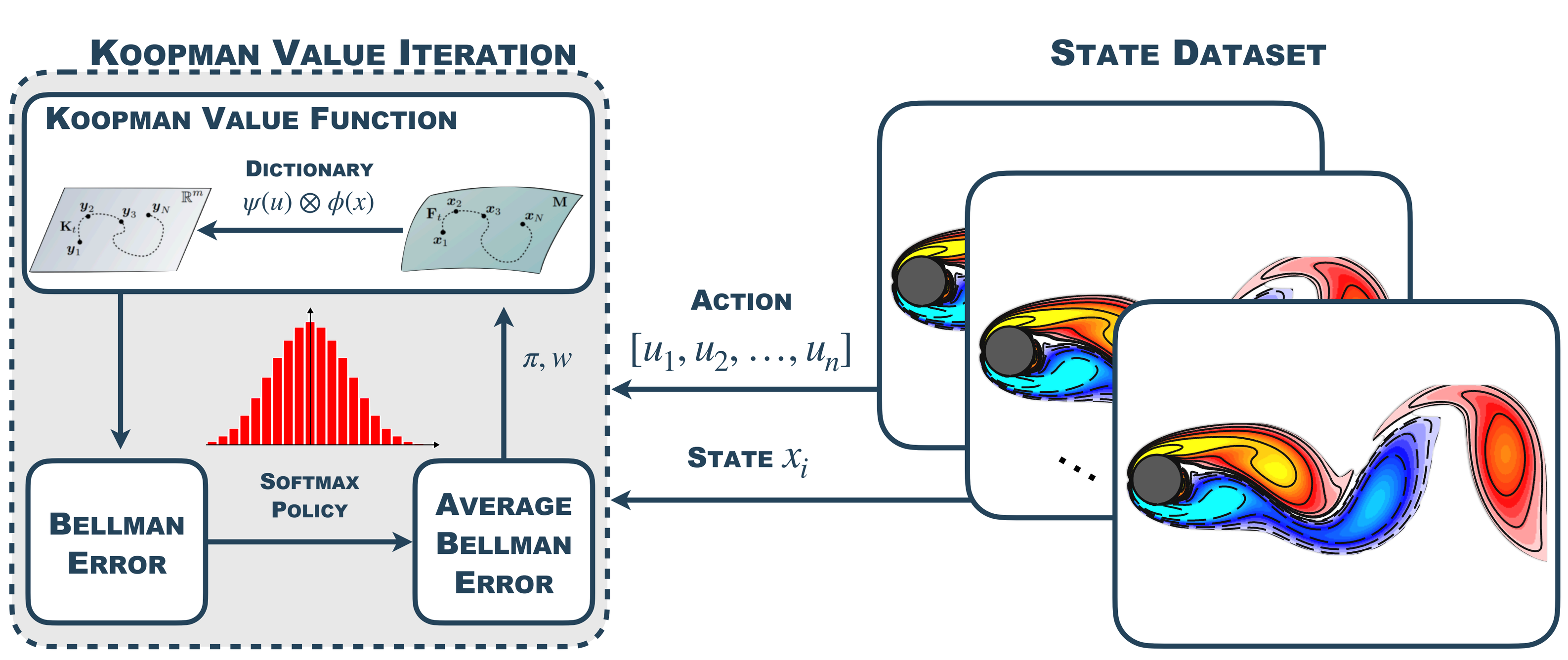
- Assume existence of Koopman invariant (function) subspace F_K on which we can represent the dynamics of the value function using only a finite dictionary
- Multiplicative separation assumption on the dictionary space in state and control

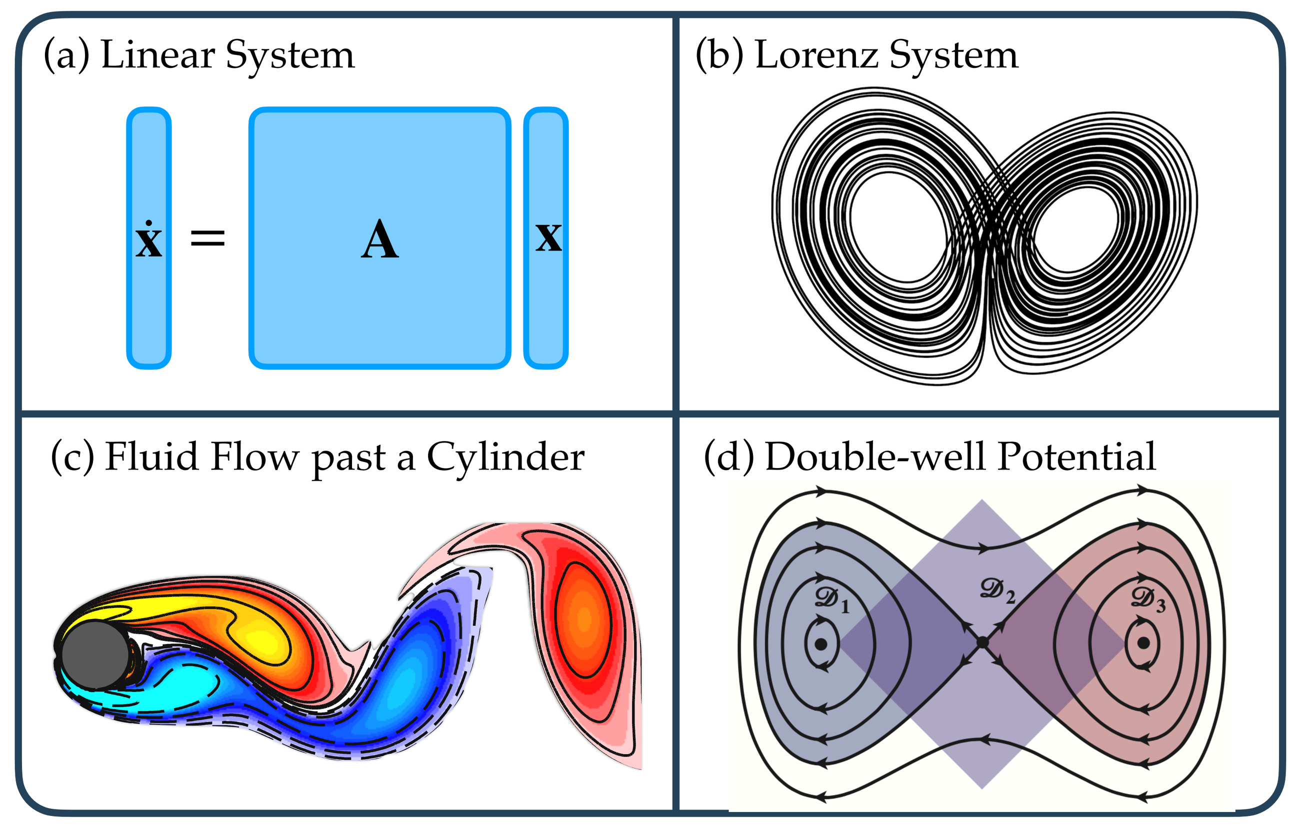


Soft Actor Koopman Critic (SAKC)



Soft Koopman Value Iteration (SKVI)

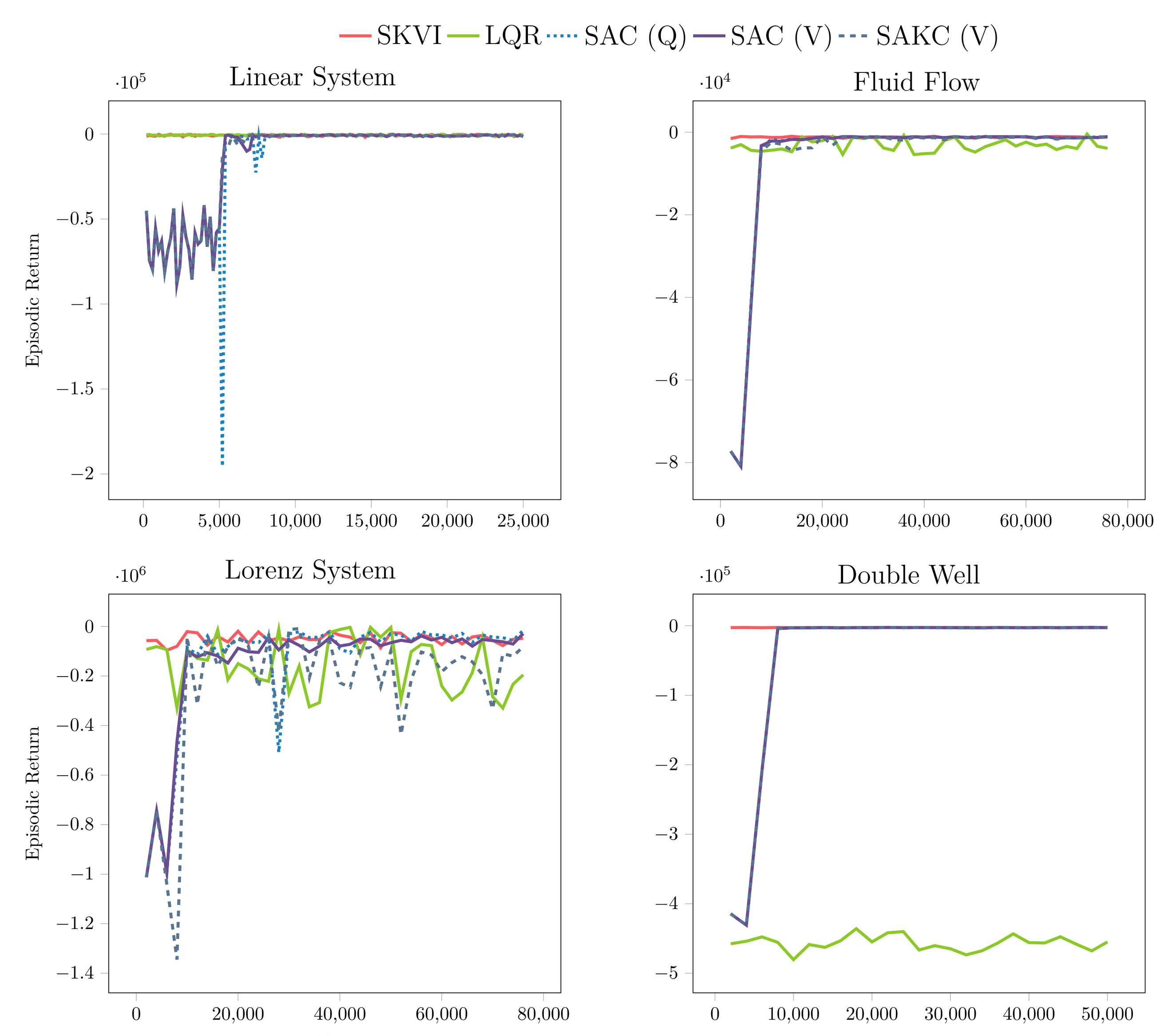




Performance of Koopman RL

- **SOTA on the linear system** after just 5,000 environment steps, outpacing Q-function-based SAC
- **SAKC consistently converges**, showcasing adaptability and closely tracking existing SAC implementations
- Pre-trained SKVI **consistently achieves optimal returns** alongside SAKC and SAC

Episodic Returns



Conclusion

- Reframe MDPs in terms of the Koopman operator, essentially recasting the Bellman equation as a dynamical system
- **Introduce example KARL methods** that are a fusion of Koopman operator methods and existing maximum entropy RL algorithms
- Proposed KARL methods **achieve SOTA** across a diverse set of dynamical system environments

Acknowledgements

We want to thank Stefan Klus for his early contributions to both the code and Koopman theory for continuous-time and stochastic systems

Project Page

Corresponding Author:
Preston Rozwood, and
Edward Mehrez
{pwr36, ejm322}@cornell.edu

Cornell University Subharmonic TUM UNIVERSITY of WASHINGTON