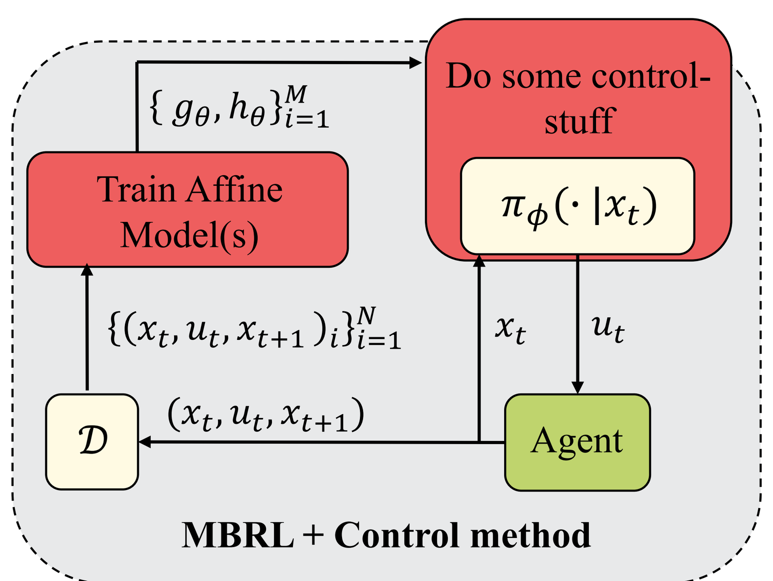
Robust Model Based Reinforcement Learning Using $\mathcal{L}_1$ Adaptive Control

Minjun Sung*, Sambhu H. Karumanchi*, Aditya Gahlawat, Naira Hovakimyan
University of Illinois Urbana-Champaign, Illinois, USA
{mjsung2, shk9, gahlawat, nhovakim} @illinois.edu
(*Equal Contribution)

Presentation **Videos** **Codes**

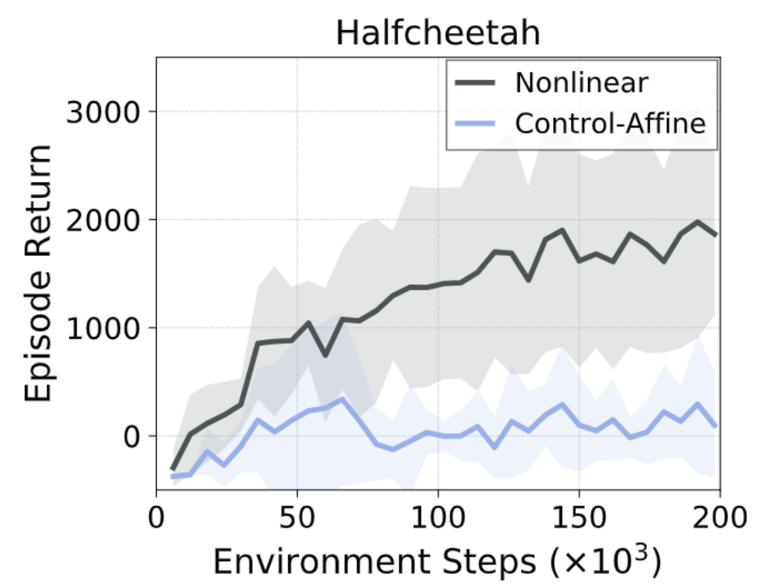
Introduction

- Model Based Reinforcement Learning (MBRL): Trains a predictive model of the system to learn a policy in a *noisy* environment (aleatoric uncertainties)
- Learning a model introduces epistemic uncertainties due to lack of sufficient data or data with insufficient information
- Robust and adaptive control theory is developed to handle uncertainties while tracking a nominal trajectory, [1].
- Consolidating control theoretic methods with MBRL is difficult due to different underlying model structures: *e.g. control-affine vs highly nonlinear (Neural Network)*



(a) Conventional MBRL + Control framework [2,3]

The model is constrained to control-affine class. Robust/Adaptive controller can be employed.

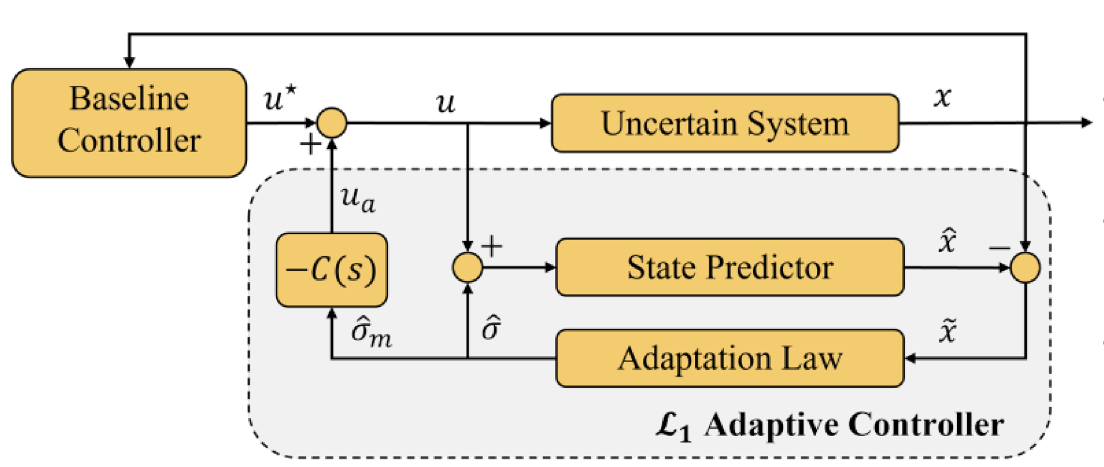


(b) Comparison of performance between NN vs control-affine model

MBRL algorithm with control-affine model class fails to learn the dynamics

- Modifying underlying NN structure to augment control-theoretic methods compromises original performance
- We propose a control-theoretic add-on module for MBRL algorithms: that offer improved robustness **without** compromising original performance

$\mathcal{L}_1$ Adaptive Control [1]



- Framework to counter uncertainties with guaranteed performance
- Uses State predictor, Adaptation law, and Low-pass filter
- Requires a control-affine model structure

Proposed Method: $\mathcal{L}_1$ - MBRL

1. Control-affine model approximation

- First order Taylor series approximation of nonlinear predictive model *around $\bar{u}$*

$$\hat{f}_\theta(x_t,u_t) \approx \hat{f}_\theta(x_t,\bar{u}) + \left(\left[\nabla_u \hat{f}_\theta(x_t,u)\right]_{u=\bar{u}}\right)^\top (u_t - \bar{u})$$
$$= \underbrace{\hat{f}_\theta(x_t,\bar{u}) - \left(\left[\nabla_u \hat{f}_\theta(x_t,u)\right]_{u=\bar{u}}\right)^\top \bar{u}}_{\triangleq\, g_\theta(x_t)} + \underbrace{\left(\left[\nabla_u \hat{f}_\theta(x_t,u)\right]_{u=\bar{u}}\right)^\top}_{\triangleq\, h_\theta(x_t)} u_t \triangleq \hat{f}^a_\theta(x_t,u_t;\bar{u}).$$

Affine in u_t

2. Model switching and *switching law*

- Switch the model (obtain a new control-affine model) when the switching condition holds

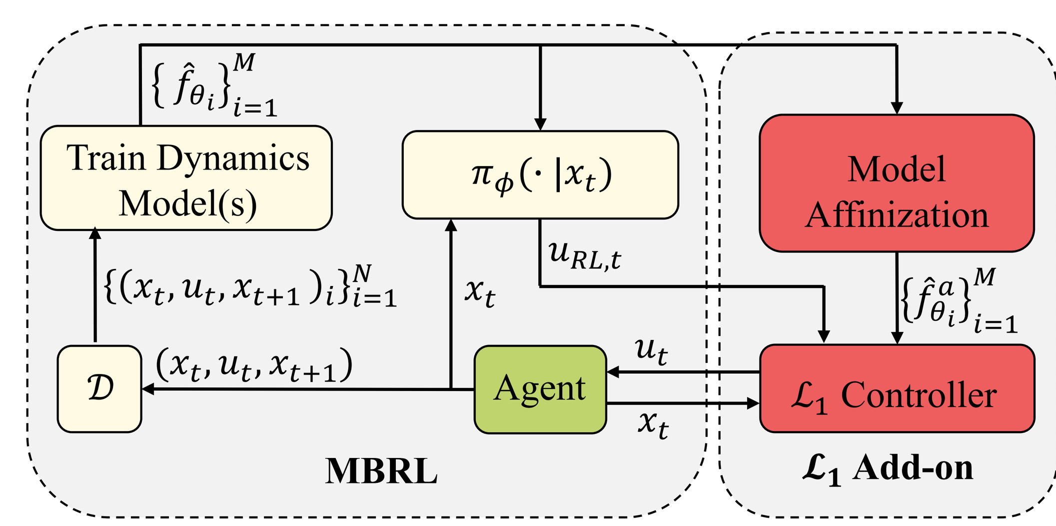
$$\|\hat{f}^a_\theta(x_t,u_t;\bar{u}) - \hat{f}_\theta(x_t,u_t)\| \geq \epsilon_a$$

Controlled parameter

- Otherwise, maintain the current model
- Setting $\epsilon_a = 0$ recovers the baseline MBRL algorithm

3. $\mathcal{L}_1$ Control input augmentation

- Use the current control-affine model to obtain $\mathcal{L}_1$ control input to cancel out uncertainties



(c) $\mathcal{L}_1$ - MBRL framework

The $\mathcal{L}_1$ add-on module adds robustness to the underlying MBRL algorithm without perturbing it. Agnostic property of the method offers wide applicability to various MBRL algorithms

4. Theoretical result

- The state predictor error is upper bounded as

$$\|e(t,x(t),u(t)\| \leq \epsilon_l + \epsilon_a, \quad \forall t \in [0,T_s)$$
$$\|e(t,x(t),u(t))\| = 2\epsilon_a + \mathcal{O}(T_s), \quad \forall t \in [T_s,t_{\max}),$$

Learning error *Affinization error*

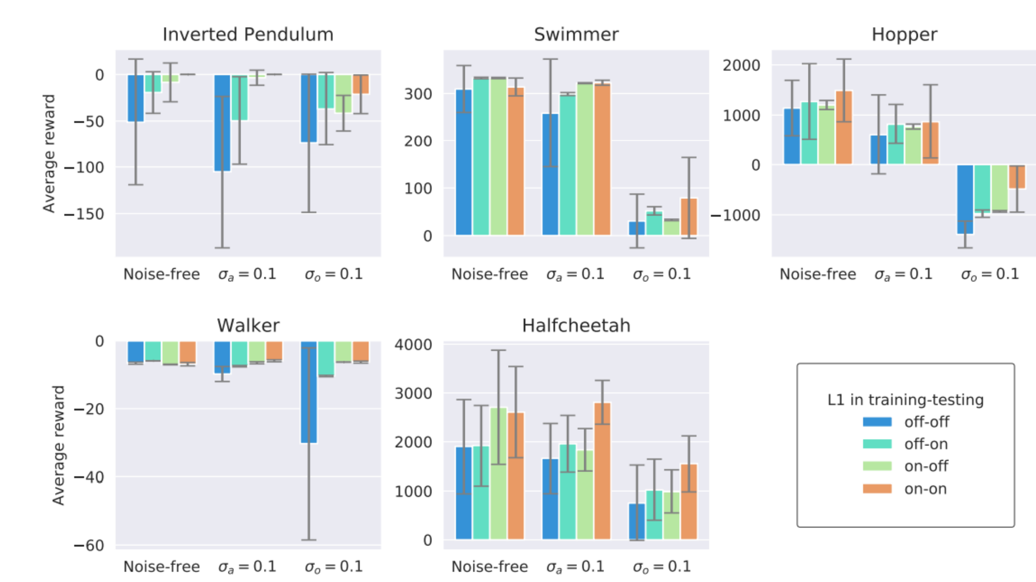
- Affinization error tends to 0 as $T_s \to 0$ for $t \geq T_s$

Experimental Results

- Performance comparison between baseline MBRL (METRPO) and $\mathcal{L}_1$ - MBRL
  - $\mathcal{L}_1$ - MBRL boosts baseline MBRL performances with enhanced robustness to external disturbance

Env.	Noise-free METRPO	Noise-free $\mathcal{L}_1$-METRPO	$\sigma_a = 0.1$ METRPO	$\sigma_a = 0.1$ $\mathcal{L}_1$-METRPO	$\sigma_o = 0.1$ METRPO	$\sigma_o = 0.1$ $\mathcal{L}_1$-METRPO
Inv. P.	-51.3 ± 67.8	$\mathbf{-0.0 \pm 0.0}$	-105.2 ± 81.6	$\mathbf{-0.0 \pm 0.0}$	-74.22 ± 74.5	$\mathbf{-21.3 \pm 20.7}$
Swimmer	309.5 ± 49.3	$\mathbf{313.8 \pm 18.7}$	258.7 ± 113.7	$\mathbf{322.7 \pm 5.3}$	30.7 ± 56.1	$\mathbf{79.2 \pm 85.0}$
Hopper	1140.1 ± 552.4	$\mathbf{1491.4 \pm 623.8}$	609.0 ± 793.5	$\mathbf{868.7 \pm 735.8}$	-1391.2 ± 266.5	$\mathbf{-486.6 \pm 459.9}$
Walker	$\mathbf{-6.6 \pm 0.3}$	-6.9 ± 0.5	-9.8 ± 2.2	$\mathbf{-5.9 \pm 0.3}$	-30.3 ± 28.2	$\mathbf{-6.3 \pm 0.3}$
Halfcheetah	2367.3 ± 1274.5	$\mathbf{2588.6 \pm 955.1}$	1920.3 ± 932.4	$\mathbf{2515.9 \pm 1216.4}$	1419.0 ± 517.2	$\mathbf{1906.3 \pm 972.7}$

- Contribution of $\mathcal{L}_1$ in training vs testing
  - $\mathcal{L}_1$ during training : Collect better data samples
  - $\mathcal{L}_1$ during testing : Uncertainty rejection
  - $\mathcal{L}_1$ ON for both training and testing shows best result



- Addressing Sim2Real gap with $\mathcal{L}_1$ - MBRL
  - Train MBRL in noise-free environment
  - Implement $\mathcal{L}_1$ - MBRL in noisy (real) environment

Env.	$\sigma_a = 0.1$ METRPO	$\sigma_a = 0.1$ $\mathcal{L}_1$-METRPO	$\sigma_o = 0.1$ METRPO	$\sigma_o = 0.1$ $\mathcal{L}_1$-METRPO	$\sigma_a = 0.1$ & $\sigma_o = 0.1$ METRPO	$\sigma_a = 0.1$ & $\sigma_o = 0.1$ $\mathcal{L}_1$-METRPO
Inv. P.	30.2 ± 45.1	$\mathbf{-0.0 \pm 0.0}$	-74.1 ± 53.1	$\mathbf{-3.1 \pm 2.0}$	-107.0 ± 72.4	$\mathbf{-6.1 \pm 4.6}$
Swimmer	250.8 ± 130.2	$\mathbf{330.5 \pm 5.7}$	$\mathbf{337.8 \pm 2.9}$	331.2 ± 8.34	248.2 ± 133.6	$\mathbf{327.3 \pm 6.8}$
Hopper	198.9 ± 617.8	$\mathbf{623.4 \pm 405.6}$	-84.5 ± 1035.8	$\mathbf{157.1 \pm 379.7}$	87.5 ± 510.2	$\mathbf{309.8 \pm 477.8}$
Walker	$\mathbf{-6.0 \pm 0.8}$	-6.3 ± 0.7	-6.4 ± 0.4	$\mathbf{-6.08 \pm 0.6}$	-6.3 ± 0.4	$\mathbf{-5.2 \pm 1.5}$
Halfcheetah	1845.8 ± 600.9	$\mathbf{1965.3 \pm 839.5}$	1265.0 ± 440.8	$\mathbf{1861.6 \pm 605.5}$	1355.0 ± 335.6	$\mathbf{1643.6 \pm 712.5}$

References

1. Hovakimyan, Naira, and Chengyu Cao. $\mathcal{L}_1$ *Adaptive Control Theory: Guaranteed Robustness with Fast Adaptation*. Society for Industrial and Applied Mathematics, 2010.
2. Khojasteh, Mohammad Javad, Vikas Dhiman, Massimo Franceschetti, and Nikolay Atanasov. "Probabilistic Safety Constraints for Learned High Relative Degree System Dynamics." In Learning for Dynamics and Control, pp. 781-792. PMLR, 2020.
3. Taylor, Andrew J., Victor D. Dorobantu, Hoang M. Le, Yisong Yue, and Aaron D. Ames. "Episodic Learning with Control Lyapunov Functions for Uncertain Robotic Systems." In IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 6878-6884. IEEE, 2019.