Robust Adaptive Dynamic Programming for Uncertain Partially Linear Systems (**r1**)

Farnaz Adib Yaghmaie (farnaz.adib.yaghmaie@liu.se), Svante Gunnarsson (svante.gunnarsson@liu.se)

Motivation

Reinforcement Learning (RL) studies Learning approaches for model-free optimal control of dynamical systems. Two basic assumptions in RL are

- The system dimension is fully known.
- All states are measurable.

We aim to relax these two assumptions.

Problem Formulation

We consider uncertain Partially Linear Systems

- Uncertain linear system

$$\dot{x} = Ax + B(u + \Delta(x, z)), \tag{1}$$

- Uncertain dynamical system

$$\begin{aligned} \dot{z} &= f(x, z) \\ \Delta &= \Delta(x, z) \end{aligned} \tag{2}$$

- The value function to be optimized

$$V(x, u) = \int_t^{\infty} x^T Q x + u^T R u \, d\tau \tag{3}$$

Assumptions

Assumption 1: x has a known dimension and it is measurable. z has an unknown dimension and it is not measurable.

Assumption 2: The function $f(x, z)$ is unknown but locally bounded, Lipschitz continuous and $f(\mathbf{0}, \mathbf{0}) = \mathbf{0}$. The output $\Delta(x, z)$ is measurable during learning.

Assumption 3: The uncertain system (2) has *strong unboundedness observability* (SUO) property with zero offset **[r2]**.

Assumption 4: The uncertain system (2) has an upper bound for the $\mathcal{L}_2$-gain γ_1.

Robust On-Policy. Extended from **[r3]**

1: **Initialize:** Select a stabilizing $K^{(0)}$ and set $k = 0$.
2: **repeat**
3: Execute $u^{(k)} = K^{(k)}x$ to collect $x,\ \Delta$ samples.
4: Find $P^{(k)}$ from

$$\begin{aligned} &x(t)^T P^{(k)} x(t) - x(t+\delta t)^T P^{(k)} x(t+\delta t) \\ &= \int_t^{t+\delta t} x^T Q x + u^{(k)T} R u^{(k)} - 2 \int_t^{t+\delta t} x^T P^{(k)} B \Delta \, d\tau. \end{aligned} \tag{4}$$

5: Improve the policy by $u^{(k+1)} = -R^{-1}B^T P^{(k)} x$.
6: **until** Convergence

Robust Off-Policy

1: **Initialize:** Select a stabilizing $K^{(0)}$ and set $k = 0$.
2: **repeat**
3: Execute $u^{(k)} = K^{(k)}x + e$ to collect $x,\ \Delta$ samples.
4: Find $P^{(k)}$ and $K^{(k+1)}$ from

$$\begin{aligned} &x^T(t) P^{(k)} x(t) - x^T(t+\delta t) P^{(k)} x(t+\delta t) \\ &= 2 \int_t^{t+\delta t} x^T K^{(k+1)T} R (u + \Delta - u^{(k)}) d\tau \\ &\quad + \int_t^{t+\delta t} x^T (Q + K^{(k)T} R K^{(k)}) x d\tau. \end{aligned} \tag{5}$$

5: **until** Convergence.

Theorem

Set $\gamma_2 < \gamma_1^{-1}$ and select $Q \geq 0,\ R > 0$ to satisfy

$$R < \eta I, \quad \eta\, \gamma_2^{-2} I < Q, \tag{6}$$

for some $\eta > 0$. Then, the uncertain system (1)-(2) is globally asymptotically stable at the origin using $u^{(k+1)} = K^{(k+1)}x,\ \forall k$ or $u^{(k+1)} = K^{(k+1)}x + e,\ \forall k$ in each iteration of the on-policy or off-policy routines.

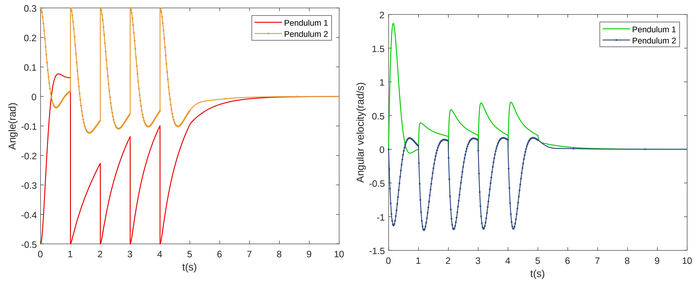
Simulation Result

Two inverted pendulums connected by a spring

$$\begin{aligned} \dot{y}_{i1} &= y_{i2}, \\ \dot{y}_{i2} &= (\frac{m_i g l}{J_i} - \frac{k r^2}{4 J_i}) y_{i1} + \frac{1}{J_i}(u_i + \frac{k l^2}{4} \sin(y_{j1})), \end{aligned} \tag{7}$$

$m_1 = 2\ kg$, $m_2 = 2.5\ kg$, $J_1 = 0.5\ kg.m^2$, $J_2 = 0.625\ kg.m^2$, $k = 100\ N/m$, $l = 0.5\ m$, $r = 1\ m$ and $g = 9.81\ m/s^2$. We consider pendulum one as the linear uncertain system, pendulum two as the dynamical uncertainty and we fix $u_2 = K^{(0)}z = \begin{bmatrix} -10 & -5 \end{bmatrix} z$. The $\mathcal{L}_2$-gain: $\gamma_1 = 7.919$.

Analytical method		Model-free Algorithm	
$P^* =$ [404.210	3.924	$P^{(5)} =$ [404.210	3.924
3.924	8.773]	3.924	8.773]
$K^* =$ [-7.848	-17.546]	$K^{(5+1)} =$ [-7.848	-17.546]



Conclusions

A simple design criterion for stability of uncertain partially linear system during on-policy and off-policy learning.

References

[r1] F. Adib Yaghmaie and S. Gunnarsson "A New Result on Robust Adaptive Dynamic Programming for Uncertain Partially Linear Systems", In 2019 Decision and Control (CDC), IEEE 58th Conference on, 2019, pp. 7480-7485.

[r2] Y. Jiang and Z.-P. Jiang, "Robust adaptive dynamic programming with an application to power systems", IEEE Transactions on Neural Networks and Learning Systems, vol. 24, no. 7, pp. 1150–1156, 2013.

[r3] D. Vrabie and F. Lewis, "Neural network approach to continuous time direct adaptive optimal control for partially unknown nonlinear systems", Neural Networks, vol. 22, no. 3, pp. 237–246, 2009.

LINKÖPING UNIVERSITY
Division of Automatic Control