Singular Perturbation-based Reinforcement Learning of Two-Point Boundary Optimal Control Systems

Vasanth Reddy Baddam[1], Hoda Eldardiry[1], Almuatazbellah Boker[2],

[1]Department of Computer Science, [2]Department of Electrical and Computer Engineering
Virginia Tech

Objectives

We propose a framework for finding the control input for the linear time-varying system. The key contributions of our proposed model are twofold

- We provide a solution for the two-boundary optimal control problem for linear time-varying systems that does not require knowledge of system models
- We introduce a reinforcement learning framework that is based on an insight from the physical dynamics. This framework is simple to implement and results in a suboptimal solution, which converges to the optimal one as the control time interval gets large.

Motivation

Two-Time scale : When the time period in which the cost function which needs to be optimized is large, the time-varying system will exhibit a two-time scale property, where the system dynamics will evolve at faster time scale relative to the time-scale of the control effort

1. Then [2] shows that the system can be reduced into two time-invariant problems. The final original state can thus be approximated by superimposing the solutions of initial and final layer problems.
2. We make further use of the scheme of offline learning [1] to estimate the control gain of two boundary problems. Therefore, by reducing the complexity of the original system into two simple problems, we will learn the controllers of both problems.

System

$$\frac{dx}{dt} = A(t)x(t) + B(t)u(t) \tag{1}$$

The control objective is to design $u(t)$ to minimize the objective function

$$J = \int_0^T x^\top(t)Q(t)x(t) + u^\top(t)R(t)u(t)\,dt \tag{2}$$

$$u(t) = -R(t)B^\top(t)p(t). \tag{3}$$

where, $p(t) \in \mathbb{R}^n$ is the co-state equation. Scaling time as:

$$\tau = \frac{t}{T}, \quad \varepsilon = 1/T \tag{4}$$

and then reconstructing (1), (2), and (3) to obtain

$$\varepsilon \begin{bmatrix} \frac{dx}{d\tau} \\ \frac{dp}{d\tau} \end{bmatrix} = \begin{bmatrix} A(\tau) & 0 \\ -Q(\tau) & -A^T(\tau) \end{bmatrix} \begin{bmatrix} x \\ p \end{bmatrix} + \begin{bmatrix} B(\tau) \\ 0 \end{bmatrix} u, \tag{5}$$

$$u(\tau) = -R(\tau)B^\top(\tau)p(\tau), \tag{6}$$

$$J = T * \int_0^1 x^\top(\tau)Q(\tau)x(t) + u^\top(\tau)R(\tau)u(\tau)\,dt, \tag{7}$$

Reduced System

Initial Regulator Problem

$$\frac{d}{d\gamma}x_a = A(0)x_a + B(0)u_a, \quad x_a(0) = x_0, \tag{8}$$

with the feedback controller u_a in the form

$$u_a(\gamma) \triangleq -K_a x_a(\gamma) = -R(0)B^\top(0)P_a(0)x_a(\gamma), \tag{9}$$

which minimizes the cost function

$$J(x_a, u_a) = \int_0^\infty x_a^\top Q(0)x_a + u_a^\top R(0)u_a d\gamma, \tag{10}$$

Final Regulator Problem

$$\frac{d}{d\beta}x_b = -A(1)x_b - B(1)u_b, \quad x_b(1) = x_T \tag{11}$$

with the feedback controller

$$u_b(\beta) \triangleq -K_b x_b(\beta) = -R(1)B^\top(1)P_b(1)x_b(\beta), \tag{12}$$

which minimizes the cost function

$$J(x_b, u_b) = \int_0^\infty x_b^\top Q(1)x_b + u_b^\top R(1)u_b d\beta. \tag{13}$$

Theorem

Under Assumptions, there exists $\varepsilon_1 > 0$ such that for all $\varepsilon \in (0, \varepsilon_1]$, the solution $x(\tau)$ of (5)-(6) satisfies

$$x(\tau) = x_a^l(\gamma) + x_b^l(\beta) + \mathcal{O}(\varepsilon), \tag{14}$$

$$u(\tau) = \triangleq u_a(\gamma) + u_b(\beta) + \mathcal{O}(\varepsilon) \tag{15}$$

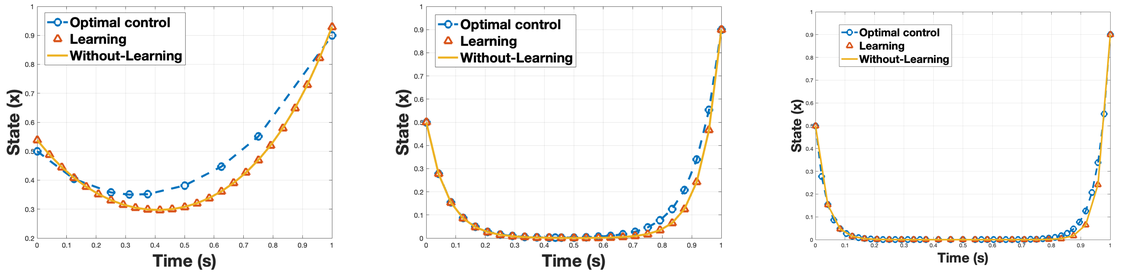

Figure 1:Plots illustrate the state space trajectory for the state $x(t)$ for different values of $\varepsilon = 0.5, 0.1$ and 0.05 respectively.

Simulation

Consider an electric circuit with the components: resistance (R) and inductor ($L(t)$) in series connection. The state space equation for the circuit is given as:

$$\frac{dx}{dt} = -\frac{R}{L(t)}x + \frac{1}{L(t)}u, \tag{16}$$

where $x(t)$ is the state (circuit current) and $u(t)$ is the control input (circuit voltage). In this case, $A(t) = -\frac{R}{L(t)}$ and $B(t) = \frac{1}{L(t)}$. We assume that the time-varying dissipating inductor L and resistor values are unknown.

Conclusion

We proposed optimal controller design using reinforcement learning for time varying systems with two boundary conditions. The proposed design leverages the fast time scale occurring at the boundary conditions to reduce the time-varying problem into two simple time-invariant problems. Furthermore, we design a learning-based control strategy that does not need knowledge of the system model. We show that the accuracy of the controller performance improves as the problem time horizon increases. We presented simulation results to support our claims using an RL circuit. In the future, we plan to extend our work in learning the controllers of nonlinear systems.

References

[1] Yu Jiang and Zhong-Ping Jiang.
Robust adaptive dynamic programming.
John Wiley & Sons, 2017.

[2] R Wilde and P Kokotovic.
A dichotomy in linear control theory.
IEEE Transactions on Automatic control, 17(3):382–383, 1972.