ASEN 5022 Project Proposal: Reinforcement Learning for System Identification

Alex Hirst and John Jackson
{cahi7388, joja7894}@colorado.edu

5 April 2019

1 Introduction

System identification is the process of identifying a dynamical model from input and output data. This problem can be difficult for an arbitrary system, with and without a prior model. We propose using model based and non-model based reinforcement learning (RL) to conduct system identification on an unknown system while minimizing the error between predicted and actual responses to a forcing function.

2 Problem Setup

Our unknown system will be the simplified wing model from homeworks 3 and 4, with unknown coefficients. The equation of motion is $\mathbf{M}\ddot{q} + \mathbf{C}\dot{q} + \mathbf{K}q = f$ with $\dot{q}(0) = \dot{q}_0$ and $q(0) = q_0$.

mass of wing	$m = 1$
moment of inertia	$I_G = 10$
linear vertical spring	$k_u = 2$
torsional spring	$k_\theta = 10$
excitation frequency	$\bar{\omega} = 2.14$
eccentricity	$e = 0.5$

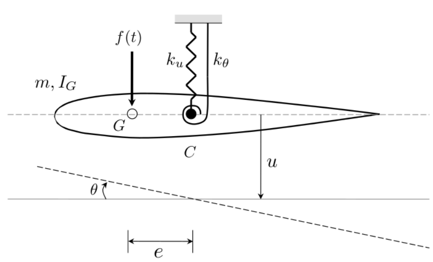

Figure 1: Simplified 2DOF aircraft wing system.

3 Problem Approach

Figure 2 shows the structure of a Markov Decision Process with unknown parameters θ_R and θ_X that specify the reward function and transition function, respectively. In our problem, we will consider finding the parameters $\theta_X = \{\mathbf{M}, \mathbf{C}, \mathbf{K}\}$ in our model-based reinforcement learning strategy. For our model-free RL strategy, we will utilize a neural network structure to predict system output.

The goal of using reinforcement learning is to determine a finite series of excitation inputs $\mathbf{f} = \{f_1, f_2, \dots\} \in A$ that 1) learns the model parameters θ_X of the MDP and 2) maximizes the reward gained by executing the sequence $\mathbf{f}$. In order to work in a discretized framework, each excitation action f_1 is treated as a single action but applied over a fixed duration of time. Given f_k, the MDP will generate a prediction of the generalized coordinate trace $\hat{h}(f_k) = \{\hat{q}_1, \hat{q}_2, \dots\}_k$. Next, the real forced output, $h(f_k) = \{q_1, q_2, \dots\}$ we used to calculate error of the predicted traces, $X_k = e(\hat{h}(f_k), h(f_k))$ where e is a function whose output is related to the statistics of $(h(\hat{f_k}) - h(f_k))$. X_k will be used to update θ_X using least-squares parameter fitting. The value of X_k and f_k will be used to calculate a reward $R(X_k, f_k)$ that will be used to guide future actions the MDP will take.

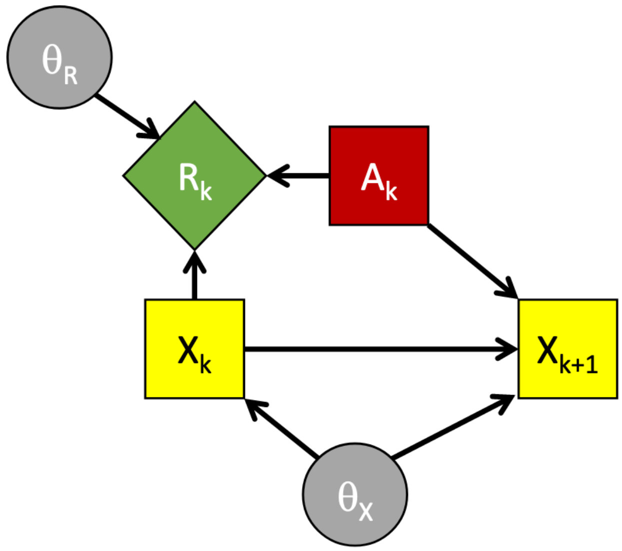

Figure 2: Overview of reinforcement learning problem on an MDP with unknown parameters θ_R and θ_X[1].

References

[1] Nisar Ahmed. ASEN 6519 Lecture 19: Introduction to Reinforcement Learning, 2019.