AE504: Homework Assignment 2

Alex Faustino

March 29, 2019

Problem 1. Consider the state space $S = \{s^1, s^2, s^3, s^4\}$ and action set $A = \{a^1, a^2\}$. Notation $(s, a) \to (s'; R)$ will denote that applying action a at state s results in the system transitioning to state s' at the next time step, while collecting the reward R. The transitions and rewards are as follows:

$$\begin{aligned}
&(s^1, a^1) \to (s^4; 4), \quad (s^1, a^2) \to (s^2; 6),\\
&(s^2, a^1) \to (s^2; 3), \quad (s^2, a^2) \to (s^3; 7),\\
&(s^3, a^1) \to (s^1; 2), \quad (s^3, a^2) \to (s^1; 5),\\
&(s^4, a^1) \to (s^3; 1), \quad (s^4, a^2) \to (s^1; 4).
\end{aligned}$$

The initial state of the system is $s_0 = s^1$. Determine the optimal control policy (i.e., optimal path) that maximizes $\sum_{t=0}^{6} R(s_t, a_t)$, if the system is required to satisfy $s_1 = s_3 = s_6 = s^2$.

Solution: Looking at the graph of the system in Figure 1 it is clear that the constraint on initial condition, $s_0 = s^1$, and the transient condition, $s_1 = s^2$, gives $a_0 = a^2$. Additionally, if $s_3 = s^2$ then $a_1 = a_2 = a^1$ since a choice of $a_1 = a^2$ would result in a s_2 where achieving $s_3 = s^2$ is impossible. Lastly, because $s_6 = s^2$ we know that $a_6 = a_2$ since this is the maximum reward that can be attained from s^2. Therefore, we can simplify the original problem to:

$$\max_a \sum_{t=3}^{5} R(s_t, a_t)$$

where $s_3 = s_6 = s^2$. Again from Figure 1 we can see there are three possible sequences of (s_t, a_t) that satisfy our constraints, $\{(s^2, a^1), (s^2, a^1), (s^2, a^1)\}$, $\{(s^2, a^2), (s^3, a^1), (s^1, a^2)\}$, and $\{(s^2, a^2), (s^3, a^2), (s^1, a^2)\}$. The total reward from each sequence is 9, 15, and 18 respectively. Therefore, the optimal control policy is $A^* = \{a^2, a^1, a^1, a^2, a^2, a^2, a^2\}$.

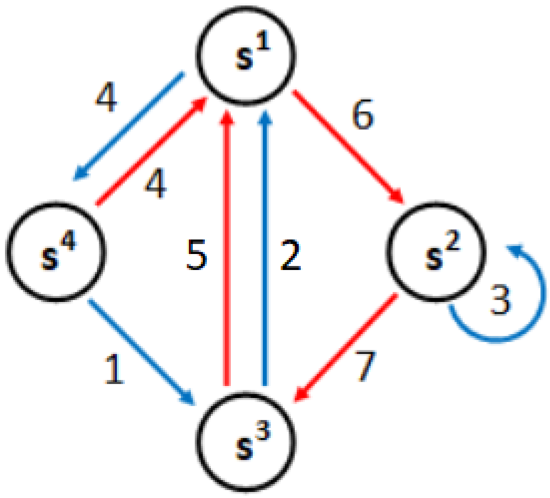

Figure 1: Graphical depcition of the state and action space described in Problem 1. Blue lines represent a^1 and red represent a^2 with their associated rewards, R.

Problem 2. Let S be a finite state space, A a finite action set, and $f : S \times A \to S$ a transition function for a discrete-time control system with initial state $s_0 \in S$. Let $\overline{R} \geq 0$, and let $R : S \times A \to [0, \overline{R}]$ describe the rewards attained for each transition (note — every reward is nonnegative, but no greater than $\overline{R}$).

It is clear that $\max_{T \geq 0} V^T(s_0)$ does not need to exist: the longer the run, the system may collect more and more rewards. However, show that

$$\max_{T \geq 0} \left(V^T(s_0) - T^2 \right)$$

does exist (and does not equal $+\infty$).

Solution: By listing the possible values of $V^T(s_0)$ for different T: $V^0(s_0) = [0, \overline{R}]$, $V^1(s_0) = [0, \overline{2R}]$, $V^2(s_0) = [0, \overline{3R}]$, etc.. It's clear that the maximum value $V^T(s_0)$ can take on for any T is $(T+1)\overline{R}$. We can then rewrite

$$\max_{T \geq 0} \left(V^T(s_0) - T^2 \right) \to \max_{T \geq 0} \left(-T^2 + \overline{R}T + \overline{R} \right)$$

which is a simple quadratic function of T. We can then use the first and second derivative test to find and confirm the global maximum.

$$\frac{d}{dT} \left(-T^2 + \overline{R}T + \overline{R} \right) = -2T + \overline{R}$$

$$\frac{d^2}{dT^2}\left(-T^2+\overline{R}T+\overline{R}\right)=-2$$

It's clear then that there exists a global maximizer at $T=\frac{\overline{R}}{2}$ and the global maximum is $\frac{1}{4}\overline{R}+\overline{R}$.

Problem 3. One model of the Boeing 747 longitudinal dynamics is given by

$$\begin{pmatrix} u_{t+1} \\ w_{t+1} \\ q_{t+1} \\ \theta_{t+1} \end{pmatrix} = \begin{pmatrix} 0.994 & 0.026 & 0 & -32.2 \\ -0.094 & 0.376 & 820 & 0 \\ 0 & -0.002 & 0.332 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} u_t \\ w_t \\ q_t \\ \theta_t \end{pmatrix} + \begin{pmatrix} 0 \\ -32.7 \\ -2.08 \\ 0 \end{pmatrix} \delta_t,$$

where u is the difference from the aircraft's nominal horizontal velocity, w its vertical velocity, q its pitch rate, θ its pitch angle, and δ the deflection from the neutral elevator position (in units of rad, ft, and sec, as appropriate).

The initial state of the aircraft at time $t=0$ is $(0, 100, 0, 0.4)$.

(a) Find the control inputs $(\delta_0, \ldots, \delta_9)$ that minimize $\sum_{t=0}^{10} w_t^2$. Note that you need to find the actual control inputs, which are real numbers, not just their relationship to the system state.

(b) Discuss why the inputs in (a) are not realistic to apply on an aircraft.

(c) Propose how the problem in (a) could be modified in such a way that the solution would generate more realistic inputs, while still aiming to minimize the aircraft's vertical speed.

Solution: (a) We can find the control inputs using the standard discrete LQR finite time horizon formula where

$$\delta_t = F_t x_t$$

$$F_t = -\left(R + B^T P_{t+1} B\right)^{-1} B^T P_{t+1} A \tag{1}$$

$$P_t = A^T P_{t+1} A - A^T P_{t+1} B \left(R + B^T P_{t+1} B\right)^{-1} B^T P_{t+1} A + Q \tag{2}$$

and $P_N = Q_N = Q$. To obtain the given cost function, $\sum_{t=0}^{10} w_t^2$, from the standard, $\sum_{t=0}^{10} x_t^T Q x_t + \delta_t^T R \delta_t$ we choose

$$R = 0 \quad \text{and} \quad Q = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

When we solve the first recursive step we immediately see that the first two terms of (2) cancel and $P = Q$ for all t. This means that F is also constant for all t so we can easily find all of our δ_t by marching our state space system forward through time. $\delta_t = \{1.15, -64.96, 3367, -1.745\text{e}{+}05, -1.745\text{e}{+}05, -4.687\text{e}{+}08, 2.429\text{e}{+}10, -1.259\text{e}{+}12, 6.525\text{e}{+}13, -3.382\text{e}{+}15\}$

(b) These inputs are unrealistic because their magnitudes are ridiculously beyond feasible bounds for elevator deflection.

(c) More realistic inputs can be achieved by increasing the value of R so that the cost function considers the magnitude of commanded inputs as well, $\sum_{t=0}^{10} w_t^2 + u_t^2$.

Problem 4. Consider the following discrete-time control system (traditionally known as *double integrator*):

$$\begin{pmatrix} x_{t+1,1} \\ x_{t+1,2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_{t,1} \\ x_{t,2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} u_t.$$

The initial state of the system is $(0, 1)$ at time $t = 0$. Find any control input (u_0, u_1) which minimizes $x_{2,2}^2$.

(Do not be discouraged if the usual formulae for the LQR method that we used in class don't work, as all matrices Q, Q_N, R are "heavily nonregular". This is a feature, not a bug.)

Solution: We begin by attempting to solve this identically to Problem 3 except here

$$Q = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

This approach immediately leads to an issue when attempting to solve the first recursive step of (1) since $R + B^T P_2 B$ evaluates to 0 and 0^{-1} is undefined. However, we can take the Moore-Penrose inverse where $0^+ = 0$ then find that $F_1 = 0$ and subsequently that $u_1 = 0$. We can then carry on solving LQR in the standard way and we find that the control input $u_t = \{-1, 0\}$ successfully minimizes $x_{2,2}^2$.