MTHE 472/MATH 872 **Winter 2022**

Computational Assignment

Problem 1: Value Iteration/Policy Iteration/Q-Learning and Linear Programming [70 Points]

Consider the following problem: Let $\mathbb{X} = \{B, G\}, \mathbb{U} = \{0, 1\}$, where $\mathbb{X}$ denotes whether a fading channel is in a good state ($x = G$) or a bad state ($x = B$). There exists an encoder who can either try to use the channel ($u = 1$) or not use the channel ($u = 0$). The goal of the encoder is send information across the channel.

Suppose that the encoder's per-stage cost (to be minimized) is given by:

$$c(x, u) = -1_{\{x=G, u=1\}} + \eta u.$$

for some $\eta \in \mathbb{R}$ to be specified below.

If you view this as a maximization problem, you can see that the goal is to maximize information transmission efficiency subject to a cost involving an attempt to use the channel; the model can be made more complicated but the idea is that when the channel state is good, $u = 1$ can represent a channel input which contains data to be transmitted and $u = 0$ denotes that the channel is not used. When $u = 1$ and $x = G$, the channel is utilized successfully.

For many channels with memory, the input also impacts the channel state. Suppose that the transition kernel is given by:

$$P(x_{t+1} = G|x_t = G, u_t = 1) = 0.3, \qquad P(x_{t+1} = B|x_t = G, u_t = 1) = 0.7$$

$$P(x_{t+1} = G|x_t = G, u_t = 0) = 0.7, \qquad P(x_{t+1} = B|x_t = G, u_t = 0) = 0.3$$

$$P(x_{t+1} = G|x_t = B, u_t = 1) = 0.5, \qquad P(x_{t+1} = B|x_t = B, u_t = 1) = 0.5$$

$$P(x_{t+1} = G|x_t = B, u_t = 0) = 0.9, \qquad P(x_{t+1} = B|x_t = B, u_t = 0) = 0.1$$

We will consider either a discounted cost criterion for some $\beta \in (0, 1)$ (you can fix an arbitrary value)

$$\inf_{\gamma} E_x^{\gamma}[\sum_{t=0}^{\infty} \beta^t c(x_t, u_t)] \tag{1}$$

or the average cost criterion

$$\inf_{\gamma} \limsup_{T\to\infty} \frac{1}{T} E_x^{\gamma}[\sum_{t=0}^{T-1} c(x_t, u_t)]. \tag{2}$$

a) Using Matlab or some other program, obtain a solution to the problem given above in (1) through the following:

(i) [15 Points] Value Iteration. Take some fixed $\beta \in (0,1)$ of your choice. Consider $\eta = \frac{2}{3}$, $\eta = 0.95$ and $\eta = 0.05$. Interpret the optimal solution for these different values of η, in view of the application.

(ii) [15 Points] Policy Iteration. With the same β as above, work again with each of the following: $\eta = \frac{2}{3}$, $\eta = 0.95$ and $\eta = 0.05$.

(iii) [20 Points] Q-Learning. Try only $\eta = \frac{2}{3}$. Note that a common way to pick α_t coefficients in the Q-learning algorithm is to take for every x, u pair:

$$\alpha_t(x,u) = \frac{1}{1 + \sum_{k=0}^{t} 1_{\{x_k = x, u_k = u\}}}$$

Compare your solutions (obtained via different methods).

b) [20 Points] Consider the criterion given in (2). Apply the convex analytic method, by solving the corresponding linear program, to find the optimal policy? In Matlab, the command *linprog* can be used to solve linear programming problems. See (7.32) in the lecture notes.

c) [Optional] Study quantized approximation methods in Section 8.3.1 and 8.3.2, for both state space quantization and action space quantization to arrive at a finite model MDP with rigorous convergence guarantees. Furthermore, study quantized Q-learning in Section 8.5. Now, revise the problem above with the following transition kernel so that $\mathbb{X} = [0,1]$ (thus the channel's quality is not binary) and for each Borel $A \in [0,1]$

$$P(x_{t+1} \in A | x_t = z_0, u_t = 1) = 2\int_A (1-x)dx, \qquad P(x_{t+1} \in A | x_t = z_0, u_t = 0) = 2\int_A x dx$$

and suppose that the encoder's per-stage cost (to be minimized) is given by:

$$c(x,u) = -xu + \eta u.$$

for some $\eta \in \mathbb{R}$. Apply quantized Q-learning by quantizing the channel state with uniform quantization of increasing granularity.

You may study the learning results for partially observed models as well (Section 8.4.3).

Problem 2: The Kalman Filter [30 Points]

Let a linear system driven by Gaussian noise be given by the following:

$$\begin{aligned} x_{t+1} &= Ax_t + w_t \\ y_t &= Cx_t + v_t \end{aligned} \tag{3}$$

Here $A = \begin{bmatrix} 0.75 & 1 & 0 \\ 0 & 0.75 & 1 \\ 0 & 0 & 0.75 \end{bmatrix}$, $C = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix}$, and the i.i.d. noise processes satisfy $w_t \sim \mathcal{N}(0, I)$, $v_t \sim \mathcal{N}(0,1)$. The initial state is also zero-mean Gaussian and suppose that $\Sigma_{0|-1} = I$ (here, we use the notation in Chapter 6 of the Lecture notes).

a) Write down the Kalman Filter update equations (including the associated Riccati recursions for the covariance matrix updates).

b) By simulating the above system in Matlab (or any other program), run the Kalman Filter from $t = 0$ until $T = 1000$. Plot $x_t, \tilde{m}_t$ and $x_t - \tilde{m}_t$.

c) Do the Riccati recursions converge to a limit; is the limit unique? Explain why or why not. By Matlab (or another program) verify your answer. For uniqueness, you can take various initial conditions and study whether the recursions converge to the same limit.

Problem 3: Q-Learning and its Convergence [Optional]

Study the proof of Theorem 8.2.1.