MATH4406/MATH7406 (Control Theory), HW Assignment #1, MDP and Physical Models.

Prepared by Yoni Nazarathy, Last Updated: Aug 7, 2016.

This homework assignment is about Markov Decision Processes and Physical Models (Units 2 and 3 of the course).

- For some of you the programming aspect may be a bit of a challenge. Make sure to allow enough time for this. Break up each programming task into well-defined sub-tasks.
- Please make sure to present your results in a clear and organised manner. Numerical output results should always be well explained and documented. Labels on graphs, diagrams, tables etc...
- Hand-in all code (preferably as an appendix).
- Follow any other general instructions for assignments as described in the subject description.

Problem 1: Cloud Seeding MDP

Consider a simplistic model of weather, operating in discrete time and exhibiting state space:

$$\mathcal{S} = \{1 = \text{'rain'},\ 2 = \text{'clouds'},\ 3 = \text{'sun'}\}.$$

At any given time, the government has the ability to seed (= 'b') or not seed (= 'a'). Hence the action space is,

$$\mathcal{A} = \{a,\ b\}.$$

When no seeding is taking place, weather evolves according to the probability transition matrix,

$$P(a) = \begin{bmatrix} 0.1 & 0.5 & 0.4 \\ 0.1 & 0.3 & 0.6 \\ 0.01 & 0.59 & 0.4 \end{bmatrix}.$$

When seeding is taking place, weather evolves according to,

$$P(b) = \begin{bmatrix} 0.15 & 0.45 & 0.4 \\ 0.3 & 0.2 & 0.5 \\ 0.1 & 0.6 & 0.3 \end{bmatrix}.$$

A reward function at any given time is obtained based on the state $x \in \mathcal{S}$ and the control $u \in \mathcal{A}$, as follows:

$$r(x,u) = \gamma \mathbf{1}\{x = 1\} - \mathbf{1}\{u = b\},$$

with γ some positive constant measuring the value of rain in comparison to the cost of seeding. Take $\gamma = 2$ for the first items. The process $x(\ell)$ controlled by $u(\ell)$ is then a Markov Decision Process.

Assume a seeding strategy wishes to maximise,

$$g = \liminf_{T\to\infty} \frac{1}{T}\mathbb{E}\sum_{\ell=0}^{T-1} r\big(x(\ell),\, u(\ell)\big).$$

1. Assume $u(\ell) \equiv' a'$ (no seeding policy). Calculate the stationary distribution vector, π, satisfying, $\pi P = \pi$ and $\sum \pi_i = 1$. Do this numerically using three different ways: (i) Taking high powers of the matrix P. (ii) Solving the system of equations. (iii) Monte-Carlo Simulation (yielding an estimate of π).

2. Using π, calculate g obtained by the no seeding policy.

3. There are a total of 8 stationary, Markov, deterministic policies. Each policy is described by a decision rule: $d : \mathcal{S} \to \mathcal{A}$. Enumerate each of the policies and write the probability transition matrix associated with each such policy (it can be obtained by interleaving rows of $P(a)$ and $P(b)$). Then for each policy calculate the stationary distribution π and the associated objective g. What is the best policy?

4. The best policy obviously varies as γ varies from 0 to ∞. Using numerical computation, investigate how this occurs. Your result should be a partition of $[0, \infty)$ into segments where on each segment there is a different optimal policy.

Problem 2: Optimal Usage of a Phone Battery

You are away on a trip with your cellular phone and without a charger. On previous days you have taken many photos and videos. You wished to share these with friends, but there was no connectivity. For 6 hours of today, you plan to be in a town with connectivity so you can upload your media. However, you can't charge your phone and hence only have limited battery life. Your goal is to try and upload as many photos as possible before the battery dies out.

You make a decision every 10 minutes if to upload more photos or not. Every 10 minutes, connectivity seems to vary between "bad state" (0) and "good state" (1) and by looking at the connectivity level you can decide if to upload or not (for the next 10 minutes). The throughput (file upload rate) achieved with both types of states is the same but the states vary as follows:

- In bad state (0), 10 minutes of upload depletes 5% of the battery and uploads $1GB$.
- In good state (1), 10 minutes of upload depletes only 1% of the battery and uploads $1GB$.

At the start of the day you have 100% of the battery. So you decision is for every 10 minute time interval (time step), if to transmit or not. With 6 hours you have 36 decision epochs.

Connectivity evolves as follows: At the first step connectivity is in bad state. After that, the chance of connectivity change is 0.1 and the chance of no connectivity change is 0.9. That is transitions occur according to a two state Markov chain with probability transition matrix

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix}.$$

A strategy is then a time-dependent function getting values in $\{a, b\}$ (don't upload vs. upload): $d_\ell(x, y)$ where $\ell \in \{0, 1, 2, \dots, 35\}$, $x \in \{0, 1\}$ (bad or good state respectively) and $y \in \{0, 1, 2, \dots, 99, 100\}$ (the precent battery remaining).

1. Assume you use the strategy of uploading as soon as possible. Compute (either analytically, numerically or by simulation) the mean number of GB you will upload.
2. Assume you use the strategy of uploading only when in good state. Compute (either analytically, numerically or by simulation) the mean number of GB you will upload.
3. Write the Bellman equation for the optimal policy and explain how you can calculate it numerically.
4. (Bonus only): Solve the Bellman equation and plot the optimal policy. What is the mean number of GB uploaded? Is there some structure to the optimal policy?

Problem 3: Dynamics of a Pendulum on a Cart

Consider the inverted pendulum example as described in class: A pendulum of mass m, length L (to the centre of gravity) and moment of inertia (with respect to centre of gravity) J, is connected to a cart of weight M pushed by a force $u(t)$ with opposing friction with a force of $-F\dot{s}(t)$. Here $s(t)$ is the displacement of the cart and $\phi(t)$ is the angle of the pendulum.

The equations describing this system are:

$$\begin{aligned}
m\frac{d^2}{dt^2}\big(s(t) + L\sin\phi(t)\big) &= H(t), && (1)\\
m\frac{d^2}{dt^2}L\cos\phi(t) &= V(t) - mg, && (2)\\
J\frac{d^2\phi(t)}{dt^2} &= LV(t)\sin\phi(t) - LH(t)\cos\phi(t), && (3)\\
M\frac{d^2 s(t)}{dt^2} &= u(t) - F\frac{ds(t)}{dt}, && (4)
\end{aligned}$$

where $H(t)$ and $V(t)$ are (respectively) the horizontal and vertical forces exerted on the pendulum at the pivot.

1. To the best of your ability, describe how these equations arise from physical first principles.

2. Setting L' as the "effective pendulum length" with

   $$L' := \frac{J + mL^2}{mL},$$

   show that these equations reduce to:

   $$\ddot{\phi}(t) - \frac{g}{L'}\sin\phi(t) + \frac{1}{L'}\ddot{s}(t)\cos\phi(t) = 0,$$
   $$M\ddot{s}(t) - u(t) + F\dot{s}(t) = 0.$$

3. Linearize the system around the solution where the pendulum is facing exactly up and is at rest. Using the state space representation,

$$x_1(t) := s(t), \quad x_2(t) := \dot{s}(t), \quad x_3(t) := s(t) + L'\phi(t), \quad x_4(t) := \dot{s}(t) + L'\dot{\phi}(t),$$

show that an (A, B, C, D) representation of the linearized system with measured output $(s(t), \phi(t))$is,

$$\begin{aligned}\dot{x}(t) &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\frac{F}{M} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{g}{L'} & 0 & \frac{g}{L'} & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \frac{1}{M} \\ 0 \\ 0 \end{bmatrix} u(t), \\ y(t) &= \begin{bmatrix} -\frac{1}{L'} & 0 & \frac{1}{L'} & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} x(t).\end{aligned}$$

4. Assume now that the pendulum is connected to a wall with a spring having spring constant k. Repeat the items above for this slightly modified situation.