Homework Set #1

1. **Dynamic Programming and MDP Equivalency**
A common DP model is given by Bertsekas in [1], page 13:
In time t, a system is said to be in state $x_t \in \mathcal{X}$. The agents/controller knows x_t and chooses a control decision u_t.
Then:

   - A random disturbance/noise (depending on the context) denoted w_t is generated according to probability distribution $p_t(w_t|x_t, u_t)$.
   - A reward $g_k(x_k, u_k, w_k)$ is generated
   - The next state evolves according to $x_{k+1} = f_k(x_k, u_k, w_k)$

   The total reward is accumulated over time:

$$g_1(x_1, u_1, w_1) + g_2(x_2, u_2, w_2) + ...$$

   Because of the presence of w_k the reward is generally a random variable. therefore we formulate the problem as an optimization of the *expected cost*:

$$E\big[g_1(x_1, u_1, w_1) + g_2(x_2, u_2, w_2) + ...\big] \tag{1}$$

   Where the expected value is taken with respect to $w_1, w_2, ...$.

   (a) Define the RL problem as given in class for discount factor $\gamma = 1$.

   (b) Assuming deterministic policy, i.e given a state S_t, action A_t is chosen deterministically. show that the formulation given for the value function in RL problem is Equivalent to the above DP formulation.
   Hint: use the following lemma:

   Lemma 1 (Functional Representation Lemma) *[2], page 2: Given 2 Random Variables, A and B with a conditional probability distribution $p(a|b)$, there exist a random variable W such that: $W \perp\!\!\!\perp B$ and $A = f(W, B)$*

2. **Only last state is relevant for decision making in MDP**
Show that for a MDP, actions at time t depend on the history only

through time t: $\pi(a_t|s_t, s_{t-1}, ...) = \pi(a_t|s_t)$.
hint: Look at arbitrary term in the expected return:

$$E_\pi\big[R_{t+1} + R_{t+2} + ...|S_t = s\big]$$

maximized over a given policy π and prove the above.

3. **MDP**
   please solve question 1 in the following link: exam-rl-questions.pdf

References

[1] Bertsekas D.P - Dynamic Programming and Optimal Control - Vol1

[2] Strong Functional Representation Lemma and Applications to Coding Theorems - C.T Li and A.E Gammal - https://arxiv.org/pdf/1701.02827.pdf