ELL729 Assignment 1 Report

Suraj Joshi 2018MT10045

Markov Decision Process Formulation

Let the time scale of the MDP be each day of the functioning of the power plant. Let the state of the power plant be described by its productivity at the start of the day before applying controls. The controls on each day are the pair of change in the temperature and the change in the heavy water that are made on each day of the operation. The noise or disturbance for each day is the categorical random variable which takes values of -0.05, 0 and +0.05 with equal probabilities and affects the next state computation.

The following symbols denote the various quantities used in describing the MDP:

k: The time scale with k belonging to the integers from 1 to 50 representing each day

x_k: The state on day k

u_k: The control on day k

ϵ_k: The disturbance on day k distributed with equal probability over -0.05, 0 and 0.05

α_k: The productivity of the plant on the start of day k, $\alpha_k \in [0, 0.1, 0.2, \ldots, 1]$

τ_k: The change in temperature of the reactor on day k, $\tau_k \in [-5, -4, \ldots, 4, 5]$.

ω_k: The change in heavy water of the reactor on day k, $\omega_k \in [-5, -4, \ldots, 4, 5]$

Some other parameters that are associated with the MDP are:

c_τ: The cost per unit increase in temperature

c_ω: The cost per unit increase in heavy water

M : The revenue generated by the plant when at full (1.0) productivity

λ : The cost per 0.1 deviation from the optimal productivity of the plant which is 0.3.

Describing the relationships among the above quantities mathematically,

The state $x_k = \alpha_k$

The control $u_k = (\tau_k, \omega_k)$

The state transition function is described by

$$x_{k+1}^* = f(x_k, u_k, \epsilon_k) = \alpha_{k+1}^* = f(\alpha_k, (\tau_k, \omega_k), \epsilon_k) = \begin{cases} \alpha_k + \tau_k\omega_k/125 + \epsilon_k & \text{if } \tau_k \geq 0, \omega_k \geq 0 \\ \alpha_k - \frac{\tau_k\omega_k}{125} + \epsilon_k & \text{if } \tau_k < 0, \omega_k < 0 \\ \alpha_k - \frac{|\tau_k - \omega_k|}{125} + \epsilon_k, & \text{otherwise} \end{cases}$$

and

$x_{k+1} = r(x_{k+1}^*)$ where r denotes the rounding function which rounds the value obtained from f according to the following rule:

- For a given x_k, τ_k and ω_k, if x_{k+1} does not belong to $\{0, 0.1, \ldots, 1\}$, round to nearest value.

- Round off middle value to lower discrete value.

The cost (revenue) per stage function is described by

$g(x_{k,}, u_k, \epsilon_k) = x_{k+1}*M - c_\tau*\max(0, \tau_k) - c_\omega*\max(0, \omega_k) – \lambda*| x_{k+1,} – 0.3|$

Impact of the Noise

The noise factor makes sure that the next iterations of the productivity will be either rounded down to the previous level or rounded up to the next level before the controls are applied. The controls, once applied, add/subtract the same values from all three possible values of the next productivity level. This ensures that the three possible iterations from a particular starting productivity cannot be all the same at the same time. For example, starting with $x_k = 0.3$, and adding the three components of noise, we get $x_{k+1}* = 0.25$ which will be rounded to 0.2, then we get $x_{k+1}* = 0.3$, and $x_{k+1}* = 0.35$ which will be rounded to 0.3. Now note that the deviation that a control pair can bring will be applied equally to all three of these outcomes. It implies that the three values of the next level productivity cannot all belong to the same productivity level before controls were applied. This leads to some interesting results. We may expect that in the case that the lambda parameter was higher than the revenue generated per day (say lambda = 20 and maximum revenue at full productivity = 10), that the optimal plant policy would be to stay in the optimal productivity state i.e. 0.3 but in fact on running the Dynamic Programming Solution it tells us that the optimal rate is in fact 0.4. This is because at 0.4, it is possible to apply a control (with (-5,-2) being one which the DP solution comes up with) such that 2 out of 3 possible noisy next state estimates are rounded to 0.3, the optimal state, which in fact leads to the maximum profit.

Dynamic Programming Solution

The dynamic programming problem to be solved is maximising the revenue over 50 days of the plant operation. The recursive relation that describes this DP problem is

$J(x_k) = \max_{(u_k)} \mathbf{E}_{(\epsilon_k)} [g(x_{k,}, u_k, \epsilon_k) + J(x_{k+1})]$

Now for each time k, the value x_k can take different values. To account for every possible x_k, multiple such DP equations have to be solved over all times k. The revenue at the last stage x_{50} depends only on the last stage so can be described as

$J(x_{50}) = \max_{(u_{50})} \mathbf{E}_{(\epsilon_{50})} [g(x_{50,}, u_{50}, \epsilon_{50})]$

Since this does not depend on any future state (or equivalently a future state where the revenue generated is identically 0) it can be used as the starting point for calculating the maximum revenue using backwards induction.

$J(x_1)$ denotes the maximum revenue starting from day 1. The exact value depends upon the starting value on the day, which can take the same range as α_k.

Particular Solution

Let us take a non-trivial example and then analyse the results on variation of input parameters.

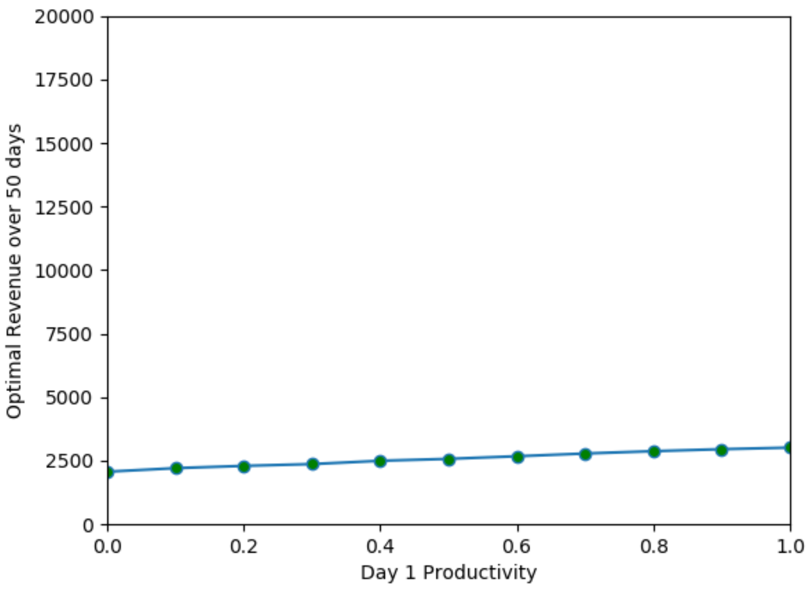
$c_\tau = 20$

$c_\omega = 20$

M = 150

$\lambda = 100$

The plot of $J(x_1)$ vs. possible x_1 values shows how the Profits vary with starting productivity of the plant on the first day. It shows that increased productivity to begin with leads to greater overall profits, which is just as expected.



Let us look at the optimal controls and the optimal next state sequences starting from

$x_1 = 0.0$

Optimal control sequence: (5,5) (5,5) (0,0)

State sequence: (0.0) (0.2) (0.4)

Hence the system tries to go to the productivity 0.4 as soon as possible and then stays there till the end of the period.

$x_1 = 0.5$

Optimal control sequence: (5,4) (1,1) (1,1) (1,1) (0,0) (-5,-1) (-5,-1), (-5,1) , (0,0) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (0,0) (-5,-1) (-5,-1) (0,0)

State sequence: (0.5) (0.7) (0.8) (0.9) (1.0) (1.0) (1.0) (1.0) (1.0) (1.0)

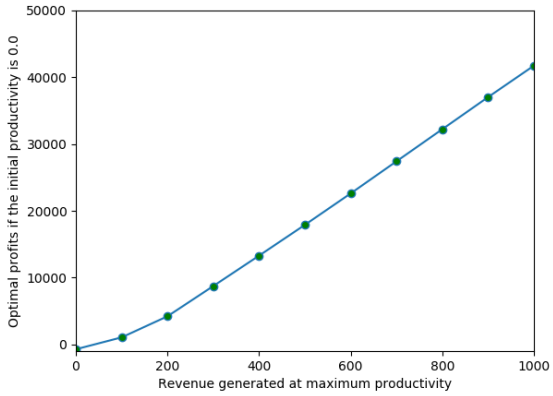
$x_1 = 1.0$

Optimal control sequence: (0,0) followed by (-5,-1) (-5,-1), (-5,1) , (0,0) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (-5, -1) (0,0) (-5,-1) (-5,-1) (0,0)

State sequence: (1.0)

So, the pattern observed is that for day 1 conditions starting at or below 0.4 productivity, for optimality the system either stays at the 0.4 state or reach there in the fewest possible days. It does not go higher because the cost of increased distance from optimal lambda along with the cost of increasing the productivity (can only be increased by positive increases of temperature and water) does not produce sufficient profits to be able to justify the change in state. However, for day 1 conditions at and above 0.5, the system does not have the pay the above-mentioned cost and the profits generated are high enough to justify even greater productivity (even at a cost) and move towards productivity 1. For initial productivity 1 the system does not deviate at all and always stays at 1.

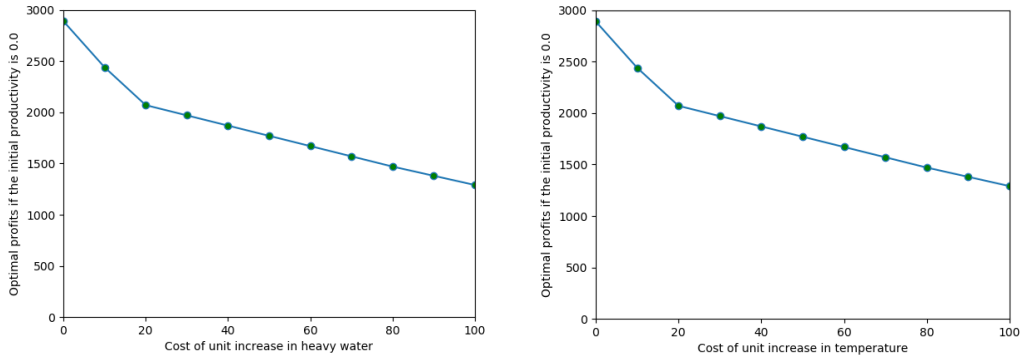
If M (revenue at max productivity) is shifted higher, say to 500, then every initial state finds it meaningful to move to 1 productivity state. This is just as expected because the extra revenue justifies bearing the extra cost earlier on during the 50 days. Of course, the overall revenue generated is also at a much higher range, from 17900 to 19500 for initial productivities from 0 to 1. The plot below shows the maximum revenue for initial state 0.0 for different values of M, keeping $c_\tau = 20$, $c_\omega = 20$ and $\lambda = 100$ as above:



A clear linear dependence is observed.

If one of the costs is increased to a very large amount, for example let $c_\omega = 400$ in the above and rest are the same as the particular solution, we observe that the optimal profits take a negative value, for $x_1 = 0.0$, the optimal action sequence is (5,3) followed by 49 (0,0) actions to get a profit of -1010. This is also profit maximisation, except that since a negative profit is being made, we can call it "loss minimisation" instead. The optimal action sequence suggests that the huge cost of changing the state once is worth it to maximise profit as opposed to remaining in a suboptimal state. Also, to note is that the non-zero control tries to take the action route that again minimises costs by having the greater increase in the cheaper control variable, i.e. temperature in this case.

Plotted below is the optimal revenue from day 1 productivity 0.0 for changing c_τ and c_ω from 0 to 100 while maintaining the other parameters the same ($\lambda = 100$, $M = 150$, $c_\tau = 20/c_\omega = 20$ when c_ω / c_τ are being changed respectively).



They are identical because keeping everything else the same, changing one parameter is equivalent to changing the other as the increase in temperature and increase and heavy water take the same range of values and the optimal controls taken in maximising the profits are changed by swapping the first entry of the tuple with the other if the costs are exchanged. Also observe the change in slope at the point where the costs are equal. This indicates that similar strategies were being used at both 0 and 10 levels while for 20 and above the strategy shifted to a more conservative one. (In this case they correspond to the initial states all moving to productivity 1 for 0 and 10 costs, while at and above 20 the initial states below and equal to 0.4 tend to go to 0.4).

If λ is increased to 1000 i.e. the penalty for being in a suboptimal productivity level is increased, then we observe that for every starting state, the system tries to move to 0.4 productivity as fast as possible, similar to how the <= 0.4 starting states moved to 0.4 in the particular solution described above. Here, observing the action sequence starting from initial productivity 1, we get

(-5, -5) (-5, -5) (-5, -5) (-5, -5) after which there are 46 actions all of which are (-5, -2). This is because the productivity sequence is as follows:

(1.0) (0.8) (0.7) (0.5) (0.4) followed by 45 states all of which are (0.4).

Also, with this modification the optimal profits for every starting state on day 1 turns out to be negative.

Plotted below is the optimal revenue from day 1 productivity 0.0 for different values of λ from 100 to 1000. Linear change is observed. Here, the other parameters are fixed as in the original example ($c_\tau = 20$, $c_\omega = 20$, $M = 150$)