Question 2

1. Each state contains the set of unfinished jobs. Each job can be finished or unfinished. So, the number of states with N=5 is the size of the power set
2. Solved by VI in the code
3. Here are the state values for all state in S using the maximal cost policy.

Here, state [1 0 0 0 0] means that job 1 still needs to be done and all other jobs are finished.

Chart, line chart

Description automatically generated

1. The plot for the initial state [1 1 1 1 1] value using PI starting from :

Chart, line chart

Description automatically generated

As we can see, it takes one iteration to the PI algorithm to find the optimal policy .

1. Let’s compare between the optimal policy achieved by PI and the policy . We observe that they are the same as expected since is known to be the optimal policy for this problem.

Moreover, we plot :

Chart

Description automatically generated with medium confidence

We observe as expected, that the CMU policy is equal to the optimal policy that we computer.

1. In the code
2. For each step size scheduler, we plot the max-norm of the error and the initial state error: .

:

Chart

Description automatically generated with low confidence

We observe that converge to 0 error is not achieved both in the Max-norm and in the Initial state values.

The idea behind this step-size is to hold the G1 requirements from the convergence theorem and thus ensures to converge to the real value function when

. However, we observe that the error doesn’t converge to 0 even after a huge number of iterations. The step size becomes too small at some point to enable convergence and a real change to the value function.

: We observe that with this step size, convergence for and a 0-error are achieved but is not stable around a 0-error. This is because the step size doesn’t tend to 0 and always updates the value function even if it achieved the minimal error.

A picture containing graphical user interface

Description automatically generated

:

A picture containing graphical user interface

Description automatically generated

For this step-size scheduler, we observe a more stable convergence of the error. This is because the step size tends to 0 with new visits and thus, new experiences don’t add noise to the computed state values. The stability occurs both in the Max-norm error and in the Initial-state error. The advantage of this step-size, is that it is higher than the

from and thus, gives more weights to more experiences such that TD(0) has the time to converge to a better value function.

In all the step sizes, we observe that the Max-norm error doesn’t converge to 0. This is because the fixed policy may not visit all the states and for these states, the values cannot converge since they’re not visited (a sufficient number of times).

1. We use several lambda values and plot the errors using step-size

. The plots show the average errors over 20 experiments/lambda.

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

In both cases, we clearly observe that higher lambda gives better convergence values.

However, we noticed that with other step-size schedulers, increasing lambda not necessarily improve the error.

For each step size scheduler, we plot the max-norm of the error and the initial state error: . The errors are computed every 100 iterations such that 1 iteration in the plots represent 100 iterations of the Q leaning steps.

:

Chart

Description automatically generated with low confidence

We observe that converge to 0 error is not achieved both in the Max-norm and in the Initial state values. However, this step-size enables stability as we already saw.

:

A picture containing histogram

Description automatically generated

This step-size enables faster learning since it doesn’t tend to 0 but the errors are not stable around some 0 error.

:

A picture containing chart

Description automatically generated

This step-size is the best combination of both worlds. Stability and fast convergence are achieved.

1. Using , and the last step-size , we obtain the following error plots.

Chart

Description automatically generated with low confidence

As we can see, the errors are approximately the same as with .