9 Dynamic Programming

9.1 Policy Improvement for a Markov Decision Process (4 units)

This project is self-contained mathematically; background information is provided in the Part II course on Optimisation and Control (see reference [1]).

1 A Car Replacement Problem

Car owners are haunted by the following problem. Every day, the operating cost for their car increases, as does the probability that the car breaks down. Even worse, when trading in the car for a different one dealers will pay less for older cars and charge more for newer ones. The problem, then, is to find an optimal policy for trading in the car.

We model the problem as a Markov decision process. Let $g_j(u)$ be the instantaneous cost incurred if one takes action u in state j and let $p_{jk}(u)$ be the probability of then moving to state k. Define sequences $\gamma^{(n)}$, $f_j^{(n)}, u_j^{(n)}$ by the recursions

$$\gamma^{(n)} + f_j^{(n)} = g_j\big(u_j^{(n)}\big) + \sum_k p_{jk}\big(u_j^{(n)}\big) f_k^{(n)} \tag{1}$$

and

$$u_j^{(n+1)} \text{ is the } u\text{-value minimising } g_j(u) + \sum_k p_{jk}(u) f_k^{(n)}. \tag{2}$$

The following exercise may help to gain some intuition.

> **Question 1** Consider the following stationary policy: for fixed n, whenever state j occurs take action $u_j^{(n)}$. What is the long-term average cost of this policy? Explain.

Note that the values $f_j^{(n)}$ determined by (1) are arbitrary up to an additive constant and can be normalised, for example by letting $f_1^{(n)} = 0$. If the matrix of transition probabilities is irreducible in every stage, then (1) will always have a solution for f. The sequence $\gamma^{(n)}$ is non-increasing, and will converge to a minimum value γ in a finite number of steps if u can take only a finite number of values. The policy $u_j^{(n)}$ will then have converged to an average optimal policy.

> **Question 2** Instantiate the above framework for the car replacement problem. You may want to introduce states representing the age of the car in appropriately chosen units of time, and an additional state in which the car is written off and has a trade-in value of zero. Describe the set of actions, and define the instantaneous costs $g_j(u)$ and the transition probabilities $p_{jk}(u)$.

> **Question 3** Write a program to find the optimal replacement policy. You are not required to write your own linear algebra routines, but you should describe any mathematical manipulations involved in bringing the equations in the desired form. Give a

j	age in years	purchase price	trade-in price	operating cost	survival probability
1	0	5000	3500	860	0.963
2	2	3150	2170	1025	0.794
3	4	2285	1500	1225	0.568
4	6	1545	900	1430	0.255
5	8	1050	590	1815	0.001
6	10	600	330	2240	0.000

Table 1: Instance of the car replacement problem with time units of two years and $N = 6$

j:	1	2	3	4	5	6	7
sell/keep:	keep	keep	keep	keep	keep	sell	sell
buy car of age:	–	–	–	–	–	2	2

Table 2: Policy for the problem of Table 1

j	age in years	purchase price	trade-in price	operating cost	survival probability
1	0	5000	3500	200	0.999
2	0.5	4285	3000	210	0.995
3	1	3750	2650	220	0.990
4	1.5	3430	2375	230	0.979
5	2	3150	2170	240	0.968
6	2.5	2900	1950	250	0.956
7	3	2645	1850	260	0.936
8	3.5	2475	1625	275	0.917
9	4	2285	1500	290	0.898
10	4.5	2130	1350	300	0.879
11	5	1970	1225	315	0.860
12	5.5	1760	1060	320	0.836
13	6	1545	900	335	0.801
14	6.5	1400	780	350	0.761
15	7	1260	700	365	0.697
16	7.5	1140	625	380	0.600
17	8	1050	590	400	0.482
18	8.5	940	520	430	0.300
19	9	830	470	465	0.129
20	9.5	720	400	520	0.020
21	10	600	330	560	0.000

Table 3: Instance of the car replacement problem with time units of six months and $N = 21$

clear and concise description of your algorithm; don't forget to mention what starting conditions you use. Run your program on the data contained in the file *table1.csv* available from the CATAM website and displayed in Table 1, and compare your results to the policy in Table 2. What is the value of γ?

Question 4 Give the optimal replacement policy for the data in the file *table2.csv* available from the CATAM website and displayed in Table 3. What is the value of γ?

Question 5 Suppose that purchase price, trade-in price, operating cost, and survival probability are all monotonically increasing or decreasing in the obvious direction. Suppose further that the optimal policy tells you to sell a car when it reaches a certain age, but that you neglect to do so. Is it possible that the same policy stipulates hanging on to the car now that it is older? Either construct an example for which the optimal policy is of this kind, or prove that this is impossible.

References

[1] R.R.Weber, *Course notes on Optimisation and Control*, Section 8.4. `http://www.statslab.cam.ac.uk/~rrw1/oc/`.

