9/14/12

Problems with an infinite planning horizon

Immediate problem: minimising cost

Suppose we have strategies which give

4, 4, 4, 4, 4, ... (1)

↓ better average cost

3, 3, 3, 3, 3, ... (11)

↓ better discounted cost

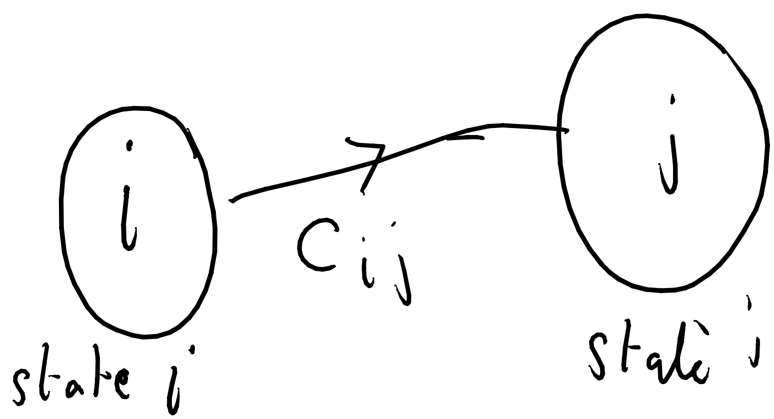
2, 4, 2, 4, 2, ... (111)

Assume we have a discount factor $\alpha < 1$

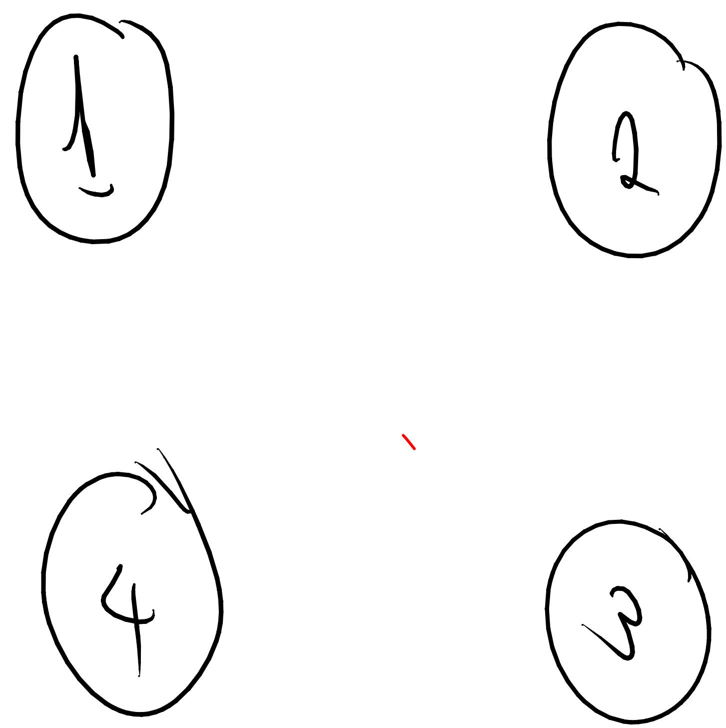
and sequence of costs $c_1, c_2, c_3, \dots$

we want to minimise $c_1 + c_2\alpha + c_3\alpha^2 + \dots < \infty$

First scenario: set of n "states" V



When in state i, I choose my next state j and pay c_{ij} and repeat forever



Policy π

When in state i go to state $j = \pi(i)$ next.

# of policies is n^n

Problem: find policy π that minimizes

Suppose we just go $1 \to 2 \to 3 \to 4 \to 1 \to$

$$\text{cost} = C_{1,2} + \alpha C_{2,3} + \alpha^2 C_{3,4} + \alpha^3 C_{4,1} + \alpha^4 C_{1,2} + \cdots$$

Algorithm to choose "best" policy:

Start with π_0 and produce $\pi_1, \pi_2, \ldots$

where

$$\pi_0 > \pi_1 > \pi_2 > \ldots$$

π_2 is "better" than π_1

Policy evaluation:

$$n=4 \qquad \begin{array}{c} \text{Costs} \\ \begin{bmatrix} 6 & 4 & 5 & 7 \\ 3 & 9 & 2 & 5 \\ 4 & 3 & 7 & 2 \\ 5 & 3 & 8 & 1 \end{bmatrix} \end{array} \qquad \alpha = 1/2$$

$$\pi = (4, 2, 3, 3)$$

y_i^π = discounted cost of following π starting at i

$$= c_{i,\pi(i)} + \alpha\, y_{\pi(i)}$$

$$y_1 = 7 + \frac{1}{2} y_4 = 29/2$$

$$y_2 = 9 + \frac{1}{2} y_2 = 18$$

$$y_3 = 7 + \frac{1}{2} y_3 = 14$$

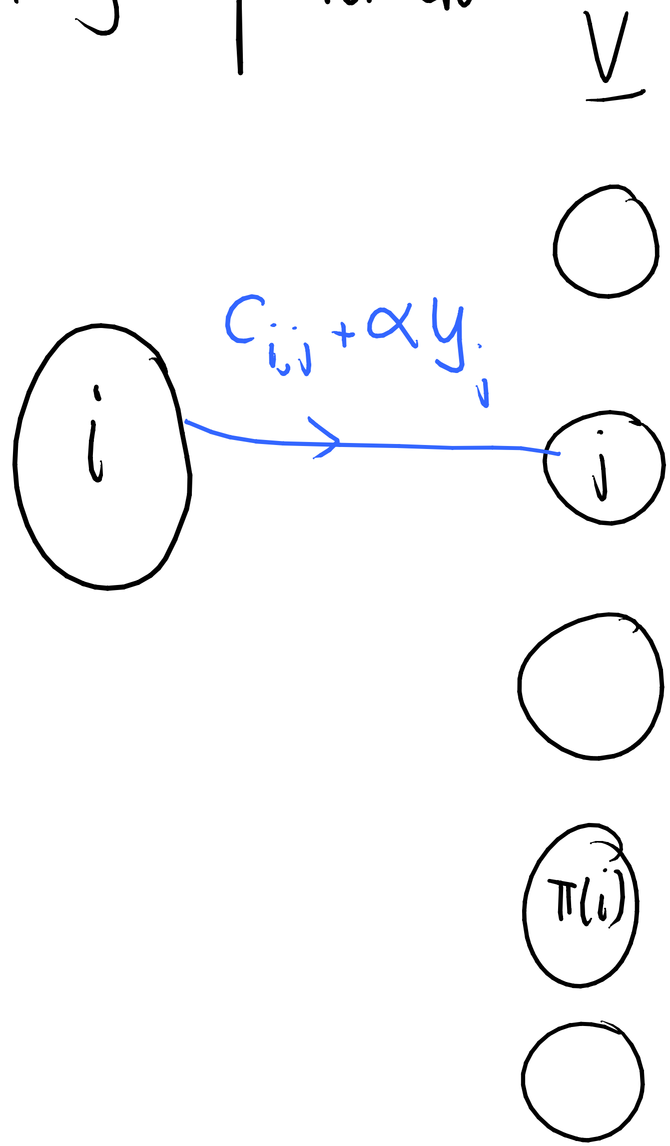
$$y_4 = 8 + \frac{1}{2} y_3 = 15$$

Now we will find a new policy that reduce <u>all</u> of these y_i's.

$$\begin{bmatrix} 6 & 4 & 5 & 7 \\ 3 & 9 & 2 & 5 \\ 4 & 3 & 7 & 2 \\ 5 & 3 & 8 & 1 \end{bmatrix} \qquad \alpha = 1/2$$

$$\pi = (4, 2, 3, 3)$$

Policy improvement



Compare going $i \to j$ and follow π for all j

For each i, find the j that minimises $c_{ij} + \alpha y_j$ and switch $\pi(i) \to j$ if $j \neq \pi(i)$.

$\underline{i = 1}$

$$6 + \frac{1}{2} \times \frac{29}{2}$$

$$4 + \frac{1}{2} \times 18$$

$$5 + \frac{1}{2} \times 14 \quad *$$

$$7 + \frac{1}{2} \times 15$$

$$\begin{bmatrix} 6 & 4 & 5 & 7 \\ 3 & 9 & 2 & 5 \\ 4 & 3 & 7 & 2 \\ 5 & 3 & 8 & 1 \end{bmatrix} \quad \alpha = 1/2$$

$$\pi = (4, 2, 3, 3)$$

$i = 2$

$3 + \frac{1}{2} \times \frac{29}{2}$

$9 + \frac{1}{2} \times 18$

$2 + \frac{1}{2} \times 14$ *

$5 + \frac{1}{2} \times 15$

$$\begin{bmatrix} 6 & 4 & 5 & 7 \\ 3 & 9 & 2 & 5 \\ 4 & 3 & 7 & 2 \\ 5 & 3 & 8 & 1 \end{bmatrix} \quad \alpha = 1/2$$

$\pi = (4, 2, 3, 3)$

$\underline{i = 3}$

$$4 + \frac{1}{2} \times \frac{29}{2}$$

$$3 + \frac{1}{2} \times 18$$

$$7 + \frac{1}{2} \times 14$$

$$2 + \frac{1}{2} \times 15 \quad *$$

$$\begin{bmatrix} 6 & 4 & 5 & 7 \\ 3 & 9 & 2 & 5 \\ 4 & 3 & 7 & 2 \\ 5 & 3 & 8 & 1 \end{bmatrix} \qquad \alpha = 1/2$$

$$\pi = (4, 2, 3, 3)$$

$\underline{i = 4}$

$$5 + \frac{1}{2} \times \frac{29}{2}$$

$$3 + \frac{1}{2} \times 18$$

$$8 + \frac{1}{2} \times 14$$

$$1 + \frac{1}{2} \times 15 \quad *$$

$$\begin{bmatrix} 6 & 4 & 5 & 7 \\ 3 & 9 & 2 & 5 \\ 4 & 3 & 7 & 2 \\ 5 & 3 & 8 & 1 \end{bmatrix} \qquad \alpha = 1/2$$

$$\pi = (4, 2, 3, 3)$$

New policy: $(3, 3, 4, 4)$