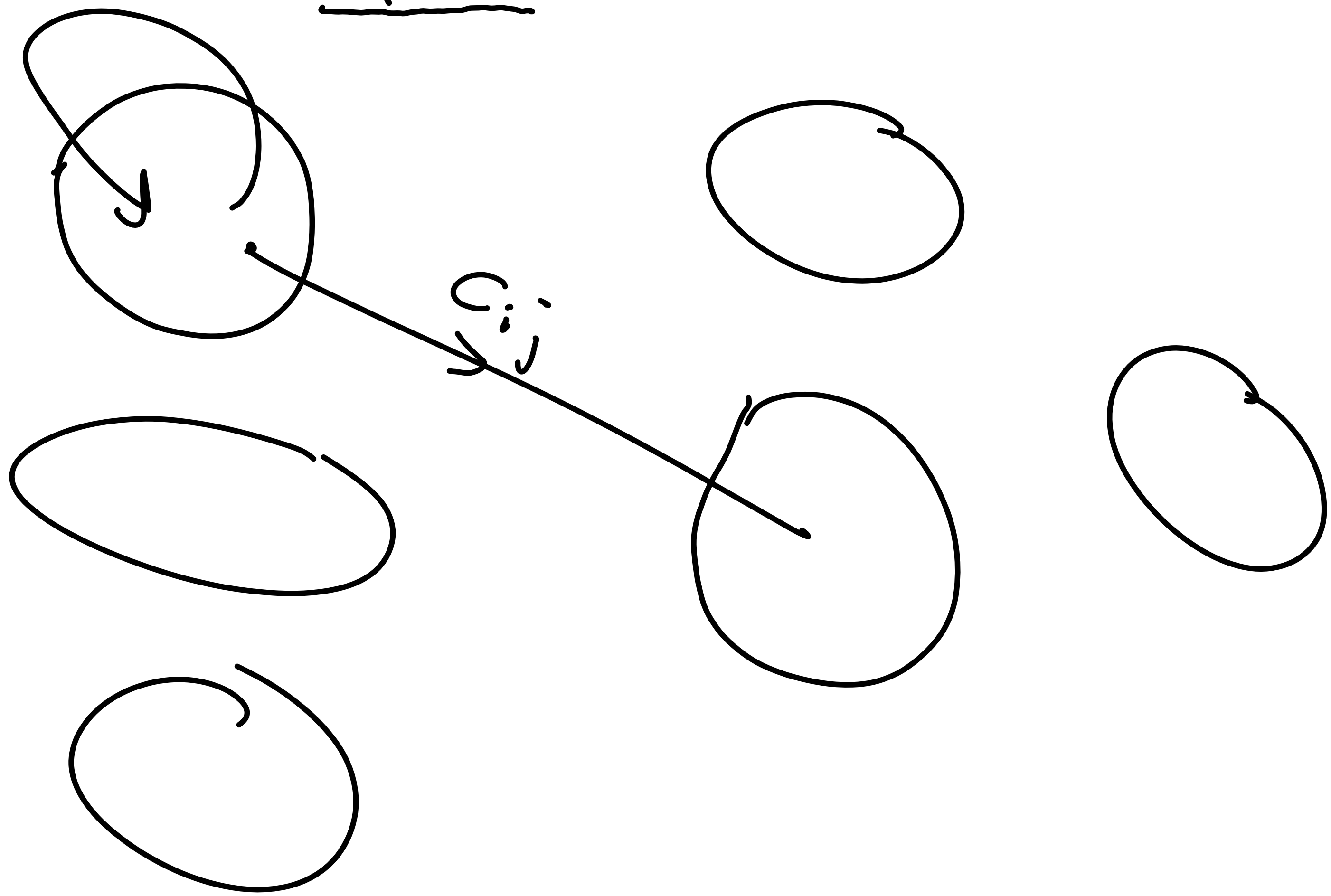
9/13/10

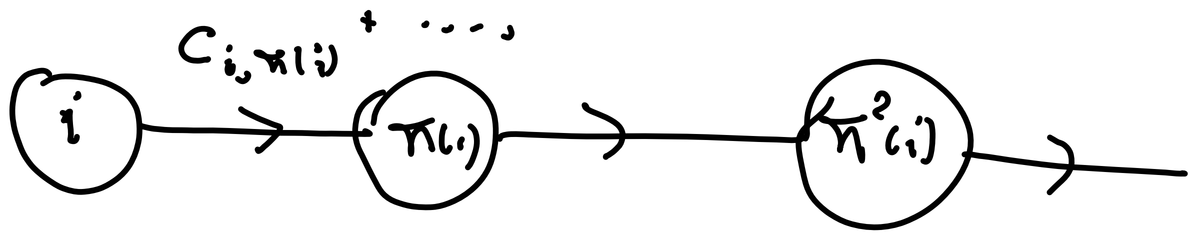


α = discount factor.

Policy π: when at i, go next to $\pi(i)$.

y_i = discounted cost, starting at i

$$= C_{i,\pi(i)} + \alpha\, y_{\pi(i)} \qquad i \in V$$



Example $n=4$

$$C \quad \begin{bmatrix} 5 & 2 & 3 & 7 \\ 4 & 1 & 2 & 5 \\ 1 & 6 & 3 & 2 \\ 4 & 8 & 2 & 5 \end{bmatrix}$$

$\alpha = 1/2$

$$\Pi_0 = \begin{array}{ccccc} i & 1 & 2 & 3 & 4 \\ \pi(i) & 4 & 4 & 2 & 2 \end{array}$$

Bad choice

Evaluate π

$$\Pi_0 = \begin{array}{ccccc} i & 1 & 2 & 3 & 4 \\ \pi(i) & 4 & 4 & 2 & 2 \end{array}$$

$$y_1 = 7 + \tfrac{1}{2} y_4 = 14$$

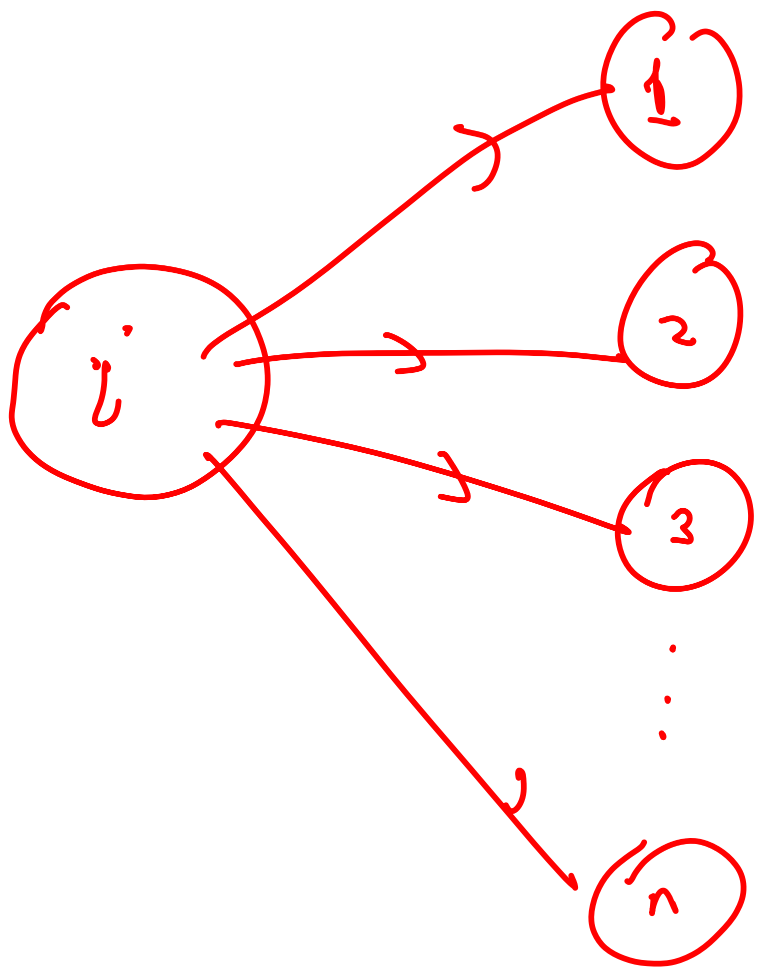
$$y_2 = 5 + \tfrac{1}{2} y_4 = 12$$

$$y_3 = 6 + \tfrac{1}{2} y_2 = 12$$

$$y_4 = 8 + \tfrac{1}{2} y_2 = 14$$

Is this optimal ?

We consider some modified policies:



follow π from now on

$$c_{i,1} + \alpha y_1$$

$$c_{i,2} + \alpha y_2$$

$$c_{i,3} + \alpha y_3$$

$$\vdots$$

$$c_n + \alpha y_n$$

Policy Improvement:

for each i compare

$$y_i \quad \text{and} \quad \min_j \left(C_{ij} + \alpha y_j \right) \leq y_i \quad \boxed{j = \pi(i)}$$

(i) If $y_i = \min, \forall i$ then π is optimal

(ii) $y_i > \min$, for $i \in U$ then switch $\pi(i)$ to j giving the min.

$$\begin{matrix} 5 & 2 & 3 & 7 \\ 4 & 1 & 2 & 5 \\ 1 & 6 & 3 & 2 \\ 4 & 8 & 2 & 5 \end{matrix}$$

i:	1	2	3	4
y_i:	14	12	12	14

Check optimality:

<u>$i = 1$</u>

$$5 + \frac{1}{2} \times 14 = 12$$

$$2 + \frac{1}{2} \times 12 = 8^{*}$$

$$3 + \frac{1}{2} \times 12 = 9$$

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

5	2	3	7
4	1	2	5
1	6	3	2
4	8	2	5

$\underline{i = 2}$

$$4 + \frac{1}{2} \times 14 = 11$$

$$1 + \frac{1}{2} \times 12 = 7^{*}$$

$$2 + \frac{1}{2} \times 12 = 8$$

$$5 + \frac{1}{2} \times 14 = 12$$

5	2	3	7
4	1	2	5
1	6	3	2
4	8	2	5

$i = 3$

$$1 + \frac{1}{2} \times 14 = 8^{*}$$

$$6 + \frac{1}{2} \times 12 = 12$$

$$3 + \frac{1}{2} \times 12 = 9$$

$$2 + \frac{1}{2} \times 14 = 9$$

5	2	3	7
4	1	2	5
1	6	3	2
4	8	2	5

$i = 4$

$$4 + \frac{1}{2} \times 14 = 11$$

$$8 + \frac{1}{2} \times 12 = 14$$

$$2 + \frac{1}{2} \times 12 = 8^{*}$$

$$5 + \frac{1}{2} \times 14 = 12$$

New policy

i	1	2	3	4
$\pi(i)$	2	2	1	3

Evaluate new policy:

$$y_1 = 2 + \tfrac{1}{2} y_2 = 3$$

$$y_2 = 1 + \tfrac{1}{2} y_2 = 2$$

$$y_3 = 1 + \tfrac{1}{2} y_1 = 5/2$$

$$y_4 = 2 + \tfrac{1}{2} y_3 = \frac{13}{4}$$

5	2	3	7
4	1	2	5
1	6	3	2
4	8	2	5

Check for optimality:

$$\begin{matrix} 5 & 2 & 3 & 7 \\ 4 & 1 & 2 & 5 \\ 1 & 6 & 3 & 2 \\ 4 & 8 & 2 & 5 \end{matrix}$$

i	1	2	3	4
y_i	3	2	$\frac{5}{2}$	$\frac{13}{4}$

$\underline{i = 1}$

$$5 + \tfrac{1}{2} \times 3 =$$

$$2 + \tfrac{1}{2} \times 2 = 3^*$$

$$3 + \tfrac{1}{2} \times \tfrac{5}{2} =$$

$$7 + \tfrac{1}{2} \times \tfrac{13}{4} =$$

$$\begin{matrix} 5 & 2 & 3 & 7 \\ 4 & 1 & 2 & 5 \\ 1 & 6 & 3 & 2 \\ 4 & 8 & 2 & 5 \end{matrix}$$

$\underline{i = 2}$

$$4 + \frac{1}{2} \times 3 =$$

$$1 + \frac{1}{2} \times 2 = 2^{*}$$

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

$$5 + \frac{1}{2} \times \frac{13}{4} =$$

5	2	3	7
4	1	2	5
1	6	3	2
4	8	2	5

$i = 3$

$$1 + \frac{1}{2} \times 3 = \frac{3}{2} *$$

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

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

$$2 + \frac{1}{2} \times \frac{13}{4} =$$

$$\begin{matrix} 5 & 2 & 3 & 7 \\ 4 & 1 & 2 & 5 \\ 1 & 6 & 3 & 2 \\ 4 & 8 & 2 & 5 \end{matrix}$$

$\underline{i = 4}$

$$4 + \frac{1}{2} \times 3 =$$

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

$$2 + \frac{1}{2} \times \frac{5}{2} = \frac{13}{4}^{*}$$

$$5 + \frac{1}{2} \times \frac{13}{4} =$$

Current policy passes test: Optimal.

Proof of optimality

(1) π satisfies

$$y_i = \min_j \left(c_{ij} + \alpha y_j \right)$$

Suppose $\hat{\pi}$ is some other policy: $\hat{y}$

$$\hat{y}_i = c_{i,\hat{\pi}(i)} + \alpha \hat{y}_{\hat{\pi}(i)}$$

$$y_i \leqslant c_{i,\hat{\pi}(i)} + \alpha y_{\hat{\pi}(i)}$$

So, $\hat{y}_i - y_i \geq \alpha \left(\hat{y}_{\hat{\pi}(i)} - y_{\hat{\pi}(i)} \right)$

$$\underbrace{\hat{y}_i - y_i}_{\xi_i} \geq \alpha\left(\hat{y}_{\hat{\pi}(i)} - y_{\hat{\pi}(i)}\right)$$

At this point we do not know if the ξ_i's are positive or negative

$$\xi_i \geq \alpha\, \xi_{\hat{\pi}(i)}$$

$$\geq \alpha^2 \xi_{\hat{\pi}^2(i)}$$

$$\vdots$$

$$\geq \alpha^k \xi_{\hat{\pi}^k(i)} \longrightarrow 0$$

So $\xi_i \geq 0$