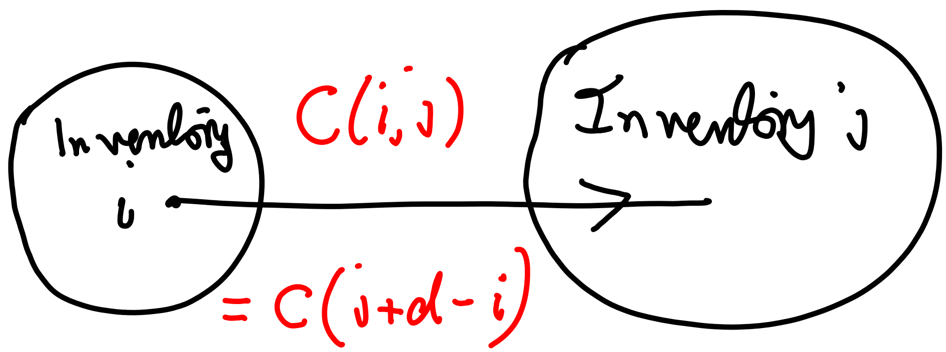
9\17\2007

n states

Discount factor α

Cost $C(i,j)$ of going from state i to state j in one step

Production Problem



We want to find optimal $\pi : V \to V$

"when in state i, go to $\pi(i)$, next step"

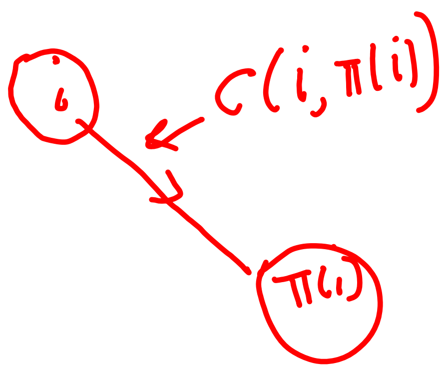
Given π: (i) Evaluate it

(ii) Check it

Done

change it

(i) $y_i = y_i(\pi)$ = discounted cost of pursuing π, starting at i.



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

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

$$\alpha = \frac{1}{2} \qquad \begin{bmatrix} 2 & 3 & 8 \\ 4 & 1 & 12 \\ 10 & 3 & 2 \end{bmatrix} \qquad \begin{aligned} \pi(1) &= 3 \\ \pi(2) &= 3 \\ \pi(3) &= 1 \end{aligned}$$

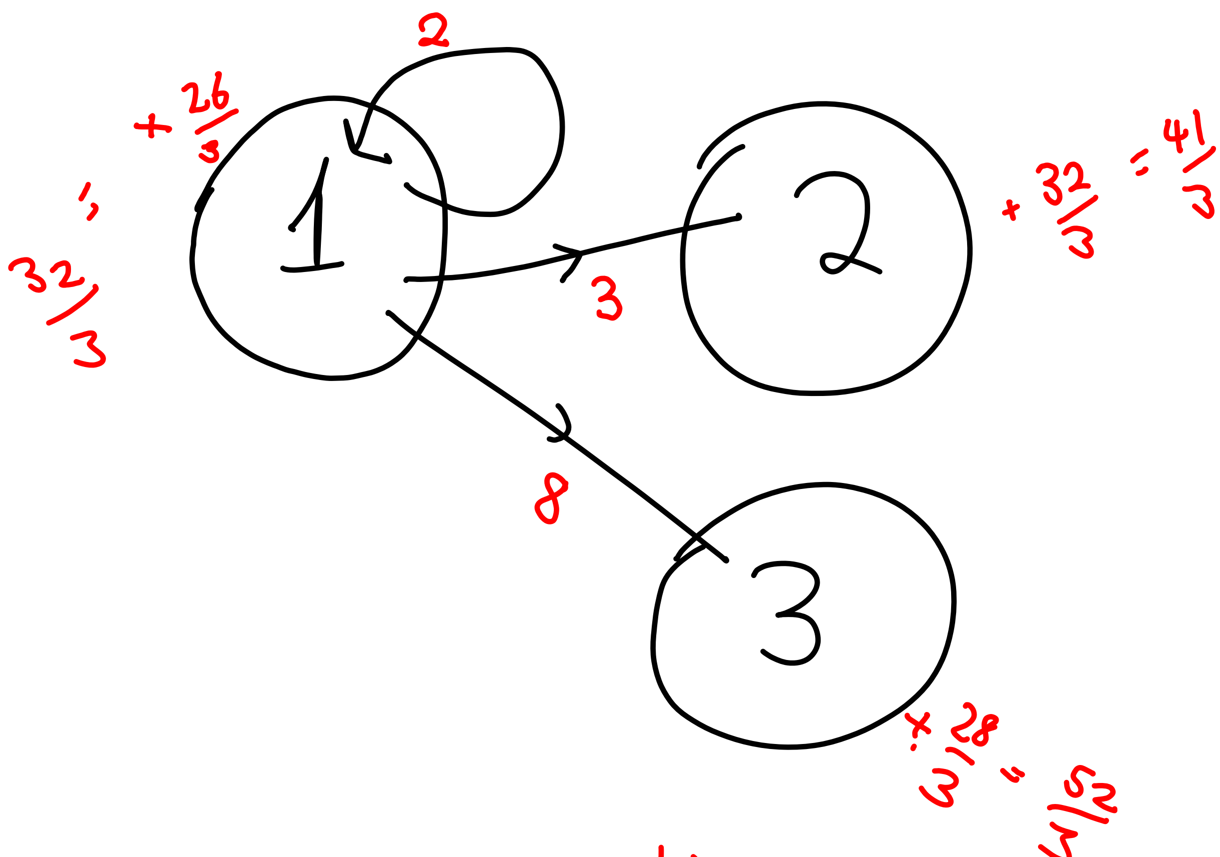
Definitely not optimal

$$y_1 = 8 + \frac{1}{2} y_3 \qquad y_1 = \frac{52}{3}$$

$$y_2 = 12 + \frac{1}{2} y_3 \qquad y_2 = \frac{64}{3}$$

$$y_3 = 10 + \frac{1}{2} y_1 \qquad y_3 = \frac{56}{3}$$

(11) Check for optimality



$y_1 = \frac{52}{3}$

$y_2 = \frac{64}{3}$

$y_3 = \frac{56}{3}$

$$\begin{bmatrix} 2 & 3 & 8 \\ 4 & 1 & 12 \\ 10 & 3 & 2 \end{bmatrix}$$

Suggest that I change policy

$$\pi(1) \longrightarrow 1$$

Policy Check:

For each i, j, compute cost of going $i \rightarrow j$ in first round and then following π. If this is better for some i, j then π is not optimal.

Compute $$\min_j \left\{ C(i,j) + \alpha y_j \right\}$$

If this is not equal to y_i, change $\pi(i)$ to where minimum occurs.

$$\begin{bmatrix} 2 & 3 & 8 \\ 4 & 1 & 12 \\ 10 & 3 & 2 \end{bmatrix}$$

$y_1 = \frac{52}{3}$

$y_2 = \frac{64}{3}$

$y_3 = \frac{56}{3}$

$\pi(1) = \pi(2) = 3$

$\pi(3) = 1$

Now Prolong

	$c(i,j)$	+	αy_j				
$i = 1$	2	+	$26/3$	=	$32/3$	*	$\pi(1) = 1$
	3	+	$32/3$	=	$41/3$		
	8	+	$28/3$	=	$52/3$		
$i = 2$	4	+	$26/3$	=	$38/3$		
	1	+	$32/3$	=	$35/3$	*	$\pi(2) = 2$
	12	+	$28/3$	=	$64/3$		
$i = 3$	10	+	$26/3$	=	$56/3$		
	3	+	$32/3$	=	$41/3$		
	2	+	$28/3$	=	$34/3$	*	$\pi(3) = 3$

Evaluate new policy:

$\pi(i) = i, \quad i = 1, 2, 3$

$$\begin{bmatrix} 2 & 3 & 8 \\ 4 & 1 & 12 \\ 10 & 3 & 2 \end{bmatrix}$$

$$y_1 = 2 + \tfrac{1}{2} y_1 = 4$$

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

$$y_3 = 2 + \tfrac{1}{2} y_3 = 4$$

$i = 1$

$2 + \tfrac{1}{2} y_1$ *

$3 + \tfrac{1}{2} y_2$ *

$8 + \tfrac{1}{2} y_3$

$i = 2$

$4 + \tfrac{1}{2} y_1$

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

$12 + \tfrac{1}{2} y_3$

$i = 3$

$10 + \tfrac{1}{2} y_1$

$3 + \tfrac{1}{2} y_2$ *

$2 + \tfrac{1}{2} y_3$ *

No incentive to change. Current policy is optimal