9\19\2007

Policy Improvement Procedure.

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

Compare

$$y_i \quad \& \quad \min_j \left[\underbrace{c(i,j) + \alpha y_j}_{\text{Cost of going to } j \text{ and then using } \pi \text{ for the rest of the time.}} \right]$$

$$y_i \;\&\; \min\left[c(i,j)+\alpha y_j\right]$$

$$U = \{i : RHS < LHS\}$$

Claim: (i) $U = \emptyset \Rightarrow \pi$ optimal

(ii) $U \neq \emptyset \Rightarrow \pi$ can be improved.

<u>Proof</u>

(i) Suppose $U = \emptyset \Rightarrow y_i = \min_j c(i,j) + \alpha y_j, \; \forall i$

Let $\hat{\pi}$ be any other policy.

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

$$y_i \leq c(i, \hat{\pi}(i)) + \alpha y_{\hat{\pi}(i)}$$

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

$$y_i \leq c(i, \hat{\pi}(i)) + \alpha\, y_{\hat{\pi}(i)}$$

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

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

$$\xi_{\hat{\pi}(i)} \leq \alpha\, \xi_{\hat{\pi}^2(i)} \qquad \hat{\pi}(\hat{\pi}(i))$$

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

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

$$\vdots$$

$$\leq \alpha^k \xi_{\hat{\pi}^k(i)}$$

$$\Downarrow$$

$0 < \alpha < 1$ $\longrightarrow 0$ as $k \to \infty$

So $\xi_i \leq 0$, $\forall i$ i.e. $y_i \leq y_{\hat{\pi}(i)}$, $\forall i$

(II) $$y_i \quad \& \quad \min_j \left[c(i,j) + \alpha y_j \right]$$

$$U = \{ i : RHS < LHS \}$$

Suppose $U \neq \emptyset$

Define a new policy.

$$\tilde{\pi}(i) = \begin{cases} \pi(i) & : i \notin U \\ j & : j \text{ minimizes RHS, if } i \in U \end{cases}$$

Compare $\tilde{y}$ & y

$$\tilde{y}_i = c(i, \tilde{\pi}(i)) + \alpha \tilde{y}_{\tilde{\pi}(i)}$$

$$y_i \geq C(i, \tilde{\pi}(i) + \alpha y_{\tilde{\pi}(i)} \qquad \forall i$$

$$> \qquad i \in U$$

$$\xi_i = y_i - \tilde{y}_i$$

$$\xi_i > \alpha \xi_{\tilde{\pi}(i)} \qquad i \in U$$

$$= \qquad i \notin U$$

$$i \in U \qquad \xi_i > \alpha \xi_{\tilde{\pi}(i)} \geq \alpha^2 \xi_{..} \geq \alpha^3 \xi_{....} \geq \cdots$$

$\xi_i > 0 \quad i \in U$

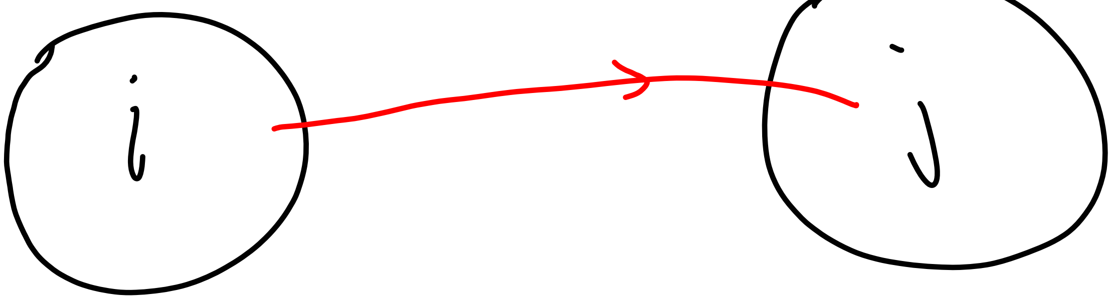
$\xi_i \geq 0 \quad i \notin U$

Adding Probability

Production problem

decide on x

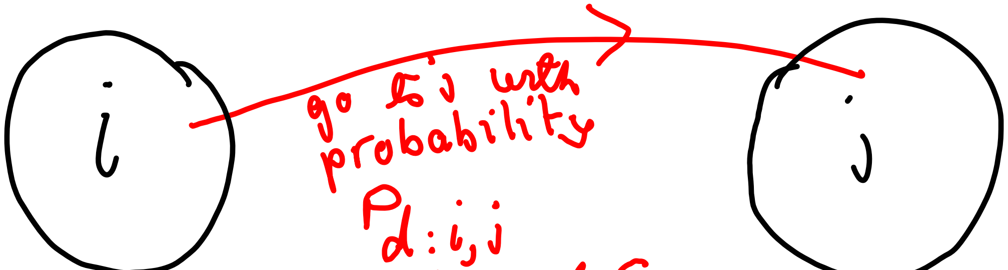
end in j with probability depending on j



state $i \equiv i$ in stock

General Model

$D_i = \{$available decisions$\}$



go to j with probability $P_{d:i,j}$ at expected cost $C_{d:i,j}$

$$C_{d,i} = \sum_j C_{d:i,j} P_{d:i,j}$$

= expected transition cost from i given decision d.

Argue that it is best to choose some $\pi(i) \in D_i$ for each i.

Same algorithm:

Fix policy π

$\downarrow$

Evaluate

$\downarrow$

Check for optimality

Given π, let y_i : expected discounted cost of using π.

$$y_i = c_{\pi(i),i} + \alpha \sum_j P_{\pi(i):i,j}\, y_j$$

make decision $\pi(i)$

Set of linear equations, which we can solve.