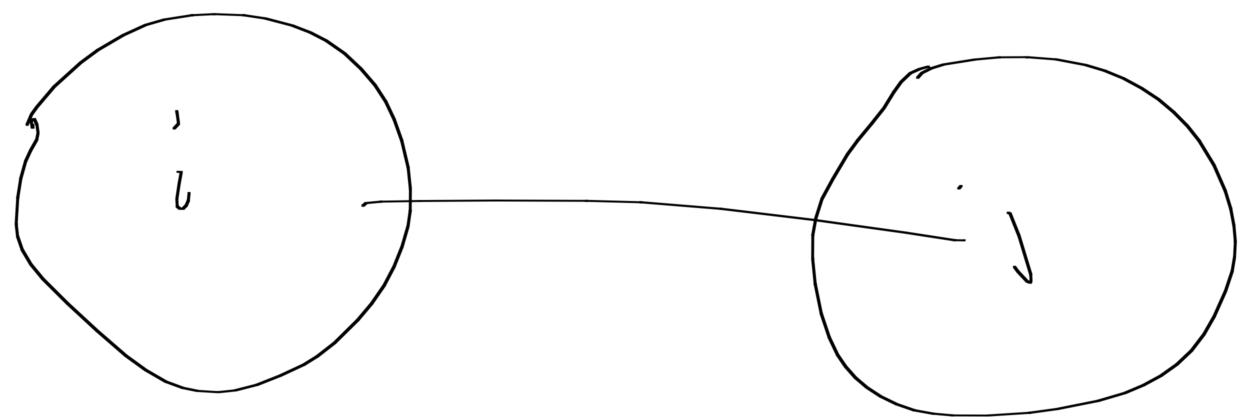
10/27/08

D_i : choose $x \in D_i$



$P[x, i, j] = Pr[i \rightarrow j]$ if you choose x.

$C[x, i]$ = expected cost from i

Policy : $\pi(i) \in D_i$ for each i

Choose policy to minimise, expected discounted cost:

$$
\begin{tikzcd}
\Pi \arrow[r, "\text{evaluate } \pi"] & y \arrow[r, "\text{check optimality}"] & \text{DONE} \arrow[ll, bend left, "\text{change } \pi"]
\end{tikzcd}
$$

Evaluate π.

$$y_i = c(\pi(i), i) + \alpha \sum_{j \in V} P(\pi(i), i, j)\, y_j$$

$\pi(i)$: decision for π

Matrix Form:

$$y = c_\pi + \alpha P_\pi y$$

y: $(y_1, y_2, \ldots, y_n)$

P_π: transition matrix of a Markov chain.

$$P_\pi(i,j) = P(\pi(i), i, j)$$

$$(I - \alpha P)\, y = c_\pi ; \quad y = (I - \alpha P_\pi)^{-1} c_\pi$$

$$(I-\alpha P)y = c_\pi; \quad y = (I-\alpha P_\pi)^{-1} c_\pi$$

$$(I-\alpha P_\pi)^{-1} = I + \alpha P_\pi + \alpha^2 P_\pi^2 + \alpha^3 P_\pi^3 + \cdots$$

$$\geq 0$$

Optimality Criterion

$y_i^{(n)}$ = min expected cost over n rounds

$$= \min_{x \in D_i} \left[c(x,i) + \alpha \sum_{j \in V} P[x,i,j] \right] y_j^{(n-1)}$$

$\boxed{n = \infty}$

$$y_i = \min_{x \in D_i} \left[c(x,i) + \alpha \sum_{j \in V} P[x,i,j] \right] y_j \quad **$$

Suppose (**) hold and $\hat{\pi}$ is any other policy.

$$\hat{y}_i = c(\hat{\pi}(i), i) + \alpha \sum_{j \in V} P[\hat{\pi}(i), i, j]\, \hat{y}_j$$

$$y_i \leq c(\hat{\pi}(i), i) + \alpha \sum_{j \in V} P[\hat{\pi}(i), i, j]\, y_j$$

$$\hat{y}_i - y_i \geq \alpha \sum_{j \in V} P[\hat{\pi}(i), i, j]\, (\hat{y}_j - y_j)$$

$$\geq 0$$

$$(I - \alpha P_{\hat{\pi}})(\hat{y} - y) \geq 0$$

$$\underline{(I - \alpha P_{\hat{\pi}})^{-1}} (I - \alpha P_{\hat{\pi}})(\hat{y} - y) \geq 0 \quad \text{or} \quad \underline{\hat{y} - y \geq 0}$$