2.4 Problem Formulation and Examples

Example ① Reaching a goal in minimum expected time.

$\alpha = 1, \quad g(i,u,j) = 1, \quad \forall i \neq 0, j, u \in U(i).$

Here, the optimal cost-to-go $J^*(i)$ is the min. expected time to termination starting from i.

$$J^*(i) = \min_{u \in U(i)} \left[1 + \sum_{j=1}^{n} p_{ij}(u) J^*(j) \right], \quad i = 1, 2, \cdots n$$

Example ② Problems with uncontrollable state components.

In many shortest path problems the evolution of a state (i, y) may not be affected by control u.

For the next state (j, z),

j is generated according $p_{ij}(u, y)$.

z is generated according to conditional probabilities $p(z|j)$ that depend on j.

Consider the cost of transition $g(i, y, u, j)$ and does not depend on uncontrollable component z,

If g depends on z, $\hat{g}(i,y,u,j) = \sum_z p(z|j) g(i,y,u,j,z)$.

Such algorithms provide a smaller state space. To see this,

$$(TJ)(i,y) = \min_{u \in U(i)} \sum_{j=0}^{n} p_{ij}(u,y) \left(g(i,y,u,j) + \sum_z p(z|j) J(j,z) \right).$$

$$(T_\mu J)(i,y) = \sum_{j=0}^{n} p_{ij}(\mu(i,y),y)\Big(g(i,y,\mu(i,y),j) + \sum_z p(z|j)J(j,z)\Big)$$

for each J,

$$\hat{J}(j) = \sum_z p(z|j)\, J(j,z).$$

To compute the optimal policy, it is sufficient to know $\hat{J}^*(j)$,

$$\underbrace{\hat{J}^*(j)}_{\text{reduced optimal cost-to-go vector}} = \sum_z p(z|j)\, J^*(j,z).$$

This leads to simplified versions,

$$(\hat{T}\hat{J})(i) = \sum_y p(y|i)\,(TJ)(i,y) = \sum_y p(y|i)\Big(\min_{u\in U(i,y)} \sum_{j=0}^{n} p_{ij}(u,y)\big(g(i,y,u,j) + \hat{J}(j)\big)\Big)$$

and $(\hat{T}_\mu \hat{J})(i) = \sum_y p(y|i)(T_\mu J)(i,y) = \sum_y p(y|i)\Big(\sum_{j=0}^{n} p_{ij}(\mu(i,y),y)\big(g(i,y,\mu(i,y),j) + \hat{J}(j)\big)\Big).$

(a) Policy Evaluation - Compute unique $\hat{J}^{\mu_k}(i)$ with $\hat{J}^{\mu_k} = \hat{T}_{\mu_k}\hat{J}^{\mu_k}$

$$\hat{J}^{\mu_k}(i) = \sum_y p(y|i)\Big(\sum_{j=0}^{n} p_{ij}(\mu_k(i,y))\big(g(i,y,\mu_k(i,y),j) + \hat{J}^{\mu_k}(j)\big)\Big).$$

(b) Policy Improvement - Compute $\mu_{k+1}(i,y)$ from $\hat{T}_{\mu_{k+1}}\hat{J}^{\mu_k} = \hat{T}\hat{J}^{\mu_k}$,

$$\mu_{k+1}(i,y) = \arg\min_{u\in U(i,y)} \sum_{j=0}^{n} p_{ij}(u,y)\big(g(i,y,u,j) + \hat{J}^{\mu_k}(j)\big).$$

x ——————— x ——————— x