Lecture 14

## * Infinite Horizon Average Cost Problem

Recall that in this problem we will aim to minimize

$$J_\pi(i) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{k=0}^{N-1} g(x_k, \mu_k(x_k)) \,\middle|\, x_0 = i\right]$$

We will first argue that for most problems of interest the average cost per stage of a policy and the optimal average cost per stage is independent of the initial state.

Consider a stationary policy $\mu$ and two states $i$ and $j$ such that under $\mu$, the system will eventually reach $j$ from $i$ with prob 1.

Let $K_{ij}(\mu)$ be the first passage time from $i$ to $j$ under $\mu$.

So, now

$$J_\mu(i) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{k=0}^{K_{ij}(\mu)-1} g(x_k, \mu(x_k))\right]$$
$$+ \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\sum_{k=K_{ij}(\mu)}^{N-1} g(x_k, \mu(x_k))\right]$$

If $\mathbb{E}[K_{ij}(\mu)] < \infty$, then by walds lemma

$$J_\mu(i) = 0 + J_\mu(j)$$

As $i, j \in S$ are arbitrary, $J_\mu(i) = J_\mu(j) \quad \forall i, j \in S$
with $\mathbb{E}[K_{ij}(\mu)] < \infty$

If there exists a stationary policy $\mu$, under which $j$ can be reached from $i$ with probability 1 then it is not possible that

$$J^*(i) > J^*(j)$$

Since when starting from $i$ we can play $\mu$ and switch to the optimal policy after reaching $j$.

Since when starting from $i$ we can play $\mu$ and switch to the optimal policy after reaching $j$.

So, $$J^*(i) = J^*(j) \quad \forall\, i,j \in S.$$

## * Associated Stochastic Shortest Path Problem

· We will associate a SSPP to the Average Cost per stage problem and then we will use the results of SSPP.

Assumption 1 : One of the states, by convention state $n$, is such that for some integer $m > 0$, and for all initial state and all policies, $n$ is visited with positive probability at least within the first $m$ stages.

* Assumption 1 will make an important connection with SSPP.

- Consider a sequence of generated states. Divide the seq by successive visits to state $n$.
  - Cycle 1 is includes transition from initial state to first visit to state $n$.
  - Cycle $k \geq 2$ includes transitions from $(k-1)^{th}$ to the $k^{th}$ visit to state $n$.

So each cycle can be viewed as trajectory of SSPP with termination state being $n$.

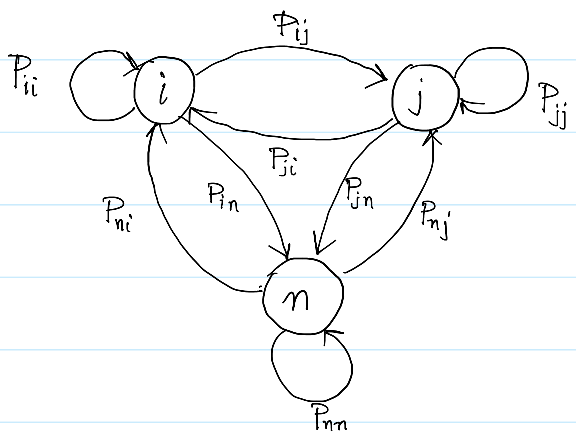
- Precisely, the problem is obtained by two changes:

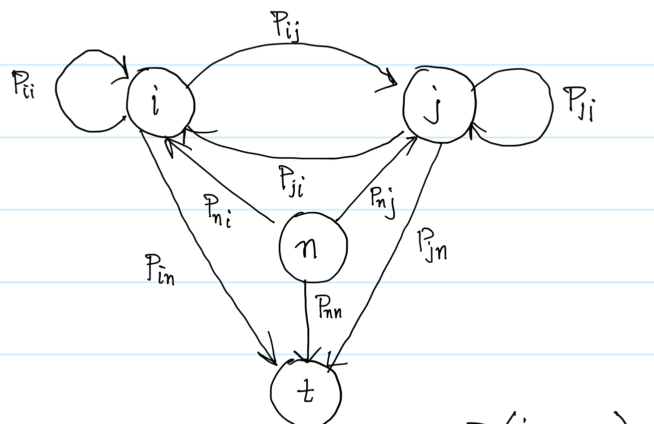
1. Change in transition probabilities:

   i. $p(i,u,j)$ are unchanged $\forall\, j \neq n$

   ii. $P(i,u,n) = 0$

   iii. Introduce new artificial termination state $t$ and $P(i,u,t) = P(i,u,n)$




$P_{ij} := P(i, u, j)$

2. <u>Change in cost per stage</u>: Now we will argue that if we fix the expected stage cost incurred at state $i$ to be $g(i,u) - \lambda^*$

where $\lambda^*$ be the optimal average cost per stage, the associated SSPP becomes equivalent to the original average cost per stage problem.

It can be shown that under a stationary policy $\mu$, average cost per stage

$$\lambda_\mu = \frac{C_{nn}(\mu)}{N_{nn}(\mu)}$$

Where for fixed $\mu$

$C_{nn}(\mu)$ : expected cost starting from $n$ up to first return to $n$

$N_{nn}(\mu)$ : expected number of stages to return to $n$ starting from $n$.

Since $\lambda^* \leq \lambda_\mu$, $\quad C_{nn}(\mu) - N_{nn}(\mu)\lambda^* \geq 0$.
with equality if $\mu$ is optimal.

* $C_{nn}(\mu) - N_{nn}(\mu)\lambda^*$ is the total expected cost of the assoc. SSPP with stage cost $g(i,u) - \lambda^*$.

Let $h^*(i)$ be the optimal cost of this SSPP when starting from state $i \in [n]$. Then the Bellman' equation is from state $i \in [n]$. Then the Bellman' equation is

$$h^*(i) = \min_{u \in U(i)} \left\{ g(i,u) - \lambda^* + \sum_{j=1}^{n-1} p(i,u,j)\, h^*(j) \right\}, \quad i \in [n]$$

If $\mu^*$ is an optimal policy then

$$C_{nn}(\mu^*) - N_{nn}(\mu^*)\, \lambda^* = 0$$

and
$$h^*(n) = C_{nn}(\mu^*) - N_{nn}(\mu^*)\, \lambda^* = 0$$

By this equation we can write Bellmans' equation

$$\lambda^* + h^*(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p(i,u,j) \cdot h^*(j) \right\}, \quad i \in [n]$$

with $\quad h^*(n) = 0.$

* **Proposition 1**: Under assumption 1 the following hold for the average cost per stage problem:

a. The optimal average cost $\lambda^*$ is the same for all initial states and together with some vector $h^* = [h^*(1), \dots, h^*(n)]$ satisfies Bellmans' equation

$$\lambda^* + h^*(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} p(i,u,j)\, h^*(j) \right\}, \quad i \in [n] \qquad (1)$$

Furthermore, if $\mu(i)$ attains the minimum in the above equation for all $i$, the stationary policy $\mu$ is optimal. In addition, out of all vectors $h^*$ satisfying (1), there is a unique vector with $h^*(n) = 0$.

b. If a scalar $\lambda$ and a vector $h = [h(1) \dots h(n)]^T$ satisfy (1), then $\lambda$ is the average optimal cost per stage for each initial state.

c. Given a stationary policy $\mu$ with corresponding average cost per stage $\lambda_\mu$, there is a unique vector $h_\mu = [h_\mu(1), \dots, h_\mu(n)]^T$ such that $h_\mu(n) = 0$ and

$$\lambda_\mu + h_\mu(i) = g(i,\mu(i)) + \sum^{n} p(i,\mu(i),j)\, h_\mu(j) \qquad i \in [n]$$

that $h_\mu(n) = 0$ and

$$\lambda_\mu + h_\mu(i) = g(i,\mu(i)) + \sum_{j=1}^{n} p(i,\mu(i),j)\, h_\mu(j) \;, \quad j \in [n] \qquad (2)$$