* **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 Bellman's equation

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

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

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

**Proof:**

a. Let us denote

$$\bar{\lambda} := \min_{\mu} \frac{C_{nn}(\mu)}{N_{nn}(\mu)} \qquad \text{— } (i)$$

Note that $C_{nn}(\mu)$ and $N_{nn}(\mu)$ are finite $\forall\, \mu$ by assumption 1. Also we have that

$$C_{nn}(\mu) - N_{nn}(\mu)\bar{\lambda} \geq 0 \qquad \text{— } (ii)$$

with equality if $\mu$ attains minimum in (i). By a previous proposition, the costs $h^*(1), \dots, h^*(n)$ solve uniquely corresponding with equality if $\mu$ attains minimum in (i). By a previous proposition, the costs $h^*(1), \dots, h^*(n)$ solve uniquely corresponding Bellman's equation

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

Since the transition probability $p(i,u,n) = 0 \;\forall\, i, u$ in the assoc. SSPP. An optimal stationary policy should minimize

$$C_{nn}(\mu) - N_{nn}(\mu)\bar{\lambda}$$

to zero, so we have that $h^*(n) = 0$.

Thus,

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

We will show that this relation implies $\bar{\lambda} = \lambda^*$.

**Let** $\pi = \{\mu_0, \mu_1, \dots\}$ be any admissible policy, let $N$ be a positive integer, and for all $k = 0, \dots, N-1$, define $J_k(i)$ using the following recursion

$$J_0(i) = h^*(i), \quad i \in [n]$$

$$J_{k+1}(i) = g(i, \mu_{N-k-1}(i)) + \sum_{j=1}^{n} p(i, \mu_{N-k-1}(i), j) \cdot J_k(j), \quad i \in [n]$$

Note that $J_N(i)$ is the $N$-stage cost of $\pi$ when the starting state is $i$ and the terminal cost function is $h^*$. By (iv)

$$\bar{\lambda} + J_0(i) \leq J_1(i), \quad i \in [n]$$

Using this relation, we have

$$g(i, \mu_{N-2}(i)) + \bar{\lambda} + \sum_{j=1}^{n} p(i, \mu_{N-2}(i), j)\, J_0(i)$$
$$\leq g(i, \mu_{N-2}(i)) + \sum_{j=1}^{n} p(i, \mu_{N-2}(i), j)\, J_1(j)$$

and as $J_0(i) = h^*(i)$

$$2\bar{\lambda} + h^*(i) \leq J_2(i)$$

Repeating this argument we obtain

$$k\bar{\lambda} + h^*(i) \leq J_k(i), \quad k = 0, \dots, N, \; i \in [n]$$

Repeating this argument we obtain

$$k\bar{\lambda} + h^*(i) \leq J_k(i) \quad , \quad k = 0, \dots, N, \ i \in [n]$$

and in particular for $k = N$,

$$\bar{\lambda} + \frac{h^*(i)}{N} \leq \frac{J_N(i)}{N}$$

Taking limit as $N \to \infty$,

$$\bar{\lambda} \leq J_\pi(i)$$

for all admissible policy $\pi$, with equality if $\pi$ is a stationary policy $\mu$ such that $\mu(i)$ attains minimum in (iv) $\forall i, k$.

It follows that

$$\bar{\lambda} = \min_\pi J_\pi(i) = \lambda^* \quad , \quad i \in [n]$$

Equation (iv) with $h^*(n) = 0$ is equivalent to equation (iii). Since the solution of (iii) is unique, the same is true for (iv) with $h^*(n) = 0$.

* b and c are home work! "▨"

## * Value Iteration :

The natural one : Initialize iterate $J_0$ arbitrarily. For $k = 0, 1, 2, \dots$ Compute $J_{k+1}$ iteratively as follows :

$$J_{k+1}(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} P(i,u,j)\, J_k(j) \right\} , \quad i \in [n]$$

It is natural to expect that $\frac{J_k(i)}{k}$ should converge to $\lambda^*$ as $k \to \infty$, i.e.,

$$\lim_{k \to \infty} \frac{J_k(i)}{k} = \lambda^* \quad \forall i.$$

To show this, let us define the recursion

$$J^*_{k+1}(i) = \min_{u \in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} P(i,u,j)\, J^*_k(j) \right\} , \quad i \in [n]$$

with $J^*_0(i) = h^*(i)$, $\forall i \in [n]$.

By induction it is easy to see that

$$J_k^*(i) = k\lambda^* + h^*(i) \quad , \quad \forall i \in [n]$$

It can also be shown that

$$|J_k(i) - J_k^*(i)| \leq \max_{j\in[n]} |J_0(j) - h^*(i)| \ , \quad i \in [n].$$

So, $|J_k(i) - k\lambda^*| \leq \max_{j\in[n]} |J_0(j) + h^*(j)| + \max_{j\in[n]} |h^*(j)|$ , $i\in[n]$

So, $\frac{J_k(i)}{k}$ converges to $\lambda^*$

* But see that $\lambda^*$ is finite hence as $k\to\infty$, $J_k(i)\to\infty$ / $-\infty$ $\forall i\in[n]$.
So there is a computational problem!
* Also we have no information about the differential cost vector $h^*$.

## * Relative Value Iteration :

– To overcome the problems in previous section we subtract a constant (cleverly) to get $h^*$ as well as $\lambda^*$.

Consider the algorithm :

$$h_k(i) = J_k(i) - J_k(s) \ , \quad i = 1,\dots,n \ , \ s\in[n] \text{ be fixed.}$$

Then

$$\begin{aligned}
h_{k+1}(i) &= J_{k+1}(i) - J_{k+1}(s) \\
&= \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} P(i,u,j)\, J_k(j) \right\} \\
&\quad - \min_{u\in U(s)} \left\{ g(s,u) + \sum_{j=1}^{n} P(s,u,j)\, J_k(j) \right\} \\
&= \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} P(i,u,j) \big(J_k(j) - J_k(s)\big) \right\} \\
&\quad - \min_{u\in U(s)} \left\{ g(s,u) + \sum_{j=1}^{n} P(s,u,j) \big(J_k(j) - J_k(s)\big) \right\} \\
&= \min_{u\in U(i)} \left\{ g(i,u) + \sum_{j=1}^{n} P(i,u,j)\, h_k(j) \right\} \\
&\quad - \min_{u\in U(s)} \left\{ g(s,u) + \sum_{j=1}^{n} P(s,u,j)\, h_k(j) \right\}
\end{aligned}$$

It can be seen that if relative value iteration converges to some vector $h$, then we have to some vector $h$, then we have

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

with $h(s) = 0$ and

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