Two Optimal Stopping Problems

Frank Massey

In these notes we look at two optimal stopping problems. Since optimal stopping problems are a special case of Markov Decision Problems, we review some basic properties of these first.

1. Markov Decision Problems

In this first section we look at Markov decision problems involving discrete time Markov processes with a finite number of states. At each time one can choose an action from a finite set of actions. Depending on the state and action a cost is incurred for that time. Then a transition is made to another state where the transition probabilities also depend on the state and action. The goal is to minimize the expected total cost where the total cost is the sum of the costs for all times. We begin with a description of the general situation. In the following

t = discrete time. $t = 0, 1, 2, \ldots$

Definition 1. A *Markov decision problem* consists of the following.

(1) $S = \{1, 2, \ldots, n\}$ = a finite set of states.

(2) A = a finite set of *actions.*

(3) For each state i and action a we are given

$c_i(a)$ = cost incurred at any particular time when in state i and action a is taken.

We assume $c_i(a) \geq 0$.

(4) For each state i and action a we are given

$p_{ij}(a)$ = probability next state is j if current state is i and action a is taken.

One has $p_{ij}(a) \geq 0$ and $\sum_{j=1}^{n} p_{ij}(a) = 1$.

We consider two types of *policies* (or strategies). In the first type the action depends only on the current state.

Definition 2. A *stationary policy f* is the choice of an action for each state, i.e.

$$f = \begin{pmatrix} f_1 \\ . \\ . \\ . \\ f_n \end{pmatrix}$$

where f_i is in A for each i.

f_i is the action one uses when the system is in state i at a particular time.

Given a stationary policy, f, the associated transition matrix and cost vector are as follows.

$$P(f) = \begin{pmatrix} p_{11}(f_1) & \cdots & p_{1n}(f_1) \\ \cdots & & \\ \cdots & & \\ p_{n1}(f_n) & \cdots & p_{nn}(f_n) \end{pmatrix}$$

$p_{i,j}(f_i)$ = probability next state is j if current state is i since action f_i is taken

$= Pr\{X_{t+1} = j \mid X_t = i\}$

where X_t, $t = 0, 1, 2, \ldots$ is the Markov chain with transition matrix $P(f)$.

$$c(f) = \begin{pmatrix} c_1(f_1) \\ \cdot \\ \cdot \\ \cdot \\ c_n(f_n) \end{pmatrix}$$

$c_i(f_i)$ = cost incurred when in state i at any time since action f_i is taken

$P(f)^t$ = transition matrix for t time steps

$[P(f)^t]_{ij}$ = probability state at time t is j if initial state is i

$= Pr\{X_t = j \mid X_0 = i\}$

$P(f)^t c(f)$ = vector of expected costs incurred at time t; the entries correspond to the various starting states

$[P(f)^t c(f)]_i$ = expected cost incurred at time t if initial state is i

$= E\{c_{X_t}(f_{X_t}) \mid X_0 = i\}$

$v(f) = \sum_{t=0}^{\infty} P(f)^t c(f)$ = vector of expected total costs; the entries correspond to the various starting states (some or all of the components of $v(f)$ may be infinite.

$v_i(f) = \sum_{t=0}^{\infty} [P(f)^t c(f)]_i$ = expected total cost when starting in state i (may be infinite)

$= E\{\sum_{t=0}^{\infty} c_{X_t}(f_{X_t}) \mid X_0 = i\}$

The main problem in Markov decision theory is to find a stationary policy that minimizes $v_i(f)$ where f varies over all stationary policies. Below (Proposition 7) we shall prove the following proposition.

Proposition 1. There is a stationary policy f^* such that $v_i(f^*) \le v_i(f)$ for all states i and stationary policies f.

A policy f^* satisfying the conditions of Proposition 1 is called an *optimal policy*. We shall also give (Proposition 8 and 9) an algorithm for finding f^*.

Since the terms in the sums defining $v(f)$ and $v_i(f)$ are non-negative we can interchange the order of summations. In particular, $v(f)$ satisfies

(5) $$v(f) = c(f) + P(f)v(f)$$

An important case is when the components of $v(f)$ are finite. The following proposition gives a condition for this to be true.

Proposition 2. The following three conditions are equivalent.

(6) $v_i(f) < \infty$ for all states i.

(7) $c_i(f_i) = 0$ for each state i that is recurrent for $P(f)$

(8) $v_i(f) = 0$ for each state i that is recurrent for $P(f)$

If any, and hence all, of these conditions are true, then

(9) $$v_T(f) = \sum_{t=0}^{\infty} [P_T(f)^t c_T(f)] = (I - P_T(f))^{-1} c_T(f)$$

where

$P_T(f)$, $c_T(f)$ and $v_T(f)$ = parts of $P(f)$, $c(f)$ and $v_T(f)$ corresponding to transient states

Proof. One has

(10) $$v_i(f) = \sum_{t=0}^{\infty} \sum_{j=1}^{n} [P(f)^t]_{ij} c_j(f_i) = \sum_{j=1}^{n} [\sum_{t=0}^{\infty} [P(f)^t]_{ij}] c_j(f_i)$$

$$= \sum_{j \text{ transient}} [\sum_{t=0}^{\infty} [P(f)^t]_{ij}] c_j(f_i) + \sum_{j \text{ recurrent}} [\sum_{t=0}^{\infty} [P(f)^t]_{ij}] c_j(f_i)$$

It is known that if j is transient then $\sum_{t=0}^{\infty} [P(f)^t]_{ij} < \infty$ and if j is recurrent and on can go from i to j then

$\sum_{t=0}^{\infty} [P(f)^t]_{ij} = \infty$. From this it follows that (6) and (7) are equivalent. Since $c_i(f) \le v_i(f)$ the condition (7) follows from (8). However, if (6) and (7) hold and i is recurrent, then it follows from (10) and the fact that one can not go from a recurrent state to a transient state that

$$v_i(f) = \sum_{j \text{ recurrent}} [\sum_{t=0}^{\infty} [P(f)^t]_{ij}] c_j(f_i) = 0$$

so (8) is true. If (6) – (8) hold, then (5) implies $v_T(f) = c_T(f) + P_T(f)v_T(f)$ and hence (9) holds. //

In order to prove Proposition 1 it is convenient to consider more general policies where the choice of the action may depend on the entire past history.

Definition 3. A policy π has the form

$$\pi = \{\pi_0, \pi_1, \ldots, \pi_t, \ldots\} = \text{a } \textit{policy}$$

where

$$\pi_t = \pi_t(i_0, i_1, \ldots, i_t) \text{ is in } A \text{ for each sequence } i_0, i_1, \ldots, i_t \text{ of states}$$

Note,

$$\pi_t(i_0, i_1, \ldots, i_t) = \text{the action taken if state at time } s \text{ is } i_s \text{ for } s = 0, 1, \ldots, t$$

If we use the policy π, then

$c_i(\pi_0(i))$ = cost at time 0 if the initial state is i

$p_{ij}(\pi_0(i))$ = probability state is j at time 1 if the initial state is i

$= Pr\{X_1 = j \mid X_0 = i\}$

$c_j(\pi_1(i,j))$ = cost at time 1 if the state is i at time 0 and j at time 1

$\sum_{j=1}^{n} p_{ij}(\pi_0(i))c_j(\pi_1(i,j))$ = expected cost at time 1 if the initial state is i

$= E\{c_{X_1}(\pi_1(i,X_1)) \mid X_0 = i\}$

$p_{ij}(\pi_0(i))\, p_{j,j_2}(\pi_1(i,j))$ = probability state is j at time 1 and j_2 at time 2 if the state is i at time 0

$= Pr\{X_1 = j, X_2 = j_2 \mid X_0 = i\}$

$c_{j_2}(\pi_2(i,j,j_2))$ = cost at time 2 if the state is i at time 0 and j at time 1 and j_2 at time 2

$\sum_{j=1}^{n} \sum_{j_2=1}^{n} p_{ij}(\pi_0(i))\, p_{j,j_2}(\pi_1(i,j))\, c_{j_2}(\pi_2(i,j,j_2))$ = expected cost at time 2 if the initial state is i

$= E\{c_{X_2}(\pi_2(i,X_1,X_2)) \mid X_0 = i\}$

$p_{ij}(\pi_0(i))\, p_{j,j_2}(\pi_1(i,j))\, p_{j_2j_3}(\pi_2(i,j,j_2))$ = probability state is j at time 1 and j_2 at time 2 and j_3 at time 3 if the initial state is i

$= Pr\{X_1 = j, X_2 = j_2, X_3 = j_3 \mid X_0 = i\}$

$c_{j_3}(\pi_3(i,j,j_2,j_3))$ = cost at time 3 if the state is i at time 0 and j at time 1 and j_2 at time 2 and j_3 at time 3

$$\sum_{j=1}^{n}\sum_{j_2=1}^{n}\sum_{j_3=1}^{n} p_{ij}(\pi_0(i))\, p_{jj_2}(\pi_1(i,j))\, p_{j_2j_3}(\pi_2(i,j,j_2))\, c_{j_3}(\pi_3(i,j,j_2,j_3))$$

$$= \text{ expected cost at time 3 if the initial state is } i$$

$$= E\{c_{X_3}(\pi_3(i,X_1,X_2,X_3)) \mid X_0 = i\}$$

$$p_{ij}(\pi_0(i))\, p_{jj_2}(\pi_1(i,j))\, p_{j_2j_3}(\pi_1(i,j,j_2))\cdots p_{j_{t-1}j_t}(\pi_{t-1}(i,j,j_2,\ldots,j_{t-1}))$$

$$= \text{ probability state is } j \text{ at time 1 and } j_s \text{ at time } s \text{ for } s = 2, \ldots, t \text{ if the initial state is } i$$

$$= Pr\{X_1 = j, X_2 = j_2, X_3 = j_3, \ldots, X_t = j_t \mid X_0 = i\}$$

$c_{j_t}(\pi_t(i,j,j_2,\ldots,j_t))$ = cost at time t if the state is i at time 0 and j at time 1 and j_s at time s for $s = 2, \ldots, t$

$$\sum_{j=1}^{n}\sum_{j_2=1}^{n}\cdots\sum_{j_t=1}^{n} p_{ij}(\pi_0(i))\, p_{jj_2}(\pi_1(i,j))\, p_{j_2j_3}(\pi_1(i,j,j_2))\cdots p_{j_{t-1}j_t}(\pi_{t-1}(i,j,j_2,\ldots,j_{t-1}))\, c_{j_t}(\pi_t(i,j,j_2,\ldots,j_t))$$

$$= \text{expected cost at time } t \text{ if the initial state is } i$$

$$= E\{c_{X_t}(\pi_t(i,X_1,X_2,X_3,\ldots,X_t)) \mid X_0 = i\}$$

$v_i(\pi)$ = expected total cost if the initial state is i and policy π is used

$$= c_i(\pi_0(i)) + \sum_{j=1}^{n} p_{ij}(\pi_0(i))c_j(\pi_1(i,j)) + \sum_{j=1}^{n}\sum_{j_2=1}^{n} p_{ij}(\pi_0(i))\, p_{jj_2}(\pi_1(i,j))\, c_{j_2}(\pi_2(i,j,j_2))$$

$$+ \sum_{j=1}^{n}\sum_{j_2=1}^{n}\sum_{j_3=1}^{n} p_{ij}(\pi_0(i))\, p_{jj_2}(\pi_1(i,j))\, p_{j_2j_3}(\pi_2(i,j,j_2))\, c_{j_3}(\pi_3(i,j,j_2,j_3)) + \cdots$$

$$+ \sum_{j=1}^{n}\sum_{j_2=1}^{n}\cdots\sum_{j_t=1}^{n} p_{ij}(\pi_0(i))\, p_{jj_2}(\pi_1(i,j))\, p_{j_2j_3}(\pi_2(i,j,j_2))\cdots p_{j_{t-1}j_t}(\pi_{t-1}(i,j,j_2,\ldots,j_{t-1}))\, c_{j_t}(\pi_t(i,j,j_2,\ldots,j_t))$$

$$+ \cdots$$

$$= E\{c_i(\pi_0(i)) + \sum_{t=1}^{\infty} c_{X_t}(\pi_t(i,X_1,X_2,X_3,\ldots,X_t)) \mid X_0 = i\}$$

We are interested in the minimum of $v_i(\pi)$ over all policies. Until we prove there is a minimum, let

$$v_i = \inf_{\pi} v_i(\pi)$$

= greatest lower bound over all policies of the expected total cost if the initial state is i

Proposition 3. Let $i \in \{1, 2, \ldots, n\}$. Let $f_i = \pi_0(i)$ and σ be the policy defined by $\sigma_0(j) = \pi_1(i,j)$ and $\sigma_1(j,j_2) = \pi_2(i,j,j_2)$ and $\sigma_2(j,j_2,j_3) = \pi_3(i,j,j_2,j_3)$ and $\sigma_t(j,j_2,\ldots,j_t,j_{t+1}) = \pi_{t+1}(i,j,j_2,\ldots,j_t,j_{t+1})$. Then

$$v_i(\pi) \;=\; c_i(f_i) \;+\; \sum_{j=1}^{n} p_{ij}(f_i)\, v_j(\sigma)$$

Proof. From the above formula for $v_i(\pi)$ one has

$$\begin{aligned}
v_i(\pi) \;=\; & c_i(\pi_0(i)) \;+\; \sum_{j=1}^{n} p_{ij}(\pi_0(i)) \Big[c_j(\pi_1(i,j)) \;+\; \sum_{j_2=1}^{n} p_{j,j_2}(\pi_1(i,j))\, c_{j2}(\pi_2(i,j,j_2)) \\
& + \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} p_{j,j_2}(\pi_1(i,j))\, p_{j_2 j_3}(\pi_2(i,j,j_2))\, c_{j_3}(\pi_3(i,j,j_2,j_3)) \;+\; \cdots \\
& + \sum_{j_2=1}^{n} \cdots \sum_{j_t=1}^{n} p_{j,j_2}(\pi_1(i,j))\, p_{j_2 j_3}(\pi_2(i,j,j_2)) \cdots p_{j_{t-1} j_t}(\pi_{t-1}(i,j,j_2,\ldots,j_{t-1}))\, c_{j_t}(\pi_t(i,j,j_2,\ldots,j_t)) \;+\; \cdots \Big] \\
= \; & c_i(f_i) \;+\; \sum_{j=1}^{n} p_{ij}(f_i) \Big[c_j(\sigma_0(j)) \;+\; \sum_{j_2=1}^{n} p_{j,j_2}(\sigma_0(j))\, c_{j_2}(\sigma_1(j,j_2)) \\
& + \sum_{j_2=1}^{n} \sum_{j_3=1}^{n} p_{j,j_2}(\sigma_0(j))\, p_{j_2 j_3}(\sigma_1(j,j_2))\, c_{j_3}(\sigma_2(j,j_2,j_3)) \;+\; \cdots \\
& + \sum_{j_2=1}^{n} \cdots \sum_{j_t=1}^{n} p_{j,j_2}(\sigma_0(j))\, p_{j_2 j_3}(\sigma_1(j,j_2)) \cdots p_{j_{t-1} j_t}(\sigma_{t-2}(j,j_2,\ldots,j_{t-1}))\, c_{j_t}(\sigma_{t-1}(j,j_2,\ldots,j_t)) \;+\; \cdots \Big] \\
= \; & c_i(f_i) \;+\; \sum_{j=1}^{n} p_{ij}(f_i)\, v_j(\sigma) \qquad\qquad //
\end{aligned}$$

Corollary 4. Let $i \in \{1, 2, \ldots, n\}$ and $f_i = \pi_0(i)$. Then $v_i(\pi) \;\geq\; c_i(f_i) \;+\; \sum_{j=1}^{n} p_{ij}(f_i)\, v_j$

Proof. This follows from Proposition 3 and the fact that $v_j = \inf_{\sigma} v_i(\sigma)$.

Corollary 5. Let $i \in \{1, 2, \ldots, n\}$. Then $v_i(\pi) \;\geq\; \min_{f \in A} \{\, c_i(f) \;+\; \sum_{j=1}^{n} p_{ij}(f)\, v_j \,\}$

Proof. This follows from Corollary 4.

Corollary 6. Let $i \in \{1, 2, \ldots, n\}$. Then $v_i \;\geq\; \min_{f \in A} \{\, c_i(f) \;+\; \sum_{j=1}^{n} p_{ij}(f)\, v_j \,\}$

Proof. This follows from Corollary 5 and the fact that $v_i = \inf_{\pi} v_i(\pi)$.

The v_i satisfy the equation (11) in the next proposition which is a type of dynamic programming equation.

Proposition 7. For each i one has

$$v_i = \min_{f \in A} \{ c_i(f) + \sum_{j=1}^{n} p_{ij}(f)\, v_j \} \tag{11}$$

Proof. By Corollary 5 one has $v_i \geq \min_{f \in A} \{ c_i(f) + \sum_{j=1}^{n} p_{ij}(f)\, v_j \}$, so we need to show $v_i \leq \min_{f \in A} \{ c_i(f) + \sum_{j=1}^{n} p_{ij}(f)\, v_j \}$. Let f_i be such that $c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j = \min_{f \in A} \{ c_i(f) + \sum_{j=1}^{n} p_{ij}(f)\, v_j \}$. We need to show $v_i \leq c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j$. It suffices to show $v_i \leq c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j + \varepsilon$ for each $\varepsilon > 0$. To this end, let $\varepsilon > 0$. For each j choose a policy $\sigma^{(j)}$ such that $v_j(\sigma^{(j)}) \leq v_j + \varepsilon$. Only the part of ${\sigma^{(j)}}_t(i_0, i_1, \ldots, i_t)$ for $i_0 = j$ enters into $v_j(\sigma^{(j)}) \leq v_j + \varepsilon$. So we can define a policy σ such that $\sigma_t(j, i_1, \ldots, i_t) = {\sigma^{(j)}}_t(j, i_1, \ldots, i_t)$ for each j and it will have the property that $v_j(\sigma) \leq v_j + \varepsilon$ for each j. In order to show $v_i \leq c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j + \varepsilon$ it suffices to show $v_i \leq c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j(\sigma)$. Let π be the policy defined by $\pi_0(i) = f_i$ and $\pi_1(i,j) = \sigma_0(j)$ and $\pi_2(i,j,j_2) = \sigma_1(j,j_2)$ and $\pi_3(i,j,j_2,j_3) = \sigma_2(j,j_2,j_3)$ and $\pi_t(i,j,j_2,\ldots,j_t) = \sigma_{t-1}(j,j_2,\ldots,j_t)$. By Proposition 2 one has $v_i(\pi) = c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j(\sigma)$. Therefore $v_i \leq c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j(\sigma)$ which is what we needed to show. //

Now we can prove Proposition 1. In fact, we shall show the following.

Proposition 8. For each i, let f_i be such that

$$c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j = \min_{f \in A} \{ c_i(f) + \sum_{j=1}^{n} p_{ij}(f)\, v_j \}$$

and let $f = \begin{pmatrix} f_1 \\ \cdot \\ \cdot \\ \cdot \\ f_n \end{pmatrix}$ be the stationary policy which uses action f_i when in state i. Then f is optimal, i.e. for each i one has

$$v_i(f) = v_i$$

Proof. We want to show $v(f) = v$ where $v = \begin{pmatrix} v_1 \\ \cdot \\ \cdot \\ \cdot \\ v_n \end{pmatrix}$. Since $v_i = \min_{f \in A} \{ c_i(f) + \sum_{j=1}^{n} p_{ij}(f)\, v_j \}$ for each i, we have $v_i = c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i)\, v_j$ or $v = c(f) + P(f)v$. If we replace v on the right by $c(f) + P(f)v$ we get

$v = c(f) + P(f)c + P(f)^2 v$. Repeating this we get $v = \sum_{s=1}^{t} P(f)^s c(f) + P(f)^{t+1} v$ for each t. Since every component of $P(f)^{t+1} v$ is non-negative we have $\sum_{s=1}^{t} (P(f)^s c(f))_i \le v_i$ for each t and i. Letting $t \to \infty$ we get $\sum_{s=1}^{\infty} (P(f)^s c(f))_i \le v_i$. Since $v_i(f) = \sum_{s=1}^{\infty} (P(f)^s c(f))_i$, we have $v_i(f) \le v_i$ which implies $v_i(f) = v_i$. So f is optimal. //

Proposition 9. Let g be a stationary policy such that for some k

$$\min_{f \in A} \{ c_k(f) + \sum_{j=1}^{n} p_{kj}(f) v_j(g) \} < v_k(g)$$

Let f be a stationary policy such that

$$c_i(f_i) + \sum_{j=1}^{n} p_{ij}(f_i) v_j(g) \le v_i(g) \tag{12}$$

for all i and

$$c_k(f_k) + \sum_{j=1}^{n} p_{kj}(f_k) v_j(g) < v_k(g) \tag{13}$$

(For example, we can take $f_i = g_i$ for $i \neq k$ and f_k be such that (12) holds.) Then $v_i(f) \le v_i(g)$ for all i and $v_k(f) < v_k(g)$.

Proof. Let $\varepsilon = v_k(g) - [c_k(f_k) + \sum_{j=1}^{n} p_{kj}(f_k) v_j]$. The inequality (12) says $c(f) + P(f)v(g) \le v(g)$. Also we have $[c(f) + P(f)v(g)]_k \le v_k(g) - \varepsilon$. Since $P(f)$ is order preserving one has $P(f)[c(f) + P(f)v(g)] \le P(f)v(g)$. Therefore, $c(f) + P(f)c(f) + P(f)^2 v(g) \le c(f) + P(f)v(g) \le v(g)$. Also we have $[c(f) + P(f)c(f) + P(f)^2 v(g)]_k \le [c(f) + P(f)v(g)]_k \le v_k(g) - \varepsilon$. Repeating this, we have for each t

$\sum_{s=1}^{t} P(f)^s c(f) + P(f)^{t+1} v(g) \le v(g)$ and $[\sum_{s=1}^{t} P(f)^s c(f) + P(f)^{t+1} v(g)]_k \le v_k(g)$. Since every component of $P(f)^{t+1} v$ is non-negative we have $\sum_{s=1}^{t} (P(f)^s c(f))_i \le v_i(g)$ for each t and i and $\sum_{s=1}^{t} (P(f)^s c(f))_k \le v_k(g) - \varepsilon$.

Letting $t \to \infty$ we get $\sum_{s=1}^{\infty} (P(f)^s c(f))_i \le v_i(g)$ and $\sum_{s=1}^{\infty} (P(f)^s c(f))_k \le v_i(g) - \varepsilon$. Since $v_i(f) = \sum_{s=1}^{\infty} (P(f)^s c(f))_i$, we have $v_i(f) \le v_i(g)$ for all i and $v_k(f) \le v_k(g) - \varepsilon$. //

Proposition 10. Suppose g is a stationary policy and f is an optimal policy. Suppose for each i one has

$$c_i(g_i) + \sum_{j=1}^{n} p_{ij}(g_i)\, v_j(g) = \min_{h \in A} \{ c_i(h) + \sum_{j=1}^{n} p_{ij}(h)\, v_j(g) \} \tag{14}$$

Furthermore, suppose that

(15) if i is a recurrent state for f then $v_i(g) = 0$.

Then g is also an optimal policy.

Proof. We first show

$$\lim_{t \to \infty} P(f)^t v(g) = 0 \tag{16}$$

One has

$$[P(f)^t v(g)]_i = \sum_{j=1}^{n} [P(f)^t]_{ij} v_j(g)$$

$$= \sum_{j \text{ transient}} [P(f)^t]_{ij}\, v_j(g) + \sum_{j \text{ recurrent}} [P(f)^t]_{ij}\, v_j(g)$$

where transient and recurrent are with respect to $P(f)$. If j is transient then $[P(f)^t]_{ij} \to 0$ as $t \to \infty$ for all . If j is recurrent then by hypothesis $v_j(g) = 0$. So $[P(f)^t v(g)]_i \to 0$ and (16) is true.

To prove the proposition, note that (14) implies $c(g) + P(g)v(g) \le c(f) + P(f)v(g)$. However, one has $v(g) = c(g) + P(g)v(g)$, so $v(g) \le c(f) + P(f)v(g)$. If we replace $v(g)$ by $c(f) + P(f)v(g)$ on the right we get $v(g) \le c(f) + P(f)c(f) + P(f)^2 v(g)$. Repeating this process we get $v(g) \le \sum_{s=1}^{t} P(f)^s c(f) + P(f)^{t+1} v(g)$. Letting $t \to \infty$ and using (16) gives $v(g) \le \sum_{s=1}^{t} P(f)^s c(f)$. However $v(f) = \sum_{s=1}^{t} P(f)^s c(f)$. So $v(g) \le v(f)$ which implies g is optimal. //

Remarks. By Proposition 2

(17) the condition (15) will be true if $c_i(g) = 0$ for each state i that is recurrent for f.

Also by Proposition 2

(18) The condition (15) will be true if $v_i(g) < \infty$ for all i and every state that is recurrent for f is recurrent for g.

The condition (14) alone is not necessary for g to be optimal. Here is a counterexample.

Example. Let $S = \{1, 2\}$, $A = \{1, 2\}$, $P(1) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $P(2) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $c(1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and , $c(2) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then $f = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is optimal since $v(f) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Consider $g = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. One has $P(g) = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}$ and $c(g) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $v(g) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. So g is not optimal. The condition (14) for g becomes

$$v_1(g) \leq c_1(1) + p_{11}(1)v_1(g) + p_{12}(1)v_2(g)$$

$$v_2(g) \leq c_2(2) + p_{21}(2)v_1(g) + p_{22}(2)v_2(g)$$

Substituting values these become

$$1 \leq 0 + (1)(1) + (0)(0)$$

$$0 \leq c_2(2) + p_{21}(2)v_1(g) + p_{22}(2)v_2(g)$$

which are true.