When To Stop Accumulating Reward/Cost

Graduate School of Economics, Kyushu University

Abstract

This paper studies an optimal stopping problem for three reward accumulation processes: terminal process, additive process and minimum process. The terminal process together with its optimal structure is well known. We show through dynamic programming that both additive process and minimum process have an optimal stopping time. The additive process admits the linearity of expectation operator. However, the minimum process does not admit the linearity. We apply an invariant imbedding approach, which expands the original state space by one dimension. A basic idea is a minimal Markovization of non-Markov process.

1 Introduction

In this paper, we consider the optimal stopping problem where the reward accumulation is terminal, additive and minimum. The theory of optimal stopping of terminal process has been studied both by dynamic programming [1] and by Snell’s envelop method [4,15,17]. It is difficult to discriminate between both approaches. The dynamic programming is methodological, and Snell’s envelop is characteristic. In fact, both are equivalent. Here we rather consider dynamic programming approach.

2 Terminal Process

Let \( \{X_n\}_{0}^{N} \) be a Markov chain on a finite state space \( X \) with a transition law \( p = \{p(\cdot|\cdot)\} \). Let \( g_n : X \to R^1 \) be a stop reward for \( 0 \leq n \leq N \). We call \( g = \{g_n\} \) a reward sequence (or stopping-reward sequence). Then a sequence of random rewards \( \{g_n(X_n)\}_{0}^{N} \) is specified. The reward process (or stopping-reward process) \( \{g_n(X_n)\}_{0}^{N} \) is called terminal. When a decision-maker stops the terminal reward process at state \( x_n \) on \( n \)-th stage, he/she will get the reward \( g_n(x_n) \). His/her problem is when to stop it. This is an optimal stopping
Let $X^k := X \times X \times \cdots \times X$ be the direct product of $k$ state spaces $X$. We take $\Omega := X^{N+1}$; the set of all paths $\omega = x_0x_1 \cdots x_N$:

$$\Omega = \{\omega = x_0x_1 \cdots x_N \mid x_n \in X, \ 0 \leq n \leq N\}.$$ 

Let $\mathcal{T}_m^n$ be the set of all subsets in $\Omega$ which are determined by random variables $\{X_m, X_{m+1}, \ldots, X_n\}$, where $X_k : \Omega \rightarrow X$ is the projection, $X_k(\omega) = x_k$. Strictly, $\mathcal{T}_m^n$ is the $\sigma$-field on $\Omega$ generated by the set of all subsets of the form

$$\{X_m = x_m, X_{m+1} = x_{m+1}, \ldots, X_n = x_n\} (\subset \Omega)$$

where $x_m, x_{m+1}, \ldots, x_n$ are all elements in state space $X$. Let us take $N = \{0, 1, \ldots, N\}$.

A mapping $\tau : \Omega \rightarrow N$ is called a stopping time if

$$\{\tau = n\} = \{x_0x_1 \cdots x_N \mid \tau(x_0x_1 \cdots x_N) = n\} \in \mathcal{T}_0^n \ \forall n \in N.$$ 

The stopping time $\tau$ is called $N$-adapted. Let $\mathcal{T}_0^N$ be the set of all such stopping times. Any stopping time $\tau \in \mathcal{T}_0^N$ generates a stopped state (random variable) $X_\tau : \Omega \rightarrow X$:

$$X_\tau(\omega) = X_{\tau(\omega)}(\omega)$$

and a stopped reward (random variable) $g_\tau : \Omega \rightarrow R^1$:

$$g_\tau(\omega) = g_{\tau(\omega)}(X_{\tau(\omega)}).$$

We remark that the expected value $E_{x_0}[g_\tau]$ is expressed by sum of multiple sums:

$$E_{x_0}[g_\tau] = \sum_{n=0}^{N} \sum_{\{\tau=n\}} g_n(x_n)P_{x_0}(X_n = x_n)$$

$$= \sum_{n=0}^{N} \sum_{\{\tau=n\}} g_n(x_n)p(x_1|x_0)p(x_2|x_1) \cdots p(x_n|x_{n-1}).$$

Now we consider the problem of maximizing an expected value of stopped process with terminal criterion [4,14,15,17]:

$$T_0(x_0) \quad \text{Max} \ E_{x_0}[g_\tau] \quad \text{s.t.} \ \tau \in \mathcal{T}_0^N.$$ 

An invariant imbedding approach begins with taking a subprocess which starts at state $x_n (\in X)$ on $n$-th stage and terminates on the final $N$-th stage:

$$T_n(x_n) \quad \text{Max} \ E_{x_n}[g_\tau] \quad \text{s.t.} \ \tau \in \mathcal{T}_n^N$$

where $\mathcal{T}_n^N$ is the set of all stopping times which take values in $\{n, n + 1, \ldots, N\}$. Let $v_n(x_n)$ be the maximum value of $T_n(x_n)$, where

$$v_N(x_N) \triangleq g_N(x_N) \quad x_N \in X.$$ 

Then we have the the backward recursive equation:
Theorem 2.1
\[
\begin{align*}
v_N(x) &= g_N(x), \quad x \in X \\
v_n(x) &= \text{Max} \left[ g_n(x), E_x[v_{n+1}(X_{n+1})] \right], \quad x \in X, \ 0 \leq n \leq N - 1
\end{align*}
\]
where \( E_x \) is the one-step expectation operator induced from the Markov transition matrix \( p(\cdot | \cdot) \):
\[
E_x(h(X_{t+1})) = \sum_{y \in X} h(y) p(y | x).
\]
Proof. The proof is done through an equivalent Markov decision problem in the following section. \( \square \)

2.1 Optimal stopping time

Theorem 2.2 The stopping time \( \tau^* \):
\[
\tau^*(\omega) = \min \{ n : v_n(x_n) = g_n(x_n) \} \quad \omega = x_0 x_1 \cdots x_N \in \Omega
\]
is optimal:
\[
E_{x_0}[g_{\tau^*}] \geq E_{x_0}[g_{\tau}] \quad \forall \tau \in T_0^N.
\]
Let two sequences of functions \( \{ f_n \}_0^N, \{ h_n \}_0^N \) on \( X \) be given. Then the process \( \{ f_n(X_n) \}_0^N \) is said to be supermartingale (resp. martingale, submartingale) if \( f_n(x) \geq \) (resp. =, \( \leq \)) \( T f_{n+1}(x) \) \( x \in X, \ 0 \leq n \leq N - 1 \), where
\[
T f_{n+1}(x) = E_x[f_{n+1}(X_{n+1})].
\]
In this case, we say that the sequence of functions \( \{ f_n \} \) is excessive.

The process \( \{ f_n(X_n) \} \) is said to dominate the process \( \{ h_n(X_n) \} \) if \( f_n(x) \geq h_n(x) \) \( x \in X, \ 0 \leq n \leq N \). We also say that \( \{ f_n \} \) is majorant of \( \{ h_n \} \).

A supermartingale \( \{ f_n(X_n) \} \) which dominates \( \{ h_n(X_n) \} \) is said to be minimal if every supermartingale which dominates \( \{ h_n(X_n) \} \) dominates \( \{ f_n(X_n) \} \). In this case, we say that \( \{ f_n \} \) is the smallest excessive majorant of \( \{ h_n \} \).

Theorem 2.3 (Characterization) The value process \( \{ v_n(X_n) \} \) is the minimal supermartingale which dominates the stopping-reward process \( \{ g_n(X_n) \} \).

It is also said that the value functions \( \{ v_n \} \) is the smallest excessive majorant of \( \{ g_n \} \).

Let a stopping time \( \tau \) and a reward sequence \( f = \{ f_n \} \) be given. Then we define a stopped process \( f^\tau = \{ f^\tau_n \} \) by
\[
f^\tau_n := f_{\tau \wedge n} \quad \text{or} \quad f_{\tau \wedge n}(X_{\tau \wedge n})(\omega) := f_{\tau(\omega) \wedge n}(X_{\tau(\omega) \wedge n}(\omega)) \quad \omega \in \Omega, \ 0 \leq n \leq N.
\]
We note that

\[
f_{n}^{\tau}(\omega) = f_{\tau(\omega) \wedge n}(X_{\tau(\omega) \wedge n}(\omega)) = \begin{cases} 
  f_{n}(X_{n}(\omega)) & \tau(\omega) \geq n \\
  f_{\tau(\omega)}(X_{\tau(\omega)}(\omega)) & \tau(\omega) < n
\end{cases}.
\]

Then \(\{f_{\tau \wedge n}(X_{\tau \wedge n})\}\) is called the stopped process for process \(\{f_{n}(X_{n})\}\) by the stopping time \(\tau\). Thus the stopped process \(\{f_{\tau \wedge n}(X_{\tau \wedge n})\}\) is supermartingale if and only if for each \(n\) \((0 \leq n \leq N - 1)\)

\[
f_{\tau(\omega) \wedge n}(x_{\tau(\omega) \wedge n}) \geq E_{x_{\tau(\omega) \wedge n}}[f_{\tau(\omega) \wedge (n+1)}(X_{\tau(\omega) \wedge (n+1)})] \quad \text{a.e.}
\]

This implies for each \(n\)

\[
\begin{align*}
  & \quad \text{on } \{\tau \geq n + 1\}, \quad f_{n}(x_{n}) \geq E_{x_{n}}[f_{n+1}(X_{n+1})] \\
  & \quad \text{on } \{\tau = n\}, \quad f_{\tau}(x_{\tau}) \geq E_{x_{\tau}}[f_{\tau}(X_{\tau})] \\
  & \quad \text{on } \{\tau \leq n - 1\}, \quad f_{\tau}(x_{\tau}) \geq E_{x_{\tau}}[f_{\tau}(X_{\tau})].
\end{align*}
\]

The latter two inequalities are satisfied with the equality. The supermartingaleness is equivalent to the first inequality.

**Theorem 2.4** (Martingale) The stopped process \(v^{\tau^{*}} = \{v_{\tau \wedge n}(X_{\tau \wedge n})\}\) of process \(\{v_{n}(X_{n})\}\) by the optimal stopping time \(\tau^{*}\) is a martingale.

**Proof.** We see that for each \(n\)

\[
\text{on } \{\tau^{*} \geq n + 1\}, \quad v_{n}(x_{n}) = E_{x_{n}}[v_{n+1}(X_{n+1})].
\]

**Theorem 2.5** (Optimality) A stopping time \(\tau\) is optimal if and only if (i) the stopped rewards are equal: \(v_{\tau} = g_{\tau}\) a.e., and (ii) the stopped process \(v^{\tau} = \{v_{\tau \wedge n}(X_{\tau \wedge n})\}\) is a martingale.

### 3 Additive Process

In this section, we assume, in addition, that \(r_{n} : X \to \mathbb{R}^{1}\) be a continuation reward for \(0 \leq n \leq N - 1\).

We consider the problem of maximizing an expected value of stopped process with additive criterion [5]:

\[
A_{0}(x_{0}) \quad \text{Max} \quad E_{x_{0}}[r_{0} + r_{1} + \cdots + r_{\tau - 1} + g_{\tau}] \\
s.t. \quad \tau \in \mathcal{T}_{0}^{N}.
\]
We note that the expected value of additive reward is the following sum of multiple sums:

\[ E_{x_0}[r_0 + \cdots + r_{\tau-1} + g_{\tau}] = \sum_{n=0}^{N} \sum_{\{\tau=n\rangle} \left[ \sum_{k=0}^{n} r_k(x_k) + g_n(x_n) \right] p(x_1|x_0)p(x_2|x_1) \cdots p(x_n|x_{n-1}). \]

Then we have the corresponding recursive equation:

**Theorem 3.1**

\[
\begin{align*}
    v_N(x) &= g_N(x) \quad x \in X \\
    v_n(x) &= \text{Max} \left[ g_n(x), E_x[r_n(x) + v_{n+1}(X_{n+1})] \right] \quad x \in X, \ 0 \leq n \leq N-1
\end{align*}
\]

(1)

Here we remark that the linearity of expectation operator admits

\[ E_x[r_n(x) + v_{n+1}(X_{n+1})] = r_n(x) + E_x[v_{n+1}(X_{n+1})]. \]

**Theorem 3.2** The stopping time \( \tau^* \):

\[ \tau^*(\omega) = \min\{n : v_n(x_n) = g_n(x_n)\} \quad \omega = x_0x_1\cdots x_N \]

is optimal:

\[ E_{x_0}[r_0 + \cdots + r_{\tau-1} + g_{\tau}] \geq E_{x_0}[r_0 + \cdots + r_{\tau-1} + g_{\tau}] \quad \forall \tau \in \mathcal{T}_0^N. \]

Let two sequences \( \{f_n\}_0^N, \{h_n\}_0^N \) be given. Then the process \( \{f_n(X_n)\}_0^N \) is said to be \( r \)-supermartingale if \( f_n(x) \geq Rf_{n+1}(x) \quad x \in X, \ 0 \leq n \leq N-1 \), where

\[ Rf_{n+1}(x) = E_x[r_n(x) + f_{n+1}(X_{n+1})]. \]

We also say that the sequence \( \{f_n\} \) is \( r \)-excessive.

**Theorem 3.3** (Characterization) The value process \( \{v_n(X_n)\} \) is the minimal \( r \)-supermartingale which dominates the stopping-reward process \( \{g_n(X_n)\} \).

It is also said that the value function \( \{v_n\} \) is the smallest \( r \)-excessive majorant of \( \{g_n\} \).

### 3.1 DP Solution

Let us consider an two-state two-stage model (2-2 model) for additive criterion

\[
\text{Max} \ E_{x_0}[r_0(X_0) + \cdots + r_{\tau-1}(X_{\tau-1}) + g_{\tau}(X_{\tau})] \\
\text{s.t.} \ (i) \ \tau \in \mathcal{T}_0^2
\]

where the continue/stop reward \( \{r_0, r_1; g_0, g_1, g_2\} \) is given in Table 1:
Let us find an optimal stopping time by solving recursive equation. First, the backward recursion (1) yields an optimal solution in Markov class $\Pi$; optimal value functions

\[ v_0 = v_0(x_0), \quad v_1 = v_1(x_1), \quad v_2 = v_2(x_2) \]

and an optimal policy

\[ \gamma^* = \{\gamma_0^*(x_0), \gamma_1^*(x_1)\}. \]

\[ v_2(s_1) = 0.7 \]
\[ v_2(s_2) = 0.6 \]

\[ v_1(s_1) = \text{Max}[0.7, 0.0 + \frac{1}{2}0.7 + \frac{1}{2}0.6] = 0.7 \]
\[ \gamma_1^*(s_1) = \text{stop} \]

\[ v_1(s_2) = \text{Max}[0.8, 0.2 + \frac{1}{2}0.7 + \frac{1}{2}0.6] = 0.85 \]
\[ \gamma_1^*(s_2) = \text{continue} \]

\[ v_0(s_1) = \text{Max}[0.9, 0.2 + \frac{1}{2}0.7 + \frac{1}{2}0.85] = 0.975 \]
\[ \gamma_0^*(s_1) = \text{continue} \]

\[ v_0(s_2) = \text{Max}[0.9, 0.1 + \frac{1}{2}0.7 + \frac{1}{2}0.85] = 0.90 \]
\[ \gamma_1^*(s_2) = \text{stop}. \]

The optimal solution is tabulated as

<table>
<thead>
<tr>
<th>$x_n$</th>
<th>$v_2(x_2)$</th>
<th>$v_1(x_1)$</th>
<th>$\gamma_1^*(x_1)$</th>
<th>$v_0(x_0)$</th>
<th>$\gamma_0^*(x_0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_1$</td>
<td>0.7</td>
<td>0.7</td>
<td>$s$</td>
<td>0.975</td>
<td>$c$</td>
</tr>
<tr>
<td>$s_2$</td>
<td>0.6</td>
<td>0.85</td>
<td>$c$</td>
<td>0.9</td>
<td>$s$</td>
</tr>
</tbody>
</table>

where $s$ and $c$ denote stop and continue, respectively.
Finally, an optimal stepping time $\tau^*$ from $x_0 = s_1$ is described through Theorem 3.2. In fact, for any path $\omega = x_0x_1x_2$, $\tau^*(\omega)$ takes the following time:

$$\tau^*(s_1s_1x_2) = 1$$
$$0.975 = v_0(s_1) > g_0(s_1) = 0.9$$
$$0.7 = v_1(s_1) = g_1(s_1) = 0.7$$

$$\tau^*(s_1s_2s_1) = 2$$
$$0.975 = v_0(s_1) > g_0(s_1) = 0.9$$
$$0.85 = v_1(s_2) > g_1(s_2) = 0.8$$
$$0.7 = v_2(s_1) = g_2(s_1) = 0.7$$

$$\tau^*(s_1s_2s_2) = 2$$
$$0.975 = v_0(s_1) > g_0(s_1) = 0.9$$
$$0.85 = v_1(s_2) > g_1(s_2) = 0.8$$
$$0.6 = v_2(s_2) = g_2(s_2) = 0.6$$

4 Minimum Process

We consider the problem of maximizing an expected value of stopped process with minimum criterion (3, 9, 16, 19). As for nonstopping but control problems, see [6-8, 10-12]):

$$\text{Max } E_{x_0} [r_0 \wedge r_1 \wedge \cdots \wedge r_{\tau-1} \wedge g_\tau]$$

$$\text{s.t. } \tau \in \mathcal{T}_0^N.$$  

The expected value of minimum reward is the sum of multiple sums as follows:

$$E_{x_0} [r_0 \wedge \cdots \wedge r_{\tau-1} \wedge g_\tau]$$

$$= \sum_{n=0}^{N} \sum_{\{\tau=n\}} [r_0(x_0) \wedge \cdots \wedge r_{n-1}(x_{n-1}) \wedge g_n(x_n)] p(x_1|x_0)p(x_2|x_1) \cdots p(x_n|x_{n-1}).$$

Here we mention that the linearity of expectation operator does not admit the equality

$$E[c \wedge Z] = c \wedge E[Z]$$

where $c$ is a constant and $Z$ is a random variable.

So, we imbed $M_0(x_0)$ into a new class of additional parametric subproblems [2, 13]. First we define the past-valued (cumulative) random variables $\{\tilde{\lambda}_n\}$ up to $n$-th stage and the past-value sets $\{\Lambda_n\}$. they take:

$$\tilde{\lambda}_0 \triangleq \tilde{\lambda}_0 \text{ where } \tilde{\lambda}_0 \text{ is larger than or equal to } g_n(x), r_n(x)$$
$$\tilde{\lambda}_n \triangleq r_0(X_0) \wedge \cdots \wedge r_{n-1}(X_{n-1}) \quad 1 \leq n \leq N,$$

$$\Lambda_0 \triangleq \{\tilde{\lambda}_0\}$$

$$\Lambda_n \triangleq \{\lambda_n \mid \lambda_n = r_0(x_0) \wedge \cdots \wedge r_{n-1}(x_{n-1}), (x_0, \ldots, x_{n-1}) \in X \times \cdots \times X \} \quad 1 \leq n \leq N.$$
The minimum criterion is terminal now:

\[ r_0(X_0) \land \cdots \land r_{\tau-1}(X_{\tau-1}) \land g_{\tau}(X_{\tau}) = \tilde{\Lambda}_{\tau} \land g_{\tau}(X_{\tau}). \]

We have

**Lemma 4.1 (Forward recursive formulae)**

\[
\begin{align*}
\bar{\Lambda}_0 &= \tilde{\lambda}_0 \\
\bar{\Lambda}_{n+1} &= \bar{\Lambda}_{n} \land r_{n}(X_{n}) \quad 0 \leq n \leq N-1, \\
\Lambda_0 &= \{\tilde{\lambda}_0\} \\
\Lambda_{n+1} &= \{\lambda \land r_{n}(x) \mid \lambda \in \Lambda_{n}, x \in X\} \quad 0 \leq n \leq N-1.
\end{align*}
\]

(2)

Let us now expand the original state space \( X \) to a direct product space:

\[ Y_n \triangleq X \times \Lambda_n \quad 0 \leq n \leq N. \]

We define a sequence of stopping reward functions \( \{G_n\}_{0}^{N} \) by

\[ G_n(x; \lambda) \triangleq \lambda \land g_{n}(x) \quad (x; \lambda) \in Y_n \]

and a nonstationary Markov transition law \( q = \{q_n\}_{0}^{N-1} \) by

\[ q_n(y; \mu | x; \lambda) \triangleq \begin{cases} p(y|x) & \text{if } \lambda \land r_{n}(x) = \mu \\ 0 & \text{otherwise}. \end{cases} \]

Then \( \{(X_n, \bar{\Lambda}_n)\}_{0}^{N} \) is a Markov process on state spaces \( \{Y_n\} \) with transition law \( q \). We consider the terminal criterion \( \{G_n\}_{0}^{N} \) on the expanded process:

\[ \overline{T}_0(y_0) \quad \text{Max} \ E_{y_0}[G_{\tau}] \quad \text{s.t. } \tau \in \tilde{\mathcal{T}}_{0}^{N} \]

where \( y_0 = (x_0; \tilde{\lambda}_0) \), and \( \tilde{\mathcal{T}}_{n}^{N} \) is the set of all stopping times which take values in \( \{n, n + 1, \ldots, N\} \) on the new Markov chain.

Now we take a subprocess which starts at state \( y_n = (x_n; \lambda_n)(\in Y_n) \) on \( n \)-th stage:

\[ \overline{T}_n(y_n) \quad \text{Max} \ E_{y_n}[G_{\tau}] \quad \text{s.t. } \tau \in \tilde{\mathcal{T}}_{n}^{N}. \]

Let \( v_n(y_n) \) be the maximum value of \( \overline{T}_n(y_n) \), where

\[ v_N(y_N) \triangleq G_N(y_N) \quad y_N \in Y_N. \]

Then we have the backward recursive equation:
Corollary 4.1

\[
\begin{cases}
  v_N(y) = G_N(y) & y \in Y_N \\
  v_n(y) = \max \left[ G_n(y), E_y[v_{n+1}(Y_{n+1})] \right] & y \in Y_n, 0 \leq n \leq N - 1
\end{cases}
\]

where \( E_y \) is the one-step expectation operator induced from the Markov transition probabilities \( q_n(\cdot|\cdot) \):

\[
E_y[h(Y_{n+1})] = \sum_{z \in Y_{n+1}} h(y)q_n(z|y).
\]

Corollary 4.2 The stopping time \( \tau^* \):

\[
\tau^*(\omega) = \min \{ n : v_n(y_n) = G_n(y_n) \} \quad \omega = y_0y_1\cdots y_N
\]
is optimal:

\[
E_{y_0}[G_{\tau^*}] \geq E_{y_0}[G_{\tau}] \quad \forall \tau \in \tilde{T}_0^N.
\]

Then we have the corresponding recursive equation for the original process with minimum reward:

Theorem 4.1

\[
\begin{cases}
  v_N(x, \lambda) = \lambda \wedge g_N(x) & x \in X, \lambda \in \Lambda_N \\
  v_n(x, \lambda) = \max \left[ \lambda \wedge g_n(x), E_x[v_{n+1}(X_{n+1}, \lambda \wedge r_n(x))] \right] & x \in X, \lambda \in \Lambda_n, 0 \leq n \leq N - 1
\end{cases}
\]

Theorem 4.2 The stopping time \( \tau^* \):

\[
\tau^*(\omega) = \min \{ n : v_n(x_n, \lambda_n) = \lambda_n \wedge g_n(x_n) \} \quad \omega = (x_0, \lambda_0)(x_1, \lambda_1)\cdots(x_N, \lambda_N)
\]
is optimal:

\[
E_{x_0}[r_0 \wedge \cdots \wedge r_{\tau-1} \wedge g_{\tau} \wedge g_{\tau}] \geq E_{x_0}[r_0 \wedge \cdots \wedge r_{\tau-1} \wedge g_{\tau}] \quad \forall \tau \in T_0^N.
\]

4.1 DP Solution

A 2-2 model is specified by:

\[
\max E_{x_0}[r_0(X_0) \wedge \cdots \wedge r_{\tau-1}(X_{\tau-1}) \wedge g_{\tau}(X_\tau)]
\]

s.t. (i) \( \tau \in T_0^2 \)

where the continue/stop reward \{\( r_0, r_1, g_0, g_1, g_2 \)\} is given in Table 2:
and the transition matrix is symmetric \((p = q = 1/2)\).

First, the forward recursion (2) generates the following past-value sets:

\[
\Lambda_0 = \{1.0\}, \quad \Lambda_1 = \{0.8, 0.9\}, \quad \Lambda_2 = \{0.6, 0.8\}.
\]

Second, the backward recursion (3) yields an optimal solution in expanded Markov class \(\bar{\Pi}\); optimal value functions

\[
v_0 = v_0(x_0; \lambda_0), \quad v_1 = v_1(x_1; \lambda_1), \quad v_2 = v_2(x_2; \lambda_2)
\]

and an optimal policy

\[
\gamma^* = \{\gamma_0^*(x_0; \lambda_0), \gamma_1^*(x_1; \lambda_1)\}.
\]

\[
v_2(s_1, 0.6) = 0.6 \wedge g_2(s_1) = 0.6 \wedge 0.7 = 0.6
\]

\[
v_2(s_2, 0.6) = 0.6 \wedge g_2(s_2) = 0.6 \wedge 0.6 = 0.6
\]

\[
v_2(s_1, 0.8) = 0.8 \wedge g_2(s_1) = 0.8 \wedge 0.7 = 0.7
\]

\[
v_2(s_2, 0.8) = 0.8 \wedge g_2(s_2) = 0.8 \wedge 0.6 = 0.6
\]

\[
v_1(s_1, 0.8) = \max[0.8 \wedge 0.7, \frac{1}{2}0.7 + \frac{1}{2}0.6] = 0.7
\]

\[
\gamma_1^*(s_1, 0.8) = \text{stop}
\]

\[
v_1(s_2, 0.8) = \max[0.8 \wedge 0.4, \frac{1}{2}0.6 + \frac{1}{2}0.6] = 0.6
\]

\[
\gamma_1^*(s_2, 0.8) = \text{continue}
\]

\[
v_1(s_1, 0.9) = \max[0.9 \wedge 0.7, \frac{1}{2}0.7 + \frac{1}{2}0.6] = 0.7
\]

\[
\gamma_1^*(s_1, 0.9) = \text{stop}
\]

\[
v_1(s_2, 0.9) = \max[0.9 \wedge 0.4, \frac{1}{2}0.6 + \frac{1}{2}0.6] = 0.6
\]

\[
\gamma_1^*(s_2, 0.9) = \text{continue}
\]

\[
v_0(s_1, 1.0) = \max[1.0 \wedge 0.5, \frac{1}{2}0.7 + \frac{1}{2}0.6] = 0.65
\]

\[
\gamma_0^*(s_1, 1.0) = \text{continue}
\]

\[
v_0(s_2, 1.0) = \max[1.0 \wedge 0.6, \frac{1}{2}0.7 + \frac{1}{2}0.6] = 0.65
\]

\[
\gamma_1^*(s_2, 0.8) = \text{continue}
\]

The optimal solution is tabulated as:

\[
\begin{array}{c|cc|cc}
 x_n & s_1 & s_2 \\
\hline
 g_0(x_0) & r_0(x_0) & 0.5 & 0.8 & 0.6 & 0.9 \\
 g_1(x_1) & r_1(x_1) & 0.7 & 0.8 & 0.4 & 0.6 \\
 g_2(x_2) & & 0.7 & 0.6 & \\
\end{array}
\]

Table 2 stop/continue reward
Finally, an optimal stopping time $\sigma^*$ from $(x_0, \lambda_0) = (s_1, 1.0)$ is described through Theorem 4.2. In fact, for any path $\tilde{\omega} = (x_0, 1.0)(x_1, \lambda_1)(x_2, \lambda_2)$, $\sigma^*(\tilde{\omega})$ takes the following time:

$$\sigma^*((s_1, 1.0)(s_1, 0.8)(x_2, \lambda_2)) = 1$$
0.65 = $v_0(s_1, 1.0) > 1.0 \wedge g_0(s_1) = 1.0 \wedge 0.5 = 0.5$
0.7 = $v_1(s_1, 0.8) = 0.8 \wedge g_1(s_1) = 0.8 \wedge 0.7 = 0.7$

$$\sigma^*((s_1, 1.0)(s_2, 0.8)(s_1, 0.6)) = 2$$
0.65 = $v_0(s_1, 1.0) > 1.0 \wedge g_0(s_1) = 1.0 \wedge 0.5 = 0.5$
0.7 = $v_1(s_2, 0.8) > 0.8 \wedge g_1(s_2) = 0.8 \wedge 0.4 = 0.4$
0.6 = $v_2(s_1, 0.6) = 0.6 \wedge g_2(s_1) = 0.6 \wedge 0.7 = 0.6$

$$\sigma^*((s_1, 1.0)(s_2, 0.8)(s_2, 0.6)) = 2$$
0.65 = $v_0(s_1, 1.0) > 1.0 \wedge g_0(s_1) = 1.0 \wedge 0.5 = 0.5$
0.7 = $v_1(s_2, 0.8) > 0.8 \wedge g_1(s_2) = 0.8 \wedge 0.4 = 0.4$
0.6 = $v_2(s_2, 0.6) = 0.6 \wedge g_2(s_2) = 0.6 \wedge 0.6 = 0.6$

Similarly the stopping time $\sigma^*$ also turns out to be optimal from $x_0 = s_2$.

References


