Journal of Information & Optimization Sciences
ISSN 0252-2667 (Print), ISSN 2169-0103 (Online)
Vol. 47 (2026), No. 4, pp. 1237–1247
DOI : 10.47974/JIOS-1329

A discrete time Markov decision process under a fractional discounted reward criterion

Teruo Tanaka
Department of Systems Engineering
Graduate School of Information Sciences
Hiroshima City University
3-4-1, Ozukahigashi
Asaminami-ku
Hiroshima 731-3194
Japan

Abstract

We consider a discrete-time Markov decision process with a state space and an action space in general setting, evaluated according to a fractional discounting reward criterion. For the sake of discussing an optimal control for this model, we transform this model to a Markov decision model under the usual discounted reward criterion by using the parametric method. We discuss the existence of an optimal policy as well as the continuity of the corresponding optimal value function for our model.

Subject Classification: *90C39, 90C40, 93E20.*

Keywords: *Markov decision process, Fractional reward, Parametric method, Bellman equation.*

1. Introduction

As the performance criteria in the fields of Markov decision processes and stochastic control problems, the finite horizon criterion, the infinite horizon discounted criterion and the time average criterion are well-known. For example, the finite horizon criterion has been studied by [2, 3, 4, 10], the infinite horizon discounted criterion has been studied by [2, 3, 4, 10], and the time average criterion has been studied by [3, 4, 10].

In this paper, we focus on the fractional criterion which has not been well studied in these fields so far.

However, in the fractional programming in the field of mathematical programming, fractional functions are treated as objective functions and

E-mail: `teruo@hiroshima-cu.ac.jp`

optimization theory for fractional functions is developed in [14]. In game theory, fractional games, which treats fractional functions as payoff functions, have been studied by [7, 12]. In optimal stopping problems and stochastic control problems, a fractional cost criterion has been treated by [8, 13, 15, 16, 17].

In completely observable Markov decision processes with a finite state space and an action space, an optimal policy and its algorithm under the fractional criterion have been studied by [1, 5]. On the other hand, in partially observable Markov decision processes, the model with a discrete (or countable) state space and a control (or action) space under a fractional time average criterion has been given and an optimal policy and its algorithm have been studied by [9, 11]. Also in [19], the most efficient escape route in the event of a fire is studied within the Markov decision process setting, and new criteria are proposed for defining an optimal evacuation policy.

Our purpose in this paper, we shall give a completely observable Markov decision process with a general state space and an action(control) space, under the fractional discounted criterion. And then we shall discuss an optimality equation and an optimal policy for our model. In partially observable Markov decision processes, the model with a state space and an action space in general setting, as well as the fractional discounted criterion has been been studied by [18].

This paper is organized as follows; in the rest of this section, we give the formulation of a completely observable discrete time Markov decision model with the general state space and action space, under the fractional discounted criterion. As for the notations and formulation of Markov decision processes, we refer to [2, 3]. In section 2, by applying a parametric method commonly used in fractional programming, we convert the discrete time Markov decision model according to a fractional discounted criterion into an equivalent model under the standard discounted criterion. In section 3, we discuss an optimality equation and an optimal policy for the model given in section 1, and give an optimal policy for our model given in section 2 by using the parametric method in section 2. In section 4, the continuity property of an optimal value function for the model given in section 1 is discussed. In section 5, the Dinkelbach algorithm for searching an optimal policy and an optimal value is discussed.

This paper follows the notations and definition presented in [3] closely. By a Borel space, we mean a Borel subset of a complete separable metric space. Let Z a Borel space. Denote by $\mathcal{B}(Z)$ the Borel σ-field of S, by $B(Z)$ the set of all bounded R-valued $\mathcal{B}(Z)$-measurable functions on Z equipped with the supremum norm $||v|| = \sup_{x\in Z}|v(x)|$, and by $C(Z)$ the set of all bounded R-valued continuous functions on Z also equipped with the supremum norm $||v||$. For Borel spaces Z and W, if, for all $w \in W$, $P(\cdot \mid w)$ is a probability measure on Z, and for all $D \in \mathcal{B}(Z)$, the map $w \mapsto P(D \mid w)$ is $\mathcal{B}(W)$-measurable from W to [0,1], then P is called a stochastic kernel on Z given W. We set $AB = A \times B$.

A discrete time Markov decision process with the general state space and action (or control) space under the fractional discounted criterion, denoted by $(S, A, p, f_i(i = 1,2))$, is characterized by the following factors.

- We denote by S the state space of our process, which is assumed to be a Borel space. A state is represented by an element $x \in S$.
- We denote by A an action (or control) space, which is assumed to be a Borel space. An action (or control) is represented by an element $a \in A$. We assume that for all $x \in S$, $A(x)$ is non-empty $\mathcal{B}(A)$-measurable, and the set K defined by
$$L = \{(x, a) \mid a \in A(x), x \in S\}$$
is $\mathcal{B}(SA)$-measurable.
- We denote by $p(dx' \mid x, a)$ a state transition law, which defines a stochastic kernel on S given L.
- We denote by $f_i \in B(L)$ $(i = 1,2)$ 1-step reward functions such that $|f_1| \leq R_1$ and $R_{21} \leq f_2 \leq R_{22}$ for some $R_1 > 0, R_{22} \geq R_{21} > 0$.
- We denote by $0 < \gamma < 1$ the discount factor.

We set $\Omega = (SA)^\infty$ which denotes a sample space and $(x_0, a_0, x_1, a_1, \cdots) \in \Omega$ a sample. Let $H_0 = X, H_{r+1} = KH_r\ (r \geq 0)$, and $h_0 = x_0 \in H_0, h_r = (x_0, a_0, \cdots, x_{r-1}, a_{r-1}, x_r) \in H_r$ be histories.

A policy (or atrategy) consists of a sequence $\delta = \{\delta_r, r \geq 0\}$ of stochastic kernels δ_r on A given H_r satisfying, for $r \geq 0$ and $h_r \in H_r$, $\delta_k(A(x_r) \mid h_r) = 1$, and Δ denotes the set of all policies. A stationary policy consists of a sequence $\xi = \{\xi, \xi, \cdots\}$ of a $\mathcal{B}(X)$-measurable function $\xi: X \to A$ satisfying, for every $x \in S$, $f(x) \in A(x)$.

The pair consisting of a policy $\delta \in \Delta$ and an initial state $x \in S$ determine a probability measure P_x^δ on Ω given by
$$\begin{aligned} &P_x^\delta(dx_0 da_0 dx_1 da_1 \cdots) \\ &= p_x(dx_0)\delta_0(da_0 \mid x_0)p(dx_1 \mid x_0, a_0)\delta_1(da_1 \mid x_0, a_0, x_1) \cdots, \end{aligned}$$
where p_x is the Dirac probability measure at x on X.

Then our control problem is, for an initial state $x \in S$, to find a policy $\delta^* = \delta^*(x) \in \Delta$ such that
$$\frac{E_x^{\delta^*}[\sum_{r=0}^\infty \gamma^t f_1(S(r), U(r))]}{E_x^{\delta^*}[\sum_{r=0}^\infty \gamma^t f_2(S(r), U(r))]} = \sup_{\delta \in \Delta} \frac{E_x^{\delta}[\sum_{r=0}^\infty \gamma^t f_1(S(r), U(r))]}{E_x^{\delta}[\sum_{r=0}^\infty \gamma^t f_2(S(r), U(r))]},$$
where $\{S(r), r \geq 0\}$, defined as the projection from Ω to S, is the stochastic process representing the state process, and $\{U(r), r \geq 0\}$, defined as the projection from Ω to A, is the stochastic process representing the control process.

2. Parametric method

For all $\phi \in R$ and $x \in S$, we denote
$$d(\phi, x) = \sup_{\delta \in \Delta} E_x^\delta[\textstyle\sum_{r=0}^\infty \gamma^t\{f_1(S(r), U(r)) - \phi f_2(S(r), U(r))\}],$$
which means the optimal value function for the discounted Markov decision process.

Proposition 2.1: For all $x \in S$, $d(\cdot, x)$ is Lipschitz continuous : for any $\phi, \theta \in R$,

$$|d(\phi, x) - d(\theta, x)| \leq \frac{R_{22}}{1-\gamma} |\phi - \theta|.$$

Proof: We get

$$\begin{aligned}
&|d(\phi, x) - d(\theta, x)| \\
&= |\sup_{\delta \in \Delta} E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{f_1(S(r), U(r)) - \phi f_2(S(r), U(r))\}] \\
&\quad -\sup_{\delta \in \Delta} E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{f_1(S(r), U(r)) - \theta f_2(S(r), U(r))\}]| \\
&\leq \sup_{\delta \in \Delta} |E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{f_1(S(r), U(r)) - \phi f_2(S(r), U(r))\}] \\
&\quad -E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{f_1(S(r), U(r)) - \theta f_2(S(r), U(r))\}]| \\
&= \sup_{\delta \in \Delta} |E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t (-\phi + \theta) f_2(S(r), U(r))\}]| \\
&\leq \sup_{\delta \in \Delta} E_x^\delta [|-\phi + \theta| \frac{R_{22}}{1-\gamma}] \\
&= \frac{R_{22}}{1-\gamma} |\phi - \theta|.
\end{aligned}$$

Proposition 2.2: For all $x \in S$, $d(\cdot, x)$ is convex.

Proof: For any $\phi_1 < \phi_2$, $\theta \in [0,1]$, we have

$$\begin{aligned}
&d(\theta\phi_1 + (1-\theta)\phi_2, x) \\
&= \sup_{\delta \in \Delta} E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{f_1(S(r), U(r)) - (\theta\phi_1 + (1-\theta)\phi_2) f_2(S(r), U(r))\}] \\
&= \sup_{\delta \in \Delta} E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{\theta f_1(S(r), U(r)) - \theta\phi_1 f_2(Y(t), B(t)) \\
&\quad +(1-\theta) f_1(S(r), U(r)) - (1-\theta)\phi_2 f_2(S(r), U(r))\}] \\
&\leq \sup_{\delta \in \Delta} E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{\theta f_1(S(r), U(r)) - \theta\phi_1 f_2(Y(t), B(t))\}] \\
&\quad +\sup_{\delta \in \Delta} E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t \{(1-\theta) f_1(S(r), U(r)) - (1-\theta)\phi_2 f_2(S(r), U(r))\}] \\
&= \theta d(\phi_1, x) + (1-\theta) d(\phi_2, x).
\end{aligned}$$

Proposition 2.3: For each $x \in S$, $d(\cdot, x)$ is strictly decreasing.

Proof: For any $\phi_1 < \phi_2$ and r, we have

$$\begin{aligned}
&f_1(S(r), U(r)) - \phi_1 f_2(S(r), U(r)) \\
&= f_1(S(r), U(r)) - \phi_2 f_2(S(r), U(r)) + (\phi_2 - \phi_1) f_2(S(r), U(r)) \\
&\geq f_1(S(r), U(r)) - \phi_2 f_2(S(r), U(r)) + (\phi_2 - \phi_1) R_{21},
\end{aligned}$$

and then

$$E_x^\delta [\textstyle\sum_{r=0}^\infty \gamma^t f_1(S(r), U(r)) - \phi_1 f_2(S(r), U(r))]$$

$\geq E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_1(S(r), U(r)) - \phi_2 f_2(S(r), U(r))] + \frac{\phi_2 - \phi_1}{1-\gamma} R_{21}.$

Hence we get

$$d(\phi_1, x) \geq d(\phi_2, x) + \frac{\phi_2 - \phi_1}{1-\gamma} R_{21} > d(\phi_2, x).$$

Proposition 2.4: For each $x \in S$, $\lim_{\phi \to \infty} d(\phi, x) = -\infty, \lim_{\phi \to -\infty} d(\phi, x) = \infty$.

Proof:

$$\begin{aligned} d(\phi, x) &= \sup_{\delta \in \Delta} E_x^{\delta}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi f_2(S(r), U(r))\}] \\ &\leq \sup_{\delta \in \Delta} E_x^{\delta}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{R_1 - \phi R_{21}\}] \\ &= \frac{R_1 - \phi R_{21}}{1-\gamma} \to -\infty \ (\phi \to \infty), \\ d(\phi, x) &\geq \sup_{\delta \in \Delta} E_x^{\delta}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{-R_1 - \phi R_{22}\}] \\ &= \frac{-R_1 - \phi R_{22}}{1-\gamma} \to +\infty \ (\phi \to -\infty). \end{aligned}$$

Proposition 2.5: For each $x \in S$, there is uniquely $\phi(x) \in R$ with $d(\phi(x), x) = 0$.

3. Optimal value and optimal policy

We impose the following additional assumptions :

- For every $x \in S$, $A(x) \subseteq A$ and $A(x)$ is compact.
- For every $x \in S$, $a \in A(x) \mapsto f_i(x, a) \in R$ is continuous.
- For every $x \in S$ and $v \in B(S)$, $a \mapsto \int_S v(y) p(dy \mid x, a)$ is a continuous function from $A(x)$ to R.

Theorem 3.1: *For any $x \in S$, there is a stationary policy $\xi^{\phi(x)}$, which depends on $\phi(x)$, such that*

$$\sup_{\delta \in \Delta} \frac{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_1(S(r), U(r))]}{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]} = \frac{E_x^{\xi^{\phi(x)}}[\sum_{r=0}^{\infty} \gamma^t f_1(S(r), U(r))]}{E_x^{\xi^{\phi(x)}}[\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]}.$$

Moreover

$$\phi(x) = \sup_{\delta \in \Delta} \frac{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_1(S(r), U(r))]}{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]}.$$

Proof: According to the results of [3], the value function $d(\phi, y)$ satisfies the following equation

$$d(\phi, y) = \max_{a \in A(y)} \{f_1(y, a) - \phi f_2(y, a) + \int_S d(\phi, z) p(dz \mid y, a)\}$$

for any $\phi \in R$ and $y \in X$. Let $x \in S$ be fixed and $\phi = \phi(x)$ which is in Proposition 2.5. Then we get

$$d(\phi(x), y) = \max_{a \in A(y)} \{f_1(y, a) - \phi(x) f_2(y, a) + \int_S d(\phi(x), z) p(dz \mid y, a)\}$$

for each $y \in S$. Let $\xi^{\phi(x)}(y)$ be the function that attains the maximum on the right-hand side of the above equation. The stationary policy $\{\xi^{\phi(x)}, \xi^{\phi(x)}, \cdots\}$ is optimal in that

$$\begin{aligned} 0 &= \sup_{\delta \in \Delta} E_x^{\delta}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi(x) f_2(S(r), U(r))\}] \\ &= E_x^{\xi^{\phi(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi(x) f_2(S(r), U(r))\}]. \end{aligned}$$

For each $\delta \in \Delta$, we have

$$\begin{aligned} 0 &= E_x^{\xi^{\phi(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi(x) f_2(S(r), U(r))\}] \\ &\geq E_x^{\delta}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi(x) f_2(S(r), U(r))\}], \end{aligned}$$

from which we get for each $\delta \in \Delta$,

$$\phi(x) = \frac{E_x^{\xi^{\phi(x)}}[\sum_{r=0}^{\infty} \gamma^t f_1(S(r), U(r))]}{E_x^{\xi^{\phi(x)}}[\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]} \geq \frac{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_1(S(r), U(r))]}{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]}.$$

4. Continuity of ϕ

Here, we introduce further supplementary assumptions :

- The multivalued map $x \in S \to A(x)$ is continuous.
- For $i = 1,2$, $r_i \in C(L)$.
- For each $v \in B(S)$, $(x, a) \mapsto \int_S v(y) p(dy \mid x, a)$ is a continuous map from L to R.

Theorem 4.1: ([3]) *For each* $\phi \in R$, $d(\phi, \cdot)$ *is bounded continuous.*

Lemma 4.1: *Let a function* $g(s, t)$ *be defined on a set* ST *which is assumed that*

- *For each* $t \in T$, $s \in S \to g(s, t) \in R$ *is a continuous function.*
- *There exists a constant* C, *which is independent of* $s \in S$, *such that, for any* $t_1, t_2 \in T$, $|g(s, t_1) - g(s, t_2)| \leq C|t_1 - t_2|$.

Then $g(s, t)$ *is continuous on* ST.

Proof: Let $(s_0, t_0) \in ST$ be fixed. Then we have

$$\begin{aligned} &|g(s_0, t_0) - g(s, t)| \\ \leq\ &|g(s_0, t_0) - g(s, t_0)| + |g(s, t_0) - g(s, t)| \\ \leq\ &|g(s_0, t_0) - g(s, t_0)| + C|t - t_0|. \end{aligned}$$

For any $\varepsilon > 0$, we can find $\delta' = \delta'(s_0, t_0, \varepsilon) > 0$ satisfying if $|s - s_0| < \delta'$, then

$$|g(s_0, t_0) - g(s, t_0)| < \frac{\varepsilon}{2}.$$

Now we set $\delta'' = \frac{\varepsilon}{2C}$. Then we get

$$\text{if } |t - t_0| < \delta'', \text{then } C|t - t_0| < \frac{\varepsilon}{2}.$$

Therefore we set $\delta = \min\{\delta', \delta''\}$ so that

$$\text{if } |s - s_0| < \delta, |t - t_0| < \delta, \text{then } |g(s_0, t_0) - g(s, t)| < \varepsilon.$$

Theorem 4.2: *Let $X = R^m$, $A = R^\ell$. Then $\phi: R^m \to R$ is continuous.*

Proof: By Proposition 2.1, Theorem 4.1 and Lemma 4.1, $d(\cdot,\cdot): R \times R^m \to R$ is continuous. From Proposition 2.3, for each $x \in R^m$, the function $\phi \to d(\phi, x)$ is injective. By the virtue of Proposition 2.5 and the arguments of [6], we have the conclusion.

5. Dinkelbach Algorithm

In what follows, we demonstrate the computational effectiveness of the Dinkelbach algorithm for both our fractional Markov decision process and a deterministic fractional optimal control problem.

Firstly we shall consider the Dinkelbach algorithm. Suppose $\varepsilon > 0$ and $x \in S$ are given.

(0.1) For every $\delta \in \Delta$, we calculate

$$\phi_1 := \frac{E_x^\delta[\sum_{r=0}^\infty \gamma^t f_1(S(r), U(r))]}{E_x^\delta[\sum_{r=0}^\infty \gamma^t f_2(S(r), U(r))]}.$$

(0.2) We compute the optimal value $d(\phi_1, x)$ and obtain the optimal policy $\xi^{\phi_1(x)}$ for this problem.

(0.3) If $d(\phi_1, x) < \varepsilon$, then we come to end. Therefore we have the optimal value ϕ_1 and policy $\xi^{\phi_1(x)}$ for our fractional problem.

(0.4) If $d(\phi_1, x) \geq \varepsilon$, then we go on. In this situation, we compute

$$\phi_2 := \frac{E_x^{\xi^{\phi_1(x)}}[\sum_{r=0}^\infty \gamma^t f_1(S(r), U(r))]}{E_x^{\xi^{\phi_1(x)}}[\sum_{r=0}^\infty \gamma^t f_2(S(r), U(r))]},$$

set k to 2 and move on to step (k.1).

(k.1) For $\xi^{\phi_{k-1}(x)}$, we calculate

$$\phi_k := \frac{E_x^{\xi^{\phi_{k-1}(x)}}[\sum_{r=0}^\infty \gamma^t f_1(S(r), U(r))]}{E_x^{\xi^{\phi_{k-1}(x)}}[\sum_{r=0}^\infty \gamma^t f_2(S(r), U(r))]}.$$

(k.2) We compute the optimal value $d(\phi_k, x)$ and obtain the optimal policy $\xi^{\phi_k(x)}$ for this problem.

(k.3) If $d(\phi_k, x) < \varepsilon$, then we come to end. In this situation, we have the optimal value ϕ_k and policy $\xi^{\phi_k(x)}$ for our fractional problem.

(k.4) If $d(\phi_k, x) \geq \varepsilon$, then we go on. In this situation, we set k to $k + 1$ and move on to step (k.1).

Secondly we shall examine the convergence of the above algorithm.

Lemma 5.1 *For each $\delta \in \Delta$, we set*

$$\phi := \frac{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_1(S(r), U(r))]}{E_x^{\delta}[\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]}.$$

Then we have $d(\phi, x) \geq 0$.

Proof: For any $\delta \in \Delta$, we have

$$\begin{aligned} d(\phi, x) &= \sup_{\delta \in \Delta} E_x^{\delta}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi f_2(S(r), U(r))\}] \\ &\geq E_x^{\delta}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi f_2(S(r), U(r))\}] \\ &= 0. \end{aligned}$$

Proposition 5.1: For every k with $d(\phi_k, x) \geq \varepsilon$, we have $\phi_k < \phi_{k+1}$.

Proof: According to the definition of ϕ_{k+1}, we get

$$\begin{aligned} 0 &< d(\phi_k, x) \\ &= E_x^{\xi^{\phi_k(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi_k f_2(S(r), U(r))\}] \\ &= (\phi_{k+1} - \phi_k) E_x^{\xi^{\phi_k(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]. \end{aligned}$$

Hence this concludes the proof.

Proposition 5.2: For $x \in S$, let $\phi = \phi(x)$ such that $d(\phi, x) = 0$. For each k with $d(\phi_k, x) > 0$, we have

$$\lim_{k \to \infty} \phi_k = \phi.$$

Proof: Let $\lim_{k \to \infty} \phi_k = \overline{\phi}$. It follows from the proof of Proposition 5.1 that

$$0 < d(\phi_k, x) \leq (\phi_{k+1} - \phi_k) \frac{R_{22}}{1-\gamma} \to 0 \ (k \to \infty).$$

Then we obtain

$$d(\phi, x) = 0 = \lim_{k \to \infty} d(\phi_k, x) = d(\lim_{k \to \infty} \phi_k, x) = d(\overline{\phi}, x).$$

By the uniqueness of $\phi(x)$, $\lim_{k \to \infty} \phi_k = \phi$.

Lemma 5.2: *For the sequence of policies* $\{f^{\phi_k(x)}, k\}$ *generated by the Dinkelbach algorithm,*

$$E_x^{f^{\phi_k(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))] \geq E_x^{f^{\phi_{k+1}(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))].$$

Proof: We have for each k,

$$\begin{aligned} & E_x^{\xi^{\phi_k(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi_k f_2(S(r), U(r))\}] \\ &\geq E_x^{f^{\phi_{k+1}(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi_k f_2(S(r), U(r))\}], \\ & E_x^{f^{\phi_{k+1}(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi_{k+1} f_2(S(r), U(r))\}] \\ &\geq E_x^{f^{\phi_k(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t \{f_1(S(r), U(r)) - \phi_{k+1} f_2(S(r), U(r))\}], \end{aligned}$$

and then

$$(\phi_{k+1} - \phi_k)\{E_x^{\xi^{\phi_k(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]$$
$$-E_x^{f^{\phi_{k+1}(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t f_2(S(r), U(r))]\} \geq 0.$$

The result is obtained by Proposition 5.1.

Lemma 5.3: *For the sequence* $\{\phi_k, k\}$ *generated by the Dinkelbach algorithm,*

$$\left|\frac{\phi(x)-\phi_{k+1}}{\phi(x)-\phi_k}\right| \leq 1 - \frac{E_x^{\delta^*(x)}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}{E_x^{f^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}.$$

Proof: We have for each k,

$$E_x^{\delta^*(x)}[\textstyle\sum_{r=0}^{\infty} \gamma^t\{f_1(S(r), U(r)) - \phi_k f_2(S(r), U(r))\}] \geq$$
$$E_x^{\xi^{\phi_k(x)}}[\textstyle\sum_{r=0}^{\infty} \gamma^t\{f_1(S(r), U(r)) - \phi_k f_2(S(r), U(r))\}].$$

Hence

$$\frac{E_x^{\delta^*(x)}[\sum_{r=0}^{\infty}\gamma^t f_1(S(r),U(r))]}{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]} - \phi_k \frac{E_x^{\delta^*(x)}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]} \leq \phi_{k+1} - \phi_k,$$

and

$$\begin{aligned}\phi(x) - \phi_{k+1} \;\leq\; & \phi(x) - \phi_k - \frac{E_x^{\delta^*(x)}[\sum_{r=0}^{\infty}\gamma^t f_1(S(r),U(r))]}{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}\\ & +\phi_k \frac{E_x^{\delta^*(x)}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}\\ = \;& (\phi(x) - \phi_k)\left(1 - \frac{E_x^{\delta^*(x)}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}\right).\end{aligned}$$

Since $\phi(x) \geq \phi_k$, we obtain the result.

Proposition 5.3: For the sequence of policies $\{\xi^{\phi_k(x)}, k\}$ generated by the Dinkelbach algorithm,

$$\phi(x) = \lim_{k\to\infty} \frac{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_1(S(r),U(r))]}{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}.$$

Proof: In the above Dinkelbach algorithm, we have

$$\phi_{k+1} := \frac{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_1(S(r),U(r))]}{E_x^{\xi^{\phi_k(x)}}[\sum_{r=0}^{\infty}\gamma^t f_2(S(r),U(r))]}.$$

By Proposition 5.1, we have the conclusion.

References

[1] Vijay Kumar Aggarwal, Ramaswamy Chandrasekaran, and Kunhiraman P. K. Nair, "Markov ratio decision processes," *Journal of Optimization Theory and Applications*, vol. 21, pp. 27–37 (1977).

[2] Dimitri P. Bertsekas and Steven E. Shreve, Stochastic Optimal Control. New York, NY, USA: Academic Press (1978).

[3] Onesimo Hernández-Lerma, Adaptive Markov Control Processes. New York, NY, USA: Springer-Verlag (1989).

[4] Onesimo Hernández-Lerma and Jean Bernard Lasserre, Discrete-Time Markov Control Processes. New York, NY, USA: Springer-Verlag (1996).

[5] Seiichi Iwamoto and Toshiharu Fujita, "Markov decision processes under fractional criterions," Kyoto University RIMS Kokyuroku, no. 1079, pp. 153–163 (1999) (in Japanese).

[6] Krisorn Jittorntrum, "An implicit function theorem," *Journal of Optimization Theory and Applications,* vol. 25, pp. 575–577 (1978).

[7] Hang-Chin Lai, "On a dynamic fractional game," Kyoto University RIMS Kokyuroku, no. 1298, pp. 151–160 (2002).

[8] Hiroaki Morimoto, "On average cost stopping time problems," *Probability Theory and Related Fields*, vol. 90, pp. 469–490 (1991).

[9] Masamitsu Ohnishi, "Optimal minimal-repair and replacement problem under average cost criterion: Optimality of(t,T)-policy," *Journal of the Operations Research Society of Japan,* vol. 40, pp. 373–389 (1997).

[10] Martin L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York, NY, USA: Wiley-Interscience (1994).

[11] Zhiyuan Ren and Bruce Krogh, "Markov decision processes with fractional costs," *IEEE Transactions on Automatic Control,* vol. 50, pp. 646–650 (2005).

[12] Yoichi Sawasaki, Yutaka Kimura, and Kensuke Tanaka, "A two-person zero-sum game with fractional loss function," *Journal of the Operations Research Society of Japan,* vol. 43, pp. 209–218 (2000).

[13] Maurice Robin, "On some impulse control problems with long run average cost," *SIAM Journal on Control and Optimization,* vol. 19, pp. 333–358 (1981).

[14] Ioan Mihai Stancu-Minasian, Fractional Programming. Dordrecht, The Netherlands: Kluwer Academic Publishers (1997).

[15] Lukasz Stettner, "On some stopping impulse control problems with a general discount rate criteria," *Probability and Mathematical Statistics*, vol. 10, pp. 223–245 (1989).

[16] Min Sun, "An optimal stopping time problem with time average cost in a bounded interval," *Systems & Control Letters*, vol. 8, pp. 173–180 (1986).

[17] Min Sun, "A multidimensional optimal stopping-time problem with time-average criterion," *Optimal Control Applications and Methods*, vol. 11, pp. 85–93 (1990).

[18] Teruo Tanaka, "A partially observable discrete-time Markov decision process with a fractional discounted reward," *Journal of Information and Optimization Sciences*, vol. 38, no. 1, pp. 21–37 (2017).

[19] Qi Wang, Masayuki Kageyama, and Jingyao Zhang, "New evaluation criteria in the Markov decision processes," *Journal of Statistics and Management Systems*, vol. 24, no. 3, pp. 625–632 (2021).

Received June, 2022

Revised November, 2022