Markov Decision Processes with Constrained Stopping Times

M. Horiguchi[†], M. Kurano[‡] and M. Yasuda[§]

[†]Division of Mathematical Sciences and Phisycs, Graduate School of Science and Technology(horiguti@math.e.chiba-u.ac.jp), [‡] Faculty of Education(kurano@math.e.chiba-u.ac.jp), [§]Faculty of Science(yasuda@math.s.chiba-u.ac.jp), Chiba University, Inage-ku, Chiba 263-8522, Japan

Abstract

The optimization problem for a stopped Markov decision process is considered to be taken over stopping times τ constrained so that $\mathbb{E}\,\tau \leqq \alpha$ for some fixed $\alpha > 0$. We introduce the concept of a randomized stationary stopping time which is a mixed extension of the entry time of a stopping region and prove the existence of an optimal constrained pair of stationary policy and stopping time by utilizing a Lagrange multiplier approach. Also, applying the idea of the one-step look ahead (OLA) policy the optimal constrained pair is sought concretely. As an example, constrained Markov deteriorating system is explained.

Key words: Markov decision process, constrained stopping time, Lagrange multiplier, OLA policy

1 Introduction

A constrained optimal stopping problem is originated by Nachman [15] and Kennedy [13], in which a Lagrangian approach has used to reduce the problem to an unconstrained stopping problem of a conventional type and the constrained optimal stopping time is characterized. Also, a constrained Markov decision process has been studied by many authors (cf. [1, 2, 3, 6, 9, 18, 19]). For the case of one constraint, Beutler and Ross [3] has formed a Lagrange method from the average expected reward and, by the corresponding parametric dynamic programming equation, has shown that there exists an optimal constrained stationary policy requiring randomization between two actions in at most one state under some ergodic conditions. This Lagrange approach was generalized to the countable state case by Sennott [18, 19]. On the other hand, a combined model of the Markov decision process and stopping problem, called a stopped decision process, has been considered by Furukawa and Iwamoto [7] in that the existence of an optimal pair of policy and stopping time associated with some optimality criteria is discussed. Hordijk [8] has considered this model from a standpoint of potential theory. Also, the general utility-treatment for stopped decision processes has been studied by Kadota et al[11, 12]. In this paper, the optimization problem for a stopped decision process is considered. Stopping times τ are forced to be constrained so that $\mathbb{E}\,\tau \leqq \alpha$ for some fixed $\alpha > 0$. We induce a randomized stationary stopping time in order to extend the entry time of a stopping region and prove the existence of an optimal constrained pair of stationary policy and stopping time utilizing a Lagrange multiplier approach. The proof is executed by applying a Lagrange multiplier method developed by Frid [6], Beutler and Ross [3] and Sennott [18]. Also, using the idea of the one-step look ahead (OLA cf.[16]) policy an optimal constrained pair is derived concretely. The constrained Markov deteriorating system is illustrated as an example.

In the reminder of this section, we shall give the problem formulation referring to Hordijk [8]. Also, an optimal constrained pair of policy and stopping time is defined. A dynamic system, at times $t = 0, 1, 2, \ldots$, is observed to be in one of a possible number of states. Let S be the countable state space, denoted by $S = \{1, 2, \ldots\}$. We denote by $\mathcal{P}(S)$ the set of all probability vectors on S, i.e.,

$$\mathcal{P}(S) := \Big\{p = (p_1, \ldots) | p_i \geqq 0 (i \geqq 1), \sum_{i=1}^{\infty} p_i \leqq 1\Big\}.$$

We allow for breaking down or disappearing of the system with positive probability, so $\sum_{i \in S} p_i \leqq 1$. For each $i \in S$, $\mathcal{P}(i)$ is a subset of $\mathcal{P}(S)$, which is assumed to be given. If at time t the system observed is in state i and the decision maker takes $p(i, \cdot) \in \mathcal{P}(i)$, then the system moves to a new state $j \in S$ selected according to the probability distribution $p(i, \cdot)$. This decision process is then repeated from the new state j.

Let $\mathcal{P}$ be the set of all stochastic matrices where i-th row vector $p(i, \cdot) \in \mathcal{P}(i)$. A notion of convergence on $\mathcal{P}$ is given as follows: a sequence $P_n = (p_n(i, j)) \in \mathcal{P}$ converges to $P = (p(i, j)) \in \mathcal{P}$ if $p_n(i, j) \to p(i, j) (n \to \infty)$ for each $i, j \in S$. In this case, we write $\lim_{n\to\infty} P_n = P$. Also, $\mathcal{P}$ with this topology forms metric space (cf. [8]). An element of $\mathcal{P}$ is called a transition matrix. The policy R for controlling the system is a sequence of transition matrices, $P_0, P_1, \cdots \in \mathcal{P}$, denoted by $R = (P_0, P_1, \ldots)$, where P_t gives the transition probability at time $t (t \geqq 0)$. Here we confine ourselves to memoryless or Markov policies, which is shown to be sufficient to our optimization problem (cf. Theorem 13.2 in [8]). We denote by $\mathcal{R}$ the set of all policies. If the policy takes at all times the same transition matrix, i.e., $P^{\infty} := (P, P, \ldots), P \in \mathcal{P}$, it is called a stationary

policy, denoted simply by P and induces a stationary Markov chain.

The sample space is the product space $\Omega = S^\infty$ such that the projection X_n on the n-th factor S describes the state at time n. For each $R \in \mathcal{R}$ and initial state $i \in S$, we can define the measure $\mathbb{P}_{i,R}$ on Ω in an obvious way. In order to solve our problem described in the sequel, we introduce randomized stopping time(cf. [5, 10, 13]). To this end, enlarging Ω to $\overline{\Omega} := \Omega \times [0,1]$, let $\mathcal{G}_n = \mathcal{F}_n \times \mathbb{B}_1$, where $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$, the σ-field induced by $\{X_0, X_1, \dots, X_n\}$, and $\mathbb{B}_1$ is Borel subsets on $[0,1] (n \geqq 0)$ and $\mathcal{G}_\infty = \mathcal{F}_\infty \times \mathbb{B}_1$, where $\mathcal{F}_\infty$ is the smallest σ-field containing all $\mathcal{F}_n (n \geqq 0)$. Let $N := \{0, 1, 2, \dots\} \cup \{\infty\}$. We call a map $\tau : \overline{\Omega} \to N$ a (randomized) stopping time with respect to $\mathcal{G} := \{\mathcal{G}_n, n \in N\}$ if $\{\tau = n\} \in \mathcal{G}_n$ for each $n \in N$. The class of stopping times with respect to $\mathcal{G}$ will be denoted by $C(\mathcal{G})$. Let $c : \mathcal{P} \times S \to \mathbb{R}$ and $r : S \to \mathbb{R}$ be running cost and terminal reward functions respectively. For simplicity, we put $c_P(i) := c(P, i) (P \in \mathcal{P}, i \in S)$. Hereafter, we assume that for $P, Q \in \mathcal{P}$ with $p(i,\cdot) = q(i,\cdot)$ $c_P(i) = c_Q(i)$. For any policy $R = (P_0, P_1, \dots) \in \mathcal{R}$ and $\tau \in C(\mathcal{G})$, we define the expected reward $J_{R,\tau}(i)$ by

$$J_{R,\tau}(i) := \mathbb{E}_{i,R}\left(\sum_{n=0}^{\tau-1} c(X_n) + r(X_\tau)\right), \tag{1.1}$$

where $\mathbb{E}_{i,R}$ is the expectation with respect to the product measure $\mathbb{P}^*_{i,R} := \mathbb{P}_{i,R} \times \mu$ on $\overline{\Omega}$ and μ is a Lebesgue measure on $\mathbb{B}_1$. Note that $\tau = \infty$ with positive probability is admissible with zero reward.

A $\tau \in C(\mathcal{G})$ is called randomized stationary if for each $n \geqq 0$, $\mathbb{P}^*_{i,R}(\tau = n | X_0, X_1, \dots, X_{n-1}, X_n = j, \tau \geqq n)$ is depending only on $j \in S$. In such a case, we can define the set $\{\delta(j), j \in S\}$ by

(1.2)

$$\delta(j) := \mathbb{P}^*_{i,R}(\tau = n | X_0, \dots, X_{n-1}, X_n = j, \tau \geqq n).$$

Then obviously

$$0 \leqq \delta(j) \leqq 1 \quad \text{for each} \quad j \in S. \tag{1.3}$$

Conversely, for any set $\{\delta(j), i \in S\}$ satisfying (1.3), we can define a randomized stationary stopping time τ through (1.2). Such a stopping time is said to be determined by $\{\delta(j)\}$. When $\delta(j) = 0$ or $\delta(j) = 1$ for all $j \in S$, the corresponding stopping time is called simply stationary, which is a entry time of $\Gamma := \{j \in S | \delta(j) = 1\}$, denoted by τ_Γ.

Let $\alpha > 0$ be given arbitrarily. Constrained optimal pairs will be defined with respect to a given initial state. So without loss of generality we may assume the initial state is "1". Let

$$\Delta(\mathcal{G}) := \{(R, \tau) \in \mathcal{R} \times C(\mathcal{G}) | \, \mathbb{E}_{1,R}(\tau) \leqq \alpha \text{ and } \mathbb{E}_R(r(X_\tau)) < \infty\},$$

where $\mathbb{E}_R(r(X_\tau))$ denotes the vector with ith component $\mathbb{E}_{i,R}(r(X_\tau))$. In this paper, we will consider the constrained optimization problem:

$$\text{maximize } J_{R,\tau}(1), \text{ subject to } (R, \tau) \in \Delta(\mathcal{G}). \tag{1.4}$$

The constrained pair $(R^*, \tau^*) \in \Delta(\mathcal{G})$ is called optimal in state $1 \in S$ if

$$J_{R^*,\tau^*}(1) \geqq J_{R,\tau}(1) \tag{1.5}$$

for all $(R, \tau) \in \Delta(\mathcal{G})$.

2 Lagrange formulation for constrained optimization

In this section, the Lagrange multiplier is introduced and the parameterized version of stopped decision process is analyzed.

Introducing the Lagrange multiplier $\lambda \geqq 0$, let

$$c^\lambda_P(i) := c_P(i) - \lambda, \quad i \in S \quad \text{and} \tag{2.1}$$

$$J^\lambda_{R,\tau}(i) := \mathbb{E}_{i,R}\left(\sum_{n=0}^{\tau-1} c^\lambda(X_n) + r(X_\tau)\right), \quad i \in S \tag{2.2}$$

for each $(R, \tau) \in \mathcal{R} \times C(\mathcal{G})$. The value function J^λ is defined as

$$J^\lambda(i) := \sup_{(R,\tau) \in \mathcal{R} \times C(\mathcal{G})} J^\lambda_{R,\tau}(i). \tag{2.3}$$

If $J^\lambda(i) = J^\lambda_{R,\tau}(i)$ for all $i \in S$, the pair (R, τ) is called λ-optimal.

We need the following assumption.

Assumption (U): The following (i)–(iii) are satisfied:

(i) $\mathcal{P}$ is compact and convex,

(ii) $c_P(i) \leqq 0$ for all $P \in \mathcal{P}$ and $i \in S$ and $c_P(i)$ is convex in $P \in \mathcal{P}$ for each $i \in S$

(iii) There exists a vector u with $u \geqq |r|\boldsymbol{e}$ such that

(2.4) $e + Pu \leqq u$, and $|c_P|\boldsymbol{e} + Pu \leqq u$,

(2.5) $\lim_{N\to\infty} P^N u = 0$ for all $P \in \mathcal{P}$ and

(2.6) $\lim_{P \to P_0} Pu = P_0 u$ for all $P_0 \in \mathcal{P}$, where $\boldsymbol{e} = (1, 1, \dots)$

For each λ, the next theorem holds, under the followings:

$$\mathcal{Q}(\lambda) := \{Q \in \mathcal{P} | \max_{P \in \mathcal{P}}(c^\lambda_P + PJ^\lambda) = c^\lambda_Q + QJ^\lambda\},$$

$$\Gamma(\lambda) := \{i \in S | J^\lambda(i) = r(i)\} \quad \text{and}$$

$$\underline{\Gamma}(\lambda) := \{i \in S | r(i) > \max_{P \in \mathcal{P}}(c^\lambda_P + PJ^\lambda)(i)\}.$$

Theorem 2.1 *(cf. chap.3,4[8] and [5]) Suppose that Assumption (U) holds. Then, for any $\lambda \geqq 0$, we have:*

(i) $\sum_{n=0}^{\infty} \mathbb{E}_R\,|c^\lambda(X_n)| < \infty$ *for all* $R \in \mathcal{R}$.

(ii) $|J^\lambda| \leqq (1+\lambda)u$ *and* J^λ *satisfies the following Bellman's optimality equation.*

$$J^\lambda = r \vee \max_{P\in\mathcal{P}}(c_P^\lambda + PJ^\lambda). \tag{2.7}$$

where $a \vee b = \max\{a, b\}$ *for real number* a, b.

(iii) $P_{i,Q}(\tau_{\underline{\Gamma}(\lambda)} < \infty) = 1$ *for all* $Q \in \mathcal{Q}(\lambda)$ *and a pair* $(Q^\infty, \tau_{\Gamma'})$ *with* $Q \in \mathcal{Q}(\lambda)$ *and* $\underline{\Gamma}(\lambda) \subset \Gamma' \subset \Gamma(\lambda)$ *is* λ*-optimal in* $i \in S$.

The following clearly holds.

Corollary 2.1 *Suppose that Assumption (U) holds. Let* $Q(\lambda), \Gamma(\lambda), \underline{\Gamma}(\lambda)$ *be as in Theorem 2.1(iii). Let* $\{\delta(i) : i \in S\}$ *be such that* $0 \leqq \delta(i) \leqq 1$ *and* $\delta(i) = 0$ *if* $i \in \Gamma(\lambda), = 1$ *if* $i \in \underline{\Gamma}(\lambda)$. *Then, for the randomized stopping time* τ *determined by* $\{\delta(i) : i \in S\}$ *through (1.2), a pair* (Q^∞, τ) *with* $Q \in \mathcal{Q}(\lambda)$ *is* λ*-optimal.*

The next three lemmas are useful in the next section, whose proofs are done by referring to the idea used in [3, 18].

Lemma 2.1 *For each* $i \in S$,$J^\lambda(i)$ *is non-increasing and continuous in* $\lambda \geqq 0$.

Proof. For any $0 < \lambda_1 < \lambda_2$ and $0 < a < 1$, Let $\lambda_3 := a\lambda_1 + (1-a)\lambda_2$. Then, we have that, for any $Q_{\lambda_k} \in \mathcal{Q}(\lambda_k)(k = 1, 2, 3)$,

$$J^{\lambda_3}(i) = \mathbb{E}_{i,Q_{\lambda_3}}\left[\sum_{n=0}^{\tau_{\Gamma(\lambda_3)}-1} c(X_n) - \lambda_3\tau_{\Gamma(\lambda_3)} + r(X_{\tau(\lambda_3)})\right]$$

$$\leqq a\,\mathbb{E}_{i,Q_{\lambda_3}}\left[\sum_{n=0}^{\tau_{\Gamma(\lambda_3)}-1} c(X_n) - \lambda_1\tau_{\Gamma(\lambda_3)} + r(X_{\tau(\lambda_3)})\right] +$$

$$(1-a)\,\mathbb{E}_{i,Q_{\lambda_3}}\left[\sum_{n=0}^{\tau_{\Gamma(\lambda_3)}-1} c(X_n) - \lambda_2\tau_{\Gamma(\lambda_3)} + r(X_{\tau(\lambda_3)})\right],$$

which implies $J^{\lambda_3}(i) \leqq aJ^{\lambda_1}(i) + (1-a)J^{\lambda_2}(i)$. This shows that $J^\lambda(i)$ is convex, so that J^λ is continuous in $\lambda(\lambda > 0)$. Let (θ, τ) be 0-optimal: $J^0(i) = J^0_{\theta,\tau}$. Then, we set, for $\lambda > 0$,

$$J^0(i) \geqq J^\lambda(i) \geqq J^\lambda_{\theta,\tau}(i) = J^0(i) - \lambda\,\mathbb{E}_{i,\theta}(\tau).$$

Since $\mathbb{E}_{i,\theta}(\tau) < \infty$, the above shows that $J^\lambda(i) \to J^0(i)$ as $\lambda \to 0$. Also, from the definition of J^λ, it follows that $J^\lambda(i)$ is non-increasing. ∎

For some λ-optimal pair $(Q_\lambda, \tau(\lambda))$ with $Q_\lambda \in \mathcal{Q}(\lambda)$, let

$$V^\lambda(i) := \mathbb{E}_{i,Q_\lambda}\left[\sum_{n=0}^{\tau(\lambda)-1} c(X_n) + r(X_{\tau(\lambda)})\right] \tag{2.8}$$

and

$$K^\lambda(i) := \mathbb{E}_{i,Q_\lambda}\,\tau(\lambda). \tag{2.9}$$

Lemma 2.2 *For each* $i \in S$, $K^\lambda(i)$ *and* $V^\lambda(i)$ *are non-increasing in* $\lambda(\lambda \geqq 0)$.

Proof. For K^λ, it suffices to show $K^\lambda(i) \geqq K^{\lambda+\delta}(i)$ for any $\lambda \geqq 0$ and $\delta > 0$.

$$\begin{aligned} -\delta K^\lambda(i) &= J^{\lambda+\delta}_{Q_\lambda,\tau(\lambda)}(i) - J^\lambda_{Q_\lambda,\tau(\lambda)} \\ &\leqq J^{\lambda+\delta}_{Q_{\lambda+\delta},\tau(\lambda+\delta)}(i) - J^\lambda_{Q_\lambda,\tau(\lambda)}(i) \\ &\leqq J^{\lambda+\delta}_{Q_{\lambda+\delta},\tau(\lambda+\delta)}(i) - J^\lambda_{Q_{\lambda+\delta},\tau(\lambda+\delta)}(i) \\ &= -\delta K^{\lambda+\delta}(i), \end{aligned}$$

which implies $K^\lambda(i) \geqq K^{\lambda+\delta}(i)$. For the latter, assume that there exists $\delta > 0$ and $\lambda \geqq 0$ with $V^{\lambda+\delta}(i) > V^\lambda(i)$. Then, from the monotonicity of K^λ, it holds that

$$\begin{aligned} J^\lambda(i) &= V^\lambda(i) - \lambda K^\lambda(i) \\ &< V^{\lambda+\delta}(i) - \lambda K^{\lambda+\delta}(i) = J^\lambda_{Q_{\lambda+\delta},\tau(\lambda+\delta)}(i), \end{aligned}$$

which leads to a contradiction. ∎

Lemma 2.3 *It holds that*

(i) *for each* $\lambda \geqq 0$, $\mathcal{Q}(\lambda)$ *is closed and convex.*

(ii) $\mathcal{Q}(\lambda)$ *is upper semi-continuous in* $\lambda \geqq 0$, *i.e., if* $Q_n \in \mathcal{Q}(\lambda_n), \lambda_n \to \lambda$ *and* $Q_n \to Q$ *as* $n \to \infty$, *then* $Q \in \mathcal{Q}(\lambda)$.

Proof. Obviously (i) holds. For (ii), let $Q_n \in \mathcal{Q}(\lambda_n), \lambda_n \to \lambda$ and $Q_n \to Q$ as $n \to \infty$. Then, by the definition, we have

$$c_{Q_n} + Q_n J^{\lambda_n} \geqq c_P^{\lambda_n} + PJ^{\lambda_n} \text{ for all } P \in \mathcal{P}.$$

Applying the generalized dominated convergence theorem(cf. [17, 20]) and Lemma 2.2, as $n \to \infty$ in the above, we get

$$c_Q + QJ^\lambda \geqq c_P^\lambda + PJ^\lambda \text{ for all } P \in \mathcal{P},$$

which implies $Q \in \mathcal{Q}(\lambda)$, as required. ∎

3 An optimal constrained pair

In this section, the existence of a constrained optimal pair is proved. The following theorem shows the validity of the Lagrangian approach to the constrained problem.

Theorem 3.1 *If there exists a non-negative number* $\overline{\lambda}$ *such that*

$$\mathbb{E}_{1,Q_{\overline{\lambda}}}(\tau(\overline{\lambda})) = \alpha \text{ for some } Q_{\overline{\lambda}} \in \mathcal{Q}(\overline{\lambda}), \tag{3.1}$$

$\overline{\lambda}$*-optimal pair* $(Q_{\overline{\lambda}}, \tau(\overline{\lambda}))$ *is an optimal constrained one.*

Proof. In case of $\overline{\lambda} = 0$, the claim holds obviously. Let $\overline{\lambda} > 0$. Then for any $(R, \tau) \in \Delta(\mathcal{G})$, we have:

$$J^{\overline{\lambda}}(1) = V^{\overline{\lambda}}(1) - \overline{\lambda} K^{\overline{\lambda}}(1)$$
$$\geqq V_{R,\tau}(1) - \overline{\lambda} K_{R,\tau}(1) \geqq V_{R,\tau}(1) - \overline{\lambda}\alpha.$$

Thus

$$V^{\overline{\lambda}}(1) \geqq V_{R,\tau}(1) - \overline{\lambda}(\alpha - K^{\overline{\lambda}}(1))$$
$$= V_{R,\tau}(1) \text{ for any } (R, \tau) \in \Delta(\mathcal{G}),$$

which shows that the constrained pair $(Q_{\overline{\lambda}}, \tau(\overline{\lambda}))$ is optimal. ∎

By Theorem 3.1, in order to show the existence of an optimal constrained pair, it is sufficient to prove that there exist the multiplier $\overline{\lambda}$ satisfying (3.1).

To this end, we introduce

$$\gamma := \inf\{\lambda | K^{\lambda}(1) \leqq \alpha\} \tag{3.2}$$

Since $K^{\lambda}(1)$ is non-increasing in $\lambda \geqq 0$, γ is well-defined in (3.2). Here, we need the following assumption.

Assumption (D): (*Slater condition cf.*[14]) There exists a pair $(R, \tau) \in \mathcal{R} \times C(\mathcal{G})$ such that

$$\mathbb{E}_{1,R}(\tau) < \alpha.$$

Lemma 3.1 *Under Assumption* (D), $\gamma < \infty$.

Proof. Suppose that $\gamma = \infty$. By Assumption (D), there exists an $\varepsilon > 0$ with $\mathbb{E}_{1,R}(\tau) < \alpha - \varepsilon$ for some $R \in \mathcal{R}$ and $\tau \in C(\mathcal{G})$. Then, we have:

$$J^{\lambda}(1) \geqq J^{\lambda}_{R,\tau}(1) \geqq V_{R,\tau}(1) - \lambda(\alpha - \varepsilon). \tag{3.3}$$

On the other hand, $\gamma = \infty$ implies $K^{\lambda}(1) > \alpha$ for all $\lambda > 0$, so that $J^{\lambda}(1) < V^{\lambda}(1) - \lambda\alpha$. This means, together with (3.3), that

$$V_{R,\tau}(1) - \lambda(\alpha - \varepsilon) < V^{\lambda}(1) - \lambda\alpha,$$

which leads to

$$V_{R,\tau}(1) + \lambda\varepsilon < V^{\lambda}(1).$$

As $\lambda \to \infty$ in the above, we have $V^{\lambda}(1) \to \infty$, which contradicts that $V^{\lambda}(1)$ is non-increasing in $\lambda \geqq 0$. ∎

Let (λ_n) and (δ_n) be any sequences such that

$$\lambda_n > \lambda_{n+1}, \delta_n < \delta_{n+1} \quad (n \geqq 1)$$
$$\text{and } \lim_{n\to\infty} \lambda_n = \lim_{n\to\infty} \delta_n = \gamma. \tag{3.4}$$

Then, since J^{λ} is non-increasing in λ, we have that

$$\Gamma(\delta_1) \subset \cdots \subset \Gamma(\delta_n) \subset \cdots \subset \Gamma(\lambda_n) \subset \cdots \subset \Gamma(\lambda_1).$$

Here, we can prove the following fact.

Lemma 3.2 *The following holds*:

(i) $\lim_{n\to\infty} \Gamma(\lambda_n) = \Gamma(\gamma)$.

(ii) $\lim_{n\to\infty} \Gamma(\delta_n) \supset \underline{\Gamma}(\gamma)$.

Proof. Clearly (i) holds. For (ii), let $i \in S$ be such that $i \notin \lim_{n\to\infty} \Gamma(\delta_n)$. Then, $r(i) < \max_{P\in\mathcal{P}}(c_P + PJ^{\delta_n})(i)$ for all $n \geqq 1$, which implies $r(i) < (c_{Q_n} + Q_n J^{\delta_n})(i)$ for any $Q_n \in \mathcal{Q}(\delta_n)$ $(n \geqq 1)$. Noting that $\mathcal{P}$ is compact, we can assume that $Q_n \to Q \in \mathcal{P}$ as $n \to \infty$. Applying Lemma 1.1, we get $r(i) \leqq (c_Q + QJ^{\lambda})(i)$. This means $i \notin \underline{\Gamma}(\gamma)$, as required. ∎

The existence of an optimal constrained pair is given in the following.

Theorem 3.2 *Suppose that Assumptions (U) and (D) hold. Then there exists an optimal constrained pair* (R^*, τ^*) *such that* R^* *is stationary policy and* τ^* *is a stationary stopping time determined by* $\{\delta(i)\}$ *with* $\delta(i) = 1$ *if* $i \notin \underline{\Gamma}(\gamma)$ *and* $\delta(i) = 0$ *if* $i \notin \Gamma(\gamma)$ *and requiring randomization in at most one state.*

Proof. For any sequences $(\lambda_n), (\delta_n)$ satisfying (3.4), there exist sequences $(\underline{Q}_n), (\overline{Q}_n)$, such that $\overline{Q}_n \in \mathcal{Q}(\lambda_n), (\underline{Q}_n) \in \mathcal{Q}(\delta_n), K^{\delta_n}(1) = \mathbb{E}_{1,\underline{Q}_n}(\tau_{\Gamma(\delta_n)}) \geqq \alpha, K^{\lambda_n}(1) = \mathbb{E}_{1,\overline{Q}_n}(\tau_{\Gamma(\lambda_n)}) < \alpha \quad (n \geqq 1)$. Noting $\mathcal{P}$ is compact, we can assume that $\underline{Q}_n \to \underline{Q}$ and $\overline{Q}_n \to \overline{Q}$ as $n \to \infty$ for some $\underline{Q}$ and $\overline{Q} \in \mathcal{P}$. By Lemma 2.4, $\underline{Q}, \overline{Q} \in \mathcal{Q}(\gamma)$. Also, from Assumption (U), $Q^N e \to 0$ as $N \to \infty$ for all $Q \in \mathcal{P}$, so that, applying the generalized dominated convergence theorem (cf.[17, 20]), by Lemma 3.2 we get

$$\mathbb{E}_{1,\underline{Q}}(\tau_{\underline{\Gamma}(\gamma)}) \geqq \alpha \qquad \text{and} \tag{3.5}$$
$$\mathbb{E}_{1,\overline{Q}}(\tau_{\Gamma(\gamma)}) \leqq \alpha. \tag{3.6}$$

If at least one of inequalities (3.5) and (3.6) holds in equality, from Theorem 3.1 it follows that there is an optimal constrained pair for state 1.

Suppose that $\mathbb{E}_{1,\underline{Q}}(\tau_{\underline{\Gamma}(\gamma)}) > \alpha$ and $\mathbb{E}_{1,\overline{Q}}(\tau_{\Gamma(\gamma)}) < \alpha$. We must investigate the following two case. In case that $\mathbb{E}_{1,\underline{Q}}(\tau_{\Gamma(\gamma)}) < \alpha$, from Corollary 2.1 there exists randomized stopping time τ determined by $\{\delta(i), i \in S\}$ with $\delta(i) = 1$ if $i \in \underline{\Gamma}(\gamma), = 0$ if $i \notin \Gamma(\gamma)$ and $0 \leqq \delta(i) \leqq 1$ if $\Gamma(\gamma) - \underline{\Gamma}(\gamma)$ and $\mathbb{E}_{1,\underline{Q}}(\tau) = \alpha$, which means from Theorem 3.1 that the constrained pair $(\underline{Q}^{\infty}, \tau)$ is optimal. For this case, obviously τ can be requiring randomization in at most one state. In case that $\mathbb{E}_{1,\underline{Q}}(\tau_{\Gamma(\gamma)}) > \alpha$, noting $\mathbb{E}_{1,\overline{Q}}(\tau_{\Gamma(\gamma)}) < \alpha$, there exists $a \in (0,1)$ such that $\mathbb{E}_{1,a\underline{Q}+(1-a)\overline{Q}}(\tau_{\Gamma(\gamma)}) = \alpha$. Since $\mathcal{Q}(\gamma)$ is convex, $a\underline{Q} + (1-a)\overline{Q} \in \mathcal{P}$, so that a constrained pair $((a\underline{Q} + (1-a)\overline{Q})^{\infty}, \tau_{\Gamma(\gamma)})$ is optimal in state 1. ∎

Using the idea of the OLA policy for the usual stopping problem, we can derive some results. For each

$\lambda \geqq 0$, let

$$\Gamma^*(\lambda) := \{i \in S | r(i) \geqq \max_{P \in \mathcal{P}} (c_P^\lambda + Pr)(i)\} \quad \text{and}$$

$$\underline{\Gamma}^*(\lambda) := \{i \in \Gamma^*(\lambda) | r(i) > \max_{P \in \mathcal{P}} (c_P^\lambda + Pr)(i)\}.$$

Here we introduce an assumption insuring the validity of the OLA stopping time.

Assumption (A_λ): For any $P = (p(i,j)) \in \mathcal{P}, p(i,j) = 0$ if $i \in \Gamma^*(\lambda)$ and $j \notin \Gamma^*(\lambda)$ or $i \in \underline{\Gamma}^*(\lambda)$ and $j \notin \underline{\Gamma}^*(\lambda)$.

Corollary 3.1 *Suppose that Assumptions in Theorem 3.1 hold and Assumption* (A_γ) *holds for* γ *as in* (3.2). *Then, we have*:

(i) $\Gamma(\gamma) = \Gamma^*(\gamma)$ *and* $\underline{\Gamma}(\gamma) = \underline{\Gamma}^*(\gamma)$.

(ii) *Let* $\{\overline{J}(i), i \in S\}$ *satisfy that* $\overline{J}(i) = \max_{P \in \mathcal{P}}(c_P^\gamma + P\overline{J})(i)$ *for* $i \in S$ *and* $\overline{J}(i) = r(i)$ *for* $i \in \Gamma^*(\gamma)$. *Then, for the initial state "1",*

$$\overline{J}(1) = \sup_{(R,\tau) \in \Delta(\mathcal{G})} J_{R,\tau}(1)$$

Here we give a simple example for a Markov deteriorating system with state space $S = \{1, 2, \dots\}$. This system is formulated as follows:

(i) $\mathcal{P} \subset \{P = (p(i,j)) | \sum_{j \in S} p(i,j) = \beta, p(i,j) \geqq 0 \text{ for } i,j \in S\}$ for some $\beta (0 < \beta < 1)$ and $\mathcal{P}$ is convex and compact.

(ii) For any $P = (p(i,j)) \in \mathcal{P}, p(i,j) = 0$ if $i > j$.

(iii) $c_P(i) = -c$ for some $c > 0$.

(iv) The reward function r on S has a property that for each $P \in \mathcal{P}, (Pr - r)(i)$ is non-increasing in $i \in S$.

Under these assumptions, we observe that Assumptions (U) and (D) hold. Also, by simple calculation we find that for $\lambda \geqq 0$ there exists non-negative integer $i_\lambda \leqq \underline{i}_\lambda$ such that $\Gamma^*(\lambda) = [i_\lambda, \infty)$ and $\underline{\Gamma}^*(\lambda) = [\underline{i}_\lambda, \infty)$, so that Assumption (A_λ) hold for all $\lambda \geqq 0$. Thus, for any $\alpha > 0$, from Corollary 3.1 we know that there exists an optimal constrained pair for this system.

References

[1] E. Altman. Denumerable constrained Markov decision process and finite approximations. *Math. Oper. Res.*, 19:169–191, 1994.

[2] E. Altman. *Constrained Markov Decision Processes.* Chapmann & Hall/CRC, 1999.

[3] F. J. Beutler and K. W. Ross. Optimal policies for controlled Markov chains with a constraint. *J. Math. Anal. Appl.*, 112:236–252, 1985.

[4] V. S. Borkar. *Topics in controlled Markov chains.* Pitman Reserch Notes in Mathematics Series, 240. Longman, Harlow, 1991.

[5] Y. S. Chow, H. Robbins, and D. Siegmund. *Great expectations: the theory of optimal stopping.* Houghton Mifflin, Boston, 1976.

[6] E. B. Frid. On optimal strategies in control problems with constraints. *T. Prob. Appl.*, 17:188–192, 1972.

[7] N. Furukawa and S. Iwamoto. Stopped decision processes on complete separable metric spaces. *J. Math. Anal. Appl.*, 31:615–658, 1970.

[8] A. Hordijk. *Dynamic programming and Markov potential theory.* Math. Centre Tracts, No. 51. Math. Centrum, Amsterdam, 1974.

[9] A. Hordijk and L. C. M. Kallenberg. Constrained undiscounted stochastic dynamic programming. *Math. Oper. Res.*, 9:276–289, 1984.

[10] A. Irle. Minimax results and randomization for certain stochastic games. *In: B. Ricceri, S, Stephen(eds.) Minimax Theory and Applications*, pages 91–103, 1998.

[11] Y. Kadota, M. Kurano, and M. Yasuda. Utility-optimal stopping in a denumerable Markov chain. *Bull. Inform. Cybernet.*, 28:15–21, 1996.

[12] Y. Kadota, M. Kurano, and M. Yasuda. On the general utility of discounted Markov decision processes. *Int. Trans. Oper. Res.*, 5:27–34, 1998.

[13] D. P. Kennedy. On a constrained optimal stopping problem. *J. Appl. Prob.*, 19:631–641, 1982.

[14] D. G. Luenberger. *Optimization by vector space methods.* John Wiley & Sons, New York, 1969.

[15] D. C. Nachman. Optimal stopping with a horizon constraint. *Math. Oper. Res.*, 5:126–134, 1980.

[16] S. M. Ross. *Applied probability Models with Optimization Applications.* Holden-Day, 1970.

[17] H. L. Royden. *Real Analysis.* Macmillan, New York, 2nd edition, 1968.

[18] L. I. Sennott. Constrained discounted Markov decision chains. *Prob. Eng. Inform. Sci.*, 5:463–475, 1991.

[19] L. I. Sennott. Constrained average cost Markov decision chains. *Prob. Eng. Inform. Sci.*, 7:69–83, 1993.

[20] L. I. Sennott. *Stochastic Dynamic Programming and the Control of Queuing Systems.* John Wiley & Sons, 1999.