Proceedings of the 30th Conference
on Decision and Control
Brighton, England • December 1991

Perturbation Theory for Semi-Markov Control Problems

Mohammed Abbad and Jerzy A. Filar
Department of Mathematics and Statistics
University of Maryland at Baltimore County
Baltimore, Maryland

1. Introduction

In the papers [1] and [2], we considered perturbation of systems undergoing Markov processes in which the times between two consecutive decision time points were equidistant. In this paper we consider perturbations of processes for which the times between transitions are random variables. These are called semi-Markov processes and they were introduced by De Cani [6], Howard [18], Jewell [19] and Schweitzer [27].

2. Definitions and Preliminaries

A Semi-Markov Control Process (SMCP, for short) is observed at decision time points $t = 0, 1, \ldots$; starting at $t = 0$. At each decision time point the system is in one of a finite number of states and an action has to be chosen.

Let $S := \{1, 2, \ldots, N\}$ be the state space, and for each $s \in S$ let $A(s)$ be the finite set of possible actions in state s.

If the system is in state $s \in S$ and an action $a \in A(s)$ is chosen, then the following occurs independently of the history of the process:

(i) The next state s' of the process is chosen according to the transition probability $p(s'|s,a)$.

(ii) Conditional on the event that the next state is s', the time until the transition from s to s' occurs is a random variable with probability distribution $F(.|s,a,s')$.

(iii) If the next decision time point falls after τ units of times, then the reward in this epoch is denoted by $r(\tau, s, a)$.

The transition law p satisfies:

$$\begin{aligned} p(s'|s,a) &\geq 0;\ s, s' \in S;\ a \in A(s);\ and \\ \sum_{s' \in S} p(s'|s,a) &= 1,\ s \in S,\ a \in A(s). \end{aligned}$$

A decision rule π^t at time t is a function which assigns a probability to the event that any particular action is taken at time t. In general π^t may depend on all realized states up to time t, and on all realized actions up to time $t-1$.

Let $h_t = (s_0, a_0, s_1, \ldots, a_{t-1}, s_t)$ be the history up to time t where $a_0 \in A(s_0), \ldots, a_{t-1} \in A(s_{t-1})$, then $\pi^t(h_t, .)$ is a probability distribution on $A(s_t)$, that is, $\pi^t(h_t, a_t)$ is the probability of selecting the action a_t at time t, given the history h_t.

A strategy π is a sequence of decision rules $\pi = (\pi^0, \pi^1, \ldots, \pi^t, \ldots)$.

A Markov strategy is one in which π^t depends only on the current state at time t.

A stationary strategy is a Markov strategy with identical decision rules.

A deterministic strategy is a stationary strategy whose single decision rule is nonrandomized.

Let C, $C(S)$ and $C(D)$ denote the sets of all strategies, all stationary strategies and all deterministic strategies respectively.

For any $t = 0, 1, \ldots$; let X_t, and Y_t, denote the observed state and the chosen action at time point t respectively.

3. Discounted Case

In this Section we shall assume that the rewards are continuously discounted, that is, a reward r incurred at time t is worth only $re^{-\alpha t}$ at time 0, where α is a fixed positive real number.

In order to insure that an infinite number of transitions does not occur in a finite interval, we shall assume throughout that the following condition holds:

For all $s, s' \in S$, and $a \in A(s)$,

$$\int_0^\infty e^{-\alpha t} dF(t|s,a,s') < 1. \qquad (3.1)$$

For any strategy $\pi \in C$ and any initial state $s \in S$, we define the expected discounted reward $V(s,\pi)$ by:

$$V(s,\pi) := E_\pi[\sum_{n=0}^{\infty} e^{-\alpha(\tau_1+\tau_2+\ldots+\tau_{n-1})} r(\tau_n, X_n, Y_n) | X_0 = s] \qquad (3.2)$$

where $\tau_1 + \tau_2 + \ldots + \tau_{n-1} := 0$ for $n = 0$, and τ_n is the time between the n-th and the (n+1)-st transition.

The discounted semi-Markov control problem is defined by the following optimization problem:

$$V(s) := max_{\pi \in C} V(s,\pi), \quad s \in S. \qquad (3.3)$$

A strategy π^0 is called optimal if for all $s \in S$,

$$V(s, \pi^0) = V(s).$$

It is well known that there exists an optimal deterministic strategy and there are a number of finite algorithms for its computation (e.g., see Denardo and Fox [9], Jewell [19], Kallenberg [20], Ross [25]).

For every $s, s' \in S$ and $a \in A(s)$, we denote by $f(t|s,a,s')$ the probability density of the probability distribution $F(t|s,a,s')$.

We define: for all s, $s' \in S$ and $a \in A(s)$

$$\begin{aligned} \bar{r}(s,a) &:= \sum_{s' \in S} p(s'|s,a) \int_0^\infty r(t,s,a) f(t|s,a,s') dt, \quad (3.4) \\ \bar{p}(s'|s,a) &:= p(s'|s,a) \int_0^\infty e^{-\alpha t} f(t|s,a,s') dt. \quad (3.5) \end{aligned}$$

The following two results can be derived (using analogous proofs) from Kallenberg [20] or Ross [25].

Lemma 3.1 *For any deterministic strategy $\pi \in C(D)$ and $s \in S$,*

$$V(s,\pi) = \bar{r}\big(s,\pi(s)\big) + \sum_{s' \in S} \bar{p}(s'|s,a)V(s',\pi).$$

Lemma 3.2 *For any state $s \in S$,*

$$V(s) = \max_{a \in A(s)} \Big\{\bar{r}(s,a) + \sum_{s' \in S} \bar{p}(s'|s,a)V(s')\Big\}.$$

Let Γ be the Markov control process defined by:

$$\Gamma =< S, \{A(s), s \in S\}, \bar{p}, \bar{r} > .$$

We define:

$$\lambda := \max\Big\{\int_0^\infty e^{-\alpha t} f(t|s,a,s')dt \mid s,s' \in S;\ a \in A(s)\Big\}.$$

Remark 3.1 *Note that from the condition (3.1), it follows that $\lambda < 1$. By using the definition (3.5) of $\bar{p}$, we have that for any $s \in S$ and $a \in A(s)$,*

$$\begin{aligned}
1 - \sum_{s' \in S} \bar{p}(s'|s,a) &= 1 - \sum_{s' \in S} p(s'|s,a) \int_0^\infty e^{-\alpha t} f(t|s,a,s')dt \\
&\geq 1 - \lambda \\
&> 0.
\end{aligned}$$

From Remark (3.1), it follows that Γ is a Markov control process with nonzero stop probabilities. This class of MCP's has been studied in a more general context (stochastic games) by Shapley [28].

From Lemma 3.2, we derive that for any $s \in S$, $V(s)$ can be interpreted as the optimal value in state s for the MCP Γ.

If we define,

$$\|r\| := sup\Big\{|r(t,s,a)| \ : t \geq 0, s \in S, a \in A(s)\Big\},$$

which is assumed to be finite, then by the definition (3.4) of $\bar{r}$ we have that

$|\bar{r}(s,a)| \leq \|r\|$ for all $s \in S$ and $a \in A(s)$.

Now, from Shapley [28], we have the following result:

Lemma 3.3 *For any $s \in S$,*

$$|V(s)| \leq \frac{\|r\|}{1-\lambda}.$$

We shall now consider the situation where the transition probabilities and the probability density of the original SMCP are perturbed slightly.

Towards this goal we shall define: for all $s, s' \in S$; $a \in A(s)$ and $t \geq 0$,

$$\begin{aligned}
p_q(s'|s,a) &:= p(s'|s,a) + q(s'|s,a), &(3.6)\\
f_g(t|s,a,s') &:= f(t|s,a,s') + g(t|s,a,s'), &(3.7)
\end{aligned}$$

where q and g are the disturbance laws on the transition probabilities and the probability density respectively.

We define:

$$\begin{aligned}
\|q\| &:= max\Big\{|q(s'|s,a)| \ : \ s,s' \in S; a \in A(s)\Big\},\\
\|g\| &:= sup\Big\{|g(t|s,a,s')| \ : \ s,s' \in S; a \in A(s); t \geq 0\Big\}.
\end{aligned}$$

We assume that there exists $\epsilon_0 > 0$ such that for all q and g satisfying $\|q\| < \epsilon_0$ and $\|g\| < \epsilon_0$, p_q and f_g define a transition law and a probability density respectively.

We now have a family of perturbed semi-Markov control processes that differ from the original SMCP only in the transition law and the probability density. Namely in the perturbed SMCP the transition law and the probability density are defined by p_q and f_g respectively.

The discounted semi-Markov control problem corresponding to the perturbed SMCP defined by the transition law p_q and the probability density f_g is the optimization problem:

$$V_{qg}(s) := max_{\pi \in C} V_{qg}(s,\pi) \ , \quad s \in S$$

where $V_{qg}(s,\pi)$ is defined in the perturbed SMCP in the same way as $V(s,\pi)$ was defined in the original SMCP.

We define the following quantities:

For all s, $s' \in S$, $a \in A(s)$,

$$\begin{aligned}
\bar{p}_{qg}(s'|s,a) &:= p_q(s'|s,a) \int_0^\infty e^{-\alpha t} f_g(t|s,a,s')dt, &(3.8)\\
\bar{r}_{qg}(s,a) &:= \sum_{s' \in S} p_q(s'|s,a) \int_0^\infty r(t,s,a) f_g(t|s,a,s')dt &(3.9)
\end{aligned}$$

From the definitions (3.9) and (3.4) of $\bar{r}_{qg}$ and $\bar{r}$ respectively, it follows that: for any $s \in S$ and $a \in A(s)$,

$$\bar{r}_{qg}(s,a) - \bar{r}(s,a) = \sum_{s' \in S} p_q(s'|s,a) \int_0^\infty r(t,s,a) f_g(t|s,a,s')dt -$$

$$\sum_{s' \in S} p(s'|s,a) \int_0^\infty r(t,s,a) f(t|s,a,s')dt = \sum_{s' \in S} p(s'|s,a) \int_0^\infty$$

$$r(t,s,a) g(t|s,a,s')dt + \sum_{s' \in S} q(s'|s,a) \int_0^\infty r(t,s,a) f_g(t|s,a,s')dt,$$

where the last equality follows from the definitions (3.6) and (3.7) of p_q and f_g respectively.

Note that for any translation of the function r, of the form $r(t,s,a) + T(s,a)$, $s \in S$ and $a \in A(s)$, the quantity $\bar{r}_{qg}(s,a) - \bar{r}(s,a)$ remains constant.

Now, we introduce the following assumption: there exists some function $T(s,a)$ such that for any $s \in S$ and $a \in A(s)$,

$$\int_0^\infty |r(t,s,a) - T(s,a)|dt \quad is \quad finite. \tag{3.10}$$

If we define $|r| := max\Big\{\int_0^\infty |r(t,s,a) - T(s,a)|dt \ : \ s \in S \ a \in A(s)\Big\}$, then it follows that for any $s \in S$ and $a \in A(s)$,

$$|\bar{r}_{qg}(s,a) - \bar{r}(s,a)| \leq \|g\||r| + N\|q\|\|r\|. \tag{3.11}$$

Remark 3.2 *Note that the assumption (3.10) is satisfied by the reward structure considered in Kallenberg [20] and Ross [25]. That is, if the process is in state s and an action $a \in A(s)$ is selected and the next transition falls after t units of times, then the reward in this epoch is given by $R(s,a) + S(s,a)t$, where $R(s,a)$ is the immediate reward and $S(s,a)$ is the reward rate. In this case, the rewards in our formulation are given by:*

$$\begin{aligned}
r(t,s,a) &= R(s,a) + S(s,a) \int_0^t e^{-\alpha x} dx\\
&= R(s,a) + \frac{1}{\alpha} S(s,a) - \frac{1}{\alpha} S(s,a) e^{-\alpha t}.
\end{aligned}$$

Hence if we define for all $s \in S$ and $a \in A(s)$, $T(s,a) := R(s,a) + \frac{1}{\alpha} S(s,a)$ then

$$\int_0^\infty |r(t,s,a) - T(s,a)|dt = \frac{1}{\alpha}|S(s,a)| \int_0^\infty e^{-\alpha t} dt,$$

which is finite.

Let $\lambda_g := \max\left\{ \int_0^\infty e^{-\alpha t} f_g(t|s,a,s')dt \mid s,s' \in S;\ a \in A(s)\right\}$.
By using the definition (3.7) of f_g, it follows that:

$$\lambda_g \leq \lambda + \frac{\|g\|}{\alpha}$$

Note that if $\|g\| < \alpha(1-\lambda)$, then $\lambda_g < 1$.
Now, from Lemma 3.3, it follows that for all $s \in S$:

$$|V_{qg}(s)| \leq \frac{\alpha\|r\|}{\alpha(1-\lambda) - \|g\|}. \tag{3.12}$$

In what follows we use the following notation:

$$\|V_{qg}(.,\pi) - V(.,\pi)\| := max_{s\in S}|V_{qg}(s,\pi) - V(s,\pi)|.$$

Lemma 3.4 *Assume that* $\|g\| < \alpha(1-\lambda)$. *Then for every* $s \in S$ *and* $\pi \in C(D)$,

$$|V_{qg}(s,\pi) - V(s,\pi)| \leq \frac{1}{1-\lambda}\left\{|r| + \frac{\|r\|}{\alpha(1-\lambda)-\|g\|}\right\}\|g\|$$

$$+\frac{N\|r\|\|q\|}{(1-\lambda)}\left\{1 + \frac{1}{\alpha(1-\lambda)-\|g\|}\right\}.$$

Proof: By Lemma 3.1, for all $s \in S$ and $\pi \in C(D)$ we have that:

$$V_{qg}(s,\pi) - V(s,\pi) = \left\{\bar{r}_{qg}\big(s,\pi(s)\big) - \bar{r}\big(s,\pi(s)\big)\right\} + \sum_{s'\in S} \bar{p}_{qg}\big(s'|s,\pi(s)\big)V_{qg}(s',\pi) - \sum_{s'\in S}\bar{p}\big(s'|s,\pi(s)\big)V(s',\pi). \tag{3.13}$$

From the definitions (3.6), (3.7) and (3.8) of p_q, f_g and $\bar{p}_{qg}$ respectively, it follows that for all $s, s' \in S$:

$$\bar{p}_{qg}\big(s'|s,\pi(s)\big) := p_q\big(s'|s,\pi(s)\big)\int_0^\infty e^{-\alpha t} f_g\big(t|s,\pi(s),s'\big)dt =$$

$$\bar{p}\big(s'|s,\pi(s)\big) + p\big(s'|s,\pi(s)\big)\int_0^\infty e^{-\alpha t} g\big(t|s,\pi(s),s'\big)dt +$$

$$q\big(s'|s,\pi(s)\big)\int_0^\infty e^{-\alpha t} f_g\big(t|s,\pi(s),s'\big)dt.$$

Hence, for all $s \in S$, we have:

$$\sum_{s'\in S}\bar{p}_{qg}\big(s'|s,\pi(s)\big)V_{qg}(s',\pi) - \sum_{s'\in S}\bar{p}\big(s'|s,\pi(s)\big)V(s',\pi) =$$

$$\sum_{s'\in S}\bar{p}\big(s'|s,\pi(s)\big)\left\{V_{qg}(s',\pi) - V(s',\pi)\right\} + \sum_{s'\in S}\left\{p\big(s'|s,\pi(s)\big)\int_0^\infty e^{-\alpha t}\right.$$

$$\left. g\big(t|s,\pi(s),s'\big)dt + q\big(s'|s,\pi(s)\big)\int_0^\infty e^{-\alpha t} f_g\big(t|s,\pi(s),s'\big)dt\right\}V_{qg}(s',\pi).$$

Now, by using (3.13), it follows that for all $s \in S$,

$$|\sum_{s'\in S}\bar{p}_{qg}\big(s'|s,\pi(s)\big)V_{qg}(s',\pi) - \sum_{s'\in S}\bar{p}\big(s'|s,\pi(s)\big)V(s',\pi)|$$

$$\leq \lambda\|V_{qg}(.,\pi) - V(.,\pi)\| + \frac{\|r\|}{\alpha(1-\lambda)-\|g\|}\left\{\|g\| + N\|q\|\right\}. \tag{3.14}$$

Finally, by using (3.11), (3.13) and (3.14) we get:

$$\|V_{qg}(.,\pi) - V(.,\pi)\| \leq \|g\||r| + N\|q\|\|r\| +$$

$$\lambda\|V_{qg}(.,\pi) - V(.,\pi)\| + \frac{\|r\|}{\alpha(1-\lambda)-\|g\|}\left\{\|g\| + N\|q\|\right\}$$

which implies that:

$$(1-\lambda)\|V_{qg}(.,\pi) - V(.,\pi)\| \leq \left\{|r| + \frac{\|r\|}{\alpha(1-\lambda)-\|g\|}\right\}\|g\| +$$

$$N\|r\|\|q\|\left\{1 + \frac{1}{\alpha(1-\lambda)-\|g\|}\right\}.$$

This completes the proof of the Lemma.

□

Theorem 3.1 *Let* π^0 *be any optimal deterministic strategy in the original SMCP. Then for all* $\beta > 0$, *there exists* $\epsilon_\beta > 0$ *such that for all* q *and* g *satisfying* $\|q\| < \epsilon_\beta$ *and* $\|g\| < \epsilon_\beta$, $\|V_{qg}(.,\pi^0) - V_{qg}(.)\| < \beta$.

Proof: For any $s \in S$, we have:

$$\begin{aligned}|V(s,\pi^0) - V_{qg}(s)| &= |max_{\pi\in C(D)}V(s,\pi) - max_{\pi\in C(D)}V_{qg}(s,\pi)| \\ &\leq max_{\pi\in C(D)}|V(s,\pi) - V_{qg}(s,\pi)|\end{aligned}$$

where the last expression converges to 0 as $\|q\|$ and $\|g\|$ go to 0 by Lemma 3.4 .
Now, by using the triangle inequality:

$$\|V_{qg}(.,\pi^0) - V_{qg}(.)\| \leq \|V_{qg}(.,\pi^0) - V(.,\pi^0)\| + \|V(.,\pi^0) - V_{qg}(.)\|,$$

the proof of the Theorem is completed.

□

4. Limiting Average Case

For any strategy $\pi \in C$ and any initial state $s \in S$, the limiting average reward $J(s,\pi)$ is defined by

$$J(s,\pi) := liminf_{T\to\infty}\frac{1}{T}J_T(s,\pi) \tag{4.1}$$

where $J_T(s,\pi)$ denotes the expected reward earned in the interval $[0,T)$ when the strategy π is used and the initial state is s. That is :

$$J_T(s,\pi) := E_\pi[\sum_{n=0}^{n(T)} r(\tau_n, X_n, Y_n)|X_0 = s],$$

where, $n(T) := max\{n \mid \tau_0 + \tau_1 + \ldots + \tau_n < T\}$.
For any $s \in S$ and $a \in A(s)$, the holding time and the immediate reward are defined respectively by:

$$\tau(s,a) := \sum_{s'\in S} p(s'|s,a)\int_0^\infty tf(t|s,a,s')dt \tag{4.2}$$

and

$$c(s,a) := r\big(\tau(s,a),s,a\big). \tag{4.3}$$

Throughout this Section, it is assumed that: for all $s \in S$ and $a \in A(s)$

$$0 < \tau(s,a) < \infty. \tag{4.4}$$

The limiting average semi-Markov control problem is defined by the following optimization problem:

$$J(s) := max_{\pi\in C}J(s,\pi)\ , \quad s \in S. \tag{4.5}$$

A strategy π^0 is called optimal if,

$$J(s,\pi^0) = J(s)\ for\ all\ s \in S.$$

It is well known that there exists an optimal deterministic strategy and there are a number of finite algorithms for its computation (e.g., see Federgruen [14], Jewell [19], Kallenberg [20], Ross [25]).

We note that any limiting average semi-Markov control problem can be described by:

$$\Lambda := < S,\ \{A(s),\ s \in S\},\ p,\ \tau,\ c\ >$$

We shall consider the following transformation which was proposed by Schweitzer [27]: for all $s,\ s' \in S$ and $a \in A(s)$

$$\tilde{r}(s,a) \ := \ \frac{c(s,a)}{\tau(s,a)}, \tag{4.6}$$

$$\tilde{p}(s'|s,a) \ := \ \delta_{s's} + \big(p(s'|s,a) - \delta_{s's}\big)\frac{\eta}{\tau(s,a)}, \tag{4.7}$$

where η is chosen such that,

$$0 < \eta < min\Big\{\frac{\tau(s,a)}{1 - p(s|s,a)} \mid s \in S,\ a \in A(s),\ p(s|s,a) < 1\Big\}.$$

To the semi-Markov control problem Λ we can associate the Markov control problem $\tilde{\Gamma}$ defined by:

$$\tilde{\Gamma} := < S,\ \{A(s),\ s \in S\},\ \tilde{p},\ \tilde{r}\ > .$$

Remark 4.1 *It is shown (e.g. Federgruen [14], Schweitzer [27]) that any optimal deterministic strategy in the SMCP Λ is also optimal in the MCP $\tilde{\Gamma}$ and vice-versa, and the optimal values in both problems Λ and $\tilde{\Gamma}$ are equal. It is also true that the values of any deterministic strategy in the SMCP Λ and the MCP $\tilde{\Gamma}$ are equal.*

Remark 4.2 *From (4.7), it follows that if $s' \neq s$ then $\tilde{p}(s'|s,a) = p(s'|s,a)\frac{\eta}{\tau(s,a)}$ for all $a \in A(s)$. Hence the SMCP Λ and the MCP $\tilde{\Gamma}$ have the same ergodicity structure, that is, for any stationary strategy π, $P(\pi)$ and $\tilde{P}(\pi)$ have the same ergodic classes, where $P(\pi)$ and $\tilde{P}(\pi)$ are the transition matrices in the SMCP Λ and the MCP $\tilde{\Gamma}$ respectively when the strategy π is used.*

Perturbation on the transition probabilities

In this case we shall consider the situation where the transition probabilities of the SMCP Λ are perturbed slightly.

Towards this goal we shall define:

$$p_\epsilon(s'|s,a) := p(s'|s,a) + \epsilon d(s'|s,a);\ s,s' \in S;\ a \in A(s). \tag{4.8}$$

We shall require that there exists $\epsilon_0 > 0$ such that for every $\epsilon \in [0,\epsilon_0]$, p_ϵ is a transition law.

We shall consider a family of perturbed processes $\big\{\Lambda_\epsilon \mid \epsilon \in [0,\epsilon_0]\big\}$ that differ from the original SMCP Λ only in the transition law, namely, in the SMCP Λ_ϵ the transition law is p_ϵ.

For each $\epsilon \in [0,\epsilon_0]$, the associated MCP $\tilde{\Gamma}_\epsilon$ to the SMCP Λ_ϵ is defined by:

$$\tilde{\Gamma}_\epsilon := < S,\ \{A(s)\ :\ s \in S\},\ \tilde{p}_\epsilon,\ \tilde{r}\ >,$$

where $\tilde{p}_\epsilon$ is defined by:

$$\tilde{p}_\epsilon(s'|s,a) := \delta_{s's} + \big(p_\epsilon(s'|s,a) - \delta_{s's}\big)\frac{\eta_\epsilon}{\tau(s,a)};\ s,s' \in S;\ a \in A(s), \tag{4.9}$$

where η_ϵ is chosen such that,

$$0 < \eta_\epsilon < min\Big\{\frac{\tau(s,a)}{1 - p_\epsilon(s|s,a)} \mid s \in S,\ a \in A(s),\ p_\epsilon(s|s,a) < 1\Big\}. \tag{4.10}$$

Note that for ϵ small, η must satisfy (4.10) and hence without lost of generality we can choose $\eta_\epsilon = \eta$.

From (4.8) and (4.9), it follows that for any $s, s' \in S$ and $a \in A(s)$,

$$\tilde{p}_\epsilon(s'|s,a) := \tilde{p}(s'|s,a) + \frac{\epsilon\eta}{\tau(s,a)}d(s'|s,a). \tag{4.11}$$

Note that, from (4.11) it results that the MCP's $\tilde{\Gamma}_\epsilon$ are exactely the perturbed MCP's of the MCP $\tilde{\Gamma}$ under the disturbance law:

$$\tilde{d}(s'|s,a) := \frac{\eta d(s'|s,a)}{\tau(s,a)};\ s,s' \in S;\ a \in A(s).$$

Now, we can use the results in [1] and [2] to the MCP $\tilde{\Gamma}$ to derive some results concerning the perturbation of the SMCP Λ.

In particular, we can derive the limit control principle (Theorem (4.1)) for the limiting average semi-Markov control processes.

For any strategy $\pi \in C$, we denote by $J_\epsilon(s,\pi)$ the limiting average reward resulting from the use of π in the SMCP Λ_ϵ and when the starting state is s.

The optimal value function J_ϵ corresponding to the SMCP Λ_ϵ is given by:

$$J_\epsilon(s) := max_{\pi \in C} J_\epsilon(s,\pi).$$

Let $\tilde{L}$ denote the limit Markov control problem corresponding to the perturbed MCP's $\tilde{\Gamma}_\epsilon$, $\epsilon \in (0,\epsilon_0]$

Theorem 4.1 *Let $\pi^0 \in C(D)$ be any optimal strategy in $\tilde{L}$. Then for all $\delta > 0$, there exists $\epsilon_\delta > 0$ such that for all $\epsilon \in (0,\epsilon_\delta)$,*

$$\|J_\epsilon(.,\pi^0) - J_\epsilon(.)\| < \delta.$$

Proof: This follows from Corollary 3.1 in [2] and Remark 4.1.

□

Remark 4.3 *From Remark 4.2, it follows that if the SMCP Λ is completely decomposable then the transformed MCP $\tilde{\Gamma}$ is also completely decomposable. Thus from the expressions (4.8) and (4.11) we can conclude that the algorithms constructed in [1] are also valid in this case.*

References

[1] M. Abbad, T. Bielecki and J.A. Filar, Algorithms for Singularly Perturbed Limiting Average Markov Control Problems, Proceedings of the 29th CDC, editor IEEE, 1990.

[2] M. Abbad and J.A. Filar, Perturbation and Stability Theory for Markov Control Problems, Technical Report 90-13, University of Maryland at Baltimore County, 1990, (Accepted by IEEE Transactions on Automatic Control).

[3] R.Aldhaheri and H.Khalil, Aggregation and Optimal Control of Nearly Completely Decomposable Markov Chains, in Proceedings of the 28th CDC, editor IEEE, 1989.

[4] T.Bielecki and J.A.Filar, Singular Perturbations of Markov Decision Chains, Proceedings of the 28th CDC, editor IEEE, 1989.

[5] M. Cordech, A. Willsky, S. Sastry and D. Castanon, Hierarchical Aggregation of Linear Systems with Multiple Time Scales, IEEE Transaction on Automatic Control, AC-28, pp. 1017-1029, 1983.

[6] J.S. De Cani, A Dynamic Programming Algorithm for Embedded Markov Chains when the Planning Horizon is at Infinity, Management Science, 10, pp. 716-733, 1964.

[7] F. Delebecque, A Reduction Process for Perturbed Markov Chains, SIAM Journal of Applied Mathematics, 48, pp.325-350, 1983.

[8] F. Delebecque and J. Quadrat, Optimal Control of Markov Chains Admitting Strong and Weak Interactions, Automatica, 17, pp. 281-296, 1981.

[9] E.V. Denardo and B. Fox, Multichain Markov Renewal Programs, SIAM J. Appl. Math., 16, pp. 468-487.

[10] E.V. Denardo, Dynamic Programming, Prentice-Hall, Eglewood Cliffs, New Jersey, 1982.

[11] C. Derman, Finite State Markovian Decision Process, Academic Press, New York, 1970.

[12] N.V. Dijk, Perturbation Theory for Unbounded Markov Reward Processes with Applications to Queueing, Adv. Appl. Prob., 20, pp. 99-111, 1988.

[13] N.V. Dijk and M. Puterman, Perturbation Theory for Markov Reward Processes with Applications to Queueing Systems, Adv. Appl. Prob., 20, pp. 79-98, 1988.

[14] A. Federgruen, Markovian Control Problems, Mathematical Centre Tracts 97, Amsterdam, 1983.

[15] B.L. Fox, Markov Renewal Programming by Linear Fractional Programming, SIAM J. Appl. Math., 14, pp. 1418-1432, 1966.

[16] V.G. Gaitsgori and A.A. Pervozvanskii, Theory of Suboptimal Decisions, Kluwer Academic Publishers, 1988.

[17] R.A. Howard, Dynamic Programming and Markov Processes, M.I.T.Press, Cambridge, Massachusetts, 1960.

[18] R.A. Howard, Semi-Markovian Decision Processes, Proceedings International Statistical Institute, Ottawa, 1963.

[19] W. Jewell, Markov Renewal Programming, Op. Res., 11, pp. 938-971, 1963.

[20] L.C.M. Kallenberg, Linear Programming and Finite Markovian Control Problems, Mathematical Centre Tracts 148, Amsterdam, 1983.

[21] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, Berlin, 1980.

[22] P. Kokotovic, Application of Singular Perturbation Techniques to Control Problems, SIAM Review, 26, pp. 501-550, 1984.

[23] R.G. Phillips and P. Kokotovic, A Singular Perturbation Approach to Modelling and Control of Markov Chains, IEEE Transactions on Automatic Control, AC-26, pp. 1087-1094, 1981.

[24] J. Rohlicek and A.Willsky, Multiple Time Scale Decomposition of Discrete Time Markov Chains, Systems and Control Letters, 11, pp. 309-314, 1988.

[25] S.M. Ross, Introduction to Stochastic Dynamic Programming, Academic Press, New York, 1983.

[26] P.J. Schweitzer, Perturbation Series Expansions for Nearly Completely Decomposable Markov Chains, Teletrafic Analysis and Computer Performance Evaluation, pp. 319-328, 1986.

[27] P.J. Schweitzer, Iterative Solution of the Functional equations of Undiscounted Markov Renewal Programming, Journal of Mathematical Analysis and Applications, 34, pp. 495-501, 1971.

[28] L.S. Shapley, Stochastic Games, Proceedings National Academy of Sciences U.S.A, 39, pp. 1095-1100, 1953.