Journal of the Operations Research
Society of Japan
Vol. 42, No. 3, September 1999

NON-DISCOUNTED OPTIMAL POLICIES IN CONTROLLED MARKOV SET-CHAINS

Masanori Hosaka Masami Kurano
Chiba University

(Received January 29, 1998; Final April 30, 1999)

Abstract In a controlled Markov set-chain with a discount factor β, we consider the case of $\beta = 1$ as a limiting case of $\beta < 1$ and find a non-discounted optimal policy which maximizes Abel-sum of rewards in time over all stationary policies under some partial order. We analyze the behavior of discounted total rewards as discount factor β approaches 1 under regularity conditions, and prove the existence of a non-discounted optimal policy, applying the Kakutani's fixed point theorem and policy improvement method. As a numerical example the Toymaker's problem is considered.

1. Introduction

Discrete-time stochastic processes, known as Markov decision processes (MDPs), have been well studied (cf.[1][2][10][11][16]). In those mathematical models the required data is assumed to be exact while in practice this data is estimated. Thus, the mathematical model of MDPs can only be viewed as approximations.

So, it may be useful that we ameliorate the model of MDPs so as to be more "robust" than the traditional MDPs in the sense that it is reasonably efficient in rough approximations. A more realistic way to consider such a problem is to use intervals which contain the required data. Hartfiel [5][6][7][8][9] applied this interval technique to Markov chains and studied Markov set-chains, where the transition matrix is allowed to change with time in a given interval.

As a model which is robust for rough approximation of the transition matrix in MDPs, Kurano et al [14] has introduced a decision model, called a controlled Markov set-chain, based on Markov set-chains developed by Hartfiel (cf.[9]), and discussed the optimization of the discounted expected rewards under some partial order. However, the case that the discount factor β approaches 1 was not treated there. The objective of this paper is to find a policy, called non-discounted optimal, which maximizes Abel-sum of rewards in time under some partial order. We analyze the behavior of discounted total rewards as β approaches 1 under regularity conditions for a Markov set-chain induced by a policy. A non-discounted optimal stationary policy is shown to exist by constructive proofs. Also, the Toymaker's problem in Howard [11] is numerically considered to explain how the theoretical results are used in a computation.

We notice Takahashi's [19][20] work. He has introduced weak D-Markov chains to consider bounds for state probabilities of aggregated chains of large scale Markov chains, and discussed its applicability to tandem queueing networks. The idea of weak D-Markov chains is essentially same as that of a series of papers by Hartfiel, and their studies have been done independently. The results of Takahashi [19][20] are closely related to ours, which will be clear in the sequel.

In the remainder of this section we shall give some notation referring to the work on Markov set-chains and interval arithmetic [15] and formulate a controlled Markov set-chain which will be examined in the sequel.

We adopt the notation in [8][14][15]. Let R, R^n and $R^{n\times m}$ be the sets of real numbers, real n-dimensional column vectors and real $n \times m$ matrices, respectively. We shall identify $n \times 1$ matrices with vectors and 1×1 matrices with real numbers, so that $R = R^{1\times 1}$ and $R^n = R^{n\times 1}$. Also, we denote by R_+, R_+^n and $R_+^{n\times m}$ the subsets of entrywise non-negative elements in R, R^n and $R^{n\times m}$, respectively.

We equip $R^{n\times m}$ with the componentwise relations $\leq, <, \geq, >$. For any $\underline{A} = (\underline{a}_{ij})$, $\overline{A} = (\overline{a}_{ij})$ in $R_+^{n\times m}$ with $\underline{A} \leq \overline{A}$, we define the set of stochastic matrices, $\langle \underline{A}, \overline{A} \rangle$, by

$$\langle \underline{A}, \overline{A} \rangle = \{A \mid A = (a_{ij}) \text{ is an } n \times m \text{ stochastic matrix with } \underline{A} \leq A \leq \overline{A}\}.$$

Let

$$\mathcal{M}_n := \{\mathcal{A} = \langle \underline{A}, \overline{A} \rangle \mid \langle \underline{A}, \overline{A} \rangle \neq \emptyset, \underline{A} \leq \overline{A} \text{ and } \underline{A}, \overline{A} \in R_+^{n\times n}\}.$$

The product of $\mathcal{A}$ and $\mathcal{B} \in \mathcal{M}_n$ is defined by

$$\mathcal{AB} = \{AB | A \in \mathcal{A}, B \in \mathcal{B}\}.$$

For any sequence $\{\mathcal{A}_i\}_{i=1}^{\infty}$ with $\mathcal{A}_i \in \mathcal{M}_n$ $(i \geq 1)$, we define the multiproduct inductively by

$$\mathcal{A}_1\mathcal{A}_2 \cdots \mathcal{A}_k := (\mathcal{A}_1 \cdots \mathcal{A}_{k-1})\mathcal{A}_k \ (k \geq 2).$$

Denote by $C(R_+)$ the set of all bounded and closed intervals in R_+. Let $C(R_+)^n$ be the set of all n-dimensional column vectors whose elements are in $C(R_+)$, i.e.,

$$C(R_+)^n = \{D = (D_1, D_2, \ldots, D_n)' | D_i \in C(R_+) \ (1 \leq i \leq n)\}.$$

where d' denotes the transpose of a vector d.

The following arithmetics are used in Section 2. For $D = (D_1, D_2, \ldots, D_n)'$, $E = (E_1, E_2, \ldots, E_n)' \in C(R_+)^n$, $h \in R_+^n$ and $\lambda \in R_+$, $D + E = \{d + e \mid d \in D, e \in E\}$, $\lambda D = \{\lambda d \mid d \in D\}$ and $h + D = \{h + d \mid d \in D\}$.

If $D = ([\underline{d}_1, \overline{d}_1], \ldots, [\underline{d}_n, \overline{d}_n])'$, D will be denoted by $D = [\underline{d}, \overline{d}]$, where $\underline{d} = (\underline{d}_1, \ldots, \underline{d}_n)'$, $\overline{d} = (\overline{d}_1, \ldots, \overline{d}_n)'$ and $[\underline{d}, \overline{d}] = \{d \mid d \in R_+^n, \underline{d} \leq d \leq \overline{d}\}$.

For any $D = (D_1, D_2, \ldots, D_n)' \in C(R_+)^n$ and subset G of $R_+^{1\times n}$ the product of G and D is defined as

$$GD = \{gd \mid g = (g_1, \ldots, g_n) \in G, \ d = (d_1, \ldots, d_n)' \in D, \ d_i \in D_i \ (1 \leq i \leq n)\}.$$

The following results are used in the sequel.

Lemma 1.1 ([5][14])

(i) *Any $\mathcal{A} \in \mathcal{M}_n$ is a polyhedral convex set in the vector space $R^{n\times n}$.*

(ii) *For any compact convex subset $G \subset R_+^{1\times n}$ and $D = (D_1, D_2, \ldots, D_n)' \in C(R_+)^n$, it holds $GD \in C(R_+)$.*

We will give a partial order $\geq, >$ on $C(R_+)$ by the definition: For $[c_1, c_2], [d_1, d_2] \in C(R_+)$,

$$[c_1, c_2] \geq [d_1, d_2] \quad \text{if} \quad c_1 \geq d_1, \ c_2 \geq d_2, \quad \text{and}$$

$$[c_1, c_2] > [d_1, d_2] \quad \text{if} \quad [c_1, c_2] \geq [d_1, d_2] \text{ and } [c_1, c_2] \neq [d_1, d_2].$$

For $v = (v_1, v_2, \ldots, v_n)'$ and $w = (w_1, w_2, \ldots, w_n)' \in C(R_+)^n$, we write

$$v \geq w \quad \text{if} \quad v_i \geq w_i,\ 1 \leq i \leq n \quad \text{and}$$

$$v > w \quad \text{if} \quad v \geq w \quad \text{and} \quad v \neq w.$$

Define a metric Δ on $C(R_+)^n$ by

$$\Delta(v, w) = \max_{i \in S} \delta(v_i, w_i)$$

for $v = (v_1, v_2, \ldots, v_n)'$, $w = (w_1, w_2, \ldots, w_n)' \in C(R_+)^n$, where δ is the Hausdorff metric on $C(R_+)$ and given by

$$\delta([a, b], [c, d]) := |a - c| \vee |b - d| \text{ for } [a, b], [c, d] \in C(R_+),$$

where $x \vee y = \max\{x, y\}$. Obviously, $(C(R_+)^n, \Delta)$ is a complete metric space (for example, [2][13]). A controlled Markov set-chain consists of four objects; S, A, $\underline{q}$, $\overline{q}$, r, where $S = \{1, 2, \ldots, n\}$ and $A = \{1, 2, \ldots, k\}$ are finite sets and for each $(i, a) \in S \times A$, $\underline{q} = \underline{q}(\cdot|i, a) \in R_+^{1 \times n}$, $\overline{q} = \overline{q}(\cdot|i, a) \in R_+^{1 \times n}$ with $\underline{q} \leq \overline{q}$ and $\langle \underline{q}, \overline{q} \rangle \neq \emptyset$ and $r = r(i, a)$ a function on $S \times A$ with $r \geq 0$. Note that A is used as a set here, different from the above. We interpret S as the set of states of some system, and A as the set of actions available at each state.

When the system is in state $i \in S$ and we take action $a \in A$, we move to a new state $j \in S$ selected according to the probability distribution on S, $q(\cdot|i, a)$, and we receive a return $r(i, a)$, where we know only that $q(\cdot|i, a)$ is arbitrarily chosen from $\langle \underline{q}(\cdot|i, a), \overline{q}(\cdot|i, a) \rangle$. This process is then repeated from the new state j.

Denote by F the set of functions from S to A.

A policy π is a sequence $(f_1, f_2, \ldots)$ of functions with $f_t \in F$, $(t \geq 1)$. Let Π denote the class of policies. We denote by f^∞ the policy $(h_1, h_2, \ldots)$ with $h_t = f$ for all $t \geq 1$ and some $f \in F$. Such a policy is called stationary, denoted simply by f, and the set of stationary policies is denoted by Π_F.

We associate with each $f \in F$ the n-dimensional column vector $r(f) \in R_+^n$ whose ith element is $r(i, f(i))$ and the set of stochastic matrices $\mathcal{Q}(f) := \langle \underline{Q}(f), \overline{Q}(f) \rangle \in \mathcal{M}_n$, where the (i, j) elements of $\underline{Q}(f)$ and $\overline{Q}(f)$ are $\underline{q}(j|i, f(i))$ and $\overline{q}(j|i, f(i))$, respectively, and $\langle \underline{Q}(f), \overline{Q}(f) \rangle$ is as defined already.

For any $\pi = (f_1, f_2, \ldots) \in \Pi$, and discount factor β $(0 < \beta < 1)$, let

$$\phi_{\beta,T}(\pi) := \Big\{ r(f_1) + \beta Q_1 r(f_2) + \cdots + \beta^T Q_1 Q_2 \cdots Q_T r(f_{T+1}) \\ \mid Q_i \in \mathcal{Q}(f_i), i = 1, 2, \ldots, T \Big\}. \tag{1.1}$$

We observe, for example, that

$$\phi_{\beta,3}(\pi) = r(f_1) + \beta \mathcal{Q}(f_1) \Big(r(f_2) + \beta \mathcal{Q}(f_2) r(f_3) \Big),$$

so that by Lemma 1.1(ii) $\phi_{\beta,T}(\pi) \in C(R_+)^n$ for all $T \geq 1$.

Also, it is shown in [14] that $\{\phi_{\beta,T}(\pi)\}_{T=1}^\infty$ is a Cauchy sequence with respect to Δ, so that the set of discounted expected total rewards from π in the infinite future can be defined by

$$\phi_\beta(\pi) := \lim_{T \to \infty} \phi_{\beta,T}(\pi). \tag{1.2}$$

Since $\phi_\beta(\pi) \in C(R_+)^n$, let denote $\phi_\beta(\pi)$ by

$$\phi_\beta(\pi) = [\underline{\phi}_\beta(\pi), \overline{\phi}_\beta(\pi)].$$

where

$$\underline{\phi}_\beta(\pi) = (\underline{\phi}_\beta(1,\pi), \underline{\phi}_\beta(2,\pi), \ldots, \underline{\phi}_\beta(n,\pi))'$$

and

$$\overline{\phi}_\beta(\pi) = (\overline{\phi}_\beta(1,\pi), \overline{\phi}_\beta(2,\pi), \ldots, \overline{\phi}_\beta(n,\pi))'.$$

In general, $\phi_\beta(\pi)$ is unbounded if $\beta \to 1^-$, so we will treat this case using Abel-sum of rewads in time. Let

$$\phi(f) := \liminf_{\beta \to 1^-} (1-\beta)\phi_\beta(f). \tag{1.3}$$

where, for a sequence $\{D_k\} \subset C(R_+)^n$,

$$\liminf_{k\to\infty} D_k = \left\{ x \in R^n \;\middle|\; \limsup_{k\to\infty} \delta_1(x, D_k) = 0 \right\}. \tag{1.4}$$

and $\delta_1(x, D) = \inf_{y \in D} \delta_2(x, y)$, δ_2 is a metric in R^n. Since $\phi(f) \in C(R_+)^n$, $\phi(f)$ is written as $\phi(f) = [\underline{\phi}(f), \overline{\phi}(f)]$.

Definition *A policy $f^* \in \Pi_F$ is called non-discounted optimal if there does not exist $f \in \Pi_F$ such that $\phi(f^*) < \phi(f)$.*

In the above definition, we confine ourselves to the stationary policies, which simplifies our discussion in the sequel.

In Section 2, a regurality condition is given for transition matrices, and several results for $\beta < 1$ are cited from [14]. In Section 3, the asymptotic behavior of $\phi_\beta(f)$ as β approaches 1 is obtained, by which the existence of a non-discounted optimal stationary policy is proved in Section 4.

2. Preliminaries

Henceforth, the following assumption will remain operative.

Assumption A *For any $f \in F$, each $Q \in \mathcal{Q}(f)$ is primitive i.e., $Q^t > 0$ for some $t \geq 1$.*

Obviously, if for any $f \in F$ $\underline{Q}(f)$ is primitive as a non-negative matrix (cf.[17]), Assumption A holds.

The following facts on Markov matrices are well-known (cf.[3][12]).

Lemma 2.1 *For any $f \in F$, let Q be any matrix in $\mathcal{Q}(f)$.*

(i) *The sequence $(I + Q + \cdots + Q^t)/(t+1)$ converges as $t \to \infty$ to a stochastic matrix Q^* with $QQ^* = Q^*, Q^* > 0$ and $\mathrm{rank}(Q^*) = 1$.*

(ii) *The stochastic matrix Q^* satisfying $QQ^* = Q^*$ and $\mathrm{rank}(Q^*) = 1$ exists uniquely.*

(iii) *$I - (Q - Q^*)$ is nonsingular, and*

$$Qh = h + \phi\mathbf{e} - r(f), \tag{2.1}$$

where I is the identity matrix, $\phi = (1/n)\mathbf{e}'Q^*r(f)$ and $h = ((I - Q + Q^*)^{-1} - Q^*)\, r(f)$ and $\mathbf{e} = (1, 1, \ldots, 1)'$.

Associated with each $f \in F$ and $\beta \in (0,1)$ is a corresponding operator $L_\beta(f)$, mapping $C(R_+)^n$ into $C(R_+)^n$, defined as follows. For $v \in C(R_+)^n$,

$$L_\beta(f)v := r(f) + \beta \mathcal{Q}(f)v. \tag{2.2}$$

Note that from Lemma 1.1, $L_\beta(f)v \in C(R_+)^n$ for each $v \in C(R_+)^n$. Putting $v = [\underline{v}, \overline{v}]$ with $\underline{v} \leq \overline{v}$, $\underline{v}, \overline{v} \in R_+^n$, (2.2) can be written as

$$L_\beta(f)v = [\underline{L}_\beta(f)\underline{v}, \overline{L}_\beta(f)\overline{v}], \tag{2.3}$$

where $\underline{L}_\beta$ and $\overline{L}_\beta$ are operators, mapping R_+^n into R_+^n, defined by :

$$\underline{L}_\beta(f)v = r(f) + \beta \min_{Q \in \mathcal{Q}(f)} Qv, \tag{2.4}$$

$$\overline{L}_\beta(f)v = r(f) + \beta \max_{Q \in \mathcal{Q}(f)} Qv. \tag{2.5}$$

and min (max) represents componentwise minimization (maximization).

The following results are given in Kurano, et al [14].

Lemma 2.2 ([14]) *For any $f \in F$, we have:*

(i) *Both $\underline{L}_\beta(f)$ and $\overline{L}_\beta(f)$ are contractions with modulus β and $\underline{\phi}_\beta(f)$ and $\overline{\phi}_\beta(f)$ are unique fixed points of $\underline{L}_\beta(f)$ and $\overline{L}_\beta(f)$, respectively, where $\phi_\beta(f) = [\underline{\phi}_\beta(f), \overline{\phi}_\beta(f)]$.*

(ii) *For any $h \in R_+^n$, $\underline{\phi}_\beta(f) = \lim_{t\to\infty} \underline{L}_\beta(f)^t h$ and $\overline{\phi}_\beta(f) = \lim_{t\to\infty} \overline{L}_\beta(f)^t h$.*

3. Asymptotic Properties of $\phi_\beta(f)$

In this section we study the asymptotic behavior of $\phi_\beta(f)$ as $\beta \to 1^-$ under Assumption A. To this end, for each $f \in F$ we consider the following interval equation

$$r(f) + \mathcal{Q}(f)\boldsymbol{h} = \boldsymbol{\psi} + \boldsymbol{h}, \tag{3.1}$$

where $\boldsymbol{\psi} = [\underline{\psi}\mathbf{e}, \overline{\psi}\mathbf{e}]$, $\boldsymbol{h} = [\underline{h}, \overline{h}] \in C(R)^n$, $\underline{\psi}$, $\overline{\psi} \in R$, $\underline{h}$, $\overline{h} \in R^n$ with $\underline{\psi} \leq \overline{\psi}$, $\underline{h} \leq \overline{h}$.

Obviously, (3.1) can be rewritten by

$$r(f) + \min_{Q \in \mathcal{Q}(f)} Q\underline{h} = \underline{\psi}\mathbf{e} + \underline{h} \tag{3.2}$$

$$r(f) + \max_{Q \in \mathcal{Q}(f)} Q\overline{h} = \overline{\psi}\mathbf{e} + \overline{h} \tag{3.3}$$

$$\text{where } \underline{\psi} \text{ , } \overline{\psi} \in R \text{ , } \underline{h} \text{ , } \overline{h} \in R^n \text{ with } \underline{\psi} \leq \overline{\psi} \text{ , } \underline{h} \leq \overline{h}. \tag{3.4}$$

Takahashi ([19][20]) has showed that upper or lower bounds of the average rewards for weak D-Markov chains satisfy the equations (3.2) to (3.4) and can be calculated with the Howard's policy improvement ([11]). Also, the existence and uniqueness of a solution of the equations (3,2) to (3,4) can be proved by a slight modification of the proofs of Theorem 2.4 in Bather [1].

However, to make the paper self-contained we give here another proof of these results which is done by applying the Kakutani's fixed point theorem and policy improvement.

For simplicity of the notation, let, for any $d \in R^n$ and $f \in F$,

$$\underline{\mathcal{Q}}(f,d) = \left\{ Q \in \mathcal{Q}(f) \;\middle|\; Qd = \min_{Q \in \mathcal{Q}(f)} Qd \right\},$$

and

$$\overline{\mathcal{Q}}(f,d)=\left\{Q\in\mathcal{Q}(f)\,\middle|\,Qd=\max_{Q\in\mathcal{Q}(f)}Qd\right\}.$$

The following fact can be easily proved by applying Theorem I.2.2. in [4].

Lemma 3.1 *Let $\{d_t\}_{t=1}^{\infty}\subset R^n$ be such that d_t converges as $t\to\infty$ to d. Then, it holds that*

$$\liminf_{t\to\infty}\underline{\mathcal{Q}}(f,d_t)\subset\underline{\mathcal{Q}}(f,d)\quad\text{and}$$

$$\liminf_{t\to\infty}\overline{\mathcal{Q}}(f,d_t)\subset\overline{\mathcal{Q}}(f,d)\quad\text{for all } f\in F.$$

Proof. For each i $(1\le i\le n)$, $f\in F$ and $d\in R^n$, let

$$\underline{\mathcal{Q}}(f,d)_i:=\left\{Q\in\mathcal{Q}(f)\,\middle|\,(Qd)_i=\min_{Q\in\mathcal{Q}(f)}\sum_{j=1}^{n}q_{ij}d_j\right\},$$

where q_{ij} is the (i,j)element of Q. Then $\underline{\mathcal{Q}}(f,d)=\bigcap_{i=1}^{n}\underline{\mathcal{Q}}(f,d)_i$. Also, for any $\{d_t\}_{t=1}^{\infty}$ such that $d_t\to d$ as $t\to\infty$, from Theorem I.2.2. in [4] , $\liminf_{t\to\infty}\underline{Q}(f,d_t)_i\subset\underline{Q}(f,d)_i$.

So we get that

$$\begin{aligned}\liminf_{t\to\infty}\underline{\mathcal{Q}}(f,d_t)&\subset\bigcap_{i=1}^{n}\liminf_{t\to\infty}\underline{Q}(f,d_t)_i\\&\subset\bigcap_{i=1}^{n}\underline{Q}(f,d)_i=\underline{Q}(f,d),\end{aligned}$$

which completes the proof.□

Theorem 3.1 *For any $f\in F$, the interval equation* (3.1) *determines $\underline{\psi}$ uniquely and $\underline{\boldsymbol{h}}$ up to an additive constant $[c_1\mathbf{e},c_2\mathbf{e}]$ with $c_1,c_2\in R$ $(c_1<c_2)$.*

Proof. Let $\underline{h}:\mathcal{Q}(f)\to R^n$ be defined by

$$\underline{h}(Q)=((I-Q+Q^*)^{-1}-Q^*)r(f)\quad\text{for } Q\in\mathcal{Q}(f),$$

where $Q^*=\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}Q^t$.

By Lemma 2.1 we observe that $\underline{h}$ is continuous.

Let $\mathcal{A}(\mathcal{Q}(f))$ denote the set of all compact and convex subsets of $\mathcal{Q}(f)$. Then, noting $\underline{\mathcal{Q}}(f,d)\in\mathcal{A}(\mathcal{Q}(f))$ for $d\in R^n$, from Lemma 3.1 the map $\underline{\mathcal{Q}}(f,\underline{h}(\cdot)):\mathcal{Q}(f)\to\mathcal{A}(\mathcal{Q}(f))$ is upper semicontinuous.

Thus, applying Kakutani's fixed point theorem (cf.[18],p.129), there exists $\underline{Q}\in\mathcal{Q}(f)$ such that $\underline{Q}\in\mathcal{Q}(f,\underline{h}(\underline{Q}))$, which implies from Lemma 2.1 that $\underline{h}:=\underline{h}(\underline{Q})$ and $\underline{\psi}:=(1/n)\mathbf{e}'\underline{Q}^*r(f)$ is a solution of (3.2).

For any $Q\in\mathcal{Q}(f)$, we have

$$Q\underline{h}\ge\underline{h}+\underline{\psi}\mathbf{e}-r(f).\tag{3.5}$$

Since $QQ^*=Q^*$ for $Q^*=\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}Q^t$, (3.5) derives that

$$\underline{\psi}\mathbf{e}\le Q^*r(f).$$

Also, for $\underline{Q} \in \mathcal{Q}(f, \underline{h})$, (3.2) implies $\underline{\psi}\mathbf{e} = \underline{Q}^* r(f)$.

Thus, letting $\mathcal{Q}^*(f) = \left\{ Q^* = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} Q^t \middle| Q \in \mathcal{Q}(f) \right\}$, we get

$$\underline{\psi}\mathbf{e} = \min_{Q^* \in \mathcal{Q}^*(f)} Q^* r(f),$$

which shows the uniqueness of $\underline{\psi}$ in (3.2).

Let $\underline{h}'$ be another solution of (3.2). Now, we will prove that $\underline{\mathcal{Q}}(f, \underline{h}) = \underline{\mathcal{Q}}(f, \underline{h}')$. To this end, let assume that $\underline{\mathcal{Q}}(f, \underline{h}) \neq \underline{\mathcal{Q}}(f, \underline{h}')$ and that there exists $Q \in \underline{\mathcal{Q}}(f, \underline{h}')$ with $Q \notin \underline{\mathcal{Q}}(f, \underline{h})$. Then, by the definition, we have

$$\begin{cases} Q\underline{h}' &= \underline{h}' + \underline{\psi}\mathbf{e} - r(f) \\ Q\underline{h} &\underset{\neq}{>} \underline{h} + \underline{\psi}\mathbf{e} - r(f), \end{cases} \tag{3.6}$$

where for $x, y \in R^n$, $x \underset{\neq}{>} y$ means $x_i \geq y_i$ $(1 \leq i \leq n)$ and $x \neq y$.

By multiplying both sides of (3.6) by Q^*, we get from $Q^* > 0$ that $\underline{\psi}\mathbf{e} = Q^* r(f)$ and $\underline{\psi}\mathbf{e} \underset{\neq}{<} Q^* r(f)$, which leads to a contradiction. Thus, we get $\underline{\mathcal{Q}}(f, \underline{h}) = \underline{\mathcal{Q}}(f, \underline{h}')$.

For any $Q \in \underline{\mathcal{Q}}(f, \underline{h})$, we have

$$\begin{cases} Q\underline{h} &= \underline{h} + \underline{\psi}\mathbf{e} - r(f) \\ Q\underline{h}' &= \underline{h}' + \underline{\psi}\mathbf{e} - r(f). \end{cases} \tag{3.7}$$

From (3.7) we get that $Q(\underline{h} - \underline{h}') = \underline{h} - \underline{h}'$, so that, by induction, we can prove that

$$Q^t(\underline{h} - \underline{h}') = \underline{h} - \underline{h}' \quad \text{for all } t \geq 1,$$

which leads that

$$Q^*(\underline{h} - \underline{h}') = \underline{h} - \underline{h}' \tag{3.8}$$

for $Q^* = \lim_{t\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} Q^t$.

We note that each low vector of Q^* is identical under Assumption A. Therefore, (3.8) implies that each element of $\underline{h} - \underline{h}'$ is identical, as required for a solution $(\underline{\psi}, \underline{h})$ of (3.2).

Similarly we can prove the assertion for a solution $(\overline{\psi}, \overline{h})$ of (3.3). This completes the proof. □

Since the unique solutions $\underline{\psi}$, $\overline{\psi}$ of (3.2) and (3.3) are depending on $f \in F$, we will denote them respectively by $\underline{\psi}(f)$ and $\overline{\psi}(f)$.

The following theorem is concerned with the asymptotic properties of $\phi_\beta(f)$ as $\beta \to 1^-$.

Theorem 3.2 *For any $f \in F$, there exists c_1, c_2, c_1' c_2' $\in R$ $(c_1 \leq c_1', c_2 \leq c_2')$ such that*

$$\left[\left(\frac{\underline{\psi}(f)}{1-\beta} + c_1' \right) \mathbf{e}, \left(\frac{\overline{\psi}(f)}{1-\beta} + c_2 \right) \mathbf{e} \right] \subset \phi_\beta(f) \subset \left[\left(\frac{\underline{\psi}(f)}{1-\beta} + c_1 \right) \mathbf{e}, \left(\frac{\overline{\psi}(f)}{1-\beta} + c_2' \right) \mathbf{e} \right] \tag{3.9}$$

where $[a, b] = \emptyset$ if $a > b$.

Proof. For any $f \in F$, let $\underline{h} = (\underline{h}_1, \underline{h}_2, \ldots, \underline{h}_n)$ and $\overline{h} = (\overline{h}_1, \overline{h}_2, \ldots, \overline{h}_n)$ be any solution of (3.2)–(3.4). We will prove by induction that, for any β $(0 < \beta < 1)$,

$$\left(\sum_{l=0}^{t-1} \beta^l \right) \underline{\psi}(f)\mathbf{e} + \underline{h} + (\beta - 1) \left(\sum_{l=0}^{t-1} \beta^l \right) \max_i \underline{h}_i \leq \underline{L}_\beta(f)^t \underline{h}$$

$$\leq \left(\sum_{l=0}^{t-1}\beta^l\right)\underline{\psi}(f)\mathbf{e}+\underline{h}+(\beta-1)\left(\sum_{l=0}^{t-1}\beta^l\right)\min_i \underline{h}_i. \tag{3.10}$$

We have:

$$\begin{aligned}\underline{L}_\beta(f)\underline{h} &= r(f)+\beta\min_{Q\in\mathcal{Q}(f)}Q\underline{h}, \quad \text{for the definition of } \underline{L}_\beta(f)\\ &= \underline{\psi}(f)\mathbf{e}+\underline{h}+(\beta-1)\min_{Q\in\mathcal{Q}(f)}Q\underline{h}, \quad \text{by (3.2)},\end{aligned}$$

which implies that (3.10) holds for $t=1$.

Now, we assume that (3.10) holds for t. Then,

$$\begin{aligned}\underline{L}_\beta(f)^{t+1}\underline{h} &= r(f)+\beta\min_{Q\in\mathcal{Q}(f)}Q\underline{L}_\beta(f)^t\underline{h}\\ &\leq \beta\left(\sum_{l=0}^{t-1}\beta^l\right)\underline{\psi}(f)\mathbf{e}+\beta(\beta-1)\left(\sum_{l=0}^{t-1}\beta^l\right)\min_i h_i+\underline{L}_\beta(f)\underline{h},\end{aligned}$$

by the hypothesis of the induction. This leads to the inequality of the right hand in (3.10) for $t+1$.

Similarly we can prove the inequality of the left hand in (3.10) for $t+1$. As $t\to\infty$ in (3.10), we get from Lemma 2.2 (ii) that

$$\frac{\underline{\psi}(f)\mathbf{e}}{1-\beta}+\underline{h}-\left(\max_i \underline{h}_i\right)\mathbf{e}\leq \underline{\phi}_\beta(f)\leq \frac{\underline{\psi}(f)\mathbf{e}}{1-\beta}+\underline{h}-\left(\min_i \underline{h}_i\right)\mathbf{e}. \tag{3.11}$$

Similarly as the above, we get

$$\frac{\overline{\psi}(f)\mathbf{e}}{1-\beta}+\overline{h}-\left(\max_i \overline{h}_i\right)\mathbf{e}\leq \overline{\phi}_\beta(f)\leq \frac{\overline{\psi}(f)\mathbf{e}}{1-\beta}+\overline{h}-\left(\min_i \overline{h}_i\right)\mathbf{e}. \tag{3.12}$$

Letting

$$\begin{aligned}c_1 &= \min_i \underline{h}_i-\max_i \underline{h}_i, \quad c_1'=-c_1,\\ c_2 &= \min_i \overline{h}_i-\max_i \overline{h}_i, \quad c_2'=-c_2,\end{aligned}$$

by (3.11) and (3.12), (3.9) follows, as required.□

Corollary 3.1 *For any $f\in F$, it holds that*

(i) $$\phi(f)=[\underline{\psi}(f)\mathbf{e},\overline{\psi}(f)\mathbf{e}] \quad , \text{and}$$

(ii) $$\underline{\psi}(f)\mathbf{e}=\min_{Q^*\in\mathcal{Q}^*(f)}Q^*r(f) \quad \text{and} \quad \overline{\psi}(f)\mathbf{e}=\max_{Q^*\in\mathcal{Q}^*(f)}Q^*r(f).$$

Proof. (i) follows from Theorem 3.2. Also, the proof of (ii) is appearing in the proof of Theorem 3.1. □

4. Non-discounted Optimal Policies

In this section, we give the existence theorem of a non-discounted optimal policy under Assumption A.

Let $\boldsymbol{q}(i,a):=\langle \underline{q}(\cdot|i,a),\overline{q}(\cdot|i,a)\rangle$ for each $i\in S$ and $a\in A$. For each $i\in S$ and $f\in F$, denote by $\underline{G}(i,f)$ the set of $a\in A$ for which

$$\underline{\psi}(f)+\underline{h}(f)_i<r(i,a)+\min_{q\in\boldsymbol{q}(i,a)}\sum_{j=1}^n q(j|i,a)\underline{h}(f)_j,$$

where $\underline{\psi}(f)$ and $\underline{h}(f) = (\underline{h}(f)_1, \ldots, \underline{h}(f)_n)$ is a solution of (3.2).

Let $g \in F$ be such that $g(i) \in \underline{G}(i,f)$ for any i with $\underline{G}(i,f) \neq \emptyset$ and $g(i) = f(i)$ for any i with $\underline{G}(i,f) = \emptyset$. Then, we have the following.

Lemma 4.1 *For any f with $\underline{G}(i,f) \neq \emptyset$ for some $i \in S$, $\underline{\psi}(f) < \underline{\psi}(g)$.*

Proof. It holds from the definition that for any $Q \in \mathcal{Q}(g)$,

$$\underline{\psi}(f)\mathbf{e} + \underline{h}(f) \underset{\neq}{\leq} r(g) + Q\underline{h}(f). \tag{4.1}$$

By multiplying both sides of (4.1) by Q^*, we get from $Q^* > 0$ that $\underline{\psi}(f)\mathbf{e} \underset{\neq}{\leq} Q^* r(g)$, which implies by Corollary 3.1 (ii) that $\underline{\psi}(f) < \underline{\psi}(g)$, as required. □

The following lemma is proved from the idea of policy improvement (cf.[11]).

Lemma 4.2 *The left-side optimality equations (4.2) below determine $\underline{\psi}^*$ uniquely and $\underline{h} \in R^n$ up to an addtive constant.*

$$\underline{\psi}^* + \underline{h}_i = \max_{a \in A} \left(r(i,a) + \min_{q \in \boldsymbol{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\underline{h}_j \right) \quad (1 \leq i \leq n). \tag{4.2}$$

Proof. Let $f_1 \in F$. If $\underline{G}(i,f_1) = \emptyset$ for all $i \in S$, $(\underline{\psi}(f_1), \underline{h}(f_1))$ is a solution of (4.2). If $\underline{G}(i,f_1) \neq \emptyset$ for some $i \in S$, we define $f_2 \in F$ by $f_2(i) \in \underline{G}(i,f_1)$ for any i with $\underline{G}(i,f_1) \neq \emptyset$ and $f_2(i) = f_1(i)$ for any i with $\underline{G}(i,f_1) = \emptyset$. Then, from Lemma 4.1, $\underline{\psi}(f_1) < \underline{\psi}(f_2)$. If $\underline{G}(i,f_2) = \emptyset$ for all $i \in S$, $(\underline{\psi}(f_2), \underline{h}(f_2))$ is a solution of (4.2). If $\underline{G}(i,f_2) \neq \emptyset$ for some $i \in S$, we define $f_3 \in F$ by the same way as we define f_2 from f_1. Then, it holds $\underline{\psi}(f_2) < \underline{\psi}(f_3)$.

Repeating this operation, we reaches the case that $\underline{G}(i,f_k) = \emptyset$ for all $i \in S$ $(k \geq 3)$. Obviously $(\underline{\psi}(f_k), \underline{h}(f_k))$ becomes a solution of (4.2).

Also, the uniqueness of (4.2) follows from Theorem 3.1.□

Let, for each i $(1 \leq i \leq n)$,

$$A_i := arg\max_{a \in A} \left(r(i,a) + \min_{q \in \boldsymbol{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\underline{h}_j \right).$$

For each $i \in S$ and $f \in F$ with $f(i) \in A_i$ for all $i \in S$, denote by $\overline{G}(i,f)$ the set of $a \in A_i$ for which

$$\overline{\psi}(f) + \overline{h}(f)_i < r(i,a) + \max_{q \in \boldsymbol{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\overline{h}(f)_j,$$

where $\overline{\psi}(f)$ and $\overline{h}(f) = (\overline{h}(f)_1, \ldots, \overline{h}(f)_n)$ is a solution of (3.3).

Using $\overline{G}(i,f)$ instead of $\underline{G}(i,f)$ and applying the same way as the proof of Lemma 4.2, we can prove the following.

Lemma 4.3 *The right-side optimality equations (4.3) below determine $\overline{\psi}^*$ uniquely and $\overline{h} \in R^n$ up to an addtive constant.*

$$\overline{\psi}^* + \overline{h}_i = \max_{a \in A_i} \left(r(i,a) + \max_{q \in \boldsymbol{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\overline{h}_j \right) \quad (1 \leq i \leq n). \tag{4.3}$$

Let, for each i $(1 \leq i \leq n)$,

$$A_i^* := arg\max_{a \in A_i} \left(r(i,a) + \max_{q \in \boldsymbol{q}(i,a)} \sum_{j=1}^{n} q(j|i,a)\overline{h}_j \right).$$

Then we have the following Theorem.

Theorem 4.1 *Let f^* be any policy with $f^*(i) \in A_i^*$ for all $i \in S$. Then f^* is non-discounted optimal and $\phi(f^*) = [\underline{\psi}^*\mathbf{e}, \overline{\psi}^*\mathbf{e}]$.*

Proof. For any $f \in F$ with $f(i) \in A_i$ $(1 \leq i \leq n)$, we get, from (4.2), that

$$\underline{\psi}^* + \underline{h} = r(f) + \min_{Q \in \mathcal{Q}(f)} Q\underline{h}. \tag{4.4}$$

By Theorem 3.1, $\underline{\psi}^* = \underline{\psi}(f)$. Also, applying Theorem 3.1 again, we can assume $\underline{h} = \underline{h}(f)$. Let $g \in F$ be any policy. Then, from (4.2) it holds

$$\begin{aligned} \underline{\psi}(f)\mathbf{e} + \underline{h}(f) &\geq r(g) + \min_{Q \in \mathcal{Q}(f)} Q\underline{h}(f) \\ &= r(g) + Q\underline{h}(f) \text{ for some } Q \in \mathcal{Q}(g). \end{aligned}$$

This inequality leads to $\underline{\psi}(f)\mathbf{e} \geq Q^* r(g)$ for $Q^* = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} Q^t$.

Thus, from Corollary 3.1, we have $\underline{\psi}(f) \geq \underline{\psi}(g)$. The above shows that

$$\underline{\psi}(f) \geq \underline{\psi}(g) \text{ for all } g \in F. \tag{4.5}$$

Now, let $f \in F$ be such that $f(i) \notin A_i$ for some $i \in S$. Then, we have, from (4.2),

$$\underline{\psi}^*\mathbf{e} + \underline{h} \underset{\neq}{\geq} r(f) + Q\underline{h} \text{ for all } Q \in \underline{\mathcal{Q}}(f, \underline{h}),$$

so that, noting $Q^{**} > 0$ with $Q^{**} = \lim_{t\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} Q^t$ we get $\underline{\psi}^*\mathbf{e} \underset{\neq}{\geq} Q^{**} r(f)$. Thus,

$$\underline{\psi}^*\mathbf{e} \underset{\neq}{\geq} \min_{Q \in \mathcal{Q}^*(f)} Qr(f) = \underline{\psi}(f)\mathbf{e},$$

which implies that (4.5) does not hold for this f.

Here, we can summarize that $f \in F$ satisfies (4.5) if and only if $f(i) \in A_i$ for all $i \in S$.

For any $f^* \in F$ which $f^*(i) \in A_i^*$ for all $i \in S$, repeating the same discussion as (4.4),(4.5) obtains that $\overline{\psi}^* = \overline{\psi}(f^*)$ and $\overline{\psi}(f) \leq \overline{\psi}(f^*)$ for all $f \in F$ with $f(i) \in A_i$ $(i \in S)$. This shows that f^* is non-discounted optimal, which completes the proof. □

Remark Theorem 3.1 provides bounds on the possible behavior of the decision process resulting from interval estimate, which can also be used for best/worst scenarios of the decision process.

It should be noted that our results do not give bounds for the accuracy of approximate Markov chains (cf.[21][22]).

Here, as a numerical example we shall solve a Markov set-chain version of the Toymaker's problem dealt with in Howard [11], whose data is presented in Table 1 with $S = \{1, 2\}$ and $A = \{1, 2\}$.

The return $r(i, a, j)$ in Table 1 is associated with the transition from state i to state j under action a, which does not affect our theoretical results with a slight modification.

Table 1 The Toymaker's Problem

State	Alternative	The interval of transition probability		Rewards	
i	a	$\boldsymbol{q}(1\|i,a)$	$\boldsymbol{q}(2\|i,a)$	$r(i,a,1)$	$r(i,a,2)$
1(Successful toy)	1(No advertising)	$[0.3,0.5]$	$[0.5,0.7]$	9	3
	2(Advertising)	$[0.5,0.8]$	$[0.3,0.5]$	4	4
2(Unsuccessful toy)	1(No research)	$[0.3,0.4]$	$[0.5,0.7]$	3	-7
	2(Research)	$[0.6,0.7]$	$[0.3,0.5]$	1	-19

Then, by checking up extreme points of the corresponding polyhedral convex set (cf.(i) of Lemma 1.1), equation (4.2) with $\underline{\psi}^*$ and $\underline{h} = (\underline{h}_1, \underline{h}_2)'$ is given as follows:

$$\underline{\psi}^* + \underline{h}_1 = \max \begin{cases} \min\{0.3\underline{h}_1 + 0.7\underline{h}_2 + 4.8,\ 0.5\underline{h}_1 + 0.5\underline{h}_2 + 6\} \\ \min\{0.5\underline{h}_1 + 0.5\underline{h}_2 + 4,\ 0.7\underline{h}_1 + 0.3\underline{h}_2 + 4\}, \end{cases}$$

$$\underline{\psi}^* + \underline{h}_2 = \max \begin{cases} \min\{0.3\underline{h}_1 + 0.7\underline{h}_2 - 4,\ 0.4\underline{h}_1 + 0.6\underline{h}_2 - 3\} \\ \min\{0.6\underline{h}_1 + 0.4\underline{h}_2 - 7,\ 0.7\underline{h}_1 + 0.3\underline{h}_2 - 5\}, \end{cases}$$

After a simple calculation, the solution of the above with $\underline{h}_1 = 0$ becomes that $\underline{\psi}^* = -1$ and $\underline{h} = (0, -10)'$. Also, we easily find $A_1 = \{2\}$ and $A_2 = \{1, 2\}$.

Similarly, by solving the equation (4.3) with $\overline{h}_1 = 0$, we get $\overline{\psi}^* = 1.3, \overline{h} = (0, -9)', A_1^* = \{2\}$ and $A_2^* = \{2\}$. So, by Theorem 4.1, f^* with $f^*(1) = 2$ and $f^*(2) = 2$ is non-discounted optimal and $\phi(f^*) = [-\mathbf{e}, 1.3\mathbf{e}]$.

In this example, we find that Abel-sum of rewards in time will be positive in best behavior but negative in worst behavior.

Acknowledgments

We thank the reviewers for several comments and corrections and showing us the references [21][22] that helped us improve the paper. Also, we are grateful to Professor Yukio Takahashi of Tokyo Institute of Technology for bringing his pioneering papers [19][20] to our attention.

Reference

[1] J. Bather: Optimal decision processes for finite Markov chains, Part II: Communicationsystems. *Advances in Applied Probability*, **5** (1973) 521–540.

[2] D. P. Bertsekas and S. E. Shreve: *Stochastic Optimal Control: the Discrete Time Case* (Academic Press, New York, 1978).

[3] D. Blackwell: Discrete dynamic programing. *Annals of Mathematical Statistics*, **33** (1962) 719–726.

[4] G. B. Dantzig, J. Folkman and N. Shapiro: On the continuity of the minimum set of a continuous function. *Journal of Mathematical Analysis and Applications*, **17** (1967) 519–548.

[5] D. J. Hartfiel: On the limiting set of stochastic product $xA_1 \cdots A_m$. *Proceedings of the American Mathematical Society*, **81** (1981) 201–206.

[6] D. J. Hartfiel: Component bounds for Markov set-chain limiting sets. *Journal of Statistical Computation and Simulation*, **38** (1991) 15–24.

[7] D. J. Hartfiel: Cyclic Markov set-chains. *Journal of Statistical Computation and Simulation*, **46** (1993) 145–167.

[8] D. J. Hartfiel and E. Seneta: On the theory of Markov set-chains. *Advances in Applied Probability*, **26** (1994) 947–964.

[9] D. J. Hartfiel: *Markov Set-Chains* (Springer-Verlag, Berlin Heidelbergn, 1998).

[10] K. Hinderer: *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter* (Springer-Verlag, New York, 1970).

[11] R. Howard: *Dynamic Programming and Markov Processes* (MIT Press, Cambridge MA, 1960).

[12] J. G. Kemeny and J. L. Snell: *Finite Markov Chains* (Van Nostrand, New York, 1960).

[13] K. Kuratowski: *Topology* (Academic, New York, 1966).

[14] M. Kurano, J. Song, M. Hosaka and Y. Huang: Controlled Markov set-chains with discounting. *Journal of Applied Probability*, **35** (1998) 293–302.

[15] A. Nenmaier: New techniques for the analyses of linear interval equations. *Linear Algebra and Applications*, **58** (1984) 273–325.

[16] M. L. Puterman: *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (John Wiley & Sons, 1994).

[17] E. Seneta: *Nonnegative Matrices and Markov Chains* (Springer-Verlag, New York, 1981).

[18] J. Stoer and C. Witzgall: *Convexity and Optimization in Finite Dimension 1* (Springer-Verlag, Berlin and New York, 1970).

[19] Y. Takahashi: Weak D-Markov chain and its application to a queueing network. G. Iazeolla, P. J. Courtois, and A. Hordijk (eds.): *Mathematical Computer Performance and Reliability* (North-Holland, Amsterdam, 1984), 153–165.

[20] Y. Takahashi: A weak D-Markov chain approach to tandem queueing networks. G. Iazeolla, P. J. Courtois, and A. Hordijk (eds.): *Computer Performance and Reliability* (North-Holland, Amsterdam, 1988), 151–159.

[21] N. M. von Dijk: On the importance of bias-terms for error bounds and comparison results. W. J. Stwart. (ed.): *Numerical Solution of Markov Chains* (New York, Marcel Dekker, 1990), 617–642.

[22] N. M. von Dijk: An error-bound theorem for approximate Markov chains. *Probability in the Engineering and Informational Sciences*, **6** (1992) 413–424.

Masanori Hosaka
Graduate School of Science
and Technology,
Chiba University
Yayoi-cho, Inage-ku, Chiba.
263-8522, Japan.

Masami Kurano
Department of Mathematics,
Faculty of Education,
Chiba University
Yayoi-cho, Inage-ku, Chiba.
263-8522, Japan.
E-mal:kurano@math.e.chiba-u.ac.jp