Sufficient conditions for value function and optimal strategy to be even and monotone

Jhelum Chakravorty and Aditya Mahajan

***Abstract*— In this paper, sufficient conditions are identified under which the value function and the optimal strategy of a Markov decision process are even in the state and monotone increasing for positive values of the state. The results are derived using a new variation of stochastic dominance which we call folded stochastic dominance. An example of a birth-death Markov chain with controlled restart is provided.**

I. Introduction

A. Motivation

Markov decision theory is often used to identify structural or qualitative properties of optimal strategies. Examples include control limit strategies in machine maintenance [1], [2], threshold-based strategies for executing call options [3], [4], and monotone strategies in queueing systems [5], [6]. In all of these models, the optimal strategy is *monotone* in the state, i.e., if $x > y$ then the action chosen at x is greater (or less) than or equal to the action chosen at y. Motivated by this, general conditions under which the optimal strategy is monotone in scalar-valued state are identified in [7]–[12]. Similar conditions for vector-valued states are identified in [13]–[15]. General conditions under which the value function is increasing and convex are established in [16].

Most of these results are motivated by queueing models where the state (i.e., the queue length) takes non-negative values. However, there are other applications in systems and control, where the state takes both positive and negative values. Motivated by the results in [7]–[15], we answer the following question: *When the state space is an interval of the form $[-a, a]$ or $(-\infty, \infty)$, under what conditions is the value function and the optimal strategy even; under what conditions are they monotone increasing (or decreasing) over $[0, a]$ or $[0, \infty)$.*

B. Model and problem formulation

Consider a Markov decision process with state space $\mathbb{X}$ and action space $\mathbb{U}$. For ease of exposition, we assume that $\mathbb{X}$ and $\mathbb{U}$ are discrete (either finite or countably infinite), though the results also hold when they are continuous. Let $X_t \in \mathbb{X}$ and $U_t \in \mathbb{U}$ denote the state and action at time t. The initial state X_1 is distributed according to P_X and the state evolves in a controlled Markov manner, i.e.,

$$\begin{aligned}\mathbb{P}(X_{t+1} = x_{t+1} \mid X_{1:t} = x_{1:t}, U_{1:t} = u_{1:t}) \\ = \mathbb{P}(X_{t+1} = x_{t+1} \mid X_t = x_t, U_t = u_t),\end{aligned}$$

The author are with Department of Electrical and Computer Engineering, McGill University, Montreal, Canada.

where $x_{1:t}$ is a short hand notation for $(x_1, \dots, x_t)$ and a similar interpretation holds of $u_{1:t}$.

For ease of exposition, we assume that the state evolves in a time-homgeneous manner, i.e.,

$$\mathbb{P}(X_{t+1} = y \mid X_t = x, U_t = u) = P_{xy}(u),$$

where $P(u)$ denotes the controlled transition probability matrix.

The system operates for a finite horizon T. For $t \in \{1, \dots, T-1\}$, $c_t \colon \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ denotes the per-step cost and $c_T \colon \mathbb{X} \to \mathbb{R}$ denotes the terminal cost.

The actions at time t are chosen according to a Markov strategy g_t, i.e.,

$$U_t = g_t(X_t), \quad t \in \{1, \dots, T-1\}.$$

The objective is to choose a decision strategy $\mathbf{g} \coloneqq (g_1, \dots, g_{T-1})$ to minimize the expected total cost

$$J(\mathbf{g}) \coloneqq \mathbb{E}^g \Big[\sum_{t=1}^{T-1} c_t(X_t, U_t) + c_T(X_T) \Big] \tag{1}$$

From Markov decision theory [8], we know that an optimal strategy is given by the solution of the following dynamic program. Recursively define $V_t \colon \mathbb{X} \to \mathbb{R}$ and $Q_t \colon \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ as follows: for all $x \in \mathbb{X}$ and $u \in \mathbb{U}$,

$$V_T(x) = c_T(x) \tag{2}$$

and for $t \in \{T-1, \dots, 1\}$,

$$\begin{aligned}Q_t(x,u) &= c_t(x,u) + \mathbb{E}[V_{t+1}(X_{t+1}) \mid X_t = x, U_t = u] \\ &= c_t(x,u) + \sum_{y \in \mathbb{X}} P_{xy}(u) V_{t+1}(y), \qquad (3) \\ V_t(x) &= \min_{u \in \mathbb{U}} Q_t(x,u). \qquad (4)\end{aligned}$$

Then, a strategy $\mathbf{g}^* = (g_1^*, \dots, g_{T-1}^*)$ defined as

$$g_t^*(x) \in \arg\min_{u \in \mathbb{U}} Q_t(x,u)$$

is optimal.

Motivated by applications in queueing theory, there has been considerable interest in identifying conditions under which V_t and g_t^* are monotone increasing (or monotone decreasing) [7]–[15].

We generalize these results to the case when $\mathbb{X}$ is an interval of the form $[-a, a]$ or $(-\infty, \infty)$ and identify sufficient conditions under which the value function and the optimal strategy is even and increasing[1], i.e., for any $x \in \mathbb{X}$, $x > 0$,

$$\begin{aligned} V_t(x) &= V_t(-x), & V_t(x) &\le V_t(x+1); \\ g_t(x) &= g_t(-x), & g_t(x) &\le g_t(x+1). \end{aligned}$$

II. Properties of value function

In the sequel, we assume that $\mathbb{X}$ is a set of integers of the form $\{-a, -a+1, \cdots, a\}$, $0 \le a < \infty$, or $\{-\infty, \cdots, \infty\}$ (the latter being a little abuse of notation, which implies the countable set of integers). Let $\mathbb{X}_{\ge 0}$ and $\mathbb{X}_{>0}$ denote, respectively, the sets $\{x \in \mathbb{X} : x \ge 0\}$ and $\{x \in \mathbb{X} : x > 0\}$.

The monotinicity results of [7]–[15] depend on stochastic dominance. The key idea of our results is to define a new stochastic order, which we call *folded stochastic dominance*, and use that to establish even and increasing properties.

Definition 1 (Stochastic Dominance) Let μ and π be two probability distributions defined over $\mathbb{X}_{\ge 0}$. Then μ is said to dominate π in the sense of stochastic dominance, which is denoted by $\mu \succeq_s \pi$, if

$$\sum_{x \ge y} \mu_x \ge \sum_{x \ge y} \pi_x, \quad \forall y \in \mathbb{X}_{\ge 0}.$$

An equivalent characterization of stochastic dominance is the following. Let M and P denote the cumulative mass function corresponding to μ and π. Then, $\mu \succeq_s \pi$ iff

$$M_x \le P_x, \quad \forall x. \tag{5}$$

A useful property of stochastic dominance is the following [8, Lemma 4.7.2]:

Lemma 1 *For any probability distributions μ and π on $\mathbb{X}_{\ge 0}$ such that $\mu \succeq_s \pi$ and for any increasing function $f \colon \mathbb{X}_{\ge 0} \to \mathbb{R}$,*

$$\sum_{x \in \mathbb{X}_{\ge 0}} f(x)\mu_x \ge \sum_{x \in \mathbb{X}_{\ge 0}} f(x)\pi_x. \tag{6}$$

To extend the notion of stochastic dominance from $\mathbb{X}_{\ge 0}$ to $\mathbb{X}$, we define the following operator:

Definition 2 (Folding Operator) Given a probability distribution π over $\mathbb{X}$, define an operator $\mathcal{F} : [\mathbb{X} \to \mathbb{R}] \to [\mathbb{X}_{\ge 0} \to \mathbb{R}]$ (called the *folding operator*) as follows:

$$(\mathcal{F}\pi)_x = \begin{cases} \pi_0, & x = 0; \\ \pi_x + \pi_{-x}, & x > 0. \end{cases}$$

Given an even function $f \colon \mathbb{X} \to \mathbb{R}$ and a distribution μ over $\mathbb{X}$, let $\tilde{\mu} = \mathcal{F}\mu$. Then, an immediate consequence of the Definition 2 is that

$$\sum_{x \in \mathbb{X}} f(x)\mu_x = \sum_{x \in \mathbb{X}_{\ge 0}} f(x)\tilde{\mu}_x. \tag{7}$$

This fact can be used to generalize the Lemma 1 as follows.

[1] Note that we use *increasing* to mean *weakly* increasing or *non-decreasing*; the other case is referred to as *strictly* increasing.

Definition 3 (Folded stochastic dominance) Let μ and π be two probability distributions defined over $\mathbb{X}$. Then μ is said to dominate π in the sense of folded stochastic dominance, which is denoted by $\mu \succeq_f \pi$, if $\mathcal{F}\mu \succeq_s \mathcal{F}\pi$.

Definition 4 (Even and increasing function (EI)) A function $f \colon \mathbb{X} \to \mathbb{R}$ is called even and increasing on $\mathbb{X}_{\ge 0}$ (abbreviated to EI) if for all $x \in \mathbb{X}_{\ge 0}$,

$$f(x) = f(-x) \quad \text{and} \quad f(x) \le f(x+1).$$

An immediate consequence of (7) and Lemma 1 is the following:

Lemma 2 *For any probability distributions μ and π on $\mathbb{X}$ such that $\mu \succeq_f \pi$ and for any EI function $f \colon \mathbb{X} \to \mathbb{R}$,*

$$\sum_{x \in \mathbb{X}} f(x)\mu_x \ge \sum_{x \in \mathbb{X}} f(x)\pi_x. \tag{8}$$

Remark 1 Stochastic dominance is sometimes defined as follows. Given two probability measures μ and π over $\mathbb{X}_{\ge 0}$, $\mu \succeq_s \pi$ if (6) holds for all increasing functions $f \colon \mathbb{X}_{\ge 0} \to \mathbb{R}$. The property of Definition 1 can be recovered by choosing f as a step function. Similarly, we could have defined folded stochastic dominance as follows. Give two probability measures μ and π over $\mathbb{X}$, $\mu \succeq_f \pi$ if (8) holds for all EI functions $f \colon \mathbb{X} \to \mathbb{R}$. The property of Definition 3 can be recovered by choosing f as follows: for a given $y \in \mathbb{X}_{\ge 0}$,

$$f(x) = \begin{cases} 1, & |x| \ge y \\ 0, & \text{otherwise.} \end{cases}$$

Now we state sufficient conditions under which the value function is EI.

Theorem 1 *Suppose the MDP satisfies the following properies:*

- (C1) *$c_T(\cdot)$ is EI and for every $u \in \mathbb{U}$, $c_t(\cdot, u)$ is EI.*
- (C2) *For every $u \in \mathbb{U}$ and any $x, y \in \mathbb{X}$, $P_{xy}(u) = P_{(-x)(-y)}(u)$.*
- (C3) *Let $P_x(u)$ denote row-x of $P(u)$. For every $u \in \mathbb{U}$ and any $x, y \in \mathbb{X}_{\ge 0}$ such that $x > y$, $P_x(u) \succeq_f P_y(u)$.*

Then, the value function is EI.

Proof We proceed by backward induction. $V_T(x) = c_T(x)$ which is EI by (C1). This forms the basis of induction. Assume that $V_{t+1}(x)$ is EI and consider $V_t(x)$.

Claim *For every $u \in \mathbb{U}$, $Q_t(x, u)$ is EI.*

Since $V_t(x)$ is the pointwise minimum of $Q_t(x, u)$ over u, $V_t(x)$ is EI. Hence, the result holds by the principle of induction.

Proof (of Claim) We prove the two properties separately:

a) For every $u \in \mathbb{U}$, $Q_t(x,u)$ is even: Consider

$$\begin{aligned} Q_t(-x,u) &= c_t(-x,u) + \sum_{y \in \mathbb{X}} P_{(-x)y}(u) V_{t+1}(y) \\ &\stackrel{(a)}{=} c_t(x,u) + \sum_{-z \in \mathbb{X}} P_{(-x)(-z)}(u) V_{t+1}(-z) \\ &\stackrel{(b)}{=} c_t(x,u) + \sum_{z \in \mathbb{X}} P_{xz}(u) V_{t+1}(z) \\ &= Q_t(x,u) \end{aligned}$$

where (a) follows from (C1) and a change of variables $z = -y$ and (b) follows from (C2) and the induction hypothesis that $V_{t+1}(\cdot)$ is even.

b) For every $u \in \mathbb{U}$, $Q_t(x,u)$ is increasing on $\mathbb{X}_{\geq 0}$: Let $x, y \in \mathbb{X}_{\geq 0}$ such that $x > y$. Consider

$$\begin{aligned} Q_t(x,u) &= c_t(x,u) + \sum_{z \in \mathbb{X}} P_{xz}(u) V_{t+1}(z) \\ &\stackrel{(c)}{\geq} c_t(y,u) + \sum_{z \in \mathbb{X}} P_{yz}(u) V_{t+1}(z) \\ &= Q_t(y,u) \end{aligned}$$

where (c) follows from (C1), (C3), the induction hypothesis that V_{t+1} is EI and then using Lemma 2.

III. Properties of optimal strategy

Definition 5 (Submodular function) A function $f \colon \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ is called submodular if for any $x, y \in \mathbb{X}$ and $u, v \in \mathbb{U}$ such that $x \geq y$ and $u \geq v$, we have

$$f(x,u) + f(y,v) \leq f(x,v) + f(y,u).$$

An equivalent characterization is that

$$f(y,u) - f(y,v) \geq f(x,u) - f(x,v)$$

or

$$f(x,v) - f(y,v) \geq f(x,u) - f(y,u),$$

which imply that the differences are decreasing.

A key property of submodular functions is the following [8, Lemma 4.7.1]

Lemma 3 *Let $f \colon \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ be a submodular function and assume that for all $x \in \mathbb{X}$, $\arg\min_{u \in \mathbb{U}} f(x,u)$ exists. Define*

$$g^*(x) = \max\left\{v \in \arg\min_{u \in \mathbb{U}} f(x,u)\right\}.$$

Then, $g^*(x)$ is (weakly) increasing in $x \in \mathbb{X}$.

Theorem 2 *Suppose that the arg min at each step of the dynamic program is attained and that (C1)–(C3) of Theorem 1 hold. In addition, we have that*

(C4) *For any EI function $w \colon \mathbb{X} \to \mathbb{R}$,*

$$W(x,u) := c_t(x,u) + \sum_{y \in \mathbb{X}} P_{xy}(u) w(y)$$

is submodular on $\mathbb{X}_{\geq 0} \times \mathbb{U}$ for all $t \in \{1, \dots, T-1\}$.

Then, $g_t^*(x)$ given by

$$g_t^*(x) = \begin{cases} \max\left\{v \in \arg\min_{u \in \mathbb{U}} Q_t(x,u)\right\}, & x \geq 0; \\ \min\left\{v \in \arg\min_{u \in \mathbb{U}} Q_t(x,u)\right\}, & x < 0; \end{cases}$$

is EI.

Proof From the Claim in the proof of Theorem 1, (C1)–(C3) imply that $Q_t(x,u)$ is even. Thus, $g_t^*(x)$ defined above is even.

Moreover, by Theorem 1, (C1)–(C3) imply that $V_{t+1}(y)$ is EI. Therefore, by (C4), $Q_t(x,u)$ is submodular on $\mathbb{X}_{\geq 0} \times \mathbb{U}$. Hence, by Lemma 3, $g_t^*(x)$ is incresing on $\mathbb{X}_{\geq 0}$.

A. Monotone dynamic programming

The monotonicity of the optimal strategy can be used to simplify the dynamic program given by (2)–(4) when the state space $\mathbb{X}$ is a set of integers form $\{-a, -a+1, \cdots, a-1, a\}$ and the action space $\mathbb{U}$ is a set of integers of the form $\{\underline{u}, \underline{u}+1, \cdots, \bar{u}-1, \bar{u}\}$.

Initialize $V_T(x)$ as in (2). Now, suppose $V_{t+1}(\cdot)$ has been calculated. Instead of computing $Q_t(x,u)$ and $V_t(x)$ according to (3) and (4), we proceed as follows:

1) Set $x = 0$ and $w_x = \underline{u}$.
2) For all $u \in [w_x, \bar{u}]$, compute $Q_t(x,u)$ according to (3).
3) Instead of (4), compute

$$V_t(x) = \min_{u \in [w_x, \bar{u}]} Q_t(x,u)$$

and set

$$g_t(x) = \max\{v \in [w_x, \bar{u}] \text{ s.t. } V_t(x) = Q_t(x,v)\}.$$

4) Set $V_t(-x) = V_t(x)$.
5) If $x = a$, then stop. Otherwise, set $w_{x+1} = g_t(x)$ and $x = x+1$. Go to step 2.

B. Sufficient conditions for (C4)

Condition (C4) can be difficult to verify. In this section, we provide easier to verify sufficient conditions for (C4).

Lemma 4 *The following conditions imply (C4):*

(C4a) *For all t, $c_t(x,u)$ is submodular on $\mathbb{X}_{\geq 0} \times \mathbb{U}$.*

(C4b) *For every $u \in \mathbb{U}$, let $\tilde{P}_x(u) = \mathcal{F} P_x(u)$.[2] Then, for all $y \in \mathbb{X}_{\geq 0}$,*

$$Q(y|x,u) := \sum_{z \geq y} \tilde{P}_{xz}(u) = \sum_{z \geq y} \left[P_{xz}(u) + P_{x(-z)}(u)\right]$$

is submodular in $(x,u) \in \mathbb{X}_{\geq 0} \times \mathbb{U}$.

Remark 2 Note that (C3) is equivalent to saying that $Q(y \,|\, x,u)$ is increasing in x.

Proof Consider $x^+, x^- \in \mathbb{X}_{\geq 0}$ and $u^+, u^- \in \mathbb{U}$ such that $x^+ \geq x^-$ and $u^+ \geq u^-$. Define two probability measures on $\mathbb{X}$ as follows: for all $x \in \mathbb{X}$,

$$\begin{aligned} \pi_x &= 0.5 P_{x^+x}(u^+) + 0.5 P_{x^-x}(u^-) \\ \mu_x &= 0.5 P_{x^+x}(u^-) + 0.5 P_{x^-x}(u^+) \end{aligned}$$

[2]Recall that $P_x(u)$ denotes row-x of $P(u)$.

Since $Q(y|x,u)$ is submodular, we have that

$$Q(y|x^+,u^+)+Q(y|x^-,u^-) \le Q(y|x^+,u^-)+Q(y|x^-,u^+).$$

Substituting the definition of $Q(y|x,u)$, we get

$$\sum_{z\ge y}\left[\tilde{P}_{x^+z}(u^+)+\tilde{P}_{x^-z}(u^-)\right] \le \sum_{z\ge y}\left[\tilde{P}_{x^+z}(u^-)+\tilde{P}_{x^-z}(u^+)\right]$$

Subsituting the definition of π and μ, we get

$$\sum_{z\ge y}\mathcal{F}\pi_z \le \sum_{z\ge y}\mathcal{F}\mu_z.$$

Hence, $\mathcal{F}\pi \preceq_s \mathcal{F}\mu$ or, equivalently, $\pi \preceq_f \mu$.

Now, suppose $w\colon \mathbb{X}\to\mathbb{R}$ is EI. Then, by Lemma 2

$$\sum_{z\in\mathbb{Z}}\pi_z w(z) \le \sum_{z\in\mathbb{Z}}\mu_z w(z).$$

or equivalently,

$$\sum_{z\in\mathbb{Z}}\left[P_{x^+z}(u^+)w(z)+P_{x^-z}(u^-)\right] \le \sum_{z\in\mathbb{Z}}\left[P_{x^+z}(u^-)w(z)+P_{x^-z}(u^+)\right] \quad (9)$$

Since the sum of two submodular functions is submodular, (C4a) and (9) imply (C4).

IV. A remark on randomized actions

Suppose $\mathbb{U}$ is a discrete set of the form $\{\underline{u},\underline{u}+1,\dots,\bar{u}\}$. In constrained optimization problems, it is often useful to consider the action space $\mathbb{W}=[\underline{u},\bar{u}]$, where for $u,u+1\in\mathbb{U}$, an action $w\in(u,u+1)$ corresponds to a randomization between the "pure" actions u and $u+1$. More precisely, let transition probability $\breve{P}$ corresponding to $\mathbb{W}$ be given as follows: for any $w\in(u,u+1)$,

$$\breve{P}(w)=(1-\theta(w))P(u)+\theta(w)P(u+1)$$

where $\theta:\mathbb{W}\to[0,1]$ is such that for any $u\in\mathbb{U}$, $\lim_{w\downarrow u}\theta(w)=0$ and $\lim_{w\uparrow u+1}\theta(w)=1$, so that $\breve{P}(w)$ is continuous at all $u\in\mathbb{U}$.

Theorem 3 *If $P(u)$ satisfies* (C2) *and* (C3)*, then so does $\breve{P}(w)$. In addition, if for any $u,u+1\in\mathbb{U}$, $\theta(w)$ is weakly increasing on $(u,u+1)$ and if $P(u)$ satifies* (C4b)*, then $\breve{P}(w)$ also satisfies* (C4b).

Proof Since $\breve{P}(w)$ is linear in $P(u)$ and $P(u+1)$, both of which satisfy (C2) and (C3), so does $\breve{P}(w)$.

To prove the last part, note that

$$\breve{Q}(y|x,w)=Q(y|x,u)+\theta(w)[Q(y|x,u+1)-Q(y|x,u)].$$

Since $Q(y|x,u+1)-Q(y|x,u)$ is decreasing in x and $\theta(w)$ is increasing in w for $w\in(u,u+1)$, $\breve{Q}(y|x,w)$ is sumodular in (x,w) on $\mathbb{X}\times(u,u+1)$. Now consider $w-\in(u-1,u)$ and $w^+\in(u,u+1)$. By continuity of $\breve{P}(w)$ (and therefore that of $\breve{Q}(y|x,w)$) at u, $\breve{Q}(y|x,w)$ is submodular on $\mathbb{X}\times\mathbb{W}$.

V. An example: Birth-death Markov chain with controlled restarts

Consider a controlled Markov chain with controlled restarts. In particular, there are two actions, i.e., $\mathbb{U}=\{0,1\}$. Under action $u=0$, the Markov chain is a birth-death Markov chain with state dependent transition probabilities given as follows:

- For $x=0$,

$$P_{0y}(0)=\begin{cases}p_0, & y=1\\ q_0, & y=-1\\ r_0, & y=0\\ 0, & \text{otherwise,}\end{cases}$$

  where p_0,q_0,r_0 are non-negative and add to one.
- For $|x|=1$

$$P_{xy}(0)=\begin{cases}p_x, & y=x+1\\ q_x, & y=x-1\\ r_x, & y=x\\ 0, & \text{otherwise,}\end{cases}$$

  where p_x,q_x,r_x are non-negative and add to one.
- For $|x|>1$,

$$P_{xy}(0)=\begin{cases}p_x, & y=x+1\\ q_x, & y=x-1\\ r_x, & y=x\\ s_x, & y=0\\ 0, & \text{otherwise}\end{cases}$$

  where p_x,q_x,r_x are non-negative and add to one.

Under action $u=1$, the state restarts at 0, i.e.,

$$P_{xy}(1)=\begin{cases}1, & y=0\\ 0, & \text{otherwise.}\end{cases}$$

Proposition 1 *In the birth-death Markov chain with controlled restarts described above:*

1) (C2) *is satisfied iff $p_x=q_{-x}$ and $r_x=r_{-x}$, and, therefore, $s_x=s_{-x}$.*
2) (C3) *and* (C4b) *are satisfied iff $r_0\ge q_1\ge s_2\ge s_3\ge\cdots$ and for all $x\in\mathbb{X}_{>0}$, $p_x+q_{x+1}+s_{x+1}\le 1$.*

See Appendix I for the proof.

In the special case when $\mathbb{X}=\mathbb{Z}$, $s_x=0$, and the transition probabilities are independent of the state, i.e.,

$$P_{xy}(0)=\begin{cases}p, & y=x+1\\ q, & y=x-1\\ r, & y=x\\ 0, & \text{otherwise;}\end{cases}$$

(C2) is satisfied if $p=q$ and (C3) and (C4b) are satisfied if $1-2p\ge p$ and $2p\le 1$, both of which are satisfied when $p\le 1/3$.

This condition is strict. In particular, if $p>1/3$, then it is possible to construct an example where the value function is not EI. For example, suppose $T = 2$, $p > 1/3$ and let $c_2(0) = 0$, $c_2(\pm 1) = 1$ and for $x \in \mathbb{X} \setminus \{-1, 0, 1\}$, $c_2(x) = 1 + k$, where k is a positive constant. Furthermore, let $c_1(x, 0) = 0$ and $c_1(x, 1) = K$, where $K > 0$ is a large number such that the action $u = 1$ is never optimal at $t = 1$. Thus,

$$V_1(0) = Q_1(0, 0) = 2p$$

and

$$V_1(1) = Q_1(1, 0) = p(1 + k) + (1 - 2p) = pk + 1 - p.$$

Now if $k < (3p - 1)/p$, then $V_1(0) > V_1(1)$ and hence the value function is not increasing on $\mathbb{X}_{\geq 0}$.

VI. Conclusion

In this paper we consider a Markov decision process with discrete state and action spaces (finite or countably infinite) and analyze the monotonicity of the optimal solutions. In particular, we provide the sufficient conditions for the even and increasing property of the optimal decision strategy and the value function corresponding to a suitably defined finite-horizon dynamic program. The proof techniques involve the theory of stochastic ordering and the concept of submodularity. For the case of a birth-death Markov chain with controlled restarts, due to two control actions, we provide easily verifiable expressions for those sufficient conditions.

Appendix I
Proof of Proposition 1

A. Conditions under which (C2) is satisfied

To satisfy (C2), for every u, $P_{xy}(u) = P_{(-x)(-y)}(u)$. This condition is always satisfied for $u = 1$ and satisfied for $u = 0$ if, for all $x \in \mathbb{X}$, $p_x = q_{-x}$, $r_x = r_{-x}$ and $s_x = s_{-x}$. Note that first two equalities implipy the third.

B. Conditions under which (C3) is satisfied

Let $\tilde{P}_x(u) = \mathcal{F}P_x(u)$. Let $\hat{P}_x(u)$ denote the cumulative mass function of $\tilde{P}_x(u)$. To satisfy (C3), for every $u \in \mathbb{U}$ and every $x, y \in \mathbb{X}_{\geq 0}$ such that $x > y$, $\tilde{P}_x(u) \succeq_s \tilde{P}_y(u)$, or equivalently,

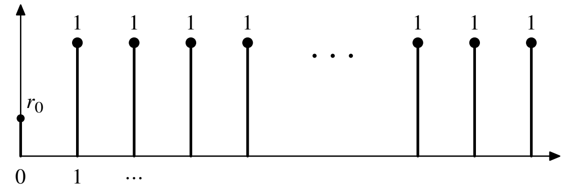
$$\hat{P}_{xz}(u) \leq \hat{P}_{yz}(u), \quad \forall z \in \mathbb{X}_{\geq 0}. \tag{10}$$

$\hat{P}_{xz}(1) = 1$ for all $x, z \in \mathbb{X}_{\geq 0}$. Hence, (C3) is satisfied for $u = 1$.

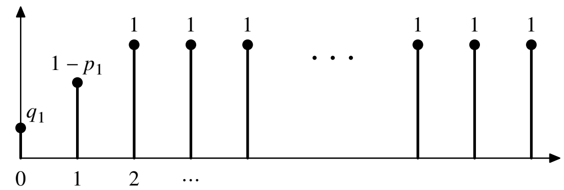
For $u = 0$, $\hat{P}_x(0)$ for $x = 0$ and $x \in \mathbb{X}_{>0}$ is shown in Fig. 1. Consider the following cases:

- $x = 1$ and $y = 0$. Then, for (10) to hold, we need $r_0 \geq q_1$.
- $x = 2$ and $y = 1$. Then for (10) to hold, we need $q_1 \geq s_2$ and $1 - p_1 \geq q_2 + s_2$ (or equivalently $p_1 + q_2 + s_2 \leq 1$).
- $x = y + 1$ and $y > 0$. Then, for (10) to hold, we need $s_y \geq s_x$ and $1 - p_y \geq q_x$, or equivalently, $s_y \geq s_{y+1}$ and $p_y + q_{y+1} \leq 1$.
- $x > y + 1$. Then, (10) always holds.

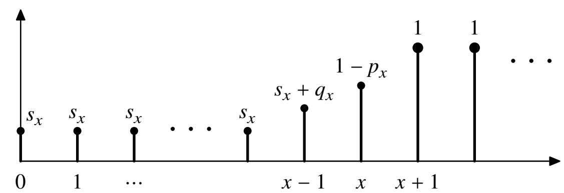
Thus, for (C3) to hold we need $r_0 \geq q_1 \geq s_2 \geq s_3 \geq \cdots$ and for all $y \in \mathbb{X}_{>0}$, $p_y + q_{y+1} \leq 1$.



(a) $\hat{P}_0(0)$, the cumulative mass function of $\mathcal{F}P_0(0)$



(b) $\hat{P}_1(0)$, the cumulative mass function of $\mathcal{F}P_1(0)$



(c) $\hat{P}_x(0)$, the cumulative mass function of $\mathcal{F}P_x(0)$, for $x > 1$

Fig. 1

C. Conditions under which (C4b) is satisfied

Note that $Q(y + 1|x, u) = \hat{P}_{xy}(u)$, where $\hat{P}_x(u)$ is the cumuluative mass function of $\mathcal{F}P_x(u)$ and is shown in Fig. 1. We want $Q(y|x, u)$ to be submodular in $(x, u) \in \mathbb{X}_{\geq 0} \times \mathbb{U}$. Thus, for $x^+ > x^-$ and $u^+ > u^-$,

$$Q(y|x^+, u^+) + Q(y|x^-, u^-) \leq Q(y|x^+, u^-) + Q(y|x^-, u^+),$$

or equivalently,

$$\hat{P}_{x+y}(u^+) + \hat{P}_{x-y}(u^-) \geq \hat{P}_{x+y}(u^-) + \hat{P}_{x-y}(u^+)$$

Since there are only two actions, the above equation simplifies to

$$\hat{P}_{x+y}(1) + \hat{P}_{x-y}(0) \geq \hat{P}_{x+y}(0) + \hat{P}_{x-y}(1).$$

For $y = 0$, $P_{xy}(1) = 1$ and for $y \neq 0$, $P_{xy}(1) = 0$. Thus, in both cases, the above equation simplifies to

$$\hat{P}_{x-y}(0) \geq \hat{P}_{x+y}(0)$$

which is the same condition as (10).

References

[1] C. Derman, "On optimal replacement rules when changes of state are Markovian," *Mathematical optimization techniques*, vol. 396, 1963.

[2] P. Kolesar, "Minimum cost replacement under Markovian deterioration," *Management Science*, vol. 12, no. 9, pp. 694–706, 1966.

[3] H. M. Taylor, "Evaluating a call option and optimal timing strategy in the stock market," *Management Science*, vol. 14, no. 1, pp. 111–120, 1967.

[4] R. C. Merton, “Theory of rational option pricing,” *The Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141–183, 1973.

[5] M. J. Sobel, “Optimal operation of queues,” in *Mathematical methods in queueing theory*. Springer, 1974, pp. 231–261.

[6] S. Stidham Jr and R. R. Weber, “Monotonic and insensitive optimal policies for control of queues with undiscounted costs,” *Operations Research*, vol. 37, no. 4, pp. 611–625, 1989.

[7] R. F. Serfozo, *Stochastic Systems: Modeling, Identification and Optimization, II*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1976, ch. Monotone optimal policies for Markov decision processes, pp. 202–215.

[8] M. Puterman, *Markov decision processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[9] C. C. White, “Monotone control laws for noisy, countable-state Markov chains,” *European Journal of Operational Research*, vol. 5, no. 2, pp. 124–132, 1980.

[10] S. M. Ross, *Introduction to Stochastic Dynamic Programming: Probability and Mathematical*. Orlando, FL, USA: Academic Press, Inc., 1983.

[11] D. P. Heyman and M. J. Sobel, *Stochastic Models in Operations Research*. New York, USA: McGraw Hill, 1984.

[12] N. L. Stokey and R. E. Lucas, Jr, *Recursive methods in economic dynamics*. Harvard University Press, 1989.

[13] D. M. Topkis, “Minimizing a submodular function on a lattice,” *Oper. Res.*, vol. 26, no. 2, pp. 305–321, 1978.

[14] ——, *Supermodularity and Complementarity*, D. M. Kreps and T. J. Sargent, Eds. Princeton University Press, 1998.

[15] K. Papadaki and W. B. Powell, “Monotonicity in multidimensional markov decision processes for the batch dispatch problem,” *Operations research letters*, vol. 35, no. 2, pp. 267–272, 2007.

[16] J. E. Smith and K. F. McCardle, “Structural properties of stochastic dynamic programs,” *Operations Research*, vol. 50, no. 5, pp. 796–809, 2002.