On evenness and monotonicity of value functions and optimal strategies in Markov decision processes

Jhelum Chakravorty, Aditya Mahajan

Electrical and Computer Engineering, McGill University, Canada

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 by using a folded representation of a Markov decision process. An example of a birth-death Markov chain with controlled restart is provided.

Keywords: Markov decision processes, stochastic dominance, submodularity

1. Introduction

1.1. 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, 8, 9, 10, 11, 12]. Similar conditions for vector-valued states are identified in [13, 14, 15]. General conditions under which the value function is increasing and convex are established in [16].

Most of the above results are motivated by queueing models where the state (i.e., the queue length) takes non-negative values. However, there are other applications, particularly those in systems and control, where the state takes both positive and negative values. Often, the system behavior is symmetric for positive and negative values, so one expects the optimal strategy to be even. In this paper, we identify sufficient conditions under which the value function and optimal strategy are even and monotone increasing on the positive values of the state space.

1.2. Model and problem formulation

Consider a Markov decision process (MDP) with state space $\mathbb{X}$ and action space $\mathbb{U}$. For ease of exposition, we assume that $\mathbb{X}$ is a symmetric subset of integers (either a finite set of the form $\{-a, \dots, a\}$ or the countably infinite set $\mathbb{Z}$) and $\mathbb{U}$ is either finite subset of integers or a compact subset of reals; under mild technical conditions the results also extend to the case when $\mathbb{X}$ is a symmetric subset of reals.

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 the probability mass function P_X and the state evolves in a controlled Markov manner, i.e.,

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

where $x_{1:t}$ is a short hand notation for $(x_1, \dots, x_t)$ and a similar interpretation holds of $u_{1:t}$. We assume that the state evolution is time-homogeneous, 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 any time $t \in \{1, \dots, T-1\}$, $c_t \colon \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ 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 $\mathbf{g} := (g_1, \dots, g_{T-1})$ to minimize the expected total cost

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

We denote such an MDP by $(\mathbb{X}, \mathbb{U}, P, c_t)$.

From Markov decision theory [8], we know that an optimal strategy is given by the solution of the following dynamic program. Recursively define value functions

$V_t \colon \mathbb{X} \to \mathbb{R}$ and value-action functions $Q_t \colon \mathbb{X} \times \mathbb{U} \to \mathbb{R}$ as follows: for all $x \in \mathbb{X}$ and $u \in \mathbb{U}$,

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

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

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

$$V_t(x) = \min_{u \in \mathbb{U}} Q_t(x,u). \tag{4}$$

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. To avoid ambiguity when the arg min is not unique, we pick

$$g_t^*(x) = \max\left\{ v \in \arg\min_{u \in \mathbb{U}} Q_t(x,u) \right\}. \tag{5}$$

1.3. Notation and terminology

For a probability distribution π on $\mathbb{X}$, π_x denotes its value at $x \in \mathbb{X}$. For a transition probability matrix P on $\mathbb{X} \times \mathbb{X}$, P_{xy} denotes its value at $x, y \in \mathbb{X}$ and (with a slight abuse of notation) P_x denotes the probability distribution corresponding to the x-th row of P. We say that a P is *even* if for all $x, y \in \mathbb{X}$, $P_{xy} = P_{(-x)(-y)}$.

We say that *an MDP is even* if for every t and every $u \in \mathbb{U}$, $V_t(x)$, $Q_t(x,u)$ and $g_t^*(x)$ are even in x.

Let $\mathbb{X}_{\geq 0}$ and $\mathbb{X}_{>0}$ denote the sets $\{x \in \mathbb{X} : x \geq 0\}$ and $\{x \in \mathbb{X} : x > 0\}$. We say that a function $f \colon \mathbb{X} \to \mathbb{R}$ is *even and increasing* if it is even and for $x \in \mathbb{X}_{>0}$, $f(x-1) \leq f(x)$. In the rest of the paper, we will identify conditions under which V_t and g_t^* are even and increasing.

1.4. Main result

The main result of our paper is the following.

Theorem 1 *Given an MDP $(\mathbb{X}, \mathbb{U}, P, u)$, define for $x, y \in \mathbb{X}_{\geq 0}$ and $u \in \mathbb{U}$,*

$$S(y|x,u) = \sum_{z \geq y} [P_{xz}(u) + P_{x(-z)}(u)].$$

Consider the following conditions:

(C1) *$c_T(\cdot)$ is even and increasing and for $t \in \{1, \dots, T-1\}$ and $u \in \mathbb{U}$, $c_t(\cdot, u)$ is even and increasing.*

(C2) *For all $u \in \mathbb{U}$, $P(u)$ is even.*

(C3) *For all $u \in \mathbb{U}$ and $y \in \mathbb{X}_{\geq 0}$, $S(y|x,u)$ is even and increasing in x.*

(C4) *For $t \in \{1, \dots, T-1\}$, $c_t(x,u)$ is submodular[1] in (x,u) on $\mathbb{X}_{\geq 0} \times \mathbb{U}$.*

(C5) *For all $y \in \mathbb{X}_{\geq 0}$, $S(y|x,u)$ is submodular in (x,u) on $\mathbb{X}_{\geq 0} \times \mathbb{U}$.*

Then, under (C1)–(C3), *$V_t(\cdot)$ is even and increasing for all $t \in \{1, \dots, T\}$ and under* (C1)–(C5), *$g_t^*(\cdot)$ is even and increasing for all $t \in \{1, \dots, T-1\}$.*

In the rest of the paper, we provide a proof of this result.

2. Folded representation of even MDPs

2.1. Sufficient condition for MDP to be even

We start by identifying sufficient conditions for an MDP to be even.

Proposition 1 *Suppose an MDP $(\mathbb{X}, \mathbb{U}, P, c)$ satisfies the following properties:*

(A1) *$c_T(\cdot)$ is even and for every $t \in \{1, \cdots, T-1\}$ and $u \in \mathbb{U}$, $c_t(\cdot, u)$ is even.*

(A2) *For every $u \in \mathbb{U}$, $P(u)$ is even.*

Then, the MDP is even.

Proof We proceed by backward induction. $V_T(x) = c_T(x)$ which is even by (A1). This forms the basis of induction. Now assume that $V_{t+1}(x)$ is even in x. For any $x \in \mathbb{X}$, we show that for every $u \in \mathbb{U}$, $Q_t(x,u)$ is even. Consider,

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

where (a) follows from (A1) and a change of variables $y = -z$ and (b) follows from (A2) and the induction hypothesis that $V_{t+1}(\cdot)$ is even. Hence, $Q_t(\cdot, u)$ is even.

Since $Q_t(\cdot, u)$ is even, Eqs. (4) and (5) imply that V_t and g_t^* are also even. Thus, the result is true for time t and, by induction, true for all time t.

2.2. Folding operator for distributions

We now show that if the value function is even, we can construct a "folded" MDP with state-space $\mathbb{X}_{\geq 0}$ such that the value function and optimal strategy of the folded MDP match that of the original MDP on $\mathbb{X}_{\geq 0}$. For that matter, we first define the following:

Definition 1 (Folding Operator) Given a probability distribution π on $\mathbb{X}$, the folding operator $\mathcal{F}\pi$ gives a distribution $\tilde{\pi}$ on $\mathbb{X}_{\geq 0}$ such that

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

[1] Submodularity is defined in Sec. 3.2

For example, if $\pi = \{0.1, 0.15, 0.5, 0.15, 0.1\}$ defined on $\mathbb{X} = \{-2, -1, 0, 1, 2\}$, then $\mathcal{F}\pi$ is the probability distribution $\{0.5, 0.3, 0.2\}$ defined on $\mathbb{X}_{\geq 0} = \{0, 1, 2\}$.

An immediate implication of the Definition 1 is the following:

Lemma 1 *If $f : \mathbb{X} \to \mathbb{R}$ is even, then for any probability distribution π on $\mathbb{X}$ and $\tilde{\pi} = \mathcal{F}\pi$, we have*

$$\sum_{x \in \mathbb{X}} f(x)\pi_x = \sum_{x \in \mathbb{X}_{\geq 0}} f(x)\tilde{\pi}_x.$$

Now, we generalize the folding operator to transition probability matrices.

Definition 2 Given a probability transition matrix P on $\mathbb{X} \times \mathbb{X}$, the folding operator $\mathcal{F}P$ gives a transition probability matrix $\tilde{P}$ on $\mathbb{X}_{\geq 0} \times \mathbb{X}_{\geq 0}$ such that for any $x \in \mathbb{X}_{\geq 0}$,

$$\tilde{P}_x = \mathcal{F}P_x.$$

For example, if $\mathbb{X} = \{-1, 0, 1\}$ and

$$P = \begin{array}{c} \\ -1 \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{ccc} -1 & 0 & 1 \end{array} \\ \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{bmatrix} \end{array} \quad \text{then} \quad \tilde{P} = \begin{array}{c} \\ 0 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} 0 & 1 \end{array} \\ \begin{bmatrix} 0.4 & 0.6 \\ 0.3 & 0.7 \end{bmatrix} \end{array}$$

Definition 3 (Folded MDP) Given an MDP $(\mathbb{X}, \mathbb{U}, P, c_t)$, define the *folded MDP* as $(\mathbb{X}_{\geq 0}, \mathbb{U}, \tilde{P}, c_t)$, where for all $u \in \mathbb{U}$, $\tilde{P}(u) = \mathcal{F}P(u)$.

Let $\tilde{Q}_t$ and $\tilde{V}_t$ and $\tilde{g}_t^*$ denote respectively the value-action function, the value function and the optimal strategy of the folded MDP. Then, we have the following.

Proposition 2 *If the MDP $(\mathbb{X}, \mathbb{U}, P, c_t)$ is even, then for any $x \in \mathbb{X}$ and $u \in \mathbb{U}$,*

$$Q_t(x, u) = \tilde{Q}_t(|x|, u), \quad V_t(x) = \tilde{V}_t(|x|), \quad g_t^*(x) = \tilde{g}_t^*(|x|). \tag{6}$$

Proof We proceed by backward induction. For $x \in \mathbb{X}$ and $\tilde{x} \in \mathbb{X}_{\geq 0}$, $V_T(x) = c_T(x)$ and $\tilde{V}_T(\tilde{x}) = c_T(\tilde{x})$. Since $V_T(\cdot)$ is even, $V_T(x) = V_T(|x|) = \tilde{V}_T(|x|)$. This is the basis of induction. Now assume that for all $x \in \mathbb{X}$, $V_{t+1}(x) = \tilde{V}_{t+1}(|x|)$. Consider $x \in \mathbb{X}_{\geq 0}$ and $u \in \mathbb{U}$. Then we have

$$\begin{aligned} Q_t(x, u) &= c_t(x, u) + \sum_{y \in \mathbb{X}} P_{xy}(u) V_{t+1}(y) \\ &\overset{(a)}{=} c_t(x, u) + \sum_{y \in \mathbb{X}_{\geq 0}} \tilde{P}_{xy}(u) V_{t+1}(y) \\ &\overset{(b)}{=} c_t(x, u) + \sum_{y \in \mathbb{X}_{\geq 0}} \tilde{P}_{xy}(u) \tilde{V}_{t+1}(y) = \tilde{Q}_t(x, u), \end{aligned}$$

where (a) uses Lemma 1 and that V_{t+1} is even and (b) uses the induction hypothesis.

Since the Q-functions match for $x \in \mathbb{X}_{\geq 0}$, equations (4) and (5) imply that the value functions and the optimal strategies also match on $\mathbb{X}_{\geq 0}$, i.e., for $x \in \mathbb{X}_{\geq 0}$,

$$V_t(x) = \tilde{V}_(x) \quad \text{and} \quad g_t^*(x) = \tilde{g}_t^*(x).$$

Since V_t and g_t^* are even, we get that (6) is true at time t. Hence, by principle of induction, it is true for all t.

3. Monotonicity of the value function and the optimal strategy

We have shown that under (A1) and (A2) the original MDP is equivalent to a folded MDP with state-space $\mathbb{X}_{\geq 0}$. Thus, we can use standard conditions to determine when the value function and the optimal strategy of the folded MDP are monotone. Translating these conditions back to the original model, we get the sufficient conditions for the original model.

3.1. Monotonicity of the value function

The results on monotonicity of value functions rely on the notion of stochastic dominance.

Definition 4 (Stochastic Dominance) Given two probability distributions μ and π defined over $\mathbb{X}_{\geq 0}$, μ is said to stochastically dominate π, which is denoted by $\mu \succeq_s \pi$, if

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

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

$$M_x \leq P_x, \quad \forall x \in \mathbb{X}_{\geq 0}. \tag{7}$$

Proposition 3 [8, **Theorem 4.7.4**] *Suppose the folded MDP $(\mathbb{X}_{\geq 0}, \mathbb{U}, \tilde{P}, c)$ satisfies the following:*

(B1) *$c_T(x)$ is increasing in x for $x \in \mathbb{X}_{\geq 0}$; for any $t \in \{1, \dots, T-1\}$ and $u \in \mathbb{U}$, $c_t(x, u)$ is increasing in x for $x \in \mathbb{X}_{\geq 0}$.*

(B2) *For any $u \in \mathbb{U}$ and any $x, y \in \mathbb{X}_{\geq 0}$ such that $x > y$, $\tilde{P}_x(u) \succeq_s \tilde{P}_y(u)$.*

Then, for any $t \in \{1, \dots, T\}$, $\tilde{V}_t(x)$ is increasing in x for $x \in \mathbb{X}_{\geq 0}$.

For ease of presentation, define for $x, y \in \mathbb{X}_{\geq 0}$ and $u \in \mathbb{U}$,

$$S(y|x, u) := \sum_{z \geq y} \tilde{P}_{xz}(u) = \sum_{z \geq y} [P_{xz}(u) + P_{x(-z)}(u)]. \tag{8}$$

From (7), (B2) is equivalent to the following:

(B2') For every $u \in \mathbb{U}$ and $x, y \in \mathbb{X}_{\geq 0}$, $S(y|x, u)$ is increasing in x.

An immediate consequence of Propositions 1, 2, and 3 is the following:

Corollary 1 *Under* (A1), (A2), (B1), *and* (B2) *(or* (B2')*), the value functions* $V_t(\cdot)$ *are even and increasing.*

Remark 1 Note that (A1) and (B1) are equivalent to (C1), (A2) is same as (C2), and (A2) and (B2) (or equivalently, (A2) and (B2')) are equivalent to (C3). Thus, Corollary 1 proves the first part of Theorem 1.

3.2. Monotonicity of the optimal strategy

Now we state sufficient conditions under which the optimal strategy is increasing. These results rely on the notion of submodularity.

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 of submodularity is that

$$\begin{aligned} & f(y,u) - f(y,v) \geq f(x,u) - f(x,v), \\ \implies\ & f(x,v) - f(y,v) \geq f(x,u) - f(y,u), \end{aligned}$$

which implies that the differences are decreasing.

Proposition 4 [8, Theorem 4.7.4] *Suppose that in addition to* (B1) *and* (B2) *(or* (B2')*), the folded MDP* $(\mathbb{X}_{\geq 0}, \mathbb{U}, \tilde{P}, c_t)$ *satisfies the following property:*

(B3) *For all* $t \in \{1, \dots, T-1\}$, $c_t(x,u)$ *is submodular in* (x,u) *on* $\mathbb{X}_{\geq 0} \times \mathbb{U}$.

(B4) *For all* $y \in \mathbb{X}_{\geq 0}$, $S(y|x,u)$ *is submodular in* (x,u) *on* $\mathbb{X}_{\geq 0} \times \mathbb{U}$, *where* $S(y|x,u)$ *is defined in* (8).

Then, for every $t \in \{1, \cdots, T-1\}$, *the optimal strategy* $\tilde{g}_t^*(x)$ *is increasing in* x *for* $x \in \mathbb{X}_{\geq 0}$.

An immediate consequence of Propositions 1, 2, and 3 is the following:

Corollary 2 *Under* (A1), (A2), (B1), (B2) *(or* (B2')*),* (B3), *and* (B4) *the optimal strategy* $g_t^*(\cdot)$ *is even and increasing.*

Remark 2 As argued in Remark 1, (A1), (A2), (B1), (B2) are equivalent to (C1)–(C3). Note that (B3), (B4) is the same as (C4), (C5). Thus, Corollary 2 proves the second part of Theorem 1.

4. Some remarks

4.1. Monotone dynamic programming

Under (C1)–(C4), the even and monotone property of the optimal strategy strategy can be used to simplify the dynamic program given by (2)–(4). For conciseness, assume that 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

$$\begin{aligned} V_t(x) &= \min_{u \in [w_x, \bar{u}]} Q_t(x,u), \quad \text{and set} \\ g_t(x) &= \max\{v \in [w_x, \bar{u}] \text{ s.t. } V_t(x) = Q_t(x,v)\}. \end{aligned}$$

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.

4.2. 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 2 *If* $P(u)$ *satisfies* (C2), (C3), *and* (C5) *then so does* $\breve{P}(w)$.

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 that $\breve{P}(w)$ satisfies (C4), note that

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

Since $S(y|x,u+1) - S(y|x,u)$ is decreasing in x and $\theta(w)$ is increasing in w for $w \in (u, u+1)$, $\breve{S}(y|x,w)$ is submodular 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 $\breve{S}(y|x,w)$) at u, $\breve{S}(y|x,w)$ is submodular on $\mathbb{X} \times \mathbb{W}$.

5. 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.,

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

Such a model arises in remote-state estimation with packet drops [17]. In such a system, X_t denotes the estimation error between the sensor and the receiver. $U_t = 0$ corresponds to the action of not transmitting and $U_t = 1$ corresponds to transmitting. The per-step cost is

$$c_t(x, u) = \begin{cases} d(x), & \text{if } u = 0 \\ \lambda, & \text{if } u = 1, \end{cases}$$

where $d(\cdot)$ denotes an estimation distortion and λ denotes the transmission cost.

Proposition 5 *In the birth-death Markov chain with controlled restarts described above:*

1. (C1) *and* (C4) *are satisfied if* $d(\cdot)$ *is even and increasing function of* x.
2. (C2) *is satisfied iff* $p_x = q_{-x}$ *and* $r_x = r_{-x}$, *and, therefore,* $s_x = s_{-x}$.
3. (C3) *and* (C5) *are satisfied iff* $r_0 \geq q_1 \geq s_2 \geq s_3 \geq \cdots$ *and for all* $x \in \mathbb{X}_{>0}$, $p_x + q_{x+1} + s_{x+1} \leq 1$.

See Appendix A 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) satisfied if $1 - 2p \geq p$ and $2p \leq 1$, both of which are satisfied when $p \leq 1/3$.

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 $\mathbb{X}_{\geq 0}$.

6. 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 relies on a folded representation of the Markov decision process. 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.

7. Ackowledgements

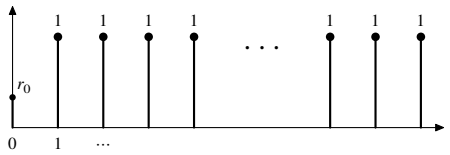
This research was funded through NSERC Discovery Accelerator Grant 493011.

Appendix A. Proof of Proposition 5

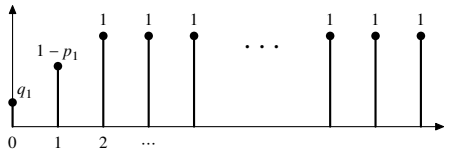
Appendix A.1. Conditions under which (C1) and (C4) are satisfied

The one-step cost can be expressed as $c_t(x, u) = \lambda u + d(x)(1 - u)$. Then, (C1) is satisfied since symmetry and monotonicity are preserved under addition with a constant.

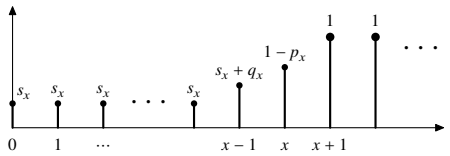
Since there are only two actions, for (C4) to hold we need that for any $x^+, x^- \in \mathbb{X}_{\geq 0}$ such that for $x^+ > x^-$, we have $\lambda - d(x^+) \leq \lambda - d(x^-)$, or equivalently, $d(\cdot)$ is increasing on $\mathbb{X}_{\geq 0}$.



(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 $x > 1$

Figure A.1

Appendix A.2. 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 imply the third.

Appendix A.3. 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), we note from Remark 1 that 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,

$$\hat{P}_{xz}(u) \leq \hat{P}_{yz}(u), \quad \forall z \in \mathbb{X}_{\geq 0}. \tag{A.1}$$

$\hat{P}_{xz}(1) = 1$ 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. A.1. Consider the following cases:

- $x = 1$ and $y = 0$. Then, for (A.1) to hold, we need $r_0 \geq q_1$.
- $x = 2$ and $y = 1$. Then for (A.1) 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 (A.1) 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, (A.1) 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 $y \in \mathbb{X}_{>0}$, $p_y + q_{y+1} \leq 1$.

Appendix A.4. Conditions under which (C5) is satisfied

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

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

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 (A.1).

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

[2] P. Kolesar, Minimum cost replacement under Markovian deterioration, Management Science 12 (9) (1966) 694–706.

[3] H. M. Taylor, Evaluating a call option and optimal timing strategy in the stock market, Management Science 14 (1) (1967) 111–120.

[4] R. C. Merton, Theory of rational option pricing, The Bell Journal of Economics and Management Science 4 (1) (1973) 141–183.

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

[6] S. Stidham Jr, R. R. Weber, Monotonic and insensitive optimal policies for control of queues with undiscounted costs, Operations Research 37 (4) (1989) 611–625.

[7] R. F. Serfozo, Stochastic Systems: Modeling, Identification and Optimization, II, Springer Berlin Heidelberg, 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 5 (2) (1980) 124–132.

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

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

[12] N. L. Stokey, 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. 26 (2) (1978) 305–321.

[14] D. M. Topkis, Supermodularity and Complementarity, Princeton University Press, 1998.

[15] K. Papadaki, W. B. Powell, Monotonicity in multidimensional markov decision processes for the batch dispatch problem, Operations research letters 35 (2) (2007) 267–272.

[16] J. E. Smith, K. F. McCardle, Structural properties of stochastic dynamic programs, Operations Research 50 (5) (2002) 796–809.

[17] J. Chakravorty, A. Mahajan, Remote-state estimation with packet drop, in: IFAC Workshop on Distributed Estimation and Control in Networked Systems, 2016.