Operations Research Letters 46 (2018) 179–184

Contents lists available at ScienceDirect

Operations Research Letters

journal homepage: www.elsevier.com/locate/orl

Reduction of total-cost and average-cost MDPs with weakly continuous transition probabilities to discounted MDPs

Eugene A. Feinberg [a,*], Jefferson Huang [b]

[a] Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794-3600, USA
[b] School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853-3801, USA

ARTICLE INFO

Article history:
Received 18 November 2017
Accepted 18 December 2017
Available online 2 January 2018

Keywords:
Markov decision process
Reduction
Total cost
Average cost
Discounted
Inventory

ABSTRACT

This note describes sufficient conditions under which total-cost and average-cost Markov decision processes (MDPs) with general state and action spaces, and with weakly continuous transition probabilities, can be reduced to discounted MDPs. For undiscounted problems, these reductions imply the validity of optimality equations and the existence of stationary optimal policies. The reductions also provide methods for computing optimal policies. The results are applied to a capacitated inventory control problem with fixed costs and lost sales.

1. Introduction

Undiscounted Markov decision processes (MDPs) are typically much more difficult to study than discounted MDPs. This is true both for models with expected total costs and for models with average costs per unit time. This paper describes conditions under which undiscounted MDPs with infinite state spaces and weakly continuous transition kernels can be transformed into discounted MDPs.

For undiscounted total costs, a classic assumption is that the expected number of visits to each state in a certain set $\mathbb{X}'$ is finite under every policy and initial state. Such an assumption is typically referred to as *transience* [3, Chapter 7], [18]. When the expected amount of time spent in $\mathbb{X}'$ (i.e., the "lifetime" of the system) is finite for every policy and initial state, the MDP is called *absorbing* [3, Chapter 7]. It is well-known that every discounted MDP can be viewed as an absorbing MDP with the lifetime of the system being geometrically distributed [3, p. 137]. We remark that every absorbing MDP is transient, and that the two conditions are equivalent when the set $\mathbb{X}'$ is finite.

For average costs per unit time, a classic approach has been to make use of results about discounted MDPs. The most general results have been obtained in [10] using the so-called vanishing discount factor approach, where the validity of optimality inequalities and existence of stationary optimal policies are obtained by considering optimality equations for discounted MDPs and letting the discount factor tend to one. Another approach, which was used early in the development of the theory of average-cost MDPs, is to transform the average-cost problem into a discounted one, and argue that optimal policies for the latter are also optimal for the former [6, Chapter 7 §10], [16,17]. One advantage of this approach is that it can be used to apply methods and algorithms developed for discounted MDPs to undiscounted MDPs. [1,8,9].

In [9], conditions were given under which undiscounted MDPs with general state and action spaces can be reduced to discounted ones. These conditions include the assumption that the transition probabilities are setwise continuous. However, for many models of interest, such as those arising in inventory control [7], the transition probabilities are only weakly continuous. In this paper, we provide conditions under which the reductions in [9] lead to optimality results for undiscounted MDPs with weakly continuous transition kernels. In particular, under these conditions the discounted MDPs to which the undiscounted MDPs are reduced have weakly continuous transition probabilities. Moreover, while sufficient conditions are provided in [5,12,15] for the validity of the optimality equations for average-cost MDPs, Assumption HT in Section 4 ensures that a solution to this optimality equation can be obtained via the optimality equation for a discounted MDP. This in turn implies that such average-cost MDPs can be solved using methods developed for discounted MDPs.

The rest of the paper is organized as follows. In Section 2, the MDP model and objective functions are described. Next, in

* Corresponding author.
E-mail addresses: eugene.feinberg@stonybrook.edu (E.A. Feinberg), jefferson.huang@cornell.edu (J. Huang).

https://doi.org/10.1016/j.orl.2017.12.007
0167-6377/© 2017 Elsevier B.V. All rights reserved.

Section 3 the results for undiscounted total-cost MDPs are presented. Section 4 contains the results for average-cost MDPs. Finally, in Section 5 we apply the preceding results to a capacitated inventory control problem with fixed ordering costs and lost sales.

2. Model description

The *state space* $\mathbb{X}$ and *action space* $\mathbb{A}$ are Borel subsets of complete separable metric spaces endowed with their respective Borel σ-algebras $\mathcal{B}(\mathbb{X})$ and $\mathcal{B}(\mathbb{A})$. When the current state is $x \in \mathbb{X}$, the decision-maker must select an action from the *set of available actions* $A(x)$, which is a nonempty Borel subset of $\mathbb{A}$. The space of all feasible state–action pairs

$$\mathrm{Gr}(A) := \{(x, a) | x \in \mathbb{X},\ a \in A(x)\}$$

is assumed to be a Borel subset of $\mathbb{X} \times \mathbb{A}$, and to contain the graph of a Borel-measurable function from $\mathbb{X}$ to $\mathbb{A}$ (these assumptions follow from Assumption WC(i)). For each $(x, a) \in \mathrm{Gr}(A)$ there is an associated *one-step cost* $c(x, a) \in [0, \infty)$ and a finite measure $q(\cdot|x, a)$ on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. We assume that the functions $(x, a) \mapsto c(x, a)$ and $(x, a) \mapsto q(B|x, a)$, for each $B \in \mathcal{B}(\mathbb{X})$, are Borel-measurable. Moreover, q is assumed to satisfy

$$\sup\{q(\mathbb{X}|x, a) : (x, a) \in \mathrm{Gr}(A)\} < \infty.$$

For possible interpretations of the values $q(B|x, a)$ for $B \in \mathcal{B}(\mathbb{X})$, which may be greater than one, see [9, Section 2.1]; in light of these interpretations, we will refer to q as the *transition kernel*.

2.1. Objective functions

Let $\mathbb{H}_0 := \mathbb{X}$, and for $n = 1, 2, \dots$ let $\mathbb{H}_n := \mathbb{X} \times \mathbb{A} \times \mathbb{H}_{n-1}$ denote the space of all *histories* of the process up to decision epoch n, endowed with the product σ-algebra. A *decision rule* for epoch $n = 0, 1, \dots$ is a mapping $\pi_n : \mathcal{B}(\mathbb{A}) \times \mathbb{H}_n \to [0, 1]$ such that for every $h_n = x_0 a_0 \cdots x_n$ the set function $\pi_n(\cdot|h_n)$ is a probability measure on $(\mathbb{A}, \mathcal{B}(\mathbb{A}))$ satisfying $\pi_n(A(x_n)|h_n) = 1$, and for every $B \in \mathcal{B}(\mathbb{A})$ the function $\pi_n(B|\cdot)$ on $\mathbb{H}_n$ is Borel-measurable.

A *policy* is a sequence $\pi = \{\pi_n\}_{n=0}^{\infty}$ of decision rules; let Π denote the set of all policies. Under a policy π, at each decision epoch $n = 0, 1, \dots$ the decision-maker observes the history $h_n = x_0 a_0 \cdots x_n \in \mathbb{H}_n$ of the process up to epoch n and selects an action $a \in A(x_n)$ according to the probability distribution $\pi_n(\cdot|h_n)$. A *stationary policy* is identified with a Borel-measurable function $\phi : \mathbb{X} \to \mathbb{A}$ satisfying $\phi(x) \in A(x)$ for all $x \in \mathbb{X}$; under such a policy, the decision-maker selects the action $\phi(x)$ if the current state is x. The set of all stationary policies is denoted by $\mathbb{F}$.

To define the objective functions under consideration, for $B \in \mathcal{B}(\mathbb{X})$ and $(x, a) \in \mathrm{Gr}(A)$ let

$$p(B|x, a) := q(B|x, a)/q(\mathbb{X}|x, a),$$

and let

$$\alpha(x, a) := q(\mathbb{X}|x, a).$$

Observe that $p(\cdot|x, a)$ is a probability measure on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ for every $(x, a) \in \mathrm{Gr}(A)$, and that $p(B|\cdot)$ is a Borel function on $\mathrm{Gr}(A)$ for every $B \in \mathcal{B}(\mathbb{X})$. Therefore, for every policy $\pi \in \Pi$ and initial state $x \in \mathbb{X}$ the Ionescu Tulcea theorem [4, pp. 140–141] uniquely defines a probability measure $\mathbb{P}_x^{\pi}$ on $((\mathbb{X} \times \mathbb{A})^{\infty}, \mathcal{B}[(\mathbb{X} \times \mathbb{A})^{\infty}])$ and its associated expectation operator $\mathbb{E}_x^{\pi}$.

When the initial state is $x \in \mathbb{X}$, under $\pi \in \Pi$ the *total cost* incurred is

$$v^{\pi}(x) := \mathbb{E}_x^{\pi} \sum_{n=0}^{\infty} \alpha(x_n, a_n) c(x_n, a_n),$$

and the *average cost* incurred is

$$w^{\pi}(x) := \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_x^{\pi} \sum_{n=0}^{N-1} \alpha(x_n, a_n) c(x_n, a_n).$$

A policy $\pi_* \in \Pi$ is *total-cost optimal* if

$$v^{\pi_*}(x) = \inf_{\pi \in \Pi} v^{\pi}(x) =: v(x) \qquad \forall x \in \mathbb{X},$$

and is *average-cost optimal* if

$$w^{\pi_*}(x) = \inf_{\pi \in \Pi} w^{\pi}(x) =: w(x) \qquad \forall x \in \mathbb{X}.$$

If there exists a constant β such that $\alpha(x, a) = \beta$ for all $(x, a) \in \mathrm{Gr}(A)$, a total-cost optimal policy is called *β-optimal*.

3. Total costs

To state Assumption T for the total-cost criterion, given $\phi \in \mathbb{F}$ and a Borel function $u : \mathbb{X} \to \mathbb{R}$ let

$$Q_{\phi} u(x) := \int_{\mathbb{X}} u(y) q(dy|x, \phi(x)), \qquad x \in \mathbb{X},$$

let $Q_{\phi}^0 u(x) := u(x)$ for $x \in \mathbb{X}$, and for $n = 1, 2, \dots$ let $Q_{\phi}^n u(x) := Q_{\phi}(Q_{\phi}^{n-1} u)(x)$ for $x \in \mathbb{X}$.

Assumption T. There exist a continuous function $V : \mathbb{X} \to [1, \infty)$ and a constant K satisfying

$$\sum_{n=0}^{\infty} Q_{\phi}^n V(x) \le K V(x) < \infty, \quad \forall \phi \in \mathbb{F},\ x \in \mathbb{X}. \tag{1}$$

The statement of Assumption WC requires several definitions. Let $\mathbb{S}$ and $\mathbb{T}$ be metric spaces endowed with their respective Borel σ-algebras $\mathcal{B}(\mathbb{S})$ and $\mathcal{B}(\mathbb{T})$. A set-valued mapping $s \mapsto \Phi(s) \subseteq \mathbb{T}$ on $\mathbb{S}$ is *compact-valued* if $\Phi(s)$ is compact for all $s \in \mathbb{S}$, and is *continuous on* $\mathbb{S}$ if for every open set $V \subseteq \mathbb{T}$ the sets $\{s \in \mathbb{S} | \Phi(s) \subseteq V\}$ and $\{s \in \mathbb{S} | \Phi(s) \cap V \neq \emptyset\}$ are open in $\mathbb{S}$.

Next, a *transition kernel* from $\mathbb{S}$ to $\mathbb{T}$ is a mapping $\kappa : \mathcal{B}(\mathbb{T}) \times \mathbb{S} \to [0, \infty)$ such that $\kappa(\cdot|s)$ is a finite measure on $(\mathbb{T}, \mathcal{B}(\mathbb{T}))$ for every $s \in \mathbb{S}$, and $\kappa(T|\cdot)$ is a Borel function on $\mathbb{S}$ for every $T \in \mathcal{B}(\mathbb{T})$. A transition kernel κ is *weakly continuous* if for every bounded continuous function $f : \mathbb{T} \to \mathbb{R}$ the mapping

$$s \mapsto \int_{\mathbb{T}} f(t) \kappa(dt|s)$$

is continuous on $\mathbb{S}$. If κ is a transition kernel such that $\kappa(\cdot|s)$ is a probability measure for every $s \in \mathbb{S}$, it is called a *transition probability kernel*.

Finally, a function $f : \mathbb{S} \to \mathbb{R}$ is *lower semicontinuous at* $s \in \mathbb{S}$ if $\liminf_{s' \to s} f(s') \ge f(s)$, and is *lower semicontinuous on* $S \subseteq \mathbb{S}$ if it is lower semicontinuous at every $s \in S$.

Assumption WC.

(i) The set-valued mapping $x \mapsto A(x)$ is compact-valued and continuous on $\mathbb{X}$.
(ii) The transition kernel q is weakly continuous.
(iii) The function $(x, a) \mapsto c(x, a)$ is lower semicontinuous on $\mathrm{Gr}(A)$.

Proposition 1. *Suppose Assumptions T and WC(i, ii) hold. Then there exists a continuous function $\mu : \mathbb{X} \to [1, \infty)$ satisfying $V(x) \le \mu(x) \le K V(x)$ for all $x \in \mathbb{X}$ and*

$$\mu(x) \ge V(x) + \int_{\mathbb{X}} \mu(y) q(dy|x, a) \tag{2}$$

for all $(x, a) \in Gr(A)$.

Proof. Consider the operator $\mathcal{U}$ defined for Borel functions $u : \mathbb{X} \to \mathbb{R}$ by

$$\mathcal{U}u(x) := \sup_{a\in A(x)} \left[V(x) + \int_{\mathbb{X}} u(y)q(dy|x,a) \right]$$

for $x \in \mathbb{X}$. Let $u_0 \equiv 0$, and for $n = 1, 2, \dots$ let $u_n := \mathcal{U}u_{n-1}$. According to the Berge maximum theorem (see e.g., [2, p. 570]), for $n = 0, 1, \dots$ the function u_n is continuous. Since $u_{n+1} \geq u_n \geq V$ pointwise for $n = 1, 2, \dots$, the sequence of continuous functions $\{u_n\}_{n=0}^{\infty}$ converges to a Borel function $\mu := \lim_{n\to\infty} u_n \geq V$. The claim that $\mu \leq KV$ can be verified using the arguments in [9, Proof of Proposition 1] and the Berge maximum theorem. Moreover, Lebesgue's monotone convergence theorem implies that $\mu = \mathcal{U}\mu$, which means (2) holds for all $(x, a) \in \mathrm{Gr}(A)$.

It remains to be shown that the function $\mu : \mathbb{X} \to \mathbb{R}$ defined above is continuous. First, observe that for any Borel functions f, g on $\mathbb{X}$,

$$f(x) \leq g(x) + \mu(x) \left(\sup_{x\in\mathbb{X}} \frac{|f(x) - g(x)|}{\mu(x)} \right), \ \forall x \in \mathbb{X},$$

which implies that for all $x \in \mathbb{X}$,

$$\begin{aligned} \mathcal{U}f(x) &\leq \mathcal{U}g(x) + (\mu(x) - V(x)) \left(\sup_{x\in\mathbb{X}} \frac{|f(x) - g(x)|}{\mu(x)} \right) \\ &\leq \mathcal{U}g(x) + \mu(x) \left(\frac{K-1}{K} \right) \left(\sup_{x\in\mathbb{X}} \frac{|f(x) - g(x)|}{\mu(x)} \right). \end{aligned}$$

By reversing the roles of f and g, it follows that

$$\frac{|\mathcal{U}f(x) - \mathcal{U}g(x)|}{\mu(x)} \leq \left(\frac{K-1}{K} \right) \sup_{x\in\mathbb{X}} \frac{|f(x) - g(x)|}{\mu(x)}, \ \forall x \in \mathbb{X}.$$

Since $V \leq \mu \leq KV$, it follows that for the sequence $\{u_n\}_{n=0}^{\infty}$ defined above,

$$\sup_{x\in\mathbb{X}} \frac{|u_{n+1}(x) - u_n(x)|}{KV(x)} \leq \left(\frac{K-1}{K} \right)^n, \ n = 0, 1, \dots,$$

which implies that for all nonnegative integers m, n satisfying $m > n$,

$$\begin{aligned} \sup_{x\in\mathbb{X}} &\frac{|u_m(x) - u_n(x)|}{KV(x)} \\ &\leq \sum_{k=0}^{m-n-1} \sup_{x\in\mathbb{X}} \frac{|u_{n+k+1}(x) - u_{n+k}(x)|}{KV(x)} \\ &\leq \sum_{k=0}^{m-n-1} \left(\frac{K-1}{K} \right)^{n+k} \\ &\leq \left(\frac{K-1}{K} \right)^n \sum_{k=0}^{\infty} \left(\frac{K-1}{K} \right)^k \\ &= K \left(\frac{K-1}{K} \right)^n. \end{aligned} \tag{3}$$

Define the *V-norm* for functions $f : \mathbb{X} \to \mathbb{R}$ by $\|f\|_V := \sup_{x\in\mathbb{X}} |f(x)|/V(x)$, and let $C_V(\mathbb{X})$ denote the space of continuous functions on $\mathbb{X}$ with finite V-norm. Then (3) implies that $\{u_n\}_{n=0}^{\infty}$ is a Cauchy sequence in $C_V(\mathbb{X})$. Since $C_V(\mathbb{X})$ is a Banach space with respect to $\|\cdot\|_V$, it follows that the sequence $\{u_n\}_{n=0}^{\infty}$ converges to a function in C_V. Since $\lim_{n\to\infty} u_n = \mu$, it follows that $\mu \in C_V$; in particular, μ is continuous. □

3.1. Hoffman–Veinott (HV) transformation

In this section, we present the HV transformation [9], which is based on ideas due to Alan Hoffman and A. F. Veinott [18]. A point s is *isolated* from a metric space $\mathbb{S}$, if there exists an $\epsilon > 0$ such that the distance between s and any element of $\mathbb{S}$ is larger than ϵ. The state space of the new MDP is $\tilde{\mathbb{X}} := \mathbb{X} \cup \{\tilde{x}\}$, where $\tilde{x} \notin \mathbb{X}$ is a cost-free absorbing state that is isolated from $\mathbb{X}$. The action space is $\tilde{\mathbb{A}} := \mathbb{A} \cup \{\tilde{a}\}$, where $\tilde{a}$ is the only action available when the current state is $\tilde{x}$. The set $\tilde{A}(x)$ of available actions is unchanged if the current state x is not $\tilde{x}$, i.e.,

$$\tilde{A}(x) := \begin{cases} A(x), & \text{if } x \in \mathbb{X}, \\ \{\tilde{a}\}, & \text{if } x = \tilde{x}. \end{cases}$$

The one-step cost function $\tilde{c}$ is defined by

$$\tilde{c}(x, a) := \begin{cases} \mu(x)^{-1} c(x, a), & \text{if } (x, a) \in \mathrm{Gr}(A), \\ 0, & \text{if } (x, a) = (\tilde{x}, \tilde{a}). \end{cases}$$

Finally, select a discount factor

$$\tilde{\beta} \in [(K-1)/K, 1),$$

and define the transition probabilities $\tilde{p}$ as follows. For $(x, a) \in \mathrm{Gr}(A)$, let

$$\tilde{p}(B|x, a) := \frac{1}{\tilde{\beta}\mu(x)} \int_B \mu(y)q(dy|x, a), \ \ B \in \mathcal{B}(\mathbb{X}),$$

$$\tilde{p}(\{\tilde{x}\}|x, a) := 1 - \frac{1}{\tilde{\beta}\mu(x)} \int_{\mathbb{X}} \mu(y)q(dy|x, a),$$

and let

$$\tilde{p}(\{\tilde{x}\}|\tilde{x}, \tilde{a}) := 1.$$

Since only one action is available in state $\tilde{x}$, and the action sets coincide otherwise, there is a one-to-one correspondence between policies for the new MDP and the original MDP.

For $x \in \tilde{\mathbb{X}}$ and $\pi \in \Pi$, let $\tilde{v}^{\pi}(x)$, be the expected total discounted cost for the new model, and let $\tilde{v}(x) := \inf_{\pi\in\Pi} \tilde{v}^{\pi}(x)$. It is well-known (see e.g., [9]) that $\tilde{v}^{\pi}(x) = \mu(x)^{-1} v^{\pi}(x)$ and $\tilde{v}(x) = \mu(x)^{-1} v(x)$ for all $x \in \mathbb{X}$.

Theorem 2. *Suppose Assumptions* T and WC*(i,ii) hold. If the function*

$$(x, a) \mapsto \int_{\mathbb{X}} V(y)q(dy|x, a) \tag{4}$$

is continuous on $\mathrm{Gr}(A)$*, then* $\tilde{p}$ *is a weakly continuous transition probability kernel. In addition, if Assumption* WC*(iii) holds, then there exists a stationary* $\tilde{\beta}$*-optimal policy for the MDP obtained from the HV transformation, and for this MDP a stationary policy* ϕ *is* $\tilde{\beta}$*-optimal if and only if for all* $x \in \mathbb{X}$,

$$\begin{aligned} \tilde{v}(x) &= \tilde{c}(x, \phi(x)) + \tilde{\beta} \int_{\mathbb{X}} \tilde{v}(y)\tilde{p}(dy|x, \phi(x)) \\ &= \min_{a\in A(x)} \left[\tilde{c}(x, a) + \tilde{\beta} \int_{\mathbb{X}} \tilde{v}(y)\tilde{p}(dy|x, a) \right]. \end{aligned} \tag{5}$$

Proof. According to Proposition 1, the function μ used in the HV transformation can be taken to be continuous. Moreover, Assumption T implies that the function V is integrable with respect to $q(\cdot|x, a)$, for all $(x, a) \in \mathrm{Gr}(A)$. Since $\mu \leq KV$, the weak continuity of $\tilde{p}$ then follows from Lemma 11 in the Appendix.

Next, recalling that $\tilde{x}$ is isolated from $\mathbb{X}$, the continuity of μ by Proposition 1 implies that the nonnegative function $\tilde{c}$ is lower semicontinuous on $\mathrm{Gr}(\tilde{A})$. Since the action sets $\tilde{A}(x)$ are compact for all $x \in \tilde{\mathbb{X}}$, it follows from [10, Theorem 2] that the value function $\tilde{v}$ for the discounted MDP defined by the HV transformation satisfies

$$\tilde{v}(x) = \min_{a\in\tilde{A}(x)} \left[\tilde{c}(x, a) + \tilde{\beta} \int_{\tilde{\mathbb{X}}} \tilde{v}(y)\tilde{p}(dy|x, a) \right]$$

for all $x \in \tilde{\mathbb{X}}$, and there exists a stationary optimal policy for this discounted problem. Moreover, since $\tilde{v}(\tilde{x}) = 0$, a stationary policy

ϕ is optimal for the discounted problem if and only if (5) holds for all $x \in \tilde{\mathbb{X}}$. The need to only consider $x \in \mathbb{X}$, for which $A(x) = \tilde{A}(x)$, follows from the fact that there is only one available action at state $\tilde{x}$. □

Corollary 3. *Suppose Assumptions T and WC hold and that the mapping (4) on Gr(A) is continuous. Then*

(i) *the value function v satisfies the optimality equation*

$$v(x) = \min_{a \in A(x)} \left[c(x,a) + \int_{\mathbb{X}} v(y) q(dy|x,a) \right]$$

for all $x \in \mathbb{X}$;

(ii) *there exists a stationary policy that is total-cost optimal;*

(iii) *a stationary policy ϕ is total-cost optimal if and only if*

$$v(x) = c_\phi(x) + Q_\phi v(x) \quad \forall x \in \mathbb{X},$$

which holds if and only if ϕ is $\tilde{\beta}$-optimal for the MDP defined by the HV transformation.

Proof. This follows from Theorem 2, the definition of the HV transformation, and the fact that $v(x) = \mu(x)\tilde{v}(x)$ for all $x \in \mathbb{X}$. □

4. Average costs per unit time

To state Assumption HT, given $\phi \in \mathbb{F}$, a Borel function $u : \mathbb{X} \to \mathbb{R}$, and a state $z \in \mathbb{X}$, let

$${}_zQ_\phi u(x) := \int_{\mathbb{X}\setminus\{z\}} u(y) q(dy|x,a), \qquad x \in \mathbb{X},$$

define ${}_zQ_\phi^0 u(x) \equiv u(x)$ for $x \in \mathbb{X}$, and for $x \in \mathbb{X}$ and $n = 1, 2, \ldots$ let ${}_zQ_\phi^n u(x) := {}_zQ_\phi({}_zQ_\phi^{n-1}u)(x)$. Also, let $\mathbf{e}(x) := 1$ for $x \in \mathbb{X}$.

Assumption HT. There exist a state $\ell \in \mathbb{X}$ and a constant K_ℓ satisfying

$$\sum_{n=0}^{\infty} {}_\ell Q_\phi^n \mathbf{e}(x) \le K_\ell < \infty, \quad \forall \phi \in \mathbb{F},\ x \in \mathbb{X}. \tag{6}$$

Proposition 4. *Suppose Assumption HT holds with a state $\ell \in \mathbb{X}$ that is isolated from $\mathbb{X}$, and Assumptions WC(i,ii) hold. Then there exists a continuous function $\mu_\ell : \mathbb{X} \to [1,\infty)$ satisfying $\mu_\ell(x) \le K_\ell$ for all $x \in \mathbb{X}$ and*

$$\mu_\ell(x) \ge 1 + \int_{\mathbb{X}\setminus\{\ell\}} \mu_\ell(y) q(dy|x,a) \tag{7}$$

for all $(x,a) \in Gr(A)$.

Proof. Consider the transition kernel q_ℓ from $\mathrm{Gr}_\ell(A) := \{(x,a) \in \mathrm{Gr}(A) | x \neq \ell\}$ to $\mathbb{X}_\ell := \mathbb{X} \setminus \{\ell\}$ where

$$q_\ell(B|x,a) := q(B \setminus \{\ell\}|x,a)$$

for $B \in \mathcal{B}(\mathbb{X}_\ell)$ and $(x,a) \in \mathrm{Gr}_\ell(A)$. Then it follows from Proposition 1 and Assumption HT that there exists a continuous function $\mu_\ell : \mathbb{X}_\ell \to [1,\infty)$ that is bounded above by

$$K_\ell^- := \sup_{x \in \mathbb{X}\setminus\{\ell\}} \left\{ \sum_{n=0}^{\infty} {}_\ell Q_\phi^n \mathbf{e}(x) \right\}$$

and satisfies (7) for all $(x,a) \in \mathrm{Gr}_\ell(A)$. Letting

$$\mu_\ell(\ell) := \sup_{a \in A(\ell)} \left[1 + \int_{\mathbb{X}\setminus\{\ell\}} \mu_\ell(y) q(dy|x,a) \right]$$

and recalling that ℓ is isolated from $\mathbb{X}$, it follows that this extension of μ_ℓ to $\mathbb{X}$ is continuous and bounded above by K_ℓ according to Assumption HT, and satisfies (7) for all $(x,a) \in \mathrm{Gr}(A)$. □

Remark 5. The function μ_ℓ that is constructed in the proof of Proposition 4 gives, for each $x \in \mathbb{X}$, the supremum $\mu_\ell(x)$ (over all policies) of the expected number of epochs before the system hits state ℓ after epoch 1. If the state ℓ is not isolated, then this function μ_ℓ may be discontinuous at ℓ despite the weak continuity of q.

To verify this, let $\ell := (\sqrt{5}-1)/2$ and consider the following MDP with only one available action a_0 for each state and a constant one-step cost function. The state space is the closed interval $\mathbb{X} := [0,\ell]$, and the transition probabilities $q(\cdot|x,a_0)$ are defined for $x \in \mathbb{X}$ as follows. Let $q(\{\ell\}|0,a_0) := 1$, $q(\{\ell\}|\ell,a_0) := 1-\ell$, and $q(\{0\}|\ell,a_0) := \ell$. In addition, for $x \in (0,\ell)$ let $q(\{x\}|x,a_0) := x^2$, $q(\{\ell\}|x,a_0) := 1-x-x^2$, and $q(\{0\}|x,a_0) := x$. Observe that Assumption HT holds because $\mu_\ell(0) = 1$, $\mu_\ell(\ell) = (\sqrt{5}+1)/2$, and $\mu_\ell(x) = 1/(1-x) \le (\sqrt{5}+3)/2$ for $x \in (0,\ell)$. Moreover, it is straightforward to verify that this MDP satisfies Assumptions WC(i,ii). On the other hand, since $\lim_{x\to\ell}\mu_\ell(x) = 1/(1-\ell) = (\sqrt{5}+3)/2 > (\sqrt{5}+1)/2 = \mu_\ell(\ell)$, the function μ_ℓ is discontinuous at ℓ.

4.1. HV-AG transformation

Suppose Assumption HT holds. We now describe the *HV-AG transformation* [9], which is based on the work of Akian & Gaubert [1]. As was the case with the HV transformation, the HV-AG transformation results in a discounted MDP, whose set of policies corresponds to the set of policies for the original MDP.

The components of the discounted MDP defined by the HV-AG transformation will be indicated by a horizontal bar. The state space is $\bar{\mathbb{X}} := \mathbb{X} \cup \{\bar{x}\}$, where $\bar{x} \notin \mathbb{X}$ is a cost-free absorbing state that is isolated from $\mathbb{X}$. The action space is $\bar{\mathbb{A}} := \mathbb{A} \cup \{\bar{a}\}$, where $\bar{a}$ is the only action available when the system is in state $\bar{x}$. The set $\bar{A}(x)$ of available actions is unchanged if the current state x is not $\bar{x}$, i.e.,

$$\bar{A}(x) := \begin{cases} A(x), & \text{if } x \in \mathbb{X}, \\ \{\bar{a}\}, & \text{if } x = \bar{x}. \end{cases}$$

The one-step cost function $\bar{c}$ is defined by

$$\bar{c}(x,a) := \begin{cases} \mu_\ell(x)^{-1} c(x,a), & \text{if } (x,a) \in \mathrm{Gr}(A), \\ 0, & \text{if } (x,a) = (\bar{x},\bar{a}). \end{cases}$$

Finally, select a discount factor

$$\bar{\beta} \in [(K_\ell - 1)/K_\ell, 1),$$

and define the transition probabilities $\bar{p}$ as follows. For $(x,a) \in \mathrm{Gr}(A)$ and $B \in \mathcal{B}(\mathbb{X} \setminus \{\ell\})$, let

$$\bar{p}(B|x,a) := \frac{1}{\bar{\beta}\mu_\ell(x)} \int_B \mu_\ell(y) q(dy|x,a),$$

and let

$$\bar{p}(\{\ell\}|x,a) := \frac{\mu_\ell(x) - 1 - \int_{\mathbb{X}\setminus\{\ell\}} \mu_\ell(y) q(dy|x,a)}{\bar{\beta}\mu_\ell(x)},$$

$$\bar{p}(\{\bar{x}\}|x,a) := 1 - \frac{\mu_\ell(x) - 1}{\bar{\beta}\mu_\ell(x)}.$$

Finally, let

$$\bar{p}(\{\bar{x}\}|\bar{x},\bar{a}) := 1.$$

Since only the action $\bar{a}$ is available at the state $\bar{x}$ and the action sets coincide otherwise, there is a one-to-one correspondence between policies for the new MDP and the original MDP.

For $x \in \bar{\mathbb{X}}$ and $\pi \in \Pi$, let $\bar{v}^\pi(x)$, be the expected total discounted cost for the new model, and let $\bar{v}(x) := \inf_{\pi \in \Pi} \bar{v}^\pi(x)$.

Theorem 6. *Suppose Assumption HT holds with a state $\ell \in \mathbb{X}$ that is isolated from $\mathbb{X}$, and Assumptions WC(i,ii) hold. Then $\bar{p}$ is a weakly continuous transition probability kernel. In addition, if Assumption WC(iii) holds, then there exists a stationary $\bar{\beta}$-optimal policy for the MDP defined by the HV-AG transformation, and for this MDP a stationary policy ϕ is $\bar{\beta}$-optimal if and only if for all $x \in \mathbb{X}$,*

$$\begin{aligned}\bar{v}(x) &= \bar{c}(x, \phi(x)) + \bar{\beta} \int_{\mathbb{X}} \bar{v}(y)\bar{p}(dy|x, \phi(x)) \\ &= \min_{a \in A(x)} \left[\bar{c}(x, a) + \bar{\beta} \int_{\mathbb{X}} \bar{v}(y)\bar{p}(dy|x, a) \right].\end{aligned} \tag{8}$$

Proof. Proposition 4 implies that the function μ_ℓ used in the HV-AG transformation can be taken to be continuous. Since $\mu_\ell \le K_\ell < \infty$, the weak continuity of $\bar{p}$ follows from Lemma 11 in the Appendix.

Next, observe that $\bar{c}$ is lower semicontinuous on $\mathrm{Gr}(\bar{A})$, and the action sets $\bar{A}(x)$ are compact for all $x \in \bar{\mathbb{X}}$. According to [10, Theorem 2], it follows that

$$\bar{v}(x) = \min_{a \in \bar{A}(x)} \left[\bar{c}(x, a) + \bar{\beta} \int_{\bar{\mathbb{X}}} \bar{v}(y)\bar{p}(dy|x, a) \right]$$

for all $x \in \bar{\mathbb{X}}$, there exists a stationary optimal policy for the discounted problem, and a stationary policy ϕ is optimal for this problem if and only if (8) holds for all $x \in \mathbb{X}$. □

Corollary 7. *Suppose Assumption HT holds with a state $\ell \in \mathbb{X}$ that is isolated from $\mathbb{X}$ and Assumption WC holds. Then*

(i) *the constant $w := \bar{v}(\ell)$ and the function $h(x) := \mu(x)[\bar{v}(x) - \bar{v}(\ell)]$, $x \in \mathbb{X}$, satisfy*

$$w + h(x) = \min_{a \in A(x)} \left[c(x, a) + \int_{\mathbb{X}} h(y)q(dy|x, a) \right]$$

for all $x \in \mathbb{X}$, and

(ii) *if the one-step cost function c is bounded, and q is a transition probability kernel, then there exists a stationary average-cost optimal policy, and any stationary policy ϕ satisfying*

$$w + h(x) = c_\phi(x) + Q_\phi h(x) \quad \forall x \in \mathbb{X},$$

where w are h are defined in (i), is average-cost optimal;

(iii) *there exists a $\bar{\beta}$-optimal stationary policy for the MDP defined by the HV-AG transformation, and under the hypotheses of (ii) every such policy is average-cost optimal for the original MDP.*

Proof. Statement (i) follows from Theorem 6 and the definition of the HV-AG transformation. Moreover, statement (ii) follows from statement (i) and [14, Proposition 5.5.5]. Finally, statement (iii) follows from Theorem 6, statement (ii), the definition of the HV-AG transformation. □

5. Capacitated inventory control with fixed ordering costs and lost sales

Consider the following single-item *capacitated periodic-review* inventory control problem with *fixed ordering costs* and *lost sales*. At each period $n = 0, 1, \ldots$, the decision-maker observes the current *inventory level* x_n and places an *order* $a_n \ge 0$. After the order is received in the same period, the demand $D_{n+1} \ge 0$ is realized. Any remaining inventory is held to the next period, and all unmet demand is lost. The demands $D_1, D_2, \ldots$ are assumed to be independent and identically distributed with distribution $G_D(\cdot)$, where $G_D(0) < 1$. Moreover, we assume that the system is *capacitated*, where the inventory level can be at most $C < \infty$ and the maximum order size is $M < \infty$.

Whenever a positive amount is ordered, a fixed cost $K \ge 0$ is incurred in addition to a per-unit cost of $\bar{c} > 0$. The cost to hold x units of inventory for one period is $h(x)$, where $h : [0, C] \to [0, \infty)$ is assumed to be continuous.

The inventory control problem described above can be formulated as an MDP as follows. The state space is $\mathbb{X} := [0, C] \cup \{0_L\}$, where 0_L is isolated from $[0, C]$. The special state 0_L, which indicates the occurrence of a lost sale, will be used to apply the results in Section 4. For every $x \in \mathbb{X}$, the set of available actions is $A(x) \equiv \mathbb{A} := [0, M]$.

Letting $0_L + y := y$ for $y \in \mathbb{R}$, the state process can be described by the stochastic equation

$$\begin{aligned} x_{n+1} &= F(x_n, a_n, D_{n+1}) \\ &:= \begin{cases} \min\{x_n + a_n - D_{n+1}, C\}, & x_n + a_n \ge D_{n+1}, \\ 0_L, & x_n + a_n < D_{n+1}. \end{cases}\end{aligned}$$

This equation defines the transition probability kernel q for the corresponding MDP, where

$$q(B|x, a) := \int_B \mathbf{1}\{F(x, a, s) \in B\} \, dG_D(s)$$

for $B \in \mathcal{B}(\mathbb{X})$ and $(x, a) \in \mathbb{X} \times \mathbb{A}$, where $\mathbf{1}\{\cdot\}$ denotes the indicator function. Since 0_L is isolated from $\mathbb{X}$ and F is continuous on $\mathbb{X} \times \mathbb{A} \times [0, \infty)$, it follows that q is weakly continuous; see e.g., [13, p. 92].

Recall that $K \ge 0$ is the *fixed ordering cost*, $\bar{c} \ge 0$ is the *per-unit ordering cost*, and $h : \mathbb{X} \to [0, \infty)$ is the per-period *holding cost function*. Letting $h(0_L) := h(0)$, it follows that the associated one-step cost function $c : \mathbb{X} \times \mathbb{A} \to [0, \infty)$ is given by $c(x, a) := K\mathbf{1}\{a > 0\} + \bar{c}a + \int_0^\infty h[F(x, a, s)] \, dG_D(s)$. Since h is continuous on $[0, C]$, c is bounded on $\mathbb{X} \times \mathbb{A}$. Moreover, for every $\lambda \in \mathbb{R}$, the set $\{(x, a) \in \mathbb{X} \times \mathbb{A} | c(x, a) \le \lambda\}$ is a compact subset of $\mathbb{X} \times \mathbb{A}$; this implies that c is lower semicontinuous on $\mathbb{X} \times \mathbb{A}$. Recalling that the action sets $A(x) \equiv \mathbb{A} = [0, M]$ for all $x \in \mathbb{X}$, it follows that Assumption WC holds.

Assumption D. With positive probability, the per-period demand D is greater than the maximum order size M, that is, $G_D(M) < 1$.

Proposition 8. *Assumption D implies that Assumption HT holds with $\ell = 0_L$.*

Proof. Let $\gamma := 1 - G_D(M) > 0$, and let $\tau_L := \inf\{n \ge 1 | x_n = 0_L\}$ denote the first epoch n when the demand D_n generated a lost sale. Since the amount of on-hand inventory is at most C, and at most M units can be ordered in a single period, it follows that $\mathbb{P}_x^\phi\{x_{\lceil C/M \rceil + 1} = 0_L\} \ge \gamma^{\lceil C/M \rceil + 1} > 0$ for all $\phi \in \mathbb{F}$ and $x \in \mathbb{X}$. Hence

$$\begin{aligned} &\sum_{n=0}^\infty {}_{0_L}Q_\phi^n \mathbf{e}(x) = \mathbb{E}_x^\phi \tau_L = \sum_{n=0}^\infty \mathbb{P}_x^\phi\{\tau_L > n\} \\ &= 1 + \sum_{n=1}^\infty \mathbb{P}_x^\phi\{x_k \ne 0_L, \ k = 1, \ldots, n\} \\ &\le 1 + \sum_{n=1}^\infty (1 - \gamma^{\lceil C/M \rceil + 1})^{\lfloor n/(\lceil C/M \rceil + 1) \rfloor} \\ &\le \frac{\lceil C/M \rceil + 1}{\gamma^{\lceil C/M \rceil + 1}} < \infty \end{aligned}$$

for all $\phi \in \mathbb{F}$ and $x \in \mathbb{X}$. □

Theorem 9. *Suppose Assumption D holds. Then there exists a $\bar{\beta}$-optimal policy for the MDP defined by the HV-AG transformation, and every such policy is average-cost optimal for the original inventory control problem.*

Proof. This follows from statements (ii) and (iii) of Corollary 7. □

Remark 10. Using the HV transformation and Corollary 3, it can be shown that, when Assumption D holds, the problem of minimizing the total cost incurred before the first lost sale can also be reduced to a discounted MDP.

Acknowledgment

Research of the first author was partially supported by NSF grant CMMI-1636193.

Appendix

Let $\mathbb{S}$ be a metric space endowed with its Borel σ-algebra $\mathcal{B}(\mathbb{S})$. A sequence $\{\nu_n\}_{n=0}^{\infty}$ of finite measures on $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$ *converges weakly* to a measure ν if, for every bounded continuous function $f : \mathbb{S} \to \mathbb{R}$,

$$\lim_{n\to\infty} \int_{\mathbb{S}} f(x)\, \nu_n(dx) = \int_{\mathbb{S}} f(x)\, \nu(dx).$$

Lemma 11 (*Dominated Convergence*). *Let $g : \mathbb{S} \to [0, \infty)$ be a continuous function, and let $\{\nu_n\}_{n=0}^{\infty}$ be a sequence of finite measures on $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$ that converges weakly to a measure ν. Suppose there exists a continuous function h on $\mathbb{S}$ such that $g \le h$ and*

$$\lim_{n\to\infty} \int_{\mathbb{S}} h(x)\, \nu_n(dx) = \int_{\mathbb{S}} h(x)\, \nu(dx) < \infty. \tag{9}$$

Then

$$\lim_{n\to\infty} \int_{\mathbb{S}} g(x)\, \nu_n(dx) = \int_{\mathbb{S}} g(x)\, \nu(dx). \tag{10}$$

Proof. According to [11, Theorem 1.1], if $f : \mathbb{S} \to [0, \infty)$ is continuous, then

$$\int_{\mathbb{S}} f(x)\, \nu(dx) \le \liminf_{n\to\infty} \int_{\mathbb{S}} f(x)\, \nu_n(dx). \tag{11}$$

The equality (10) then follows by applying (9) and (11) to the nonnegative continuous functions $h - g$ and $h + g$. □

References

[1] M. Akian, S. Gaubert, Policy iteration for perfect information stochastic mean payoff games with bounded first return times is strongly polynomial, 2013. ArXiv:1310.4953.

[2] C.D. Aliprantis, K. Border, Infinite Dimensional Analysis: A Hitchhiker's Guide, Springer Science & Business Media, 2006.

[3] E. Altman, Constrained Markov Decision Processes, CRC Press, 1999.

[4] D.P. Bertsekas, S.E. Shreve, Stochastic Optimal Control: The Discrete Time Case, Academic Press, New York, 1978.

[5] O.L.V. Costa, F. Dufour, Average control of Markov decision processes with Feller transition probabilities and general action spaces, J. Math. Anal. Appl. 396 (1) (2012) 58–69.

[6] E.B. Dynkin, A.A. Yushkevich, Controlled Markov Processes, Springer New York, 1979.

[7] E.A. Feinberg, Optimality Conditions for Inventory Control, in: Optimization Challenges in Complex, Networked and Risky Systems, INFORMS Tutorials in Operations Research, INFORMS, 2016, pp. 14–45.

[8] E.A. Feinberg, J. Huang, Strong polynomiality of policy iterations for average-cost MDPs modeling replacement and maintenance problems, Oper. Res. Lett. 41 (3) (2013) 249–251.

[9] E.A. Feinberg, J. Huang, On the reduction of total-cost and average-cost MDPs to discounted MDPs, Nav. Res. Logist. (2017). http://dx.doi.org/10.1002/nav.21743.

[10] E.A. Feinberg, P.O. Kasyanov, N.V. Zadoianchuk, Average cost Markov decision processes with weakly continuous transition probabilities, Math. Oper. Res. 37 (4) (2012) 591–607.

[11] E.A. Feinberg, P.O. Kasyanov, N.V. Zadoianchuk, Fatou's lemma for weakly converging probabilities, Theory Probab. Appl. 58 (4) (2014) 683–689.

[12] E.A. Feinberg, Y. Liang, On the optimality equation for average cost Markov decision processes and its validity for inventory control, Ann. Oper. Res. (2017). http://dx.doi.org/10.1007/s10479-017-2561-9.

[13] O. Hernández-Lerma, Adaptive Markov Control Processes, Springer, 1989.

[14] O. Hernández-Lerma, J.B. Lasserre, Discrete-Time Markov Control Processes: Basic Optimality Criteria, Springer New York, 1995.

[15] A. Jaśkiewicz, A.S. Nowak, On the optimality equation for average cost Markov control processes with Feller transition probabilities, J. Math. Anal. Appl. 316 (2) (2006) 495–509.

[16] S.M. Ross, Arbitrary state Markovian decision processes, Ann. Math. Statist. 39 (6) (1968) 2118–2122.

[17] S.M. Ross, Non-discounted denumerable Markovian decision models, Ann. Math. Statist. 39 (2) (1968) 412–423.

[18] A.F. Veinott, Discrete dynamic programming with sensitive discount optimality criteria, Ann. Math. Statist. 40 (5) (1969) 1635–1660.