On the Optimality of (s, S) Policies

Eugene A. Feinberg

Department of Applied Mathematics and Statistics

State University of New York at Stony Brook

Stony Brook, NY 11794-3600

efeinberg@notes.cc.sunysb.edu

631-632-7189

Mark E. Lewis

School of Operations Research and Information Engineering

Cornell University

221 Rhodes Hall, Ithaca, NY 14853-3801

mark.lewis@cornell.edu

607-255-0757

July 21, 2015

Keywords: Markov decision process, average cost per unit time, optimality inequality, optimal policy, inventory control

MSC 2000 Subject Classifications: Primary: 90C40; Secondary: 90B05

OR/MS subject classifications: Primary: Dynamic Programming/optimal control/Markov/Infinite state; Secondary: Inventory/production/Uncertainty/Stochastic

Abstract

This paper describes results on the existence of optimal policies and convergence properties of optimal actions for discounted and average-cost Markov Decision Processes with weakly continuous transition probabilities. It is possible that cost functions are unbounded and action sets are not compact. The results are applied to stochastic periodic-review inventory control problems, for which they imply the existence of stationary optimal policies and certain optimality properties. The optimality of (s, S) policies is proved by using dynamic programming equations for discounted costs and the vanishing discount factor approach for average costs per unit time.

1 Introduction

Since Scarf [22] proved optimality of (s, S) policies for finite-horizon problems with continuous demand, there has been significant efforts to describe when the original insights generalize. Arthur P. Veinott [27, 28] was one of the pioneers in this exploration and he combined a deep understanding of Markov decision processes with a passion for the study of inventory control. It is a great pleasure to dedicate this paper to him.

We provide results on Markov Decision Processes (MDPs) with infinite state spaces, weakly continuous transition probabilities, one-step costs, that can be unbounded, and possibly noncompact action sets with discounted and average-cost criteria. These results are applied to the prove optimality of (s, S) policies for stochastic periodic-review inventory control problems.

Since 1950s, inventory control has been one of the major motivations for studying MDPs. However, there has been a gap between the available results in the MDP theory and the results needed to analyze inventory control problems. Even now most work on inventory control assumes that the demand is either discrete or continuous. Moreover, the proofs are often problem-specific and do not use general results on MDPs.

With such a long history, the inventory control literature is far too expansive to attempt a complete literature review. The reader is pointed to the survey article by Porteus [20]. There are also treatments of both Markov decision processes and their relationship to inventory control in Heyman and Sobel [16] and Bertsekas [1]. In the case of inventory control, under the average cost criterion the optimality of (s, S) policies was proved by Iglehart [18] and Veinott and Wagner [29] in the continuous and discrete demand cases, respectively. The latter proof was simplified significantly by Zheng [30], where the author proved in the discrete case the existence of a solution to the average cost optimality equations by construction. Beyer and Sethi [3] reconsidered the continuous demand model of Iglehart [18], compared the results in [18] with the results by Veinott and Wagner [29] on problems with discrete demand, and provided rigorous proofs for the results in [18]. Almost all studies on inventory control deal with either discrete or continuous demand. Without such assumptions, the optimality of (s, S) policies for average cost inventory control problems follows from Chen and Simchi-Levi [6] where coordinated price-inventory control is studied and methods specific to inventory control are used. Huh et al. [17] developed additional problem-specific methods for inventory control problems with arbitrary distributed demands.

Early studies of MDPs dealt with finite-state problems and infinite-state problems with bounded costs. The case of average costs per unit time is more difficult than the case of total discounted costs. Sennott [25] developed the theory for the average-cost criterion for countable-state problems with unbounded costs. Schäl [23, 24] developed the theory for uncountable state problems with the discounted and average-cost criteria with compact action sets. In particular, Schäl [23, 24] identifies two groups of assumptions on transition probabilities: weak continuity and setwise continuity. As explained in Feinberg and Lewis [13, Section 4], models with weakly continuous transition probabilities are more natural for inventory control than models with setwise continuous transition probabilities. Hernández-Lerma and Lasserre [15] developed the theory for problems with setwise continuous transition probabilities, unbounded costs, and possibly noncompact action sets. Luque-Vasques and Hernández-Lerma [19] identified an additional technical difficulty in dealing with problems with weakly continuous transition probabilities even for finite-horizon problems by demonstrating that Berge's theorem, that ensures semi-continuity of the value function, does not hold for problems with noncompact action sets. Feinberg and Lewis [13] investigated the total discounted costs for inf-compact cost functions and obtained sufficient conditions for average costs. Compared to Schäl [24] these results required an additional local boundedness assumption (LB) that holds for inventory control problems, but its verification is not easy. Feinberg et al. [10, 11] introduced a natural class of $\mathbb{K}$-inf-compact costs functions, extended Berge's theorem to noncompact action sets, and developed the theory of MDPs with weakly continuous transition probabilities, unbounded costs and with both criteria; average and discounted costs. In particular, the results from [10] do not require the validity of the local boundedness assumption (LB), and they are applicable to inventory control problems. Such applications are considered in Section 6 below.

Section 2 of this paper describes an MDP with an infinite state space, weakly continuous transition probabilities, possibly unbounded one-step costs, and possibly noncompact action sets. Sections 3 and 4 provide the results for discounted and total-reward criteria. In particular, in addition to reviewing some results from [11], new results on convergence of optimal actions are provided for the following two situations: (i) convergence of finite-horizon discounted problems to infinite-horizon ones (Theorem 3.4), and (ii) convergence of discounted MDPs to average-cost ones (Theorem 4.3). Section 5 relates MDPs to the problems, whose dynamics is defined by stochastic equations, as this takes place for inventory control. Section 6 describes the inventory control problem with backorders, setup costs, linear ordering costs, and convex holding costs and provides two results on the existence of discounted and average-cost optimal (s, S) policies. The first result, Theorem 6.11, states the existence of optimal (s, S) policies for large discount factors and average costs. It does not use any additional assumptions, and the proof is based on adding additional terminal costs to finite-horizon problems. The second result, Theorem 6.13, states the existence of optimal (s, S) policies for all discount factors under an additional assumption often used in the literature including in Bertsekas [1], Chen and Simchi-Levi [5, 6], and Huh et al. [17].

2 Definition of MDPs with Borel State and Action Sets

Consider a discrete-time Markov decision process with the state space $\mathbb{X}$ and action space $\mathbb{A}$, one-step costs c, and transition probabilities q. The state space $\mathbb{X}$ and action space $\mathbb{A}$ both assumed to be Borel subsets of Polish (complete separable metric) spaces. If an action $a \in \mathbb{A}$ is selected at a state $x \in \mathbb{X}$, then a cost $c(x, a)$ is incurred, where $c : \mathbb{X} \times \mathbb{A} \to \overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$, and the system moves to the next state according to the probability distribution $q(\cdot|x, a)$ on $\mathbb{X}$. The function c is assumed to be bounded below and Borel measurable, and q is a transition probability, that is, $q(B|x, a)$ is a Borel function on $\mathbb{X} \times \mathbb{A}$ for each Borel subset B of $\mathbb{X}$, and $q(\cdot|x, a)$ is a probability measure on the Borel σ-field of $\mathbb{X}$ or each $(x, a) \in \mathbb{X} \times \mathbb{A}$.

The decision process proceeds as follows: at time n the current state of the system, x, is observed. A decision-maker decides which action, a, to choose, the cost $c(x, a)$ is accrued, the system moves to the next state according to $q(\cdot \mid x, a)$, and the process continues. Let $H_n = (\mathbb{X} \times \mathbb{A})^n \times \mathbb{X}$ be the set of histories for $n = 0, 1, \ldots$. A (randomized) decision rule at epoch $n = 0, 1, \ldots$ is a regular transition probability π_n from H_n to $\mathbb{A}$ concentrated on $A(x_n)$. In other words, (i) $\pi_n(\cdot|h_n)$ is a probability distribution on $\mathbb{A}$, where $h_n = (x_0, a_0, x_1, \ldots, a_{n-1}, x_n)$ and (ii) for any measurable subset $B \subseteq \mathbb{A}$, the function $\pi_n(B|\cdot)$ is measurable on H_n. A policy π is a sequence $(\pi_0, \pi_1, \ldots)$ of decision rules. Moreover, π is called non-randomized if each probability measure $\pi_n(\cdot|h_n)$ is concentrated at one point. A non-randomized policy is called Markov if all decisions depend only on the current state and time. A Markov policy is called stationary if all decisions depend only on the current state. Thus, a Markov policy ϕ is defined by a sequence $\phi_0, \phi_1, \ldots$ of measurable mappings $\phi_n : \mathbb{X} \to \mathbb{A}$ such that $\phi_n(x) \in A(x)$ for all $x \in \mathbb{X}$. A stationary policy ϕ is defined by a measurable mapping $\phi : \mathbb{X} \to \mathbb{A}$ such that $\phi(x) \in A(x)$ for all $x \in \mathbb{X}$.

The Ionescu–Tulcea theorem (see [2, p. 140-141] or [15, p. 178]) yields that an initial state x and a policy π define a unique probability distribution $\mathbb{P}_x^\pi$ on the set of all trajectories $H_\infty = (\mathbb{X} \times \mathbb{A})^\infty$ endowed with the product σ-field defined by Borel σ-fields of $\mathbb{X}$ and $\mathbb{A}$. Let $\mathbb{E}_x^\pi$ be the expectation with respect to this distribution. For a finite horizon $N = 0, 1, \dots$ and a bounded below measurable function $T : \mathbb{X} \to \overline{\mathbb{R}}$ called the terminal value, define the expected total discounted costs

$$v_{N,C,\alpha}^\pi(x) := \mathbb{E}_x^\pi \left[\sum_{n=0}^{N-1} \alpha^n c(x_n, a_n) + T(x_N) \right], \tag{2.1}$$

where $\alpha \in [0, 1)$, $v_{0,T,\alpha}^\pi(x) = T(x)$, $x \in \mathbb{X}$. When $T(x) = 0$ for all $x \in \mathbb{X}$, we shall write $v_{N,\alpha}^\pi(x)$ instead of $v_{N,T,\alpha}^\pi(x)$. When $N = \infty$ and $T(x) = 0$ for all $x \in \mathbb{X}$, (2.1) defines the infinite horizon expected total discounted cost of π denoted by $v_\alpha^\pi(x)$ instead of $v_{\infty,\alpha}^\pi(x)$. The average costs per unit time are defined as

$$w^\pi(x) := \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_x^\pi \sum_{n=0}^{N-1} c(x_n, a_n). \tag{2.2}$$

For each function $V^\pi(x) = v_{N,T,\alpha}^\pi(x)$, $v_{N,\alpha}^\pi(x)$, $v_\alpha^\pi(x)$, or $w(x)$, define the optimal cost

$$V(x) := \inf_{\pi \in \Pi} V^\pi(x), \tag{2.3}$$

where Π is the set of all policies. A policy π is called *optimal* for the respective criterion if $V^\pi(x) = V(x)$ for all $x \in \mathbb{X}$.

We remark that the definition of an MDP usually includes the sets of available actions $A(x) \subseteq \mathbb{A}$, $x \in \mathbb{X}$. We do not do this explicitly because we allow $c(x, a)$ to be equal $+\infty$. In other words, a feasible pair (x, a) is modeled as a pair with the finite costs. To transform this model to a model with feasible action sets, it is sufficient to consider the set of available actions $A(x)$ such that $A(x) \supseteq A_c(x)$, where $A_c(x) = \{a \in \mathbb{A} : c(x, a) < +\infty\}$, $x \in \mathbb{X}$. In order to transform an MDP with action sets $A(x)$ to a MDP with action sets $\mathbb{A}$, $x \in \mathbb{X}$, it is sufficient to set $c(x, a) = +\infty$ when $a \in \mathbb{A} \setminus A(x)$. Of course, certain measurability conditions should hold, but this is not an issue when the function c is measurable. We remark that early works on MDPs by Blackwell [4] and Strauch [26] considered models with $A(x) = \mathbb{A}$ for all $x \in \mathbb{X}$. This approach caused some problems with the generality of the results because the boundedness of the cost function c was assumed and therefore $c(x, a) \in \mathbb{R}$ for all (x, a). If the cost function is allowed to take infinitely large values, models with $A(x) = \mathbb{A}$ are as general as models with $A(x) \subseteq \mathbb{A}$, $x \in \mathbb{X}$.

3 Optimality Results for Discounted Cost MDPs with Borel State and Action Sets

It is well-known (see e.g. [2, Proposition 8.2]) that $v_{n,\alpha}(x)$ satisfies the following *optimality equations*,

$$v_{n+1,T,\alpha}(x) = \inf_{a \in A(x)} \{c(x, a) + \alpha \int v_{n,T,\alpha}(y) q(dy|x, a)\}, \qquad x \in \mathbb{X},\ n = 0, 1, \dots . \tag{3.1}$$

In addition, a Markov policy ϕ_α, defined at the first N steps by the mappings $\phi_0, \ldots, \phi_{N-1}$ that satisfies the following equations for all $x \in \mathbb{X}$ and all $n = 1, \ldots, N$

$$v_{n,T,\alpha}(x) = c(x, \phi_{N-n}(x)) + \alpha \int v_{n-1,T,\alpha}(y) q(dy|x, \phi_{N-n,\alpha}(x)), \qquad x \in \mathbb{X}, \tag{3.2}$$

is optimal for the horizon N; see e.g. [2, Lemma 8.7].

It is also well-known (see e.g. [2, Proposition 9.8]) that $v_\alpha(x)$ satisfies the following *discounted cost optimality equations* (DCOE),

$$v_\alpha(x) = \inf_{a \in A(x)} \{c(x, a) + \alpha \int v_\alpha(y) q(dy|x, a)\}, \qquad x \in \mathbb{X}. \tag{3.3}$$

If a stationary policy ϕ_α satisfies

$$v_\alpha(x) = c(x, \phi_\alpha(x)) + \alpha \int v_\alpha(y) q(dy|x, \phi_\alpha(x)), \qquad x \in \mathbb{X}, \tag{3.4}$$

then ϕ_α is optimal; [2, Proposition 9.8].

However, additional conditions on cost functions and transition probabilities are needed to ensure the existence of optimal policies. Earlier conditions required compactness of action sets. They were introduced and studied by Schäl [23] and consisted of two sets of conditions that required either weak or setwise continuity assumptions. For setwise continuous transition probabilities, Hernandez-Lerma and Lasserre [15] extended these conditions to MDPs with general action sets and cost functions $c(x, a)$ that are inf-compact in the action variable a. Feinberg and Lewis obtained results for weakly continuous transition probabilities and inf-compact cost functions. Feinberg et al. [10] generalized and unified the results by Schäl [23] and Feinberg and Lewis [13] for weakly continuous transition probabilities to more general cost functions by using the notion of a $\mathbb{K}$-inf-compact function. $\mathbb{K}$-inf-compacted functions were originally introduced in [10, Assumption **W***] without using the term $\mathbb{K}$-inf-compact, and formally introduced and studied in [11, 9]. As explained in Feinberg and Lewis [13, Section 4], weak continuity holds for periodic review inventory control problems. The setwise continuity assumption may not hold, but it holds for problems with continuous or discrete demand distributions. Since this paper concerns with inventory control applications, it focuses on essentially more general case of weak continuous transition probabilities.

For an $\overline{\mathbb{R}}$-valued function f, defined on a metric space $\mathbb{U}$, consider the level sets

$$\mathcal{D}_f(\lambda; \mathbb{U}) = \{y \in \mathbb{U} : f(y) \leq \lambda\}, \qquad \lambda \in \mathbb{R}. \tag{3.5}$$

We recall that a function f is *lower semi-continuous on* $\mathbb{U}$ if all the level sets $\mathcal{D}_f(\lambda; \mathbb{U})$ are closed, and a function f is *inf-compact* on U if all these sets are compact. For a set U, let 2^U denote the set of its subsets. For three sets U, V, and W, where $U \subset V$, and two functions $g : U \to W$ and $f : V \to W$, the function g is called the restriction of f to U if $g(x) = f(x)$ when $x \in U$.

Definition 3.1 (cp. Feinberg et al. [11, 9], Feinberg and Kasyanov [8]) *A function* $f : \mathbb{X} \times \mathbb{A} \to \overline{\mathbb{R}}$ *is called* $\mathbb{K}$*-inf-compact if, for any nonempty compact subset* $K \subseteq \mathbb{X}$, *the* restriction *of this function to* $K \times \mathbb{A}$ *is inf-compact .*

In particular, according to Feinberg et al. [11, Lemma 2.1], the following two facts hold:

(i) an inf-compact function is $\mathbb{K}$-inf-compact;

(ii) if $A : \mathbb{X} \to 2^{\mathbb{A}} \setminus \{\emptyset\}$ is a compact-valued semi-continuous set-valued mapping and $c : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$ is a lower semi-continuous function such that $c(x,a) = +\infty$ for $x \in \mathbb{X}$ and for $a \in \mathbb{A} \setminus A(x)$, then the function c is $\mathbb{K}$-inf-compact.

For each pair $(x,a) \in \mathrm{Gr}_{\mathbb{X}}(A) := \{(x,a) : x \in \mathbb{X}, a \in A(x)\}$, the probability measure $q(\cdot|x,a)$ is defined on the Borel σ-field on $\mathbb{X}$. The transition probability q is called weakly continuous if for every bounded continuous function $f : \mathbb{X} \to \mathbb{R}$ and for each sequence $\{(x^{(n)}, a^{(n)})\}_{n=1,2\ldots}$ on $\mathrm{Gr}_{\mathbb{X}}(A)$ converging to $(x^{(0)}, a^{(0)}) \in \mathrm{Gr}_{\mathbb{X}}(A)$,

$$\int_{\mathbb{X}} f(x) q(dx|x^{(n)}, a^{(n)}) \to \int_{\mathbb{X}} f(x) q(dx|x^{(0)}, a^{(0)}) \qquad \text{as } n \to \infty. \tag{3.6}$$

Assumption W*. The following conditions hold:

(i) the cost function c is bounded below and $\mathbb{K}$-inf-compact;

(ii) If $(x^{(0)}, a^{(0)})$ is a limit of a converging sequence $\{(x^{(n)}, a^{(n)})\}_{n=1,2\ldots}$ of elements of $\mathbb{X} \times \mathbb{A}$ such that $c(x^{(n)}, a^{(n)}) < +\infty$ for all $n = 0, 1, 2, \ldots,$ then the sequence $\{q(\cdot|(x^{(n)}, a^{(n)}))\}_{n=1,2\ldots}$ converges weakly to $q(\cdot|(x^{(0)}, a^{(0)}))$. That is, (3) holds for every bounded continuous function f on $\mathbb{X}$.

For example, Assumption **W***(ii) holds if the transition probability $q(\cdot|x,a)$ is weakly continuous on $\mathbb{X} \times \mathbb{A}$. The following theorem describes the structure of optimal policies, continuity properties of value functions, and convergence of value iterations.

Theorem 3.2 (Feinberg et al. [10, Theorem 2]) *Suppose Assumption* **W*** *holds and let* $T(x) = 0$. *Then*

(i) the functions $v_{n,\alpha}$, $n = 0, 1, 2, \ldots,$ *and* v_α *are lower semi-continuous on* $\mathbb{X}$, *and* $v_{n,\alpha}(x) \to v_\alpha(x)$ *as* $n \to +\infty$ *for all* $x \in \mathbb{X}$;

(ii) the functions $v_{n,\alpha}$ *satisfy,*

$$v_{n+1,\alpha}(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_{\mathbb{X}} v_{n,\alpha}(y) q(dy|x,a) \right\}, \quad x \in \mathbb{X},\ n = 0, 1, \ldots \tag{3.7}$$

The nonempty sets $A_{n,\alpha}(x) := \{a \in \mathbb{A} : v_{n+1,\alpha}(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_{n,\alpha}(y) q(dy|x,a)\}$, $x \in \mathbb{X}$, $n = 0, 1, \ldots,$ *satisfy the following properties:*

(a) the graph $\mathrm{Gr}_{\mathbb{X}}(A_{n,\alpha}) = \{(x,a) : x \in \mathbb{X}, a \in A_{n,\alpha}(x)\}$, $n = 0, 1, \ldots,$ *is a Borel subset of* $\mathbb{X} \times \mathbb{A}$, *and*

(b) if $v_{n+1,\alpha}(x) = +\infty$, *then* $A_{n,\alpha}(x) = \mathbb{A}$ *and, if* $v_{n+1,\alpha}(x) < +\infty$, *then* $A_{n,\alpha}(x)$ *is compact;*

(iii) for each $N = 1, 2, \ldots,$ *there exists a Markov optimal* N*-horizon policy* $(\phi_0, \ldots, \phi_{N-1})$.

(iv) *if for an* N*-horizon Markov policy* $(\phi_0, \dots, \phi_{N-1})$ *the inclusions* $\phi_{N-1-n}(x) \in A_{n,\alpha}(x)$*,* $x \in \mathbb{X}$*,* $n = 0, \dots, N-1$*, hold then this policy is* N*-horizon optimal;*

(v) *for* $\alpha \in [0, 1)$

$$v_\alpha(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) q(dy|x,a) \right\}, \qquad x \in \mathbb{X}, \tag{3.8}$$

and the nonempty sets $A_\alpha(x) := \{a \in \mathbb{A} : \ v_\alpha(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) q(dy|x,a)\}$*,* $x \in \mathbb{X}$*, satisfy the following properties:*

(a) *the graph* $\mathrm{Gr}_{\mathbb{X}}(A_\alpha) = \{(x,a) : \ x \in \mathbb{X}, a \in A_\alpha(x)\}$ *is a Borel subset of* $\mathbb{X} \times \mathbb{A}$*, and*

(b) *if* $v_\alpha(x) = +\infty$*, then* $A_\alpha(x) = \mathbb{A}$ *and, if* $v_\alpha(x) < +\infty$*, then* $A_\alpha(x)$ *is compact.*

(vi) *for an infinite-horizon there exists a stationary discount-optimal policy* ϕ_α*, and a stationary policy is optimal if and only if* $\phi_\alpha(x) \in A_\alpha(x)$ *for all* $x \in \mathbb{X}$*.*

(vii) (Feinberg and Lewis [13, Proposition 3.1(iv)]) *if the cost function* c *is inf-compact, then the functions* $v_{n,\alpha}$*,* $n = 1, 2, \dots$*, and* v_α *are inf-compact on* $\mathbb{X}$*.*

The following corollary is useful for the analysis of inventory control problems.

Corollary 3.3 *Let Assumption* **W*** *hold. Consider a bounded below, lower semi-continuous function* $T : \mathbb{X} \to \overline{\mathbb{R}}$*. Then*

(i) *the bounded below functions* $v_{n,T,\alpha}$*,* $n = 0, 1, 2, \dots,$ *are lower semi-continuous;*

(ii)

$$v_{n+1,T,\alpha}(x) = \min_{a \in A(x)} \left\{ c(x,a) + \alpha \int_{\mathbb{X}} v_{n,T,\alpha}(y) q(dy|x,a) \right\}, \quad x \in \mathbb{X}, \ n = 0, 1, ..., \tag{3.9}$$

where $v_{0,T,\alpha}(x) = T(x)$ *for all* $x \in \mathbb{X}$*, and the nonempty sets*

$$A_{n,T,\alpha}(x) := \{a \in \mathbb{A} : \ v_{n+1,T,\alpha}(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_{n,T,\alpha}(y) q(dy|x,a)\}, x \in \mathbb{X}, n = 0, 1, \dots,$$

satisfy the following properties:

(a) *the graph* $\mathrm{Gr}_{\mathbb{X}}(A_{n,T,\alpha}) = \{(x,a) : \ x \in \mathbb{X}, a \in A_{n,T,\alpha}(x)\}$*,* $n = 0, 1, \dots,$ *is a Borel subset of* $\mathbb{X} \times \mathbb{A}$*, and*

(b) *if* $v_{n+1,T,\alpha}(x) = +\infty$*, then* $A_{n,T,\alpha}(x) = \mathbb{A}$ *and, if* $v_{n+1,T,\alpha}(x) < +\infty$*, then* $A_{n,T,\alpha}(x)$ *is compact;*

(iii) *for a problem with the terminal value function* T*, for each* $N = 1, 2, \dots,$ *there exists a Markov optimal* N*-horizon policy* $(\phi_0, \dots, \phi_{N-1})$ *and if, for an* N*-horizon Markov policy* $(\phi_0, \dots, \phi_{N-1})$ *the inclusions* $\phi_{N-1-n}(x) \in A_{n,T,\alpha}(x)$*,* $x \in \mathbb{X}$*,* $n = 0, \dots, N-1$*, hold then this policy is* N*-horizon optimal;*

(iv) *if* $T(x) \le v_\alpha(x)$ *for all* $x \in \mathbb{X}$*, then* $v_{n,T,\alpha}(x) \to v_\alpha(x)$ *as* $n \to +\infty$ *for all* $x \in \mathbb{X}$*;*

(v) *if the cost function* c *is inf-compact, the functions* $v_{n,T,\alpha}$*,* $n = 1, 2, \dots,$ *are inf-compact.*

Proof. Statements (i)-(iii) are corollaries from statements (i)-(iii) of Theorem 3.2. Indeed, the statements of Theorem 3.2, that deal with the finite horizon N, hold when one-step costs at different time epochs vary. In particular, if the one-step cost at epoch $n = 0, 1, \ldots, N-1$ is defined by a bounded below, measurable cost function c_n rather than by the function c, this case can be reduced to the single function c by replacing the state space $\mathbb{X}$ with the state space $\mathbb{X} \times \{0, 1, \ldots, N-1\}$, setting $c((x,n),a) = c_n(x,a)$, and applying the corresponding statements of Theorem 3.2. In our case, $c_n(x,a) = c(x,a)$ for $n = 0, 1, \ldots, N-1$, and $c_{N-1}(x,a) = c(x,a) + \int_{\mathbb{X}} T(y)p(dy|x,a)$. The function c_{N-1} is bounded below and lower semi-continuous.

To prove (iv) and (v), consider first the case when the functions c and T are nonnegative. In this case,

$$v_{n,\alpha}(x) \le v_{n,T,\alpha}(x) \le v_{n,v_\alpha,\alpha}(x) = v_\alpha(x), \qquad x \in \mathbb{X},\ n = 0, 1, \ldots \tag{3.10}$$

Therefore, for nonnegative cost functions, Statement (iv) follows from Theorem 3.2(i). Statement (v) follows from (iv), Theorem 3.2(vii), and the fact that $v_{n,T,\alpha} \ge v_{n,\alpha}$ since T is nonnegative. In a general case, consider a finite positive constant K such that the functions c and T are bounded below by $(-K)$. If the cost functions c and T are increased by K then the new cost functions are nonnegative, each finite-horizon value function $v_{n,T,\alpha}$ is increased by the constant $d_n = K(1-\alpha^{n+1})/(1-\alpha)$, and the infinite-horizon value function v_α is increased by the constant $d = K/(1-\alpha)$. Since $d_n \le d$ and $d_n \to d$ as $n \to \infty$, the general case follows from the case of non-negative cost functions. ∎

While Theorem 3.2 and Corollary 3.3 state convergence of value functions and describe the structure of optimal sets of actions, the following theorem describes convergence properties of optimal actions.

Theorem 3.4 *Let Assumption* **W*** *hold. Let $T : \mathbb{X} \to \overline{\mathbb{R}}$ be bounded below, lower semi-continuous, and such that for all $x \in \mathbb{X}$*

$$T(x) \le v_\alpha(x) \qquad \text{and} \qquad v_{1,T,\alpha}(x) \ge T(x). \tag{3.11}$$

Then for $x \in \mathbb{X}$, such that $v_\alpha(x) < \infty$, each sequence $\{a^{(n)} \in A_{n,T,\alpha}(x)\}_{n=1,2,\ldots}$, where the sets $A_{n,T,\alpha}(x)$ and $A_\alpha(x)$ are defined in Corollary 3.3 and Theorem 3.2 respectively, $n = 1, 2, \ldots,$ has a limit point, and, if a^* is a limit point of such sequence, then $a^* \in A_\alpha(x)$.

In order to prove Theorem 3.4, we need the following lemma, which is a simplified version of [15, Lemma 4.6.6].

Lemma 3.5 *Let A be a compact subset of $\mathbb{A}$ and $f,\ f_n : A \to \overline{\mathbb{R}}$, $n = 1, 2, \ldots,$ be nonnegative, lower semi-continuous, real-valued functions such that $f_n(a) \uparrow f(a)$ as $n \to \infty$ for all $a \in A$. Let $f_n(a^{(n)}) = \min_{a \in A} f_n(a)$ for $a^{(n)} \in A$, $n = 1, 2, \ldots,$ and a^* be a limit point of the sequence $\{a^{(n)}\}_{n=1,2,\ldots}$. Then $f(a^*) = \min_{a \in A} f(a)$.*

Proof. Let $a' \in A$ and $f(a') = \min_{a \in A} f(a)$. Then $f(a') \ge f_n(a^{(n)}) \ge f_k(a^{(n)})$ for all $n \ge k$. Lower semi-continuity of f and the previous inequalities imply $f_k(a^*) \le \liminf_{n \to \infty} f_n(a^{(n)}) \le f(a')$. Thus $f(a') \ge f_k(a^*) \uparrow f(a^*)$. ∎

Proof of Theorem 3.4. We assume without loss of generality that the bounded below functions c and T are nonnegative. We can do this because of the arguments provided at the end of the proof of Corollary 3.3

and the additional argument that, if the one-step cost functions c and terminal cost functions are shifted by constants then the set of optimal finite-horizon action $A_{n,T,\alpha}(\cdot)$ and infinite-horizon actions $A_\alpha(\cdot)$ remain unchanged.

Fix $x \in \mathbb{X}$ and define the compact set $A := \{a \in \mathbb{A} : c(x,a) \le v_\alpha(x)\}$. Since the function $v_{n,T,\alpha}$ takes nonnegative values and, in view of (3.10), $v_{n+1,T,\alpha}(x) \le v_\alpha(x)$,

$$A_{n,T,\alpha}(x) = \{a \in \mathbb{A} : c(x,a) + \alpha \int_{\mathbb{X}} v_{n,T,\alpha}(y) q(dy|x,a) = v_{n+1,T,\alpha}(x)\} \subseteq A, \qquad n = 1, 2, \dots .$$

Since A is compact, a sequence $\{a^{(n)} \in A_{n,T,\alpha}(x)\}_{n=1,2,\dots}$ has a limit point. The theorem follows from Lemma 3.5 applied to the functions

$$\begin{aligned} f(a) &= c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) q(dy|x,a), && a \in A, \\ f_n(a) &= c(x,a) + \alpha \int_{\mathbb{X}} v_{n,T,\alpha}(y) q(dy|x,a), && a \in A,\ n = 0, 1, \dots . \end{aligned}$$

For all $z \in \mathbb{X}$,

$$v_\alpha(z) = v_{n,v_\alpha,\alpha}(z) \ge v_{n,T,\alpha}(z) \ge v_{n,\alpha}(z) \uparrow v_\alpha(z),$$

where the first equality follows from the optimality equation, the first and the second inequalities follow from $v_\alpha(\cdot) \ge T(\cdot) \ge 0$, and the convergence is stated in Theorem 3.2(i); this convergence is monotone because c and T are nonnegative functions. The inequality $v_{1,T,\alpha}(\cdot) \ge T(\cdot)$ in (3.11), equality (3.9), and standard induction arguments imply $v_{n+1,T,\alpha}(\cdot) \ge v_{n,T,\alpha}(\cdot)$, $n = 0, 1, \dots$. Thus Assumption (3.11) implies that $v_{n,T,\alpha} \uparrow v_\alpha$, and the monotone convergence theorem implies $f_n \uparrow f$ as $n \to \infty$. ∎

4 Average-Cost MDPs with Borel State and Action Sets

The average cost case is more subtle than the case of expected total discounted costs. The following assumption was introduced by Schäl [24]. Without this assumption the problem is trivial because $w(x) = \infty$ for all $x \in \mathbb{X}$, and therefore any policy is optimal.

Assumption G. $w^* := \inf_{x \in \mathbb{X}} w(x) < +\infty$.

Assumption **G** is equivalent to the existence of $x \in \mathbb{X}$ and $\pi \in \Pi$ with $w^\pi(x) < \infty$. Define the following quantities for $\alpha \in [0,1)$:

$$m_\alpha = \inf_{x \in \mathbb{X}} v_\alpha(x), \quad u_\alpha(x) = v_\alpha(x) - m_\alpha,$$

$$\underline{w} = \liminf_{\alpha \uparrow 1} (1-\alpha) m_\alpha, \quad \overline{w} = \limsup_{\alpha \uparrow 1} (1-\alpha) m_\alpha.$$

Observe that $u_\alpha(x) \geq 0$ for all $x \in \mathbb{X}$. According to Schäl [24, Lemma 1.2], Assumption **G** implies

$$0 \leq \underline{w} \leq \overline{w} \leq w^* < +\infty. \tag{4.1}$$

According to Schäl [24, Proposition 1.3], if there exists a measurable function $u : \mathbb{X} \to [0, \infty)$ and a stationary policy ϕ such that

$$\underline{w} + u(x) \geq c(x, \phi(x)) + \int u(y) q(dy|x, \phi(x)), \qquad x \in \mathbb{X}, \tag{4.2}$$

then ϕ is average cost optimal and $w(x) = w^*$ for all $x \in \mathbb{X}$. The following condition plays an important role for the validity of (4.2).

Assumption B. Assumption **G** holds and $\sup_{\alpha<1} u_\alpha(x) < \infty$ for all $x \in \mathbb{X}$.

We note that the second part of Assumption **B** is Condition **B** in Schäl [24]. Thus, under Assumption **G**, which is assumed throughout [24], Assumptions **B** are equivalent to Condition **B** in [24].

For $x \in \mathbb{X}$ and for a nonnegative lower semi-continuous function $u : \mathbb{X} \to [0, \infty)$, define the set

$$A_u^*(x) := \left\{ a \in A(x) \, : \, \underline{w} + u(x) \geq c(x, a) + \int_{\mathbb{X}} u(y) q(dy|x, a) \right\}, \; x \in \mathbb{X}. \tag{4.3}$$

If there exists a stationary policy ϕ satisfying (4.2), then for $x \in \mathbb{X}, \phi(x) \in A_u^*(x)$ and $A_u^*(x) \neq \emptyset$.

Theorem 4.1 (Feinberg et al. [10, Theorem 4]). *Suppose Assumptions* **W*** *and* **B** *hold. There exists a nonnegative lower semi-continuous function u and a stationary policy ϕ satisfying (4.2), that is, $\phi(x) \in A_u^*(x)$ for all $x \in \mathbb{X}$. Furthermore, every stationary policy ϕ, for which (4.2) holds, is optimal for the average costs per unit time criterion,*

$$w^\phi(x) = w(x) = w^* = \underline{w} = \overline{w} = \lim_{\alpha \uparrow 1} (1 - \alpha) v_\alpha(x) = \lim_{N \to \infty} \frac{1}{N} v_{N,1}^\phi(x), \qquad x \in \mathbb{X}. \tag{4.4}$$

Moreover, the following statements hold:

(a) the nonempty sets $A_u^*(x), x \in \mathbb{X}$, satisfy the following properties:

(a1) the graph $\mathrm{Gr}_{\mathbb{X}}(A_u^*) = \{(x, a) : x \in \mathbb{X}, a \in A_u^*(x)\}$ is a Borel subset of $\mathbb{X} \times \mathbb{A}$;

(a2) for each $x \in \mathbb{X}$ the set $A_u^*(x)$ is compact;

(b) there exists a stationary policy ϕ with $\phi(x) \in A_u^*(x)$ for all $x \in \mathbb{X}$.

As shown in Feinberg et al. [10, Formula (21)], the function u can be defined as

$$u(x) := \liminf_{\alpha \uparrow 1, \; y \to x} u_\alpha(y), \quad x \in \mathbb{X}. \tag{4.5}$$

Alternatively, for each sequence $\alpha_n \to 1-$, it can be defined as

$$\tilde{u}(x) := \liminf_{n \to \infty, \; y \to x} u_{\alpha_n}(y), \quad x \in \mathbb{X}. \tag{4.6}$$

It follows from these definitions that $u(x) \le \tilde{u}(x)$, $x \in \mathbb{X}$. However, the question whether $u = \tilde{u}$ has not been investigated. If the cost function c is inf-compact, then the functions v_α, u, and $\tilde{u}$ are inf-compact as well; see Theorem 3.2 for the proof of this fact for v_α and Feinberg et al. [10, Theorem 4(e) and Corollary 2] for u and $\tilde{u}$. In addition, if the one-step cost function c is inf-compact, the minima of functions v_α possess additional properties.

Set

$$X_\alpha := \{x \in \mathbb{X} : v_\alpha(x) = m_\alpha\}, \qquad \alpha \in [0,1). \tag{4.7}$$

Since $X_\alpha = \{x \in \mathbb{X} : v_\alpha(x) \le m_\alpha\}$, this set is closed. If the function c is inf-compact then inf-compactness of v_α implies that the sets X_α are nonempty and compact. The following fact is useful for verifying the validity of Assumption **B**; see Feinberg and Lewis [13, Lemma 5.1] and the references therein.

Theorem 4.2 *(Feinberg et al. [10]). Let Assumptions* **G** *and* **W*** *hold. If the function c is inf-compact, then there exists a compact set $\mathcal{K} \subseteq \mathbb{X}$ such that $X_\alpha \subseteq \mathcal{K}$ for all $\alpha \in [0,1)$.*

According to Feinberg et al. [10, Theorem 5], certain average cost optimal policies can be approximated by discount optimal policies with vanishing discount factor. The following theorem describes sufficient conditions when approximations take place. Recall that, for the function $u(x)$ defined in (4.5), for each $x \in \mathbb{X}$ there exist sequences $\{\alpha_n \uparrow 1\}$ and $\{x^{(n)} \to x\}$, where $x^{(n)} \in \mathbb{X}$, $n = 1, 2, \ldots$, such that $u(x) = \lim_{n\to\infty} u_{\alpha_n}(x^{(n)})$. Similarly, for a sequence $\{\alpha_n \uparrow 1\}$ consider the function $\tilde{u}$ defined in (4.6). Then for each $x \in \mathbb{X}$ there exist a sequence $\{x^{(n)} \to x\}$ of points in $\mathbb{X}$ and a subsequence $\{\alpha_n^*\}_{n=1,2,\ldots}$ of the sequence $\{\alpha_n\}_{n=1,2,\ldots}$ such that $\tilde{u}(x) = \lim_{n\to\infty} u_{\alpha_n^*}(x^{(n)})$.

Theorem 4.3 *Let Assumptions* **W*** *and* **B** *hold. For $x \in \mathbb{X}$ and $a^* \in \mathbb{A}$, the following two statements hold:*

(i) if there exists a sequence $(\alpha_n^*, x^{(n)})$ with $\alpha_n^* \uparrow 1$ and $x^{(n)} \to x$, $n = 1, 2, \ldots$, such that $u_{\alpha_n}(x^{(n)}) \to u(x)$ as $n \to \infty$, and there are sequences of natural numbers $\{n_k\}_{k=1,2,\ldots}$ and actions $\{a^{(n_k)}\}_{k=1,2,\ldots}$, from $\mathbb{A}$ such that $a^{(n_k)} \in A_{\alpha_{n_k}}(x^{(n_k)})$ and $a^{(n_k)} \to a^*$ as $k \to \infty$, then $a^* \in A_u^*(x)$ with the function u defined in (4.5);

(ii) if for a sequence $\{\alpha_n \uparrow 1\}$ there are a subsequence $\{\alpha_{(n_k)}\}$ and a sequence $\{x^{(n)} \to x\}$ of points from $\mathbb{X}$, such that $u_{\alpha_{n_k}}(x^{(n_k)}) \to \tilde{u}(x)$ as $k \to \infty$, where the function $\tilde{u}$ is defined in (4.6) for the sequence $\{\alpha_n\}_{n=1,2,\ldots}$ of discount factors, and, if there exist actions $a^{(n_k)} \in \mathbb{A}$, $k = 1, 2, \ldots$, such that $a^{(n_k)} \in A_{\alpha_{n_k}}(x^{(n_k)})$ and $a^{(n_k)} \to a^*$ as $k \to \infty$, then $a^* \in A_{\tilde{u}}^*(x)$.

Proof. To show (i), consider sequences whose existence is assumed in the theorem. Then

$$v_{\alpha_{n_k}}(x^{(n_k)}) = c(x^{(n_k)}, a^{(n_k)}) + \alpha \int_{\mathbb{X}} v_{\alpha_{n_k}}(y) q(dy|x^{(n_k)}, a^{(n_k)}),$$

which implies

$$u_{\alpha_{n_k}}(x^{(n_k)}) + (1 - \alpha_{n_k}) m_{\alpha_{n_k}} = c(x^{(n_k)}, a^{(n_k)}) + \alpha \int_{\mathbb{X}} u_{\alpha_{n_k}}(y) q(dy|x^{(n_k)}, a^{(n_k)}).$$

Fatou's lemma for weakly converging measures (see e.g., Feinberg et al. [12, Theorem 1.1]), the choice of the sequence $x^{(n_k)}$, and Theorem 4.1 imply that

$$\underline{w} + u(x) \geq c(x, a^*) + \int_{\mathbb{X}} u(y) q(dy|x, a^*).$$

Thus $a^* \in A_u^*(x)$. The proof of Statement (ii) is similar. ■

Corollary 4.4 *Let Assumptions* **W*** *and* **B** *hold. For $x \in \mathbb{X}$ and $a^* \in \mathbb{A}$, the following two statements hold:*

(i) if each sequence $\{(\alpha_n^*, x^{(n)}\}_{n=1,2,\dots})$ with $\alpha_n^* \uparrow 1$ and $x^{(n)} \to x$, $n = 1, 2, \dots,$ contains a subsequence $(\alpha_{n_k}, x^{(n_k)})$ such that there exist actions $a^{(n_k)} \in A_{\alpha_{n_k}}(x^{(n_k)})$ such that $a^{(n_k)} \to a^*$ as $k \to \infty$, then $a \in A_u^*(x)$ with the function u defined in (4.5);

(ii) if there is a sequence $\{\alpha_n \uparrow 1\}$ such that conditions of statement (i) hold for each subsequence $\{\alpha_n^*\}_{n=1,2,\dots}$ of this sequence and for each sequence $\{x_n \to x\}$ of states in $\mathbb{X}$, then $a^* \in A_{\tilde{u}}^*(x)$ with the function $\tilde{u}$ defined in (4.6).

Proof. Statement (i) follows from Theorem 4.3(i) applied to a sequence $\{(\alpha_n^*, x^{(n)})\}_{n=1,2,\dots}$ with the property $u(x) = \lim_{n\to\infty} u_{\alpha_n^*}(x^{(n)})$. Statement (ii) follows from Theorem 4.3(ii) applied to a sequence $\{(\alpha_n^*, x^{(n)})\}_{n=1,2,\dots}$ with the property $\tilde{u}(x) = \lim_{n\to\infty} u_{\alpha_n^*}(x^{(n)})$ and the sequence $\{\alpha_n^*\}_{n=1,2,\dots}$ being a subsequence of the sequence $\{\alpha_n\}_{n=1,2,\dots}$. ■

5 MDPs Defined by Equations

Let $\mathbb{S}$ be a metric space, $\mathcal{B}(\mathbb{S})$ be its Borel σ-field, and μ be a probability measure on $(\mathbb{S}, \mathcal{B}(\mathbb{S}))$. Consider a stochastic sequence x_n, whose dynamics are defined by the stochastic equation

$$x_{n+1} = F(x_n, a_n, \xi_n) \qquad n = 0, 1, \dots, \tag{5.1}$$

where $\xi_0, \xi_1, \dots$ are independent and identically distributed random variables in $\mathbb{S}$ whose distributions are defined by μ and $F : \mathbb{X} \times \mathbb{A} \times \mathbb{S} \to \mathbb{X}$ is a continuous mapping. This equation defines the transition probability

$$q(B|x, a) = \int_{\mathbb{S}} \mathbf{I}\{F(x, a, s) \in B\} \mu(ds), \qquad B \in \mathcal{B}(\mathbb{S}), \tag{5.2}$$

from $\mathbb{X} \times \mathbb{A} \to \mathbb{X}$.

Lemma 5.1 *The transition probability q is weakly continuous in $(x, a) \in \mathbb{X} \times \mathbb{A}$.*

Proof. For a closed subset B of $\mathbb{X}$ and for two sequences $x_n \to x$ and $a_n \to a$ defined on $\mathbb{X}$ and $\mathbb{A}$ respectively

$$\begin{aligned}\limsup_{n\to\infty} q(B|x_n, a_n) &= \limsup_{n\to\infty} \int_{\mathbb{S}} \mathbf{I}\{F(x_n, a_n, s) \in B\} \mu(ds) \\ &\leq \int_{\mathbb{S}} \limsup_{n\to\infty} \mathbf{I}\{F(x_n, a_n, s) \in B\} \mu(ds) \leq q(B|x, a),\end{aligned}$$

where the first inequality follows from Fatou's lemma and the second follows from (5.2) and upper semi-continuity of the function $\mathbf{I}\{F(x_n, a_n, s) \in B\}$ for a closed set B. ■

Corollary 5.2 *Consider an MDP $\{\mathbb{X}, \mathbb{A}, q, c\}$ with the transition function q defined in (5.2) and nonnegative $\mathbb{K}$-inf compact cost function c. This MDP satisfies Assumption* **W*** *and therefore the conclusions of Theorem 3.2 hold.*

Proof. Assumption **W***3 holds in view of Lemma 5.1. ■

For inventory control problems, MDPs are usually defined by particular forms of (5.1). In addition, the cost function c has the form

$$c(x, a) = C(a) + H(x, a), \tag{5.3}$$

where $C(a)$ is the ordering cost and $H(x, a)$ is either holding cost or expected holding cost at the following step. For simplicity we assume that the functions take nonnegative values. These functions are typically inf-compact. If C is lower semi-continuous and H is inf-compact, then c is inf-compact because C is lower semi-continuous as a function of two variables $x \in \mathbb{X}$ and $a \in \mathbb{A}$, and a sum of a nonnegative lower semi-continuous function and an inf-compact function is an inf-compact function. However, as stated in the following theorem, for discounted problems the validity of Assumption **W*** and therefore the validity of the optimality equations, existence of optimal policies, and convergence of value iteration, takes place even under weaker assumptions.

Theorem 5.3 *Consider an MDP $\{\mathbb{X}, \mathbb{A}, q, c\}$ with the transition function q defined in (5.2) and cost function c defined in (5.3). If either of the following two assumptions holds:*

1. *the function $C : \mathbb{A} \to [0, \infty]$ is lower semi-continuous and the function $H : \mathbb{X} \times \mathbb{A} \to [0, \infty]$ is $\mathbb{K}$-inf-compact,*
2. *the function $C : \mathbb{A} \to [0, \infty]$ is inf-compact and the function $H : \mathbb{X} \times \mathbb{A} \to [0, \infty]$ is lower semi-continuous*

then Assumption **W*** *holds and therefore the conclusions of Theorem 3.2(i)-(vi) hold. Furthermore, if the function H is inf-compact, then function c is inf-compact and therefore the discounted value function v_α is inf-compact for all $\alpha \in [0, 1)$.*

Proof. Lemma 5.1 implies the weak continuity of the transition function q. The definition of a $\mathbb{K}$-inf-compact function implies directly that the function $C^*(x, a) := C(a)$ is $\mathbb{K}$-inf-compact on $\mathbb{X} \times \mathbb{A}$. Thus in the either case c is a $\mathbb{K}$-inf-compact function because it is a sum a nonnegative low semi-continuous function and a $\mathbb{K}$-inf-compact function. In addition, if the function H is inf-compact then, as explained in the paragraph preceding the formulation of the theorem, the one-step cost function c is inf-compact. The rest of the proof follows from Theorem 3.2. ■

Remark 5.4 *It is also possible to derive sufficient conditions for the validity of Assumptions* **G** *and* **B** *and therefore for the existence of stationary optimal policies. However, this is more subtle than for Assumption* **W***, *and in this paper we verify Assumptions* **G** *and* **B** *directly for the periodic review inventory control problems.*

6 Optimality of (s, S) Policies for Inventory Control Problems

In this section we consider a discrete-time periodic-review inventory control problem with back orders and prove the existence of an optimal (s, S) policy. For this problem the dynamics are defined by the following stochastic equation

$$x_{n+1} = x_n + a_n - D_{n+1}, \quad n = 0, 1, 2, \ldots, \tag{6.1}$$

where x_n is the inventory at the end of period n, a_n is the decision how much should be ordered, and D_n is the demand during period n. The demand is assumed to be i.i.d. In other words, if we change the notation ξ_n to D_{n+1}, the dynamics are defined by equation (5.1) with $F(x, a, D) = x + a - D$. Of course, this function is continuous.

The model has the following decision-making scenario: a decision-maker views the current inventory of a single commodity and makes an ordering decision. Assuming zero lead times, the products are immediately available to meet demand. Demand is then realized, the decision-maker views the remaining inventory, and the process continues. Assume the unmet demand is backlogged and the cost of inventory held or backlogged (negative inventory) is modeled as a convex function. The demand and the order quantity are assumed to be non-negative. The dynamics of the system are defined by (6.1). Let

- $\alpha \in (0, 1)$ be the discount factor,
- $K \geq 0$ be a fixed ordering cost,
- $c > 0$ be the per unit ordering cost,
- D be a nonnegative random variable with the same distribution as D_n, and $P(D > 0) > 0$,
- $h(\cdot)$ denote the holding/backordering cost per period. It is assumed that $h : \mathbb{R} \to [0, \infty)$ is a convex function, $h(x) \to \infty$ as $|x| \to \infty$, and $\mathbb{E}\, h(x - D) < \infty$ for all $x \in \mathbb{R}$.

Without loss of generality, assume that $h(0) = 0$. The fact that $P(D > 0) > 0$ avoids the trivial case. For example, if $D = 0$ almost surely then the policy that never orders when the inventory level is non-negative and orders up to zero when the inventory level is negative, is optimal under the average cost criterion. Note that $\mathbb{E}\, D < \infty$ since, in view of Jensen's inequality, $h(x - \mathbb{E}\, D) \leq \mathbb{E}\, h(x - D) < \infty$.
The cost function for the model is

$$c(x, a) = C(a) + H(x)$$

with $C(a) := K 1_{\{a \neq 0\}} + ca$ and $H(x, a) := \mathbb{E}\, h(x + a - D)$. The function $C(a)$ is inf-compact on $[0, \infty)$ and takes finite values. In fact, it is continuous at $a > 0$ and lower semi-continuous at $a = 0$. The function is inf-compact because the function $\mathbb{E}\, h(z-D)$, $z\in\mathbb{R}$, is convex, takes finite values, and is not a constant. Thus, the function $c(x,a)$ takes finite values and is inf-compact.

The problem is posed with $\mathbb{X}=\mathbb{R}$ and $\mathbb{A}=[0,\infty)$. However, if the demand and action sets are integer or lattice, the model can be restated with $\mathbb{X}=\mathbb{Z}$, where $\mathbb{Z}$ is the set of integer numbers, and $\mathbb{A}=\{0,1,\dots\}$; see Remark 6.14.

Corollary 6.1 *For the inventory control model, Assumption* ***W**** *holds and the one-step cost function c is inf-compact. Therefore, the conclusions of Theorems 3.2, 3.4 and Corollary 3.2 hold.*

Proof. The validity of Assumption **W*** and inf-compactness of c follow from Theorem 5.3. ∎

Consider the renewal process

$$N(t) := \sup\{n|\boldsymbol{S}_n \le t\}, \tag{6.2}$$

where $\boldsymbol{S}_0=0$ and $\boldsymbol{S}_n=\sum_{j=1}^n D_j$ for $n>0$. Observe that $\mathbb{E}\,N(t)<\infty$ for each $0\le t<\infty$; Resnick [21, Theorem 3.3.1]. Thus, Wald's identity yields that for any $0\le y<\infty$

$$\mathbb{E}\,\mathbf{S}_{N(x)+1} = \mathbb{E}(N(y)+1)\,\mathbb{E}\,D<\infty. \tag{6.3}$$

We next state a useful lemma.

Lemma 6.2 *For fixed initial state x*

$$E_y(x) := \mathbb{E}\,h(x-\boldsymbol{S}_{N(y)+1})<\infty, \tag{6.4}$$

where $0\le y<\infty$.

Proof. Define

$$h^*(x) := \begin{cases} h(x) & \text{for } x\le 0,\\ 0 & \text{for } x>0.\end{cases}$$

Observe that it suffices to show that

$$E^*_y(x) := \mathbb{E}\,h^*(x-\boldsymbol{S}_{N(y)+1})<\infty. \tag{6.5}$$

Indeed, for $Z=x-\boldsymbol{S}_{N(y)+1}$,

$$E_y(x)=\mathbb{E}\,1\{Z\le 0\}h^*(Z)+\mathbb{E}\,1\{Z>0\}h(Z)\le E^*_y(x)+h(x).$$

To show that $E^*_y(x)$ we shall prove the inequality

$$\mathbb{E}\,h^*(x-\boldsymbol{S}_{N(y)+1})\le(1+\mathbb{E}\,N(y))\,\mathbb{E}\,h^*(x-y-D_1). \tag{6.6}$$

If (6.6) holds, the assumptions on h imply (6.5). Define the function $f(z) = h^*(x - y - z)$. This function is nondecreasing and convex. Since f is convex, its derivative exists almost everywhere. Denote the excess of $N(y)$ by $R(y) := \boldsymbol{S}_{N(y)+1} - y$. According to [14, p. 59]

$$P\{R(y) > t\} = 1 - F(y + t) + \int_0^y (1 - F(y + t - s))dU(s),$$

where $U(s) = \mathbb{E}\, N(s)$ is the renewal function. Thus,

$$\mathbb{E}\, h^*(x - \boldsymbol{S}_{N(y)+1}) = \mathbb{E}\, h^*(x - y - R(y)) = \mathbb{E}\, f(R(y)) = \int_0^\infty f'(t)P\{R(y) > t\}dt = J_1 + J_2, \quad (6.7)$$

where $J_1 = \int_0^\infty f'(t)(1 - F(y + t))dt$, $J_2 = \int_0^\infty f'(t) \left(\int_0^y (1 - F(y + t - s))dU(s)\right) dt$, and the third equality in (6.7) holds according to [7, p. 263]. Note that since F is non-decreasing,

$$J_1 \leq \int_0^\infty f'(t)(1 - F(t))dt = \mathbb{E}\, f(D_1) = \mathbb{E}\, h^*(x - y - D_1), \quad (6.8)$$

where the first equality follows from [7, p. 263]. Similarly, by applying Fubini's theorem

$$\begin{aligned} J_2 &= \int_0^y \left(\int_0^\infty f'(t)(1 - F(y + t - s))dt\right) dU(s) \\ &\leq \int_0^y \left(\int_0^\infty f'(t)(1 - F(t))dt\right) dU(s) = \mathbb{E}\, f(D_1)\, \mathbb{E}\, U(y) = \mathbb{E}\, h^*(x - y - D_1)\, \mathbb{E}\, N(y). \quad (6.9) \end{aligned}$$

Combining (6.7)-(6.9) yields (6.6) and the lemma is proven. ∎

The following result applies the results on average-cost MDPs to inventory control.

Proposition 6.3 *The inventory control model satisfies Assumption* ***B****. Therefore, the conclusions of Theorems 4.1–4.3 hold.*

Proof. Consider the policy ϕ that orders up to the level 0 if the inventory level is non-positive and does nothing otherwise. Then $w^\phi(x) \leq K + c\,\mathbb{E}\, D + \mathbb{E}\, h(-D) < \infty$ for $x \leq 0$. That is, Assumption **G** holds.

In view of Corollary 6.1, Theorem 4.2 implies that $\{x_\alpha : \alpha \in [0, 1]\} \subseteq \mathcal{K} := [x_L, x_U]$. Fix the initial state x. Since increasing x_U only expands $\mathcal{K}$, without loss of generality assume that $x_U > x$. For any $\alpha \in [0, 1)$ consider two cases: $x \leq x_\alpha$ and $x > x_\alpha$. For $x \leq x_\alpha$, suppose ϕ is a stationary policy that immediately orders up to level x_α plus orders whatever amount a stationary optimal policy for the discount factor α would order in x_α. From then on it proceeds to follow the optimal policy. We have the following sequence of inequalities

$$v_\alpha(x) - m_\alpha \leq v_\alpha^\phi(x) - m_\alpha \leq K + c(x_U - x). \quad (6.10)$$

Suppose now that $x > x_\alpha$ and that ϕ does not order until the total demand is greater than $x - x_L$. In this case, the difference in costs between a process (Process 1) starting in x that follows ϕ and one that starts in x_α (and follows the optimal policy) can be broken into 3 parts; the holding costs accrued before the inventory of Process 1 moves below x_L, the holding cost accrued in the step that takes the inventory position below x_L and the ordering costs accrued to move the position to x_α.

Since h is convex, $\max\{h(x_L), h(x_U)\} \geq h(y)$ for all $y \in [x_L, x_U]$ so that the expected total discounted holding costs accrued before the inventory position falls below x_L is bounded by $\mathbb{E}\, N(x_U - x_L) \max\{h(x_L), h(x_U)\}$. The inventory position immediately prior to the order being placed is then $x - \boldsymbol{S}_{N(x-x_L)+1}$. Since $x_U > x$ and h is convex, the expected total discounted holding cost is bounded by

$$E(x) = \max\{E_{x_U - x_L}(x), h(x_U)\},$$

where $E_{x_U - x_L}(x)$ is defined in (6.4) and the finiteness of $E(x)$ follows from Lemma 6.4. The expected discounted order cost is bounded by

$$K + c(x_\alpha - [x - \mathbb{E}(N(x - x_L) + 1)\, \mathbb{E}\, D_1]) \leq K + c(x_U + \mathbb{E}(N(x_U - x_L) + 1)\, \mathbb{E}\, D_1).$$

Combining these upper bounds yields

$$\begin{aligned} v_\alpha(x) - m_\alpha \leq \mathbb{E}\, N(x_U - x_L) \max\{h(x_L), h(x_U)\} + E(x) \\ + K + c(x_U + \mathbb{E}(N(x_U - x_L) + 1)\, \mathbb{E}\, D_1) < \infty. \end{aligned} \tag{6.11}$$

■

Consider a nonnegative, real-valued, lower semi-continuous terminal value T. For $n = 0, 1, \ldots, x \in \mathbb{R}$, and $a \in \mathbb{R}^+$, define

$$J_{n,T,\alpha}(x, a) := K 1_{\{a \neq 0\}} + ca + \mathbb{E}[h(x + a - D) + \alpha v_{n,T,\alpha}(x + a - D)], \tag{6.12}$$

$$J_\alpha(x, a) := K 1_{\{a \neq 0\}} + ca + \mathbb{E}[h(x + a - D) + \alpha v_\alpha(x + a - D)]. \tag{6.13}$$

These functions are inf-compact because the summand $\mathbb{E}\, h(x + a - D)$ is an inf-compact function, and the remaining summands are nonnegative lower semi-continuous functions. In view of Theorem 3.2, equations (3.7) and (3.8) can be written as

$$v_{n+1,T,\alpha}(x) = \min_{0 \leq a < \infty} \{J_{n,T,\alpha}(x, a)\}, \tag{6.14}$$

$$v_\alpha(x) = \min_{0 \leq a < \infty} \{J_\alpha(x, a)\}. \tag{6.15}$$

Similarly, Theorem 4.1 and inequality (4.2) imply that for a function u, whose existence is stated in Theorem 4.1,

$$w + u(x) \geq \min_{0 \leq a < \infty} \left\{K 1_{\{a \neq 0\}} + ca + \mathbb{E}\, h(x + a - D) + \mathbb{E}\, u(x + a - D)\right\}. \tag{6.16}$$

The sets of equations (6.15 – 6.16) can be rewritten as

$$v_{n+1,\alpha}(x) = \min\{\min_{a \geq 0}[K + G_{n,T,\alpha}(x + a)], G_{n,T,\alpha}(x)\} - cx, \tag{6.17}$$

$$v_\alpha(x) = \min\{\min_{a \geq 0}[K + G_\alpha(x + a)], G_\alpha(x)\} - cx, \tag{6.18}$$

$$w + u(x) \geq \min\{\min_{a \geq 0}[K + H(x + a)], H(x)\} - cx, \tag{6.19}$$

where

$$G_{n,T,\alpha}(x) := cx + \mathbb{E}\, h(x-D) + \alpha\, \mathbb{E}\, v_{n,T,\alpha}(x-D), \tag{6.20}$$

$$G_{\alpha}(x) := cx + \mathbb{E}\, h(x-D) + \alpha\, \mathbb{E}\, v_{\alpha}(x-D), \tag{6.21}$$

$$H(x) := cx + \mathbb{E}\, h(x-D) + \mathbb{E}\, u(x-D). \tag{6.22}$$

We explain the correctness of (6.17). The explanations for (6.18) and (6.19) are similar. Optimality equation (6.12) is equivalent to $v_{n+1,T,\alpha}(x) = \min\{\inf_{a>0}[K + G_{n,T,\alpha}(x+a)], G_{n,T,\alpha}(x)\} - cx$, and the internal infimum can be replaced with the minimum in (6.17) because of the following two arguments:

(i) the function $K + G_{n,T,\alpha}(y)$ is lower semi-continuous on $[x,\infty)$ and $G_{n,T,\alpha}(y) \to \infty$ as $y \to \infty$, and

(ii) $K + G_{n,T,\alpha}(x) \geq G_{n,T,\alpha}(x)$ since $K \geq 0$.

Corollary 6.4 *The functions $G_{n,T,\alpha}$, G_α and H are lower semi-continuous, $n = 0,1,\dots,$ $\alpha \in [0,1)$.*

Proof. In view of (6.20)–(6.22), each of these functions is a sum of several functions, two of which are continuous and the third is lower semi-continuous, as follows from Corollary 6.1 and from Proposition 6.3. ∎

Lemma 6.5 *Let $n = 0,1,\dots,$ $\alpha \in [0,1)$, and $0 \leq T(x) \leq v_\alpha(x)$ for all $x \in \mathbb{X}$. Then $G_\alpha(x) < \infty$ and $G_{\alpha,T,n}(x) < \infty$ for all $x \in \mathbb{X}$.*

Proof. Since $G_{\alpha,T,n} \leq G_\alpha$, in view of (6.21), the lemma follows from $Ev_\alpha(x-D) < \infty$. To prove this inequality, consider the policy ϕ that orders up to the level 0 if the inventory level is non-positive and orders nothing otherwise. For $x \leq 0$

$$v_\alpha(x) \leq v_\alpha^\phi(x) \leq K - cx + \frac{\alpha(K + c\,\mathbb{E}\, D + \mathbb{E}\, h(-D))}{1-\alpha}. \tag{6.23}$$

Letting $B_\alpha := \frac{\alpha(K+c\,\mathbb{E}\, D+\mathbb{E}\, h(-D))}{1-\alpha}$, we have $\mathbb{E}\, v_\alpha(x-D) \leq K - c\,\mathbb{E}(x-D) + B_\alpha < \infty$. For $x > 0$,

$$\begin{aligned} v_\alpha(x) \leq v_\alpha^\phi(x) &= \mathbb{E}\Big[\sum_{n=1}^{N(x)+1} \alpha^n h(x - \mathbf{S}_n) + \alpha^{N(x)+1} v_\alpha^\phi(x - \mathbf{S}_{N(x)+1})\Big] \\ &\leq h(x)\,\mathbb{E}\, N(x) + \mathbb{E}\, h(x - \mathbf{S}_{N(x)+1}) + K - c(x - \mathbb{E}\,\mathbf{S}_{N(x)+1}) + B_\alpha < \infty, \end{aligned}$$

where the second inequality follows from the facts that $\alpha^n < 1$ for $n \geq 1$, $0 \leq h(x - \mathbf{S}_n) \leq h(x)$ for $n = 1,\dots,N(x)$, and (6.23). The second inequality holds because $\mathbb{E}\, N(x) < \infty$, Lemma 6.2, and (6.3). Since $v_\alpha^\phi(x) = \mathbb{E}\, h(x-D) + \alpha\,\mathbb{E}\, v_\alpha^\phi(x-D) < \infty$, then $\mathbb{E}\, v_\alpha(x-D) \leq \mathbb{E}\, v_\alpha^\phi(x-D) < \infty$. The result follows. ∎

Recall the following classic definition.

Definition 6.6 *A function $f : \mathbb{R} \to \mathbb{R}$ is called K-convex, $K \geq 0$, if for each $x \leq y$ and for each $\lambda \in (0,1)$,*

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y) + \lambda K.$$

The next result is a version of Bertsekas [1, Lemma 4.2.1(c-d)] for lower semi-continuous K-convex functions.

Proposition 6.7 *The following results hold:*

1. *If $g(y)$ is a measurable K-convex function and D is a random variable, then $\mathbb{E}\, g(y-D)$ is also K-convex provided $\mathbb{E}\,|g(y-D)| < \infty$ for all y.*

2. *Suppose g is a lower semi-continuous K-convex function. such that $g(x) \to \infty$ as $|x| \to \infty$. Let*

$$S \in \mathrm{argmin}_{x\in\mathbb{R}}\{g(x)\}, \tag{6.24}$$
$$s = \inf\{x \le S \mid g(x) \le K + g(S)\}. \tag{6.25}$$

   Then

   (a) $g(S) \le g(x)$ for all $x \in \mathbb{R}$,

   (b) $g(S) + K < g(x)$ for all $x < s$,

   (c) $g(x)$ is decreasing on $(-\infty, s)$ and $g(s) < g(x)$ for all $x < s$,

   (d) $g(x) \le g(S) + K$ for all x such that $s \le x \le S$,

   (e) $g(x) \le g(z) + K$ for all $S < x \le z$.

Proof. We prove only statements 3(d). The others follow in the same way as Lemma 4.2.1 in Bertsekas [1]. If $x = s$ or S, then 3(d) is trivial. Suppose $s < x < S$. By the definition of lower semi-continuity there exists $\delta > 0$ such that $g(x+\delta) > g(x) - \epsilon\left(\frac{\delta}{S-x}\right)$ for arbitrary $\epsilon > 0$. However, K-convexity implies

$$K + g(S) \ge g(x) + \frac{S-x}{\delta}[g(x+\delta) - g(x)] > g(x) - \epsilon.$$

Since ϵ is arbitrary, the result follows. ∎

Consider the discounted cost problem and suppose G_α is K-convex, lower semi-continuous and approaches infinity as $|x| \to \infty$. If we define S_α and s_α by (6.24) and (6.25) with g replaced by G_α, Proposition 6.7 parts 3(b) and (c), along with the DCOE imply that it is optimal to order up to S_α when $x < s_\alpha$. Parts 3(d) and (e) imply that it is optimal not to order when $s_\alpha \le x$.

Lower semi-continuity of G_α (recall (6.20)) follows from the convexity of h, the lower semi-continuity of v_α, and the weak continuity of the transition probabilities. In order to show that G_α is K-convex, note that v_α is K-convex since it is a limit of K-convex functions $v_{n,\alpha}$; see Bertsekas [1, Section 4.2]. The next result along with the first result of Proposition 6.7 completes the proof that G_α is K-convex.

The following statement is needed to establish K-convexity of the functions $G_{n,T,\alpha}$, G_α, and H for certain T and α.

Proposition 6.8 *There exists $\alpha^* \in [0,1)$ such that $G_\alpha(x) \to \infty$ as $|x| \to \infty$ for all $\alpha \in [\alpha^*, 1)$ and for all setup costs $K \ge 0$.*

Proof. For a fixed ordering cost $K \geq 0$ we write v_α^K and G_α^K instead of v_α and G_α respectively. Obviously, $G^K(x) \geq G^0(x)$ and $G^K(x) \to \infty$ as $x \to \infty$ for all $K \geq 0$. To complete the proof, we need to show that there exists $\alpha^* \in [0,1)$ such that for all $\alpha \in [\alpha^*, 1)$

$$G_\alpha^0(x) \to \infty \qquad \text{as} \qquad x \to -\infty. \tag{6.26}$$

It is well-known that the function G_α^0 is convex. This holds since, if we consider $T(x) = 0$ for all $x \in \mathbb{R}$ then the function $v_{0,\alpha}^0 = 0$ is convex. Equations (6.12), (6.14), Heyman and Sobel [16], and induction imply that the functions $v_{n,\alpha}^0$, $n = 1, 2, \ldots$, are convex. Convergence of value iterations, stated in Theorem 3.2(i), implies the convexity of the functions v_α^0. The convexity of G_α^0 follows from (6.21).

We show by contradiction that there exists $\alpha^* \in [0,1)$ such that G_α^0 is decreasing on an interval $(-\infty, M_\alpha]$ for some $M_\alpha > -\infty$ when $\alpha \in [\alpha^*, 1)$. Suppose this is not the case. For $K = 0$, (6.18) can be written as

$$v_\alpha^0(x) = \inf_{a \geq 0} \{G_\alpha^0(x+a)\} - cx. \tag{6.27}$$

If a constant M_α does not exist for some $\alpha \in (0,1)$ then the convexity and nonnegativity of $G_\alpha^0(x)$ imply that the policy ψ that never orders is optimal for the discount factor α. If there is no α^* with the described property, Corollary 4.4 implies that the policy ψ is average-cost optimal. This is impossible because $w^\psi(x) \leq h(x) \to \infty$ as $x \to \infty$, but, in view of Theorem 4.1, $w(x)$ is a finite constant. Since G_α^0 is a convex function, (6.26) holds. ∎

Lemma 6.9 *Consider the constant $\alpha^* \in [0,1)$ whose existence is stated in Proposition 6.8. Then the functions $G_{n,v_\alpha^0,\alpha}$ and G_α are K-convex and tend to $+\infty$ as $x \to -\infty$ for all $n = 0, 1, \ldots$ and $\alpha \in [\alpha^*, 1]$.*

Proof. Let $\alpha \in [\alpha^*, 1]$. According to Proposition 6.8, $G_\alpha^0(x) \to \infty$ and $G_\alpha(x) \to \infty$ as $\alpha \to \infty$. In addition, these functions are lower semi-continuous. As explained in the proof of Proposition 6.8, the function $G_\alpha^0(x) \to \infty$ is convex and therefore it is K-convex. Formulae (6.17), (6.20) of Heyman and Sobel [16, Lemma7-2, p. 312] and induction arguments imply that the functions $G_{n,v_\alpha^0,\alpha}$ and $v_{n+1,v_\alpha^0,\alpha}$, $k = 1, 2, \ldots$ are K-convex. In addition, $v_{n,v_\alpha^0,\alpha}(x) \uparrow v_\alpha(x)$ and $G_{n,v_\alpha^0,\alpha}(x) \uparrow G_\alpha(x)$ as $n \to \infty$ in view of Corollary 3.3 and since all the costs are nonnegative. Thus, all these functions are K-convex and tend to $+\infty$ as $x \to -\infty$. ∎

Definition 6.10 *Let s_n and S_n be real numbers such that $s_n \leq S_n$, $n = 0, 1, \ldots$. Suppose x_n denotes the current inventory level at decision epoch n. A policy is called an (s_n, S_n) policy at step n if it orders up to the level S_n if $x_n < s_n$ and does not order when $x_n \geq s_n$. A Markov policy is called an (s_n, S_n) policy if it is an (s_n, S_n) policy at all steps $n = 0, 1, \ldots$. A policy is called an (s, S) policy if it is stationary and it is an (s, S) policy at all steps $n = 0, 1, \ldots$.*

The following theorem is the main result of this section.

Theorem 6.11 *Consider $\alpha^* \in [0,1)$ whose existence is stated in Proposition 6.8. The following statements hold for the inventory control problem.*

(i) For $\alpha \in [\alpha^*, 1)$ and $n = 0, 1, \dots,$ define $g(x) := G_{n,v_\alpha^0,\alpha}(x)$, $x \in \mathbb{R}$. Consider real numbers $S^*_{n,\alpha}$ satisfying (6.24) and $s^*_{n,\alpha}$ defined in (6.25). Then for each $N = 1, 2, \dots,$ the $(s^*_{N-n,\alpha}, S^*_{N-n,\alpha})$ policy, $n = 1, 2, \dots, N$, is optimal for the N-horizon problem with the terminal values $T(x) = v^0_\alpha(x)$, $x \in \mathbb{R}$;

(ii) For the infinite-horizon expected total discounted cost criterion with a discount factor $\alpha \in [\alpha^*, 1)$, define $g(x) := G_\alpha(x)$, $x \in \mathbb{R}$. Consider real numbers S_α satisfying (6.24) and s_α defined in (6.25). Then the (s_α, S_α) policy is optimal for the discount factor α. Furthermore, each sequence of pairs $\{(s^*_{n,\alpha}, S^*_{n,\alpha})\}_{n=0,1,\dots}$ is bounded and, for each its limit point (s^*_α, S^*_α), the (s^*_α, S^*_α) policy is optimal for the discount factor α.

(iii) Consider the infinite-horizon average cost criterion. For each $\alpha \in [\alpha^*, 1)$, consider an optimal (s'_α, S'_α) policy for the discounted cost criterion with the discount factor α, whose existence follows from Statement (ii). Let $\alpha_n \uparrow 1$, $n = 1, 2, \dots,$ with $\alpha_1 \geq \alpha^*$. Every sequence $\{(s'_{\alpha_n}, S'_{\alpha_n})\}_{n=1,2,\dots}$ is bounded and each limit point (s', S') defines an average-cost optimal (s', S') policy.

Proof. First, for Statements (i) and (ii) note that for $\alpha \in [\alpha^*, 1)$, the functions $G_{n,v^0_\alpha,\alpha}$ are K-convex, lower semi-continuous and $G_{n,v^0_\alpha,\alpha} \to \infty$ as $n \to \infty$; see Corollary 6.4 and Lemma 6.9. The proofs for finite and infinite horizon discounted problems follows directly from optimality equations (6.17), (6.18), Proposition 6.7 with $g = G_{N,v^0_\alpha,\alpha}$ and $g = G_\alpha$ respectively, and Theorem 3.2.

Since $G^0_\alpha(x) \leq G_{n,v^0_\alpha,\alpha}(x) \leq G_{n+1,v^0_\alpha,\alpha}(x) \leq G_\alpha(x)$, $x \in \mathbb{R}$, then the points $s^*_{n,\alpha}$ and $S^*_{n,\alpha}$ belong to the compact set $\{x \in \mathbb{R} : G^0_\alpha(x) \leq K + \min_{x\in\mathbb{R}} G_\alpha(x)\}$. Therefore, the sequence $\{(s^*_{n,\alpha}, S^*_{n,\alpha})\}_{n=0,1,\dots}$ has a limit point (s^*_α, S^*_α). The function $T(x) = v^0_\alpha(x)$ satisfies inequalities in (3.11) so that the results in Theorem 3.4 hold. Theorem 3.4 implies that no inventory should be ordered for $x > s^*_\alpha$ and the inventory up to the level S^*_α should be ordered for $x < s^*_\alpha$. This means that $G_\alpha(x) \leq K + G_\alpha(S^*_\alpha)$ for $x > s^*_\alpha$. Lower semi-continuity of $G_\alpha(x)$ implies that $G_\alpha(s^*_\alpha) \leq K + G_\alpha(S^*_\alpha)$. Thus, the decision that inventory should not be ordered is optimal at $x = s^*_\alpha$. That is, the (s^*_α, S^*_α) policy is optimal for the infinite-horizon problem with the discount factor α.

(iii) We start with the proof of the boundedness of sequences $\{(s'_{\alpha_n}, S'_{\alpha_n})\}_{n=1,2,\dots}$. This means that for an arbitrary selected sequence $\alpha_n \uparrow 1$, $n = 1, 2, \dots,$

$$\liminf_{n\to\infty} s_{\alpha_n} > -\infty \qquad \text{and} \qquad \limsup_{n\to\infty} S_{\alpha_n} < +\infty. \tag{6.28}$$

First, we prove the first inequality in (6.28). If it does not hold, then there is a sequence $\alpha_n \uparrow 1$ such that $\lim_{n\to\infty} s_{\alpha_n} = -\infty$. This means that for each $x \in \mathbb{R}$ there is a sequence $\alpha_n \uparrow 1$, $n = 1, 2, \dots,$ such that $s_{\alpha_n} \leq x$, $n = 1, 2, \dots$. This implies that $0 \in \mathbb{A}_{\alpha_n}(y)$, $n = 1, 2, \dots$ for all $y < x$. Corollary 4.4(ii) implies that the action $0 \in A^*_{\tilde{u}}(y)$ for all $y > x$. Since x is arbitrary, $0 \in A^*_{\tilde{u}}(y)$ for all $y \in \mathbb{R}$. This means that the policy ψ that never places an order is optimal. However, $w^\psi(x) \leq h(x) \to \infty$ as $x \to -\infty$. In view of Assumption **G**, that holds for the inventory control problem, the average cost for an optimal policy is equal to a finite constant w^*. The left inequality in (6.28) is proved.

Second, we prove the second inequality in (6.28). To do this, we formally allow infinite orders if infinite costs are paid. After such an order is placed, the system moves to the absorbing state $+\infty$ and stops there.

Let us add the state $+\infty$ to $\mathbb{R}$ as an isolated point. Of course, optimal policies do not place infinitely large orders. Formally speaking, we consider the state space $\mathbb{R} \cup \{\infty\}$, where ∞ is an isolated point in this case. The action set $\mathbb{A} := [0, \infty]$ and $c(x, \infty) := +\infty$. Consider the metric ρ on $\mathbb{A} = [0, \infty]$ with $\rho(u, u') = |\arctan u - \arctan u'|$ for $u, u' \in [0, \infty$, where $\arctan +\infty = \frac{\pi}{2}$. This metric measures in radians an angle between two vectors connecting on a plane the point $(0, 1)$ with the points $(u, 0), (u', 0)$. On $[0, \infty)$ this metric is equivalent to the Euclidean metric, but any sequence $u_n \to \infty$ converging in this metric to ∞. The metric space $([0, \infty], \rho)$ is Polish (complete and separable).

Since the left inequality in (6.28) holds and the right one does not, there is a sequence $\alpha_n \uparrow 1$ and $x^* \in \mathbb{R}$ such that $x^* < s_{\alpha_n}$, $n = 1, 2, \ldots$, and $S_{\alpha_n} \to \infty$. Corollary 4.4(ii) implies that there exists an average-cost optimal policy ψ such that $\psi(x) = \infty$ when $x < x^*$. This implies that $w^\psi(x) = -\infty$ when $x < x^*$. This is impossible. Thus, (6.28) is proved. Consider a subsequence $\alpha_{n_k} \uparrow 1$ such that $(s'_{\alpha_{n_k}}, S'_{\alpha_{n_k}}) \to (s', S')$ for optimal (s', S') policies for discount factors α_n, $n = 1, 2, \ldots$. Corollary 4.4(ii) implies that $0 \in A^*_{\tilde{u}}(x)$, if $x > s'$ and $S' - x \in A^*_{\tilde{u}}(x)$, if $x < s'$, where the function $\tilde{u}$ is defined in (4.6) for the sequence $\{\alpha_{n_k}\}_{k=1,2,\ldots}$ of discount factors. The last step is to prove that $0 \in A^*_{\tilde{u}}(s')$. To do this, consider a subsequence $\{\alpha^*_n \to 1\}$ of the sequence $\{\alpha_{n_k}\}_{k=1,2,\ldots}$ and a sequence $\{x^{(n)} \to s'\}$ such that $\tilde{u}(s') = \lim_{n\to\infty} u_{\alpha^*_n}(x^{(n)})$. First, consider the case when there is a sequence $\ell_k \to \infty$ such that $x^{\ell_k} \geq s'_{\ell_k}$ for all $k = 1, 2, \ldots$. In this case, $0 \in A_{\alpha^*_{\ell_k}}(x^{(\ell_k)})$, and Corollary 4.4(ii) implies that $0 \in A^*_{\tilde{u}}(s')$. Second, consider the complimentary case when there exists a number N such that $x^{n_k} < s'_{\alpha^*_{\ell_k}}$ when $n_k \geq N$. In view of Proposition 6.7(c), $G_{\alpha^*_{\ell_k}}(x^{n_k}) \geq G_{\alpha^*_{\ell_k}}(s'_{n_k})$. Therefore, for $\ell_k \geq N$

$$\begin{aligned} u_{\alpha^*_{\ell_k}}(x_{\ell_k}) &= v_{\alpha^*_{\ell_k}}(x_{\ell_k}) - m_{\alpha^*_{\ell_k}} = K + G_{\alpha^*_{\ell_k}}(S'_{\alpha^*_{\ell_k}}) - cx_{\ell_k} - m_{\alpha^*_{\ell_k}} \geq G_{\alpha^*_{\ell_k}}(s'_{\alpha^*_{\ell_k}}) - cx_{\ell_k} - m_{\alpha^*_{\ell_k}} \\ &\geq v_{\alpha^*_{\ell_k}}(s'_{\alpha^*_{\ell_k}}) + cs'_{\alpha^*_{\ell_k}} - cx_{\ell_k} - m_{\alpha^*_{\ell_k}} = u_{\alpha^*_{\ell_k}}(s'_{\alpha^*_{\ell_k}}) + c(s'_{\alpha^*_{\ell_k}} - x_{\ell_k}), \end{aligned}$$

where the first and the last equalities follow from the definition of the functions u_α, the second equality follows from (6.18) and from the optimality of the $(s'_{\alpha^*_{\ell_k}}, S'_{\alpha^*_{\ell_k}})$ policies for discount factors $\alpha^*_{\ell_k}$, the first inequality follows from Proposition 6.7(d), and the last inequality follows from (6.18). Since $s'_{\alpha^*_{\ell_k}} \to s'$ and $x_{\ell_k} \to s'$,

$$\tilde{u}(s') = \lim_{k\to\infty} u_{\alpha^*_{\ell_k}}(x_{\ell_k}) = \lim_{k\to\infty} u(s'_{\alpha^*_{\ell_k}}) \geq \tilde{u}(s'),$$

where the last inequality follows from the definition of $\tilde{u}$ in (4.6). Thus, $u(s'_{\alpha^*_{\ell_k}}) \to \tilde{u}(s')$ as $k \to \infty$. Since $0 \in A_{\alpha^*_{\ell_k}}(s'_{\alpha^*_{\ell_k}})$, Theorem 4.4(ii) implies that $0 \in A^*_{\tilde{u}}(s')$. Thus, the (s', S') policy is average-cost optimal. ∎

For $N = 1, 2, \ldots$, we shall write $G_{N,\alpha}$ instead of $G_{N,T,\alpha}$ if $T(x) = 0$ for all $x \in \mathbb{R}$.

Lemma 6.12 *Suppose there exist $z, y \in \mathbb{R}$ such that $z < y$ and*

$$\frac{\mathbb{E}[h(y - D) - h(z - D)]}{y - z} < -c. \tag{6.29}$$

Then $G_\alpha(x) \to \infty$ and $G_{N,\alpha}(x) \to \infty$ as $|x| \to \infty$ for all $\alpha \in [0, 1)$ and for all $N \geq 0$, and these functions are K-convex.

Proof. The proof is based on induction. Obviously, $G_{N,\alpha}(x) \to \infty$ as $x \to \infty$. We show that the result continues to hold when $x \to -\infty$. Suppose $z < y$ satisfy (6.29). Inequality (6.29) can be rewritten as

$$cy + \mathbb{E}\, h(y - D) < cz + \mathbb{E}\, h(z - D).$$

Thus, $G_{0.\alpha}(z) > G_{0,\alpha}(y)$. Since $G_{\alpha,0}$ is convex, then $G_{0,\alpha}(x) \to \infty$ as $x \to -\infty$ and the result holds for $N = 0$. Assume that it holds for N. Since $v_{N,\alpha}$ is lower semi-continuous and q is weakly continuous, $G_{N,\alpha}(x)$ is lower semi-continuous. This together with the inductive hypothesis implies the existence of a minimum of $G_{N,\alpha}(x)$, say $S_{N,\alpha}$. Thus, there exists L_N such that $v_{N+1,\alpha}(x) = K + G_{N,\alpha}(S_{N,\alpha}) - cx$ for all $x \leq L_N$. That is, $v_{N+1,\alpha}(x) \to \infty$ as $x \to -\infty$. Since

$$G_{N+1,\alpha}(x) = G_{0,\alpha}(x) + \mathbb{E}\, v_{N+1,\alpha}(x - D),$$

the result holds for all N. Since $G_{N,\alpha}$ is non-decreasing in N, letting $N \to \infty$ yields the result for $G_\alpha(x)$. ■

Theorem 6.13 *Suppose that the condition stated in Proposition 6.12 holds. Then:*

(i) For $\alpha \in [0, 1)$ and $n = 0, 1, \ldots,$ consider real numbers $S_{n,\alpha}$ and $s_{n,\alpha}$ defined by formulae (6.24) and (6.25) respectively with $g(x) = G_{n,\alpha}(x)$, $x \in \mathbb{R}$. Then for every $N = 1, 2, \ldots$ the $(s^*_{N-n,\alpha}, S^*_{N-n,\alpha})$ policy, $n = 1, 2, \ldots, N$, is an optimal policy for the N-horizon problem with the zero terminal values.

(ii) The discount optimal (s_α, S_α) defined in Theorem 6.11 is optimal for each discount factor $\alpha \in [0, 1)$. Furthermore, $s_\alpha = \lim_{n\to\infty} s_{n,\alpha}$ and any (s_α, S'_α) polic y, where S'_α is a limit point of the sequence $\{S_{n,\alpha}\}_{n=0,1,\ldots}$, is optimal for the discount factor α.

Proof. Observe that $G_{0,\alpha}(x) = cx + \mathbb{E}\, h(x - D)$. This function is convex and, in view of Lemma 6.12, $G_{0,\alpha}(x) \to \infty$ as $|x| \to \infty$. The rest of the proof coincides with the proof of Theorem 6.11 with the functions $G_{n,v^0_\alpha,\alpha}$ replaced with the functions $G_{n,\alpha}$. ■

Remark 6.14 *For the inventory control problem, we have considered an MDP with $\mathbb{X} = \mathbb{R}$ and $A(x) = \mathbb{R}^+ = [0, \infty)$ for each $x \in \mathbb{X}$. However, if the demand takes only integer values, for many problems it is natural to consider $\mathbb{X} = \mathbb{Z}$ and $A(x) = \mathbb{Z}^+$. Therefore, if the demand is integer, we have two MDPs for the inventory control problems: an MDP with $\mathbb{X} = \mathbb{R}$ and an MDP with $\mathbb{X} = \mathbb{Z}$. Though the first MDP yields potentially lower costs, its implementation may not be reasonable for some applications because it may prescribe to order up to a non-integer inventory level. However, all of the results of this paper hold for the second representation, when the state space is integer, with a minor modification that the action sets are integer as well. In fact the case $\mathbb{X} = \mathbb{Z}$ is slightly easier because every function is continuous on it and therefore it is lower semi-continuous. One more note is that the case when the possible demand is proportional to some number $d \in \mathbb{R}$ is similar to the integer demand case. One can consider an MDP with an integer state space for this case as well.*

Acknowledgement. Research of the first author was partially supported by NSF grant CMMI-1335296.

References

[1] D.P. Bertsekas, *Dynamic Programming and Optimal Control*, second edition, vol. 1, Athena Scientific, Belmont, MA, 2000. 2, 3, 19

[2] D.P. Bertsekas and S.E. Shreve, *Stochastic Optimal Control: the Discrete-Time Case*, Athena Scientific, Belmont, MA, 1996. 4, 5

[3] D. Beyer and S.P. Sethi, *The classical average-cost inventory models of Iglehart and Veinott-Wagner revisited*, Journal of Optimization and Applications **101** (1999), no. 3, 523–555. 2

[4] D. Blackwell, *Discounted dynamic programming*, Ann. Math. Statist. **36** (1965), 226–235. 4

[5] X. Chen and D. Simchi-Levi, *Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The finite horizon case*, Oper. Res. **52** (2004), 887–896. 3

[6] X. Chen and D. Simchi-Levi, *Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The infinite horizon case*, Math. Oper. Res. **29** (2004), 698–723. 2, 3

[7] E.A. Feinberg, *A generalization of 'expectation equals reciprocal of intensity' to non-stationary exponential distributions*, J. Appl. Prob. **31** (1994), 262–267. 16

[8] E.A. Feinberg, P.O. Kasyanov, Continuity of minima: local results, Set-Valued Var. Anal., DOI 10.1007/s11228-015-0318-7, Published online on Feb. 8, 2015. 5

[9] E.A. Feinberg, P.O. Kasyanov, M. Voorneveld, Berge's maximum theorem for noncompact image sets, J. Math. Anal. Appl., 413, pp. 1040–1046 (2014). 5

[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. 3, 5, 6, 10, 11

[11] E.A. Feinberg, P.O. Kasyanov, N.V. Zadoianchuk, Berge's theorem for noncompact image sets, J. Math. Anal. Appl., **397**(1) (2013), 255–259. 3, 5, 6

[12] 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

[13] E.A. Feinberg, M. E. Lewis, Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem, Math. Oper. Res. **32**(4), (2007), 769–783. 2, 5, 7, 11

[14] A. Gut, *Stopped Random Walks*, Springer-Verlag, New York, 1988. 16

[15] O. Hernández-Lerma and J.B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer, New York, 1996. 2, 4, 5, 8

[16] D.P. Heyman and M.J. Sobel, *Stochastic Models in Operations Research*, vol. II, McGraw-Hill, New York, 1984. 2, 20

[17] W.T. Huh, G. Janakiraman, M. Nagarajan, *Average cost single-stage inventory models: an analysis using a vanishing discount approach,* Oper. Res. **59** (2011), 143–155. 2, 3

[18] D.L. Iglehart, *Dynamic programming and stationary analysis of inventory problems,* Office of Naval Research Monographs on Mathematical Methods in Logistics, ch. 1, pp. 1–31, Stanford University Press, Stanford, California, 1963. 2

[19] F. Luque-Vasques, O. Hernández-Lerma, *A counterexample on the semicontinuity lemma,* Proc. Amer. Math. Soc. **123** (10) (1995), 3175–3176. 2

[20] E. Porteus, *Stochastic Inventory Theory*, Handbooks in Operations Research and Management Science (Daniel Heyman and Matthew Sobel, eds.), vol. 2, Elsevier Science Publishers, Amsterdam, 1990, pp. 605–652. 2

[21] S.I. Resnick, *Adventures in Stochastic Processes*, Birkhäuser, Boston, 1992. 15

[22] H. Scarf, *The optimality of (s, S) Policies in the Dynamic Inventory Problem*, Mathematical Methods in the Social Sciences 1959, Sranford University Press, Stanford, CA (1960). 1

[23] M. Schäl, *Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal*, Z. Wahrsch. Verw. Gebiete **32** (1975), 179–196. 2, 5

[24] M. Schäl, *Average optimality in dynammic programming with general state space*, Mathematics of Operations Research **18** (1993), no. 1, 163–172. 2, 9, 10

[25] L.I. Sennott, *Stochastic Dynamic Programming and the Control of Queueing Systems*, John Wiley and Sons, Inc., New York, 1999. 2

[26] R. Strauch, *Negative dynamic programming*, Ann. Math. Statist. **37** (1966), 871–890. 4

[27] A.F. Veinott, Jr. *On the optimality of (s, S) inventory policies: new conditions and new proof,* SIAM J. Appl. Math. **14** (1966), 1067–1083. 1

[28] A.F. Veinott, Jr. *The status of mathematical inventory theory,* Manag. Sci. **12** (1966), 745–777. 1

[29] A.F. Veinott, Jr. and H.M. Wagner, *Computing optimal (s,S) policies*, Management Science **11** (1965), no. 5, 525–552. 2

[30] Y. Zheng, *A simple proof for the optimality of (s,S) policies in infinite horizon inventory systems*, Journal of Applied Probability **28** (1991), 802–810. 2