On the Convergence of Optimal Actions for Markov Decision Processes and the Optimality of (*s*, *S*) Inventory Policies

Eugene A. Feinberg,[1] **Mark E. Lewis**[2]

[1] *Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, New York 11794-3600*

[2] *School of Operations Research and Information Engineering, Cornell University, 221 Rhodes Hall, Ithaca, New York 14853-3801*

Received 30 May 2015; revised 20 March 2017; accepted 19 July 2017
DOI 10.1002/nav.21750
Published online 00 00 2017 in Wiley Online Library (wileyonlinelibrary.com).

Abstract: This article studies convergence properties of optimal values and actions for discounted and average-cost Markov decision processes (MDPs) with weakly continuous transition probabilities and applies these properties to the stochastic periodic-review inventory control problem with backorders, positive setup costs, and convex holding/backordering costs. The following results are established for MDPs with possibly non-compact action sets and unbounded cost functions: (i) convergence of value iterations to optimal values for discounted problems with possibly non-zero terminal costs, (ii) convergence of optimal finite-horizon actions to optimal infinite-horizon actions for total discounted costs, as the time horizon tends to infinity, and (iii) convergence of optimal discount-cost actions to optimal average-cost actions for infinite-horizon problems, as the discount factor tends to 1. Being applied to the setup-cost inventory control problem, the general results on MDPs imply the optimality of (s, S) policies and convergence properties of optimal thresholds. In particular this article analyzes the setup-cost inventory control problem without two assumptions often used in the literature: (a) the demand is either discrete or continuous or (b) the backordering cost is higher than the cost of backordered inventory if the amount of backordered inventory is large.

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

1. INTRODUCTION

Since Scarf [29] proved the optimality of (s, S) policies for finite-horizon problems with continuous demand, there have been significant efforts to extend this result to other models. Arthur F. Veinott [35, 36] was one of the pioneers in this exploration, and he combined a deep understanding of Markov decision processes (MDPs) with a passion for the study of inventory control. It is a great pleasure to dedicate this article to him.

This article introduces new results on MDPs with infinite state spaces, weakly continuous transition probabilities, one-step costs that can be unbounded, and possibly non-compact action sets under the discounted and average-cost criteria. The results on MDPs are applied to the stochastic periodic-review setup-cost inventory control problem. We show that this problem satisfies general conditions sufficient for the existence of optimal policies, the validity of the optimality equations, and the convergence of value iterations. In particular, these results are used to show the optimality of (s, S) policies for finite-horizon problems, and for infinite-horizon problems with the discounted and long-term average-cost criteria.

Since the 1950s, inventory control has been one of the major motivations for studying MDPs. However, until recently 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, which often provide additional insight. For example, Theorem 6.10 below states convergence properties of optimal thresholds in addition to the existence of optimal (s, S) policies, and the proof of this theorem is based on Theorem 3.6 and Corollary 4.4 established for 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 books by Bensoussan [1], Beyer et al. [4], Heyman and Sobel [23], Porteus [27], Simchi-Levi et al. [33], and Zipkin [40]. Applications of

Correspondence to: Eugene A. Feinberg (efeinberg@notes.cc.sunysb.edu)

MDPs to inventory control are also discussed in Bertsekas [2]. In the case of inventory control, under the average cost criterion the optimality of (s, S) policies was established by Iglehart [25] and Veinott and Wagner [37] in the continuous and discrete demand cases, respectively. As explained in Beyer and Sethi [5, p. 526] in detail, the analysis in Iglehart [25] assumes the existence of a demand density. The proofs for discrete demand distributions were significantly simplified by Zheng [39]. Zabel [38] corrected Scarf's [29] results on finite-horizon inventory control. Beyer and Sethi [5] described and corrected gaps in the proofs in Refs. 25 and 37. Almost all studies of infinite-horizon inventory control deal with either discrete or continuous demand. In some cases, the choice between the use of discrete and continuous distributions depends on a particular application. There is an important practical reason why many studies use discrete demand. In operations management practice, the overwhelming majority of information systems record integer quantities of demand and stock level. Without assumptions that the demand is discrete or continuous, the optimality of (s, S) policies for average cost inventory control problems follows from Chen and Simchi-Levi [9], where under some technical assumptions coordinated price-inventory control is studied and methods specific to inventory control are used. Huh et al. [24] developed additional problem-specific methods for inventory control problems with arbitrary distributed demands. Under some additional assumptions, including the assumption that holding costs are bounded above by polynomial functions, the optimality of (s, S) policies also takes place when the demand evolves according to a Markov chain; see Beyer et al. [4] and the references therein.

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 [32] developed the theory for the average-cost criterion for countable-state problems with unbounded costs. Schäl [30, 31] developed the theory for uncountable state problems with discounted and average-cost criteria when action sets are compact. In particular, Schäl [30, 31] identified two groups of assumptions on transition probabilities: weak continuity and setwise continuity. As explained in Feinberg and Lewis [17, 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 [22] developed the theory for problems with setwise continuous transition probabilities, unbounded costs, and possibly non-compact action sets. Luque-Vasques and Hernández-Lerma [26] 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, which ensures semi-continuity of the value function, does not hold for problems with non-compact action sets. Feinberg and Lewis [17] investigated total discounted costs for inf-compact cost functions and obtained sufficient optimality conditions for average costs. Compared to Schäl [31] these results required an additional local boundedness assumption that holds for inventory control problems, but its verification is not easy. Feinberg et al. [14, 15] introduced a natural class of $\mathbb{K}$-inf-compact cost functions, extended Berge's theorem to non-compact action sets, and developed the theory of MDPs with weakly continuous transition probabilities, unbounded costs, and with the criteria of total discounted costs and long-term average costs. In particular, the results from Ref. 14 do not require the validity of the local boundedness assumption. This simplifies their applications to inventory control problems. Such applications are considered in Section 6 below. The tutorial by Feinberg [11] describes in detail the applicability of recent results on MDPs to inventory control.

Section 2 of this article describes an MDP with an infinite state space, weakly continuous transition probabilities, possibly unbounded one-step costs, and possibly non-compact action sets. Sections 3 and 4 provide the results for discounted and average cost criteria. In particular, new results are provided on the following topics: (i) convergence of value iterations for discounted problems with possibly non-zero terminal values (Corollary 3.5), (ii) convergence of optimal finite-horizon actions to optimal infinite-horizon actions for total discounted costs, as the time horizon tends to infinity (Theorem 3.6), and (iii) convergence of optimal discount-cost actions to optimal average-cost actions for infinite-horizon problems, as the discount factor tends to 1 (Theorems 4.3 and 4.5). Studying the convergence of value iterations and optimal actions for discounted costs with non-zero terminal values in this article is motivated by inventory control. As was understood by Veinott and Wagner [37], without additional assumptions (s, S) policies may not be optimal for problems with discounted costs, but they are optimal for large values of discount factors. Even for large discount factors, (s, S) policies may not be optimal for finite-horizon problems with discounted cost criteria and zero terminal costs. However, (s, S) policies are optimal for such problems with the appropriately chosen non-zero terminal costs, and this observation is useful for proving the optimality of (s, S) policies for infinite-horizon problems.

Section 5 relates MDPs to problems whose dynamics are defined by stochastic equations, as this takes place for inventory control. Section 6 studies 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 6.10, 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 terminal costs to finite-horizon problems. The second result, Theorem 6.12, states the existence of optimal (s, S) policies for all discount factors under an additional assumption that it is expensive to keep a large backordered amount of inventory. Such assumptions are often used in the literature; see Bertsekas [2], Beyer et al. [4], Chen and Simchi-Levi [8, 9], Huh et al. [24], and Veinott and Wagner [37]. Theorems 6.10 and 6.12 also describe the convergence properties of optimal thresholds for the following two cases: (i) the horizon length tends to infinity, and (ii) the discount factor tends to 1.

In the conclusion of the introduction, we would like to mention that the results on MDPs with weakly continuous transition probabilities, non-compact action sets and unbounded costs presented in this article are broadly applicable to a wide range of engineering and managerial problems. Potential applications include resource allocation problems, control of workload in queues, and a large variety of inventory control problems. In particular, the presented results should be applicable to combined pricing-inventory control and to supply chain management; see Refs. 8, 9 and 33. Moreover, as mentioned above, the results for MDPs presented below significantly simplify the analysis of the stochastic cash balance problem investigated in Ref. 17 because the current results do not require verifying the local boundedness assumption introduced in Ref. 17. Instead Theorem 4.1 below can be employed. The periodic-review setup-cost inventory control problem was selected as an application in this article mainly because it is probably the most highly studied inventory control model. We provide new results for this classic problem.

2. DEFINITION OF MDPS WITH BOREL STATE AND ACTION SETS

Consider a discrete-time MDPs with the state space $\mathbb{X}$, action space $\mathbb{A}$, one-step costs c, and transition probabilities q. The state space $\mathbb{X}$ and action space $\mathbb{A}$ are 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 $t = 0, 1, \dots$ the current state of the system, x_t, 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 $q(\cdot|x, a)$, and the process continues. Let $H_t = (\mathbb{X} \times \mathbb{A})^t \times \mathbb{X}$ be the set of histories for $t = 0, 1, \dots$. A (randomized) decision rule at epoch $t = 0, 1, \dots$ is a regular transition probability π_t from H_t to $\mathbb{A}$. In other words, (i) $\pi_t(\cdot|h_t)$ is a probability distribution on $\mathbb{A}$, where $h_t = (x_0, a_0, x_1, \dots, a_{t-1}, x_t)$ and (ii) for any measurable subset $B \subseteq \mathbb{A}$, the function $\pi_t(B|\cdot)$ is measurable on H_t. A policy π is a sequence $(\pi_0, \pi_1, \dots)$ of decision rules. Moreover, π is called non-randomized if each probability measure $\pi_t(\cdot|h_t)$ 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 $\phi_0, \phi_1, \dots$ of measurable mappings $\phi_t : \mathbb{X} \to \mathbb{A}$. A stationary policy ϕ is defined by a measurable mapping $\phi : \mathbb{X} \to \mathbb{A}$.

The Ionescu Tulcea theorem (see Ref. 3, p. 140–141 or 22, p. 178) implies 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 $\mathbf{F} : \mathbb{X} \to \overline{\mathbb{R}}$ called the terminal value, define the expected total discounted costs

$$v_{N,\mathbf{F},\alpha}^\pi(x) := \mathbb{E}_x^\pi \left[\sum_{t=0}^{N-1} \alpha^t c(x_t, a_t) + \alpha^N \mathbf{F}(x_N) \right], \qquad (2.1)$$

where $\alpha \in [0, 1)$, $v_{0,\mathbf{F},\alpha}^\pi(x) = \mathbf{F}(x)$, $x \in \mathbb{X}$. When $\mathbf{F}(x) = 0$ for all $x \in \mathbb{X}$, we shall write $v_{N,\alpha}^\pi(x)$ instead of $v_{N,\mathbf{F},\alpha}^\pi(x)$. When $N = \infty$ and $\mathbf{F}(x) = 0$ 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

$$w^\pi(x) := \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_x^\pi \sum_{t=0}^{N-1} c(x_t, a_t). \qquad (2.2)$$

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

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

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 to $+\infty$. In other words, a feasible pair (x, a) is modeled as a pair with finite costs. To transform this model to one with feasible action sets, it is sufficient to consider the sets 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 particular, it is possible to set $A(x) := A_c(x)$, $x \in \mathbb{X}$. 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 $a \in A \setminus A(x)$. 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 [7] and Strauch [34] 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. Ref. 3, Proposition 8.2) that $v_{t,\mathbf{F},\alpha}(x)$ satisfies the following *optimality equation*,

$$v_{t+1,\mathbf{F},\alpha}(x) = \inf_{a\in\mathbb{A}} \left\{ c(x,a) + \alpha \int v_{t,\mathbf{F},\alpha}(y) q(dy|x,a) \right\}, \quad x \in \mathbb{X},\ t = 0,1,\dots . \tag{3.1}$$

In addition, a Markov policy ϕ, defined at the first $N+1$ steps by the mappings $\phi_0,\dots,\phi_N$ satisfying the following equations for all $x \in \mathbb{X}$ and all $t = 0,\dots,N$,

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

is optimal for the horizon $N+1$; see e.g. Ref. 3, Lemma 8.7.

It is also well-known (see e.g. Ref. 3, Proposition 9.8) that $v_\alpha(x)$ satisfies the following *discounted cost optimality equation*,

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

According to Ref. 3, Proposition 9.12, a stationary policy ϕ^α is optimal if and only if

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

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 by Schäl [30] and consisted of two sets of conditions that required either weak or setwise continuity assumptions. For setwise continuous transition probabilities, Hernandez-Lerma and Lasserre [22] 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 [17] obtained results for weakly continuous transition probabilities and inf-compact cost functions. Feinberg et al. [14] generalized and unified the results by Schäl [30] and Feinberg and Lewis [17] for weakly continuous transition probabilities to more general cost functions by using the notion of a $\mathbb{K}$-inf-compact function. $\mathbb{K}$-inf-compact functions were originally introduced in Ref. 14, Assumption **W*** without using the term $\mathbb{K}$-inf-compact, and formally introduced and studied in Feinberg et al. [13, 15]. As explained in Feinberg and Lewis [17, 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. This article focuses on the essentially more general case of weakly continuous transition probabilities.

Let $\mathbb{U}$ be a metric space and $U \subseteq \mathbb{U}$. Consider a function $f : U \to \overline{\mathbb{R}}$. For $V \subseteq U$ define the level sets

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

A function $f : U \to \overline{\mathbb{R}}$ is called *lower semi-continuous* at a point $y \in U$ if $f(y) \le \liminf_{n\to\infty} f(y^{(n)})$ for every sequence $\left\{y^{(n)} \in U\right\}_{n=1,2,\dots}$ converging to y. A function $f : U \to \overline{\mathbb{R}}$ is called lower semi-continuous if it is lower semi-continuous at each $y \in U$. A function $f : U \to \overline{\mathbb{R}}$ is called *inf-compact* if all the level sets $\mathcal{D}_f(\lambda; U)$ are compact. Inf-compact functions are lower semi-continuous. For three sets U, V, and W, where $V \subset U$, and two functions $g : V \to W$ and $f : U \to W$, the function g is called the *restriction* of f to V if $g(x) = f(x)$ when $x \in V$.

DEFINITION 3.1: (cf. Feinberg et al. [13, 15], Feinberg and Kasyanov [12]) Let $\mathbb{S}^{(i)}$ be metric spaces and $S^{(i)} \subseteq \mathbb{S}^{(i)}$ be their non-empty Borel subsets, $i = 1,2$. A function $f : S^{(1)} \times S^{(2)} \to \overline{\mathbb{R}}$ is called $\mathbb{K}$-inf-compact if, for any non-empty compact subset K of $S^{(1)}$, the restriction of f to $K \times S^{(2)}$ is inf-compact.

We are mainly interested in applying this definition to the function $f = c$, where c is the one-step cost. In this case, $\mathbb{X}$ and $\mathbb{A}$ are Borel subsets of the Polish spaces $\mathbb{S}^{(1)}$ and $\mathbb{S}^{(2)}$ mentioned in the definition of an MDP. Inventory control applications often deal with $\mathbb{S}^{(1)} = \mathbb{X}$ and $\mathbb{S}^{(2)} = \mathbb{A}$. However, other applications are possible. For example, assumption (ii) of Theorem 5.3 deals with $\mathbb{S}^{(1)} = \mathbb{A}$ and $\mathbb{S}^{(2)} = \mathbb{X}$.

The next proposition, which follows directly from Feinberg et al. [15, Lemma 2.1], demonstrates that $\mathbb{K}$-inf-compact cost functions are natural generalizations of inf-compact cost functions considered in Feinberg and Lewis [17] and lower semi-continuous cost functions considered in the literature on MDPs with compact action sets, see e.g., Schäl [30, 31].

PROPOSITION 3.2: The following two statements hold:

(i) an inf-compact function $f : \mathbb{X} \times \mathbb{A} \to \overline{\mathbb{R}}$ is $\mathbb{K}$-inf-compact;
(ii) if $A : \mathbb{X} \to 2^{\mathbb{A}} \setminus \{\emptyset\}$ is a compact-valued upper semi-continuous set-valued mapping and $f : \mathbb{X} \times \mathbb{A} \to \overline{\mathbb{R}}$ is a lower semi-continuous function such that $f(x,a) = +\infty$ for $x \in \mathbb{X}$ and for $a \in \mathbb{A} \setminus A(x)$, then the function f is $\mathbb{K}$-inf-compact, where 2^U denotes the set of all subsets of a set U.

DEFINITION 3.3: The transition probability q is called weakly continuous, if

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

for every bounded continuous function $f : \mathbb{X} \to \mathbb{R}$ and for each sequence $\{(x^{(n)},a^{(n)}), n = 1,2,\dots\}$ on $\mathbb{X} \times \mathbb{A}$ converging to $(x^{(0)},a^{(0)}) \in \mathbb{X} \times \mathbb{A}$.

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 convergent sequence $\{(x^{(n)},a^{(n)}), n = 1,2,\dots\}$ of elements of $\mathbb{X}\times\mathbb{A}$ such that $c(x^{(n)},a^{(n)}) < +\infty$ for all $n = 0,1,2,\dots$, then the sequence $\{q(\cdot|(x^{(n)},a^{(n)})), n = 1,2,\dots\}$ converges weakly to $q(\cdot|(x^{(0)},a^{(0)}))$; that is, (3.6) 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 iteration.

THEOREM 3.4: (Feinberg et al. [14, Theorem 2]) Suppose Assumption **W*** holds. For $t = 0,1,\dots,N = 0,1,\dots$, and $\alpha \in [0,1)$, the following statements hold:

(i) the functions $\{v_{t,\alpha}, t \geq 0\}$ and v_α are lower semi-continuous on $\mathbb{X}$, and $v_{t,\alpha}(x) \to v_\alpha(x)$ as $t \to +\infty$ for each $x \in \mathbb{X}$;
(ii) the value functions $\{v_{t,\alpha}, t \geq 0\}$ satisfy the optimality equations

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

and the non-empty sets $A_{t,\alpha}(x) := \{a \in \mathbb{A} : v_{t+1,\alpha}(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_{t,\alpha}(y)q(dy|x,a)\}$, $x \in \mathbb{X}$, satisfy the following properties:
  (a) the graph $\mathrm{Gr}_{\mathbb{X}}(A_{t,\alpha}) = \{(x,a) : x \in \mathbb{X}, a \in A_{t,\alpha}(x)\}$ is a Borel subset of $\mathbb{X} \times \mathbb{A}$;
  (b) the following hold:
  (b1) if $v_{t+1,\alpha}(x) = +\infty$, then $A_{t,\alpha}(x) = \mathbb{A}$;
  (b2) if $v_{t+1,\alpha}(x) < +\infty$, then $A_{t,\alpha}(x)$ is compact;
(iii) for each horizon $(N+1)$, there exists a Markov optimal policy $(\phi_0,\dots,\phi_N)$;
(iv) if for an $(N+1)$-horizon Markov policy $(\phi_0,\dots,\phi_N)$ the inclusions $\phi_{N-t}(x) \in A_{t,\alpha}(x)$, $x \in \mathbb{X}$, $t = 0,\dots,N$ hold, then this policy is $(N+1)$-horizon optimal;
(v) the value function v_α satisfies the optimality equation

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

(vi) the non-empty 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 $\mathbb{X} \times \mathbb{A}$;
  (b) if $v_\alpha(x) = +\infty$, then $A_\alpha(x) = \mathbb{A}$ and, if $v_\alpha(x) < +\infty$, then $A_\alpha(x)$ is compact;
(vii) for the 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 $x \in \mathbb{X}$;
(viii) (Feinberg and Lewis [17, Proposition 3.1(iv)]) if the cost function c is inf-compact, then the functions $v_{t,\alpha}$, $t = 1,2,\dots$, and v_α are inf-compact on $\mathbb{X}$

The following corollary extends the previous theorem to non-zero terminal values **F**. This extension is useful for the analysis of inventory control problems.

COROLLARY 3.5: Let Assumption **W*** hold. Consider a bounded below, lower semi-continuous function $\mathbf{F} : \mathbb{X} \to \overline{\mathbb{R}}$. The following statements hold for $t = 0,1,2,\dots$, $N = 0, 1,2,\dots$, and $\alpha \in [0,1)$:

(i) the functions $v_{t,\mathbf{F},\alpha}$ are bounded below and lower semi-continuous;

(ii) the value functions $v_{t+1,\mathbf{F},\alpha}$ satisfy the optimality equations

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

where $v_{0,\mathbf{F},\alpha}(x)=\mathbf{F}(x)$ for all $x\in\mathbb{X}$;

(iii) the non-empty sets

$$A_{t,\mathbf{F},\alpha}(x) := \left\{a\in\mathbb{A}:\ v_{t+1,\mathbf{F},\alpha}(x)=c(x,a) + \alpha\int_{\mathbb{X}} v_{t,\mathbf{F},\alpha}(y)q(dy|x,a)\right\}, \quad x\in\mathbb{X},$$

satisfy the following properties:

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

  (b) the following hold:

  (b1) if $v_{t+1,\mathbf{F},\alpha}(x)=+\infty$, then $A_{t,\mathbf{F},\alpha}(x)=\mathbb{A}$;

  (b2) if $v_{t+1,\mathbf{F},\alpha}(x)<+\infty$, then $A_{t,\mathbf{F},\alpha}(x)$ is compact;

(iv) for an $(N+1)$-horizon problem with the terminal value function $\mathbf{F}$, there exists a Markov optimal policy $(\phi_0,\dots,\phi_N)$ and if, for an $(N+1)$-horizon Markov policy $(\phi_0,\dots,\phi_N)$ the inclusions $\phi_{N-t}(x)\in A_{t,\mathbf{F},\alpha}(x)$, $x\in\mathbb{X}$, $t=0,\dots,N$, hold then this policy is $(N+1)$-horizon optimal;

(v) if $\mathbf{F}(x)\le v_\alpha(x)$ for all $x\in\mathbb{X}$, then $v_{t,\mathbf{F},\alpha}(x)\to v_\alpha(x)$ as $n\to+\infty$ for all $x\in\mathbb{X}$;

(vi) if the cost function c is inf-compact, then each of the functions $v_{t,\mathbf{F},\alpha}$, $t=1,2,\dots$, is inf-compact.

PROOF: Statements (i)–(iv) are corollaries from statements (i)–(iii) of Theorem 3.4. Indeed, the statements of Theorem 3.4, that deals with the finite horizon N, hold when one-step costs at different time epochs vary. In particular, if the one-step cost at epoch $t=0,1,\dots,N$ is defined by a bounded below, measurable cost function c_t 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 $\mathbb{X}\times\{0,1,\dots,N\}$, setting $c((x,t),a)=c_t(x,a)$, and applying the corresponding statements of Theorem 3.4. In our case, $c_t(x,a)=c(x,a)$ for $t=0,1,\dots,N$, and $c_N(x,a)=c(x,a)+\int_{\mathbb{X}}\mathbf{F}(y)q(dy|x,a)$. The function c_N is bounded below and lower semi-continuous.

To prove (v) and (vi), consider first the case when the functions c and $\mathbf{F}$ are non-negative. In this case,

$$v_{t,\alpha}(x)\le v_{t,\mathbf{F},\alpha}(x)\le v_{t,v_\alpha,\alpha}(x)=v_\alpha(x), \quad x\in\mathbb{X},\ t=0,1,\dots. \tag{3.10}$$

Therefore, for non-negative cost functions, Statement (v) follows from Theorem 3.4(i). Statement (vi) follows from (v), Theorem 3.4(viii), and the fact that $v_{t,\mathbf{F},\alpha}\ge v_{t,\alpha}$ since $\mathbf{F}$ is non-negative. In a general case, consider a finite positive constant K such that the functions c and $\mathbf{F}$ are bounded below by $(-K)$. If the cost functions c and $\mathbf{F}$ are increased by K, then the new cost functions are non-negative, each finite-horizon value function $v_{t,\mathbf{F},\alpha}$ is increased by the constant $d_t=K(1-\alpha^{t+1})/(1-\alpha)$, and the infinite-horizon value function v_α is increased by the constant $d=K/(1-\alpha)$. Since $d_t\le d$ and $d_t\to d$ as $t\to\infty$, the general case follows from the case of non-negative cost functions. □

While Theorem 3.4 and Corollary 3.5 state the convergence of value functions and describe the structure of optimal sets of actions, the following theorem describes convergence properties of optimal actions. For $x\in\mathbb{X}$ and $\alpha\in[0,1)$, define the sets $D^*_\alpha(x):=\{a\in\mathbb{A}: c(x,a)\le v_\alpha(x)\}$.

THEOREM 3.6: Let Assumption **W*** hold and $\alpha\in[0,1)$. Suppose $\mathbf{F}:\mathbb{X}\to\overline{\mathbb{R}}$ is bounded below, lower semi-continuous, and such that for all $x\in\mathbb{X}$

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

For $x\in\mathbb{X}$, such that $v_\alpha(x)<\infty$, the following two statements hold:

(i) the set $D^*_\alpha(x)$ is compact, and $A_{t,\mathbf{F},\alpha}(x)\subseteq D^*_\alpha(x)$ for all $t=1,2,\dots$, where the sets $A_{t,\mathbf{F},\alpha}(x)$ are defined in Corollary 3.5(iii);

(ii) each sequence $\left\{a^{(t)}\in A_{t,\mathbf{F},\alpha}(x), t=1,2,\dots\right\}$ is bounded, and all its limit points belong to $A_\alpha(x)$.

In particular, if $c(x,a)\ge 0$ for all $x\in\mathbb{X}$, $a\in\mathbb{A}$, then the function $\mathbf{F}(x)\equiv 0$ satisfies conditions (3.11). In order to prove Theorem 3.6, we need the following lemma, which is a simplified version of Ref. 22, Lemma 4.6.6.

LEMMA 3.7: Let A be a compact subset of $\mathbb{A}$ and f, $f_n: A\to\overline{\mathbb{R}}$, $n=1,2,\dots$, be non-negative, lower semi-continuous, real-valued functions such that $f_n(a)\uparrow f(a)$ as $n\to\infty$ for all $a\in A$. Let $a^{(n)}\in\mathrm{argmin}_{a\in A}f_n(a)$, $n=1,2,\dots$, and a^* be a limit point of the sequence $\left\{a^{(n)}, n=1,2,\dots\right\}$. Then $a^*\in\mathrm{argmin}_{a\in A}f(a)$.

PROOF: Let $a'\in\mathrm{argmin}_{a\in A}f(a)$. Then $f(a')\ge f_n(a^{(n)})\ge f_k(a^{(n)})$ for all $n\ge k$. Since A is compact, then $a^*\in\mathbb{A}$. 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 $f(a')\ge f_k(a^*)\uparrow f(a^*)$. Since $f(a^*)\le f(a')$, then $a^*\in\mathrm{argmin}_{a\in A}f(a)$. □

PROOF OF THEOREM 3.6: We assume without loss of generality that the bounded below functions c and $\mathbf{F}$ are non-negative. We can do this because of the arguments provided at the end of the proof of Corollary 3.5 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 $A_{t,\mathbf{F},\alpha}(\cdot)$ and infinite-horizon actions $A_\alpha(\cdot)$ remain unchanged.

Fix $x \in \mathbb{X}$. Since the function $v_{t,\mathbf{F},\alpha}$ is non-negative and, in view of (3.10), $v_{t+1,\mathbf{F},\alpha}(x) \le v_\alpha(x)$,

$$A_{t,\mathbf{F},\alpha}(x) = \left\{ a \in \mathbb{A} : c(x,a) + \alpha \int_{\mathbb{X}} v_{t,\mathbf{F},\alpha}(y) q(dy|x,a) = v_{t+1,\mathbf{F},\alpha}(x) \right\} \subseteq D^*_\alpha(x), \qquad t = 1,2,\dots .$$

Statement (i) is proved. Since $D^*_\alpha(x)$ is compact, every sequence $\{a^{(t)} \in A_{t,\mathbf{F},\alpha}(x)\}_{t=1,2,\dots}$ is bounded and has a limit point. The theorem follows from Lemma 3.7 applied to the set $A := D^*_\alpha(x)$ and functions

$$f(a) = c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) q(dy|x,a), \quad a \in A,$$

$$f_t(a) = c(x,a) + \alpha \int_{\mathbb{X}} v_{t,\mathbf{F},\alpha}(y) q(dy|x,a), \quad a \in A,\ t = 0,1,\dots .$$

To verify the conditions of Lemma 3.7, observe that for all $z \in \mathbb{X}$

$$v_\alpha(z) = v_{t,v_\alpha,\alpha}(z) \ge v_{t,\mathbf{F},\alpha}(z) \ge v_{t,\alpha}(z) \uparrow v_\alpha(z),$$

where the first equality follows from the optimality equation, the first and the second inequalities follow from $v_\alpha(\cdot) \ge \mathbf{F}(\cdot) \ge 0$, and the convergence is stated in Theorem 3.4(i); this convergence is monotone because c and $\mathbf{F}$ are non-negative functions. The inequality $v_{1,\mathbf{F},\alpha}(\cdot) \ge \mathbf{F}(\cdot)$ in (3.11), equality (3.9), and standard induction arguments imply $v_{t+1,\mathbf{F},\alpha}(\cdot) \ge v_{t,\mathbf{F},\alpha}(\cdot)$, $t = 0,1,\dots$. Thus Assumption (3.11) implies that $v_{t,\mathbf{F},\alpha} \uparrow v_\alpha$, and the monotone convergence theorem implies $f_t \uparrow f$ as $t \to \infty$. □

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 [31]. Without this assumption the problem is trivial because $w(x) = +\infty$ for all $x \in \mathbb{X}$, and therefore every 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 $\alpha \in [0,1)$:

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

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

Observe that $u_\alpha(x) \ge 0$ for all $x \in \mathbb{X}$. According to Schäl [31, Lemma 1.2], Assumption **G** implies

$$0 \le \underline{w} \le \bar{w} \le w^* < +\infty. \tag{4.1}$$

Moreover, Schäl [31, Proposition 1.3], states that, if there exist a measurable function $u : \mathbb{X} \to \mathbb{R}^+$, where $\mathbb{R}^+ := [0,+\infty)$, and a stationary policy ϕ such that

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

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 $\sup_{\alpha\in[0,1)} u_\alpha(x) < \infty$ for all $x \in \mathbb{X}$.

We note that the second part of Assumption **B** is Condition **B** in Schäl [31]. Thus, under Assumption **G**, which is assumed throughout Ref. 31, Assumption **B** is equivalent to Condition **B** in Ref. 31.

For $x \in \mathbb{X}$ and for a non-negative lower semi-continuous function $u : \mathbb{X} \to \mathbb{R}^+$, define the set

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

A stationary policy ϕ satisfies (4.2) if and only if $A^*_u(x) \ne \emptyset$ and $\phi(x) \in A^*_u(x)$ for all $x \in \mathbb{X}$.

Following Feinberg et al. [14, Formula (21)], define

$$u(x) := \liminf_{(y,\alpha)\to(x,1-)} u_\alpha(y), \quad x \in \mathbb{X}. \tag{4.4}$$

In words, $u(x)$ is the largest number such that $u(x) \le \liminf_{n\to\infty} u_{\alpha_n}(y_n)$ for all sequences $\{y_n, n \ge 1\}$ and $\{\alpha_n, n \ge 1\}$ such that $y_n \to x$ and $\alpha_n \to 1$.

Following Schäl [31, Page 166], where the notation $\underline{w}$ is used instead of $\tilde{u}$, and Feinberg et al. [14, Formula (38)], for a particular sequence $\alpha_n \to 1-$, define

$$\tilde{u}(x) := \liminf_{(y,n)\to(x,\infty)} u_{\alpha_n}(y), \quad x \in \mathbb{X}. \tag{4.5}$$

In words, $\tilde{u}(x)$ is the largest number such that $\tilde{u}(x) \leq \liminf_{n\to\infty} u_{\alpha_n}(y_n)$ for all sequences $\{y_n, n \geq 1\}$ converging to x.

It follows from these definitions that $u(x) \leq \tilde{u}(x)$, $x \in \mathbb{X}$. However, the questions, whether $u = \tilde{u}$ and whether the values of $\tilde{u}$ depend on a particular choice of the sequence α_n, have not been investigated. If Assumption **B** holds, then $\tilde{u}(x) < +\infty$, $x \in \mathbb{X}$. If Assumption **B** holds and the cost function c is inf-compact, then the functions v_α, u, and $\tilde{u}$ are inf-compact as well; see Theorem 3.4(i) for this fact for v_α and Feinberg et al. [14, Theorem 4(e) and Corollary 2] for u and $\tilde{u}$.

THEOREM 4.1: (Feinberg et al. [14, Theorem 4 and Corollary 2]). Suppose Assumptions **W*** and **B** hold. The following two properties hold for the function u defined in (4.4) and for $u = \tilde{u}$, where $\tilde{u}$ is defined in (4.5) for a sequence $\{\alpha_n, n \geq 1\}$ such that $\alpha_n \uparrow 1$:

(a) for each $x \in \mathbb{X}$ the set $A_u^*(x)$ is non-empty and compact;
(b) the graph $\mathrm{Gr}_{\mathbb{X}}(A_u^*) = \{(x,a) : x \in \mathbb{X}, a \in A_u^*(x)\}$ is a Borel subset of $\mathbb{X} \times \mathbb{A}$.

Furthermore, the following statements hold:

(i) there exists a stationary policy ϕ satisfying (4.2);
(ii) every stationary policy ϕ satisfying (4.2) is optimal for the average cost per unit time criterion, and

$$w^\phi(x) = w(x) = w^* = \underline{w} = \bar{w} = \lim_{\alpha\uparrow 1}(1-\alpha)v_\alpha(x) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}_x^\pi \sum_{t=0}^{N-1} c(x_t, a_t), \qquad x \in \mathbb{X}. \tag{4.6}$$

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 $X_\alpha = \{x \in \mathbb{X} : v_\alpha(x) \leq m_\alpha\}$, this set is closed if Assumptions **G** and **W*** hold. If the function c is inf-compact then inf-compactness of v_α implies that the sets X_α are non-empty and compact. The following fact is useful for verifying the validity of Assumption **B**; see Feinberg and Lewis [17, Lemma 5.1] and the references therein.

THEOREM 4.2: (Feinberg et al. [14, Theorem 6]). 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. [14, Theorem 5 and Corollary 3], certain average cost optimal policies can be approximated by discount optimal policies with a vanishing discount factor. The following theorem describes particular constructions of such approximations. Recall that, for the function u(x) defined in (4.4), for each $x \in \mathbb{X}$ there exist sequences $\{\alpha_n, n \geq 1\}$ and $\{x^{(n)}, n \geq 1\}$ such that $\alpha_n \uparrow 1$ and $x^{(n)} \to x$, where $x^{(n)} \in \mathbb{X}$, such that $u(x) = \lim_{n\to\infty} u_{\alpha_n}(x^{(n)})$. Similarly, for a sequence $\{\alpha_n, n \geq 1\}$ such that $\alpha_n \uparrow 1$ consider the function $\tilde{u}$ defined in (4.5). For each $x \in \mathbb{X}$ there exist a sequence $\{x^{(n)}, n \geq 1\}$ of points in $\mathbb{X}$ converging to x and a subsequence $\{\alpha_n^*, n \geq 1\}$ of the sequence $\{\alpha_n, n \geq 1\}$ 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) Consider a sequence $\{(x^{(n)}, \alpha_n), n \geq 1\}$ with $0 \leq \alpha_n \uparrow 1$, $x^{(n)} \in \mathbb{X}$, $x^{(n)} \to x$, and $u_{\alpha_n}(x^{(n)}) \to u(x)$ as $n \to \infty$. If there are sequences of natural numbers $\{n_k, k \geq 1\}$ and actions $\{a^{(n_k)} \in A_{\alpha_{n_k}}(x^{(n_k)}), k \geq 1\}$, such that $n_k \to \infty$ and $a^{(n_k)} \to a^*$ as $k \to \infty$, then $a^* \in A_u^*(x)$, where the function u is defined in (4.4);
(ii) Suppose $\{\alpha_n, n \geq 1\}$ is a sequence of discount factors such that $\alpha_n \uparrow 1$. Let $\{\alpha_n^*, n \geq 1\}$ be its subsequence and $\{x^{(n)}, n \geq 1\}$ be a sequence of states such that $x^{(n)} \to x$ and $u_{\alpha_n^*}(x^{(n)}) \to \tilde{u}(x)$ as $n \to \infty$, where the function $\tilde{u}$ is defined in (4.5) for the sequence $\{\alpha_n, n \geq 1\}$. If there are actions $a^{(n)} \in A_{\alpha_n^*}(x^{(n)})$ such that $a^{(n)} \to a^*$ as $n \to \infty$, then $a^* \in A_{\tilde{u}}^*(x)$.

PROOF: To show (i), consider sequences whose existence is assumed in the theorem. We have

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

This implies (with a little algebra)

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

Fatou's lemma for weakly converging measures (see e.g., Feinberg et al. [16, Theorem 1.1]), the choice of the sequence $x^{(n_k)}$, and Theorem 4.1 yield

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

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 hold:

(i) if each sequence $\{(\alpha_n^*, x^{(n)}), n \geq 1\}$, with $0 \leq \alpha_n^* \uparrow 1$, $x^{(n)} \in \mathbb{X}$, and $x^{(n)} \to x$, contains a subsequence $(\alpha_{n_k}, x^{(n_k)})$, such that there exist actions $a^{(n_k)} \in A_{\alpha_{n_k}}(x^{(n_k)})$ satisfying $a^{(n_k)} \to a^*$ as $k \to \infty$, then $a^* \in A_u^*(x)$ with the function u defined in (4.4);

(ii) if there exists a sequence $\{\alpha_n, n \geq 1\}$ such that $\alpha_n \uparrow 1$ and for every sequence of states $\{x_n \to x\}$ from $\mathbb{X}$ there are actions $a^{(n)} \in A_{\alpha_n}(x^{(n)})$, $n = 1, 2, \ldots$, satisfying $a^{(n)} \to a^*$ as $n \to \infty$, then $a^* \in A_{\tilde{u}}^*(x)$, where the function $\tilde{u}$ is defined in (4.5) for the sequence $\{\alpha_n, n \geq 1\}$.

PROOF: Statement (i) follows from Theorem 4.3(i) applied to a sequence $\{(\alpha_n^*, x^{(n)}), n \geq 1\}$ 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 $\{x^{(n)}, n \geq 1\}$ and a subsequence $\{\alpha_n^*, n \geq 1\}$ of $\{\alpha_n, n \geq 1\}$ such that $\tilde{u}(x) = \lim_{n\to\infty} u_{\alpha_n^*}(x^{(n)})$. □

The following theorem is useful for proving asymptotic properties of optimal actions for discounted problems when the discount factor tends to 1.

THEOREM 4.5: Let Assumptions **W*** and **B** hold. For $x \in \mathbb{X}$ the following hold:

(i) There exists a compact set $D^*(x) \subseteq \mathbb{A}$ such that $A_\alpha(x) \subseteq D^*(x)$ for all $\alpha \in [0, 1)$;

(ii) If $\{\alpha_n, n \geq 1\}$ is a sequence of discount factors $\alpha_n \in [0, 1)$, then every sequence of infinite-horizon α_n-discount cost optimal actions $\{a^{(n)}, n \geq 1\}$, where $a^{(n)} \in A_{\alpha_n}(x)$, is bounded and therefore has a limit point $a^* \in \mathbb{A}$.

PROOF: For each x, the set of optimal actions $A_\alpha(x)$ in state x does not change if a constant is added to the cost function c. Therefore, we assume without loss of generality that the cost function c is non-negative. Fix $x \in \mathbb{X}$ and $\varepsilon^* > 0$. Since x is fixed, we sometimes omit it. For $\alpha \in [0, 1)$ and $a \in \mathbb{A}$ define

$$U(x) := \sup_{\alpha\in[0,1)} u_\alpha(x),$$

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

$$g_\alpha(a) := c(x, a) + \alpha \int_{\mathbb{X}} u_\alpha(y) q(dy|x, a).$$

Observe that $g_\alpha(a) = f_\alpha(a) - \alpha m_\alpha$ and

$$A_\alpha(x) = \left\{a \in \mathbb{A} : f_\alpha(a) = \min_{b\in\mathbb{A}} f_\alpha(b)\right\}$$

$$= \left\{a \in \mathbb{A} : g_\alpha(a) = \min_{b\in\mathbb{A}} g_\alpha(b)\right\}.$$

Assumption **B** implies that $U(x) < +\infty$, and Theorem 4.1 implies that $\lim_{\alpha\uparrow 1}(1 - \alpha)m_\alpha = w^*$. As shown in Feinberg et al. [14, the first displayed formula on p. 602], there exists $\alpha_0 \in [0, 1)$ such that for $\alpha \in [\alpha_0, 1)$,

$$w^* + \varepsilon^* + U(x) \geq (1 - \alpha)m_\alpha + u_\alpha(x) = \min_{a\in\mathbb{A}} g_\alpha(a),$$

This implies that for $\alpha \in [\alpha_0, 1)$

$$A_\alpha(x) \subseteq \mathcal{D}_{g_\alpha}(\lambda_1; \mathbb{A}) \subseteq \mathcal{D}_{g_0}(\lambda_1; \mathbb{A}),$$

where the definition of the level sets $\mathcal{D}_\cdot(\cdot; \cdot)$ is given in (3.5), $\lambda_1 := w^* + \varepsilon^* + U(x)$, and the second inclusion holds because the function u_α takes non-negative values. Recall that $f_0(a) = g_0(a) = c(x, a)$, $a \in \mathbb{A}$, and the function $c(x, \cdot) : \mathbb{A} \to \overline{\mathbb{R}}$ is inf-compact. Therefore, $\mathcal{D}_{f_0}(\lambda; \mathbb{A}) = \mathcal{D}_{g_0}(\lambda; \mathbb{A})$ and this set is compact for each $\lambda \in \mathbb{R}$. In addition, for all $\alpha \in [0, \alpha_0)$,

$$v_{\alpha_0}(x) \geq v_\alpha(x) = \min_{a\in\mathbb{A}} f_\alpha(a),$$

where the inequality holds because one-step costs c are non-negative. The equality is simply the optimality Eq. (3.1). This implies that for $\alpha \in [0, \alpha_0)$

$$A_\alpha(x) \subseteq \mathcal{D}_{f_\alpha}(v_{\alpha_0}(x); \mathbb{A}) \subseteq \mathcal{D}_{f_0}(v_{\alpha_0}(x); \mathbb{A}).$$

Let $D^*(x) := \mathcal{D}_{g_0}(\lambda_1; \mathbb{A}) \cup \mathcal{D}_{f_0}(v_{\alpha_0}(x); \mathbb{A})$. This set is compact as it is the union of two compact sets, and $A_\alpha(x) \subseteq D^*(x)$ for all $\alpha \in [0, 1)$. Statement (i) is proved, and it implies Statement (ii). □

5. MDPS DEFINED BY STOCHASTIC 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 $\{x_t, t \geq 0\}$ whose dynamics are defined by the stochastic equation

$$x_{t+1} = \mathbf{f}(x_t, a_t, \xi_{t+1}), \qquad t = 0, 1, \ldots, \tag{5.1}$$

where $\{\xi_t, t \geq 1\}$ are independent and identically distributed random variables with values in $\mathbb{S}$, whose distributions are defined by the probability measure μ, and $\mathbf{f} : \mathbb{X}\times\mathbb{A}\times\mathbb{S} \to \mathbb{X}$ is a continuous mapping. This equation defines the transition probability

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

from $\mathbb{X} \times \mathbb{A}$ to $\mathbb{X}$, where $\mathbf{I}$ is the indicator function.

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 $a^{(n)} \to a$ as $n \to +\infty$ defined on $\mathbb{X}$ and $\mathbb{A}$ respectively,

$$\begin{aligned}\limsup_{n\to\infty} q(B|x^{(n)}, a^{(n)}) \\ &= \limsup_{n\to\infty} \int_{\mathbb{S}} \mathbf{I}\left\{\mathbf{f}(x^{(n)}, a^{(n)}, s) \in B\right\} \mu(ds) \\ &\le \int_{\mathbb{S}} \limsup_{n\to\infty} \mathbf{I}\left\{\mathbf{f}(x^{(n)}, a^{(n)}, s) \in B\right\} \mu(ds) \le q(B|x, a),\end{aligned}$$

where the first inequality follows from Fatou's lemma and the second follows from (5.2) and upper semi-continuity of the function $\mathbf{I}\{\mathbf{f}(x, a, s) \in B\}$ on $\mathbb{X} \times \mathbb{A} \times \mathbb{S}$ for a closed set B. The weak continuity of q follows from Billingsley [6, Theorem 2.1] □

COROLLARY 5.2: Consider an MDP $\{\mathbb{X}, \mathbb{A}, q, c\}$ with the transition function q defined in (5.2) for the continuous function **f** introduced in (5.1) and with the non-negative $\mathbb{K}$-inf compact cost function c. This MDP satisfies Assumption **W*** and therefore the conclusions of Theorem 3.4 hold.

PROOF: Assumption **W***(i) is assumed in the corollary. Assumption **W***(ii) 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/backordering cost or expected holding/back-ordering cost at the following step. For simplicity we assume that the functions take non-negative 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 non-negative 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 take place even under weaker assumptions on the functions $C(a)$ and $H(x, a)$.

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 Theorems 3.4(i)–(vii), 3.6 and Corollary 3.5(i)–(v) hold. Furthermore, if either of the following two assumptions holds:

(i) 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 inf-compact;
(ii) the function $C : \mathbb{A} \to [0, \infty]$ is inf-compact and the function $H^* : \mathbb{A} \times \mathbb{X} \to [0, \infty]$ is $\mathbb{K}$-inf-compact, where $H^*(a, x) := H(x, a)$ for all $x \in \mathbb{X}$ and all $a \in \mathbb{A}$;

then the function c is inf-compact and therefore the conclusions of Theorems 3.4, 3.6 and Corollary 3.5 hold.

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 $\mathbb{X} \times \mathbb{A}$, if the function $C : \mathbb{A} \to [0, \infty]$ is inf-compact. Thus under assumptions (1) or (2), c is a $\mathbb{K}$-inf-compact function because it is a sum of a non-negative lower semi-continuous function and a $\mathbb{K}$-inf-compact function. In addition, under assumption (i), as explained in the paragraph preceding the formulation of the theorem, the one-step cost function c is inf-compact. Under assumption (ii), the function c is inf-compact because of the following arguments. Let $c^*(a, x) := C(a) + H^*(a, x)$ for all $(a, x) \in \mathbb{A} \times \mathbb{X}$. The function $c^* : \mathbb{A} \times \mathbb{X} \to [0, \infty]$ is lower semi-continuous if either Assumption (1) or Assumption (2) holds. Since $c(x, a) = c^*(a, x)$ for all $x \in \mathbb{X}$ and $a \in \mathbb{A}$, the function $c : \mathbb{X} \times \mathbb{A} \to [0, \infty]$ is inf-compact if and only is the function $c^* : \mathbb{A} \times \mathbb{X} \to [0, \infty]$ is inf-compact. The function c^* is a sum of the non-negative lower semi-continuous function C and the $\mathbb{K}$-inf-compact function H^*. Therefore, c^* is $\mathbb{K}$-inf-compact. Consider an arbitrary $\lambda \in \mathbb{R}$. Since $c^*(a, x) \ge C(a) > \lambda$ for $a \notin \mathcal{D}_C(\lambda; \mathbb{A})$, then $\mathcal{D}_{c^*}(\lambda; \mathbb{A} \times \mathbb{X}) = \mathcal{D}_{c^*}(\lambda; \mathcal{D}_C(\lambda; \mathbb{A}) \times \mathbb{X})$, and this set is compact because the set $\mathcal{D}_C(\lambda; \mathbb{A})$ is compact and the function c^* is $\mathbb{K}$-inf-compact. Thus the functions c^* and c are inf-compact. □

REMARK 5.4: In view of Theorem 3.4, Assumption **W*** implies the existence of optimal policies for the expected total discounted cost criterion. 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 for the average costs per unit time criterion. However, this is more subtle than for Assumption **W***, and in this article we verify Assumptions **G** and **B** directly for the periodic review inventory control problems.

6. OPTIMALITY OF (s, S) POLICIES FOR SETUP-COST INVENTORY CONTROL PROBLEMS

In this section we consider a discrete-time periodic-review inventory control problem with backorders, prove the existence of an optimal (s, S) policy, and establish several relevant results. For this problem, the state space is $\mathbb{X} := \mathbb{R}$, the action space is $\mathbb{A} := \mathbb{R}^+$, and the dynamics are defined by the following stochastic equation

$$x_{t+1} = x_t + a_t - D_{t+1}, \quad t = 0, 1, 2, \dots, \tag{6.1}$$

where x_t is the inventory at the end of period t, a_t is the decision on how much should be ordered, and D_t is the demand during period t. The demand is assumed to be i.i.d. In other words, if we change the notation ξ_t to D_{t+1}, the dynamics are defined by Eq. (5.1) with $\mathbf{f}(x, a, D) = x + a - D$. 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,
- $\bar{c} > 0$ be the per unit ordering cost,
- D be a non-negative random variable with the same distribution as D_t,
- $h(\cdot)$ denote the holding/backordering cost per period. It is assumed that $h : \mathbb{R} \to \mathbb{R}^+$ is a convex function, $h(x) \to \infty$ as $|x| \to \infty$, and $\mathbb{E}h(x - D) < \infty$ for all $x \in \mathbb{R}$.

Note that $\mathbb{E}D < \infty$ since, in view of Jensen's inequality, $h(x - \mathbb{E}D) \leq \mathbb{E}h(x - D) < \infty$. Without loss of generality, assume that h is non-negative, $h(0) = 0$, and $h(x) > 0$ for $x < 0$. Otherwise, let $x^* \in \mathbb{R}$ be a point, at which the function h reaches its minimum value on $\mathbb{R}$. Define the variable $\bar{x} := x - x^*$ and the function $\bar{h}(\bar{x}) := h(\bar{x} + x^*) - h(x^*)$, $\bar{x} \in \mathbb{R}$. Then $\bar{h}$ is a non-negative convex function with $\bar{h}(\bar{x}) \to \infty$ as $|\bar{x}| \to \infty$, $\bar{h}(0) = 0$, and $\bar{h}(\bar{x}) > 0$ for $\bar{x} < 0$.

The cost function c for this model is defined in (5.3) with $C(a) := K1_{\{a>0\}} + \bar{c}a$ and $H(x, a) := \mathbb{E}h(x + a - D)$. The function $C : \mathbb{A} \to \mathbb{R}^+$ is inf-compact. In fact, it is continuous at $a > 0$ and lower semi-continuous at $a = 0$. The function $H^* : \mathbb{A} \times \mathbb{X} \to \mathbb{R}^+$, where $H^*(a, x) := H(x, a)$ for all $(a, x) \in \mathbb{A} \times \mathbb{X}$, is $\mathbb{K}$-inf-compact because of the properties of the function h. Theorem 5.3 (case (ii)) implies that the function c is inf-compact. Therefore, in view of Proposition 3.2, the function c is $\mathbb{K}$-inf-compact.

The problem is posed with $\mathbb{X} = \mathbb{R}$ and $\mathbb{A} = \mathbb{R}^+$. 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.

Consider the following corollary from Theorems 3.4, 3.6, and 5.3.

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 3.4, 3.6 and Corollary 3.5 hold.

PROOF: The validity of Assumption **W*** and inf-compactness of c follow from Theorem 5.3 (case (ii)). □

Consider the renewal process

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

where $t \in \mathbb{R}^+$, $\boldsymbol{S}_0 = 0$ and $\boldsymbol{S}_n = \sum_{j=1}^{n} D_j$ for $n > 0$. Observe that, if $P(D > 0) > 0$, then $\mathbb{E}\mathbf{N}(t) < \infty$ for each $t \in \mathbb{R}^+$; Resnick [28, Theorem 3.3.1]. Thus, Wald's identity yields that for all $y \in \mathbb{R}^+$

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

We next state a useful lemma.

LEMMA 6.2: For fixed initial state x, if $P(D > 0) > 0$, then

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

where $0 \leq y < +\infty$.

PROOF: Define

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

Observe that it suffices to show that

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

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

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

To show that $E^*_y(x) < +\infty$, we shall prove that

$$\mathbb{E}h^*(x - \mathbf{S}_{\mathbf{N}(y)+1}) \le (1 + \mathbb{E}\mathbf{N}(y))\mathbb{E}h^*(x - y - D) < +\infty. \quad (6.6)$$

Define the function $f(z) = h^*(x - y - z)$. This function is non-decreasing and convex. Since f is convex, its derivative exists almost everywhere. Denote the excess of $\mathbf{N}(y)$ by $R(y) := \mathbf{S}_{\mathbf{N}(y)+1} - y$. According to Gut [21, p. 59], for $t, y \in \mathbb{R}^+$

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

where F_D is the distribution function of D and $U(s) = \mathbb{E}\mathbf{N}(s)$ is the renewal function. Thus,

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

where $J_1 = \int_0^\infty f'(t)(1 - F_D(y + t))dt$, $J_2 = \int_0^\infty f'(t)\left(\int_0^y (1 - F_D(y + t - s))dU(s)\right)dt$, and the third equality in (6.7) holds according to Feinberg [10, p. 263]. Note that since F_D is non-decreasing,

$$J_1 \le \int_0^\infty f'(t)(1 - F_D(t))dt = \mathbb{E}f(D) = \mathbb{E}h^*(x - y - D) \le \mathbb{E}h(x - y - D) < +\infty, \quad (6.8)$$

where the first equality follows from Ref. 10, p. 263. Similarly, by applying Fubini's theorem,

$$J_2 = \int_0^y \left(\int_0^\infty f'(t)(1 - F_D(y + t - s))dt\right)dU(s) \le \int_0^y \left(\int_0^\infty f'(t)(1 - F_D(t))dt\right)dU(s) = \mathbb{E}f(D)\mathbb{E}U(y) = \mathbb{E}h^*(x - y - D)\mathbb{E}\mathbf{N}(y). \quad (6.9)$$

Combining (6.7)–(6.9) yields (6.6). □

The following proposition is useful for the average-cost criterion. In addition to this proposition, observe that the case $D=0$ almost surely is trivial for this criterion. In this case, the policy ϕ, ordering up to the level 0, if $x < 0$, and doing nothing otherwise, is average-cost optimal. For this policy $w(x) = w^\phi(x) = 0$, if $x \le 0$, and $w(x) = w^\phi(x) = h(x)$, if $x > 0$. Observe that ϕ is the (0, 0) policy. Since $w(x)$ depends on x, then Theorem 4.1 implies that Assumption **B** does not hold when $D=0$ almost surely.

PROPOSITION 6.3: The inventory control model satisfies Assumption **G** and, therefore, the conclusions of Theorem 4.2 hold. Furthermore, if $P(D > 0) > 0$, then Assumption **B** is satisfied and the conclusions of Theorems 4.1, 4.3 and 4.5 hold.

PROOF: Consider the policy ϕ that orders up to the level 0, if the inventory level is less than 0, and does nothing otherwise. Then $w^\phi(0) = KP(D > 0) + \bar{c}\mathbb{E}D + \mathbb{E}h(-D) < +\infty$. That is, Assumption **G** holds.

In view of Corollary 6.1, Theorem 3.4 implies that for every discount factor $\alpha \in [0, 1)$ there exists a stationary discount-optimal policy ϕ^α. Theorem 4.2 implies that $\cup_{\alpha\in[0,1)}X_\alpha \subseteq \mathcal{K}$ for some $\mathcal{K} \subseteq \mathbb{R}$. Let $[x^*_L, x^*_U]$ be a bounded interval in $\mathbb{R}$ such that $\mathcal{K} \subseteq [x^*_L, x^*_U]$. Thus,

$$\cup_{\alpha\in[0,1)}X_\alpha \subseteq [x^*_L, x^*_U].$$

For a discount factor $\alpha \in [0, 1)$, fix a stationary optimal policy ϕ^α and a state $x^\alpha \in [x^*_L, x^*_U]$ such that $v_\alpha(x^\alpha) = m_\alpha$. Observe that $\phi^\alpha(x^\alpha) = 0$. Indeed, let $\phi^\alpha(x^\alpha) = a > 0$. We have

$$v_\alpha(x^\alpha) = K + ca + h(x^\alpha + a - D) + \alpha\mathbb{E}v_\alpha(x^\alpha + a - D) > K + c\left(\frac{a}{2}\right) + h\left(\left(x^\alpha + \frac{a}{2}\right) + \frac{a}{2} - D\right) + \alpha\mathbb{E}v_\alpha\left(\left(x^\alpha + \frac{a}{2}\right) + \frac{a}{2} - D\right) \ge v_\alpha\left(x^\alpha + \frac{a}{2}\right),$$

where the second inequality follows since the optimal action in state $x^\alpha + \frac{\alpha}{2}$ may not be to order $\frac{a}{2}$. The inequality $v_\alpha(x^\alpha) > v_\alpha(x^\alpha + \frac{a}{2})$ contradicts $v_\alpha(x^\alpha) = m_\alpha$.

Let σ be the policy defined by the following rules depending on the initial state x : (i) if $x < x^\alpha$, then at the initial time instance σ orders up to a level x^α and then switches to ϕ^α, and (ii) if $x \ge x^\alpha$, the policy σ does not order as long as the inventory level is greater than or equal to x^α and starting from the time, when the inventory level becomes smaller than to x^α, the policy σ behaves as described in (i) starting from time 0.

For $x < x^\alpha$,

$$v^\sigma_\alpha(x) = K + \bar{c}(x^\alpha - x) + v_\alpha(x^\alpha) \le K + \bar{c}(x^*_U - x) + m_\alpha. \quad (6.10)$$

The inequality in (6.10) yields for $x < x^\alpha$,

$$v_\alpha(x) - m_\alpha \le v^\sigma_\alpha(x) - m_\alpha \le K + \bar{c}(x^*_U - x) < +\infty. \quad (6.11)$$

For $x \geq x^\alpha$,

$$v_\alpha(x) \leq v_\alpha^\sigma(x) = \mathbb{E}\Bigg[\sum_{t=1}^{\mathbf{N}(x-x^\alpha)+1} \alpha^{t-1} h(x_t) + \alpha^{\mathbf{N}(x-x^\alpha)+1} \times [K + \bar{c}(x^\alpha - x_{\mathbf{N}(x-x^\alpha)+1}) + v_\alpha(x^\alpha)]\Bigg]. \quad (6.12)$$

Let $E(x) := h(x) + E_{x-x_L^*}(x) < \infty$, where the function $E_y(x)$ is defined in (6.4) and its finiteness is stated in Lemma 6.2. Since the non-negative function h is convex, then for $x_t = x - \mathbf{S}_t$, $t = 1, \ldots, \mathbf{N}(x - x_L^*) + 1$,

$$\begin{aligned} 0 &\leq h(x_t) \\ &\leq \max\left\{h(x - \mathbf{S}_{\mathbf{N}(x-x_L^*)+1}), h(x)\right\} \\ &\leq h(x) + h(x - \mathbf{S}_{\mathbf{N}(x-x_L^*)+1}) \end{aligned} \quad (6.13)$$

and

$$\mathbb{E}h(x_t) \leq h(x) + \mathbb{E}h(x - \mathbf{S}_{\mathbf{N}(x-x_L^*)+1}) = E(x). \quad (6.14)$$

Observe that

$$\mathbb{E}\left[\sum_{t=1}^{\mathbf{N}(x-x^\alpha)+1} \alpha^{t-1} h(x_t)\right] \leq \mathbb{E}\left[\sum_{t=1}^{\mathbf{N}(x-x_L^*)+1} h(x_t)\right] \leq E(x)(1 + \mathbb{E}\mathbf{N}(x - x_L^*)), \quad (6.15)$$

where the first inequality follows from $x_L^* \leq x^\alpha$ and $\alpha \in [0,1)$; the second inequality follows from $x_L^* \leq x^\alpha$, (6.13),(6.14), and Wald's identity. In addition,

$$\begin{aligned} &\mathbb{E}[\alpha^{\mathbf{N}(x-x^\alpha)+1}[K + \bar{c}(x^\alpha - x_{\mathbf{N}(x-x^\alpha)+1}) + v_\alpha(x^\alpha)] \\ &\quad \leq K + \bar{c}(x^\alpha - x + \mathbb{E}\mathbf{S}_{\mathbf{N}(x-x^\alpha)+1}) + m_\alpha \\ &\quad \leq K + \bar{c}(1 + \mathbb{E}\mathbf{N}(x - x_L^*))\mathbb{E}D + m_\alpha, \end{aligned} \quad (6.16)$$

where the first inequality follows from $\alpha \in [0,1)$, $x_t = x - \mathbf{S}_t$, and $v_\alpha(x^\alpha) = m_\alpha$; the second inequality follows from $x \geq x^\alpha \geq x_L^*$ and Wald's identity. Formulae (6.12), (6.15), and (6.16) imply that for $x \geq x^\alpha$

$$v_\alpha(x) - m_\alpha \leq K + (E(x) + \bar{c}\mathbb{E}D)(1 + \mathbb{E}\mathbf{N}(x - x_L^*)) < +\infty. \quad (6.17)$$

Inequalities (6.11) and (6.17) imply that Assumption **B** holds. □

Consider a non-negative, real-valued, lower semi-continuous terminal value **F**. In view of Corollaries 3.5, 6.1, Theorems 3.4, 4.1, and Proposition 6.3, Eqs. (3.9), (3.8) and, for the case $P(D > 0) > 0$, inequality (4.2) can be rewritten as

$$v_{t+1,\mathbf{F},\alpha}(x) = \min\left\{\min_{a\geq 0}[K + G_{t,\mathbf{F},\alpha}(x+a)], G_{t,\mathbf{F},\alpha}(x)\right\} - \bar{c}x, \quad (6.18)$$

$$v_\alpha(x) = \min\left\{\min_{a\geq 0}[K + G_\alpha(x+a)], G_\alpha(x)\right\} - \bar{c}x, \quad (6.19)$$

$$w + u(x) \geq \min\left\{\min_{a\geq 0}[K + H(x+a)], H(x)\right\} - \bar{c}x, \quad (6.20)$$

where $t = 0, 1, \ldots$ and $w := w(x) = w^* = \underline{w} = \bar{w}$, $x \in \mathbb{X}$, and the last three equalities hold in view of (4.6), and

$$G_{t,\mathbf{F},\alpha}(x) := \bar{c}x + \mathbb{E}h(x - D) + \alpha\mathbb{E}v_{t,\mathbf{F},\alpha}(x - D), \quad (6.21)$$

$$G_\alpha(x) := \bar{c}x + \mathbb{E}h(x - D) + \alpha\mathbb{E}v_\alpha(x - D), \quad (6.22)$$

$$H(x) := \bar{c}x + \mathbb{E}h(x - D) + \mathbb{E}u(x - D). \quad (6.23)$$

We explain the correctness of (6.18). The explanations for (6.19) and (6.20) are similar. For this particular problem, optimality Eq. (3.9) is equivalent to $v_{t+1,\mathbf{F},\alpha}(x) = \min\left\{\inf_{a>0}[K + G_{t,\mathbf{F},\alpha}(x+a)], G_{t,\mathbf{F},\alpha}(x)\right\} - \bar{c}x$, and the internal infimum can be replaced with the minimum in (6.18) because of the following two arguments:

(i) the function $K + G_{t,\mathbf{F},\alpha}(y)$ is lower semi-continuous on $[x, \infty)$ and $G_{t,\mathbf{F},\alpha}(y) \to \infty$ as $y \to \infty$, and
(ii) $K + G_{t,\mathbf{F},\alpha}(x) \geq G_{t,\mathbf{F},\alpha}(x)$ since $K \geq 0$.

We remark that, in general, while Eqs. (6.18) and (6.19) are the necessary and sufficient conditions of optimality, inequality (6.20) is the sufficient condition of optimality. Also, if $P(D = 0) = 1$, then inequality (6.20) does not hold because $w(x)$ is not a constant function, as explained before Proposition 6.3.

COROLLARY 6.4: Let $\alpha \in [0,1)$. The following statements hold:

(a) the function $G_\alpha(\cdot)$ is lower semi-continuous,
(b) if **F** is non-negative, real-valued, and lower semi-continuous, then the functions $\left\{G_{t,\mathbf{F},\alpha}(\cdot)\right\}_{t=0,1,\ldots}$ are lower semi-continuous, and
(c) if $P(D > 0) > 0$, then H is lower semi-continuous.

PROOF: In view of (6.21)–(6.23), each of these functions is a sum of several functions, two of which are continuous and the third one is lower semi-continuous, as follows from Corollary 6.1 and from Proposition 6.3. □

LEMMA 6.5: Let $\alpha \in [0,1)$. Then $G_\alpha(x) < +\infty$ for all $x \in \mathbb{X}$. Furthermore, if $0 \le \mathbf{F}(x) \le v_\alpha(x)$ for all $x \in \mathbb{X}$, then $G_{\alpha,\mathbf{F},t}(x) < +\infty$ for all $x \in \mathbb{X}$ and for all $t = 0, 1, \ldots$.

PROOF: Since $G_{\alpha,\mathbf{F},t} \le G_\alpha$, in view of (6.22), 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 $x \le 0$

$$v_\alpha(x) \le v_\alpha^\phi(x) \le K - \bar{c}x + \frac{\alpha(K + \bar{c}\mathbb{E}D + \mathbb{E}h(-D))}{1-\alpha}. \tag{6.24}$$

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

$$\begin{aligned} v_\alpha(x) &\le v_\alpha^\phi(x) \\ &= \mathbb{E}\left[\sum_{t=1}^{\mathbf{N}(x)+1} \alpha^t h(x-\mathbf{S}_t) + \alpha^{\mathbf{N}(x)+1} v_\alpha^\phi(x-\mathbf{S}_{\mathbf{N}(x)+1})\right] \\ &\le h(x)\mathbb{E}\mathbf{N}(x) + \mathbb{E}h(x-\mathbf{S}_{\mathbf{N}(x)+1}) \\ &\quad + K - \bar{c}(x - \mathbb{E}\mathbf{S}_{\mathbf{N}(x)+1}) + B_\alpha < +\infty, \end{aligned}$$

where the second inequality follows from the facts that $\alpha^t < 1$ for $t \ge 1$, $0 \le h(x-\mathbf{S}_t) \le h(x)$ for $t = 1, \ldots, \mathbf{N}(x)$, and (6.24). The second inequality holds because $\mathbb{E}\mathbf{N}(x) < \infty$, Lemma 6.2, and (6.3). Let $\alpha \in (0,1)$. Since $v_\alpha^\phi(x) = \mathbb{E}h(x-D) + \alpha\mathbb{E}v_\alpha^\phi(x-D) < +\infty$, then $\mathbb{E}v_\alpha(x-D) \le \mathbb{E}v_\alpha^\phi(x-D) < +\infty$. In addition, $v_0^\phi(x-D) \le v_\alpha^\phi(x-D) < +\infty$. The result follows. □

Recall the following classic definition.

DEFINITION 6.6: For a real number $K \ge 0$, a function $f : \mathbb{R} \to \overline{\mathbb{R}}$ is called K-convex, if for each $x \le y$ and for each $\lambda \in (0,1)$,

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

The following lemma summarizes some properties of K-convex functions.

LEMMA 6.7: The following statements hold for a K-convex function $g : \mathbb{R} \to \mathbb{R}$:

1. If the function g is measurable and D is a random variable, then $\mathbb{E}g(x-D)$ is also K-convex provided $\mathbb{E}|g(x-D)| < \infty$ for all $x \in \mathbb{R}$.
2. Suppose g is inf-compact (that is, lower semi-continuous and $g(x) \to \infty$ as $|x| \to \infty$). Let

$$S \in \text{argmin}_{x\in\mathbb{R}} \{g(x)\}, \tag{6.25}$$

$$s = \inf\{x \le S \mid g(x) \le K + g(S)\}. \tag{6.26}$$

Then

(a) $g(S) + K < g(x)$ for all $x < s$,
(b) $g(x)$ is decreasing on $(-\infty, s]$ and, therefore, $g(s) < g(x)$ for all $x < s$,
(c) $g(x) \le g(z) + K$ for all x such that $s \le x \le z$,

PROOF: See Bertsekas [2, Lemma 4.2.1] and Simchi-Levi et al. [33, Lemma 8.3.2] for the case of a continuous function g. The proofs there with minor adjustments cover the case when g satisfies the measurability and continuity properties stated in the lemma. □

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 (6.25) and (6.26) with g replaced by G_α, Statement 2(a) of Lemma 6.7, along with the optimality Eq. 6.19, imply that it is optimal to order up to S_α when $x < s_\alpha$. Statement 2(c) of Lemma 6.7 imply that it is optimal not to order when $x \ge s_\alpha$. Our next goal is the established these properties of the function G_α and of some relevant functions.

For a fixed ordering cost $K \ge 0$ we sometimes write v_α^K, $v_{t,\alpha}^K$, $v_{t,\mathbf{F},\alpha}^K$, G_α^K, and $G_{t,\mathbf{F},\alpha}^K$, instead of v_α, $v_{t,\alpha}$, $v_{t,\mathbf{F},\alpha}$, G_α, and $G_{t,\mathbf{F},\alpha}$, respectively. Consider the terminal value $\mathbf{F}(x) = v_\alpha^0(x)$, $x \in \mathbb{X}$. According to Theorem 3.4(viii) and Corollary 3.5(vi), the functions v_α, v_α^0, $v_{t,\alpha}$, and $v_{t,v_\alpha^0,\alpha}$, $t = 1, 2, \ldots$, are inf-compact.

LEMMA 6.8: The following statements hold:

(i) the functions v_α and $v_{t,v_\alpha^0,\alpha}$, $t = 0, 1, \ldots$, are inf-compact;
(ii) the functions G_α and $G_{t,v_\alpha^0,\alpha}$, $t = 0, 1, \ldots$, are lower semi-continuous, and

$$\lim_{x\to+\infty} G_\alpha(x) = \lim_{x\to+\infty} G_{t,v_\alpha^0,\alpha}(x) = +\infty, \quad t = 0, 1, \ldots;$$

(iii) there exists $\alpha^* \in [0,1)$ such that $G_\alpha^0(x) \to \infty$ as $x \to -\infty$ for all $\alpha \in [\alpha^*, 1)$;
(iv) for $\alpha \in [\alpha^*, 1)$, where α^* is the constant $\alpha^* \in [0,1)$ whose existence is stated in Statement (iii), the functions $G_\alpha(x)$ and $G_{t,v_\alpha^0,\alpha}(x)$, $t = 0, 1, \ldots$, are K-convex and tend to $+\infty$ as $x \to -\infty$, and therefore, in view of Statement (ii), these functions are inf-compact. Furthermore, the functions v_α and $v_{t,v_\alpha^0,\alpha}(x)$, $t = 0, 1, \ldots$, are K-convex.

PROOF: In view of Corollary 6.1, Statement (i) follows from Theorem 3.4(viii) and Corollary 3.5(vi). Statement

(ii) follows from Statement (i), non-negativity of costs, and definitions (6.21) and (6.22).

To prove Statement (iii) note that it is well-known that the function G^0_α is convex, where $\alpha \in [0,1)$. Indeed, the function $v^0_{0,\alpha} = 0$ is convex. For $K = 0$, Eqs. (6.18), (6.21) and induction based on Heyman and Sobel [23, Proposition B-4] imply that the functions $v^0_{t,\alpha}$, $t = 1, 2, \ldots$, are convex. Convergence of value iterations, stated in Theorem 3.4(i), implies the convexity of the functions v^0_α. The convexity of G^0_α follows from (6.22).

We show by contradiction that there exists $\alpha^* \in [0,1)$ such that G^0_α 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 $K = 0$, (6.19) can be written as

$$v^0_\alpha(x) = \inf_{a\geq 0}\left\{G^0_\alpha(x+a)\right\} - \bar{c}x. \qquad (6.27)$$

If a constant M_α does not exist for some $\alpha \in (0,1)$, then the convexity of $G^0_\alpha(x)$ implies 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, if x is small enough that the convex function $h(x)$ is decreasing at x, then $w^\psi(x) \geq \; Eh(x - D) > h(x) \to +\infty$ as $x \to -\infty$, but, in view of Theorem 4.1, $w(x)$ is a finite constant. This contradiction implies that for $\alpha \in [\alpha^*, 1)$ the functions G^0_α decreases when $x \in (-\infty, M_\alpha]$, where $M_\alpha > -\infty$. The convexity of G^0_α implies that $G^0_\alpha(x) \to \infty$ as $x \to -\infty$.

Let us prove Statement (iv). The convergence of the functions to $+\infty$, as $x \to -\infty$, follows from Statement (iii) and the inequalities $G^K_\alpha(x) \geq G^0_\alpha(x)$ and $G^K_{t,v^0_\alpha,\alpha}(x) \geq G^0_\alpha(x)$, which hold for all $x \in \mathbb{X}$. Indeed, the first inequality follows from $v^K_\alpha(x) \geq v^0_\alpha(x)$, $x \in \mathbb{X}$, and (6.22). The second inequality follows from $v^K_{t,v^0_\alpha,\alpha}(x) \geq v^0_{t,v^0_\alpha,\alpha}(x) = v^0_\alpha(x)$, $x \in \mathbb{X}$, and (6.21).

Now let $\alpha \in [\alpha^*, 1)$. As explained in the proof of (iii), the function G^0_α is convex and therefore it is K-convex. Formulae (6.18), (6.21), Heyman and Sobel [23, Lemma 7–2, p. 312], and induction arguments imply that the functions $G_{t,v^0_\alpha,\alpha}$ and $v_{t+1,v^0_\alpha,\alpha}$, $t = 1, 2, \ldots$ are K-convex. In addition, $v_{t,v^0_\alpha,\alpha}(x) \uparrow v_\alpha(x)$ as $t \to \infty$ in view of Corollary 3.5(v) and since all the costs are non-negative. Formulae (6.21), (6.22) and the monotone convergence theorem imply that $G_{t,v^0_\alpha,\alpha}(x) \uparrow G_\alpha(x)$ as $t \to \infty$. Thus, the functions v_α and G_α are K-convex. □

DEFINITION 6.9: Let s_t and S_t be real numbers such that $s_t \leq S_t$, $t = 0, 1, \ldots$. Suppose x_t denotes the current inventory level at decision epoch t. A policy is called an (s_t, S_t) policy at step t if it orders up to the level S_t if $x_t < s_t$ and does not order when $x_t \geq s_t$. A Markov policy is called an (s_t, S_t) policy if it is an (s_t, S_t) policy at all steps $t = 0, 1, \ldots$. A policy is called an (s, S) policy if it is stationary and it is an (s, S) policy at all steps $t = 0, 1, \ldots$.

The following theorem is the main result of this section.

THEOREM 6.10: Consider $\alpha^* \in [0,1)$ whose existence is stated in Lemma 6.8. The following statements hold for the inventory control problem.

(i) For $\alpha \in [\alpha^*, 1)$ and $t = 0, 1, \ldots$, define $g(x) := G_{t,v^0_\alpha,\alpha}(x)$, $x \in \mathbb{R}$. Consider real numbers $S^*_{t,\alpha}$ satisfying (6.25) and $s^*_{t,\alpha}$ defined in (6.26). For each $N = 1, 2, \ldots$, the (s_t, S_t) policy with $s_t = s^*_{N-t-1,\alpha}$ and $S_t = S^*_{N-t-1,\alpha}$, $t = 0, 1, \ldots, N-1$, is optimal for the N-horizon problem with the terminal values $\mathbf{F}(x) = v^0_\alpha(x)$, $x \in \mathbb{R}$.

(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 (6.25) and s_α defined in (6.26). The (s_α, S_α) policy is optimal for the discount factor α. Furthermore, a sequence of pairs $\left\{(s^*_{t,\alpha}, S^*_{t,\alpha})\right\}_{t=0,1,\ldots}$ is bounded, where $s^*_{t,\alpha}$ and $S^*_{t,\alpha}$ are described in Statement (i), $t = 0, 1, \ldots$. If (s^*_α, S^*_α) is a limit point of this sequence, then the (s^*_α, S^*_α) policy is optimal for the infinite-horizon problem with 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 $\alpha_n \uparrow 1$, $n = 1, 2, \ldots$, with $\alpha_1 \geq \alpha^*$. Every sequence $\left\{(s'_{\alpha_n}, S'_{\alpha_n}), n \geq 1\right\}$ is bounded and each its limit point (s, S) defines an average-cost optimal (s, S) policy. Furthermore, if $P(D > 0) > 0$, this policy satisfies the optimality inequality (6.20) with $u = \tilde{u}$, where the function $\tilde{u}$ is defined in (4.5) for an arbitrary subsequence $\left\{\alpha_{n_k}\right\}_{k=1,2,\ldots}$ of $\{\alpha_n, n \geq 1\}$ satisfying $(s, S) = \lim_{k\to\infty}(s'_{\alpha_{n_k}}, S'_{\alpha_{n_k}})$.

PROOF: To prove Statements (i) and (ii), let $\alpha \in [\alpha^*, 1)$. In view of Lemma 6.8(iv), the functions G_α and $G_{t,v^0_\alpha,\alpha}$, $t = 0, 1, \ldots$, are K-convex and inf-compact. The optimality of (s_t, S_t) policies and (s, S) policies with $s = s_\alpha$ and $S = S_\alpha$ stated in (i) and (ii) follows from optimality Eqs. (6.18), (6.19), Lemma 6.7 with $g = G_{N,v^0_\alpha,\alpha}$ and $g = G_\alpha$ respectively, and Theorem 3.4.

Consider now the remaining claims in (ii). Since $G^0_\alpha(x) \leq G_{t,v^0_\alpha,\alpha}(x) \leq G_{t+1,v^0_\alpha,\alpha}(x) \leq G_\alpha(x)$, $x \in \mathbb{R}$, the points $s^*_{t,\alpha}$ and $S^*_{t,\alpha}$ belong to the compact set $\left\{x \in \mathbb{R} : G^0_\alpha(x) \leq K + \min_{x\in\mathbb{R}} G_\alpha(x)\right\}$. Therefore, the sequence $\left\{(s^*_{t,\alpha}, S^*_{t,\alpha})\right\}_{t=0,1,\ldots}$ is bounded and has a limit point (s^*_α, S^*_α). The function $\mathbf{F}(x) = v^0_\alpha(x)$ satisfies inequalities in

(3.11), and therefore the assumptions of Theorem 3.6 hold. Theorem 3.6 implies that, for the infinite-horizon problem with the discount factor α, the following decisions are optimal for the corresponding states: 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 implies that $G_\alpha(x) \le K + G_\alpha(S_\alpha^*)$ for $x \in (s_\alpha^*, S^*\alpha)$. Lower semi-continuity of $G_\alpha(x)$ implies that $G_\alpha(s_\alpha^*) \le K + G_\alpha(S_\alpha^*)$. 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.

It remains to prove Statement (iii). Let $P(D > 0) > 0$. We start with the proof that the sequence $\{(s'_{\alpha_n}, S'_{\alpha_n})\}_{n=1,2,\ldots}$ is bounded. First, we prove that the sequence $\{s'_{\alpha_n}, n \ge 1\}$ is bounded below. If this is not true, then $\lim_{k\to\infty} s'_{\alpha_{n_k}} = -\infty$ for some $n_k \to \infty$ as $k \to \infty$. This means that for each $x \in \mathbb{R}$ there is a natural number $k(x)$ such that $x > s'_{\alpha_{n_k}}$ for $k \ge k(x)$. Therefore, $0 \in \mathbb{A}_{\alpha_{n_k}}(y)$, $k \ge k(x)$, for all $y \ge x$. Corollary 4.4(ii) implies that the action $0 \in A_{\tilde{u}}^*(y)$ for all $y > x$, where $\tilde{u}$ is defined in (4.5) for the sequence of discount factors $\{\alpha_{n_k}, k \ge 1\}$. Since $x \in \mathbb{R}$ is arbitrary, $0 \in A_{\tilde{u}}^*(y)$ for all $y \in \mathbb{R}$. This means that the policy ψ, that never orders inventory, is optimal for average costs per unit time. However,

$$w^\psi(x) \ge \mathbb{E}h(x - \mathbf{S}_n) \ge h(x - n\mathbb{E}D).$$

Letting $n \to \infty$ on the right hand side yields $w^\psi(x) = +\infty$ for all $x \in \mathbb{R}$. In view of Assumption **G** that holds for the inventory control problem, $w^\psi(x) < +\infty$ for some $x \in \mathbb{R}$. Thus the sequence $\{s'_{\alpha_n}, n \ge 1\}$ is bounded.

Second, we prove that the sequence $\{S'_{\alpha_n}, n \ge 1\}$ is also bounded. Let $x \in \mathbb{R}$ be a lower bound for $\{s'_{\alpha_n}, n \ge 1\}$. Thus, $a^{(n)} := (S'_{\alpha_n} - x) \in A_{\alpha_n}(x)$. In view of Theorem 4.5, the sequence $\{a^{(n)}, n \ge 1\}$ is bounded. This implies that the sequence $\{S'_{\alpha_n}, n \ge 1\}$ is bounded as well.

Consider a subsequence $\alpha_{n_k} \uparrow 1$ such that $(s'_{\alpha_{n_k}}, S'_{\alpha_{n_k}}) \to (s', S')$ as $k \to \infty$. 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 (4.5) for the sequence of discount factors $\{\alpha_{n_k}, k \ge 1\}$. The last step is to prove that $0 \in A_{\tilde{u}}^*(s')$. To do this, consider a subsequence $\{\alpha_n^*, n \ge 1\}$ such that $\alpha_n^* \to 1$ of the sequence $\{\alpha_{n_k}, k \ge 1\}$ and a sequence $\{x^{(n)}, n \ge 1\}$ with $x^{(n)} \to s'$ 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 $x^{(\ell_k)} \ge s'_{\alpha^*_{\ell_k}}$ 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 $0 \in A_{\tilde{u}^*}^*(s')$, where the function $\tilde{u}^*$ is defined in (4.5) for the sequence of discount factors $\{\alpha^*_{n_k}\}_{k=1,2,\ldots}$. Observe that $\tilde{u}^*(s') = \tilde{u}(s')$ and $\tilde{u}^*(x) \ge \tilde{u}(x)$ for all $x \in \mathbb{R}$. This implies $A_{\tilde{u}^*}^*(s') \subseteq A_{\tilde{u}}^*(s')$. Thus $0 \in A_{\tilde{u}}^*(s')$.

Second, consider the complimentary case, when there exists a number N such that $x^{(n)} < s'_{\alpha_n^*}$ for $n \ge N$. Let $n \ge N$. In view of Statement 2(b) of Lemma 6.7, $G_{\alpha_n^*}(x^{(n)}) \ge G_{\alpha_n^*}(s'_{\alpha_n^*})$. Therefore,

$$\begin{aligned} u_{\alpha_n^*}(x^{(n)}) &= v_{\alpha_n^*}(x^{(n)}) - m_{\alpha_n^*} = K + G_{\alpha_n^*}(S'_{\alpha_n^*}) - \bar{c}x^{(n)} - m_{\alpha_n^*} \\ &\ge G_{\alpha_n^*}(s'_{\alpha_n^*}) - \bar{c}x^{(n)} - m_{\alpha_n^*} \\ &\ge v_{\alpha_n^*}(s'_{\alpha_n^*}) + \bar{c}s'_{\alpha_n^*} - \bar{c}x^{(n)} - m_{\alpha_n^*} \\ &= u_{\alpha_n^*}(s'_{\alpha_n^*}) + \bar{c}(s'_{\alpha_n^*} - x^{(n)}), \end{aligned}$$

where the first and the last equalities follow from the definition of the functions u_α, the second equality follows from (6.19) and from the optimality of the $(s'_{\alpha_n^*}, S'_{\alpha_n^*})$ policies for discount factors α_n^*, the first inequality follows from Statement 2(c) of Lemma 6.7, and the last inequality follows from (6.19). Since $s'_{\alpha_n^*} \to s'$ and $x^{(n)} \to s'$,

$$\tilde{u}(s') = \lim_{n\to\infty} u_{\alpha_n^*}(x^{(n)}) = \lim_{n\to\infty} u_{\alpha_n^*}(s'_{\alpha_n^*}).$$

Moreover, since $0 \in A_{\alpha_n^*}(s'_{\alpha_n^*})$ for all $n = 1, 2, \ldots$, Theorem 4.3(ii) implies that $0 \in A_{\tilde{u}}^*(s')$. Thus, the (s', S') policy is average-cost optimal.

Now let $D = 0$ almost surely. As explained in the paragraph preceding Proposition 6.3, the (0, 0) policy ϕ is average-cost optimal. Let us prove that

$$\lim_{\alpha\uparrow 1} s_\alpha = \lim_{\alpha\uparrow 1} S_\alpha = 0. \qquad (6.28)$$

Let $\alpha \in (0, 1)$. Consider an arbitrary policy σ. Since $v^\sigma(x) \ge \frac{h(x)}{1-\alpha} = v_\alpha^\phi(x)$, when $x \ge 0$, then $v_\alpha(x) = \frac{h(x)}{1-\alpha}$ for all $x \ge 0$. This formula and (6.22) imply $G_\alpha(x) = \bar{c}x + h(x)/(1-\alpha)$ for $x \ge 0$. Thus, the function $G_\alpha(x)$ is increasing, when $x \in [0, \infty)$. This implies $S_\alpha \le 0$. Since $s_\alpha \le S_\alpha$, then $s^* = \liminf_{\alpha\uparrow 1} s_\alpha \le 0$. To complete the proof of (6.28), we need to show that $s^* = 0$. Indeed, let us assume that $s^* < 0$. Fix an arbitrary $x \in (s^*, 0)$. Then there exists a sequence $\alpha_n \uparrow 0$ such that $s_{\alpha_n} \to s^*$ as $n \to \infty$ and $s_{\alpha_n} < x$, $n = 1, 2, \ldots$. The $(s_{\alpha_n}, S_{\alpha_n})$ policy ϕ^n is optimal for the discount factor α_n, and this policy does not order at the state x, $n = 1, 2, \ldots$. Therefore $v_{\alpha_n}^{\phi^n}(x) = h(x)/(1-\alpha_n) \to +\infty$ as $n \to \infty$. However, $v_{\alpha_n}^\phi(x) = K - \bar{c}x$. This implies that the $(s_{\alpha_n}, S_{\alpha_n})$ policy ϕ^n cannot be optimal for a discount factor $\alpha_n > (K - \bar{c}x)/(K - \bar{c}x - h(x))$. □

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

LEMMA 6.11: Suppose there exist $z, y \in \mathbb{R}$ such that $z < y$ and

$$\frac{h(y) - h(z)}{y - z} < -\bar{c}. \qquad (6.29)$$

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: Observe that the assumption in Lemma 6.11 is equivalent to the existence of $z, y \in \mathbb{R}$ such that $z < y$ and

$$\frac{\mathbb{E}[h(y - D) - h(z - D)]}{y - z} < -\bar{c}. \tag{6.30}$$

Indeed, since h is convex, $h(y-d) - h(z-d) \leq h(y) - h(z)$, and (6.29) implies (6.30). Also, (6.30) implies that for some $d \geq 0$ inequality (6.29) holds for $y := y - d$ and $z := z - d$.

According to (6.21), $G_{N,\alpha}(x) \to \infty$ as $x \to \infty$ for all $N = 0, 1, \ldots$. We show that the result continues to hold when $x \to -\infty$. Suppose $z < y$ satisfy (6.30). Inequality (6.30) can be rewritten as

$$\bar{c}y + \mathbb{E}h(y - D) < \bar{c}z + \mathbb{E}h(z - D).$$

Thus, $G_{0,\alpha}(z) > G_{0,\alpha}(y)$. Since $G_{0,\alpha}$ is convex, then $G_{0,\alpha}(x) \to \infty$ as $x \to -\infty$. According to (6.21),

$$G_{N,\alpha}(x) = G_{0,\alpha}(x) + \alpha\mathbb{E}v_{N,\alpha}(x - D) \geq G_{0,\alpha}(x), \quad N = 1, 2, \ldots,$$

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

where$G_{0,\alpha}(x) \to +\infty$as$x \to -\infty$. □

THEOREM 6.12: Under the condition stated in Lemma 6.11, the following statements hold for each discount factor $\alpha \in [0, 1)$:

(i) For $t = 0, 1, \ldots$ consider real numbers $S_{t,\alpha}$ satisfying (6.25) and $s_{t,\alpha}$ defined in (6.26) with $g(x) = G_{t,\alpha}(x)$, $x \in \mathbb{R}$. Then for every $N = 1, 2, \ldots$ the (s_t, S_t) policy with $s_t = s_{N-t-1,\alpha}$ and $S_t = S_{N-t-1,\alpha}$, $t = 0, 1, \ldots, N - 1$, is optimal for the N-horizon problem with the zero terminal values.

(ii) Consider real numbers S_α satisfying (6.25) and s_α defined in (6.26) for $g(x) := G_\alpha(x)$, $x \in \mathbb{R}$. Then the (s_α, S_α) policy is optimal for the infinite-horizon problem with the discount factor α. Furthermore, a sequence of pairs $\left\{(s_{t,\alpha}, S_{t,\alpha})\right\}_{t=0,1,\ldots}$ considered in statement (i) is bounded, and, if (s^*_α, S^*_α) is a limit point of this sequence, then the (s^*_α, S^*_α) policy is optimal for the infinite-horizon problem with the discount factor α.

PROOF: Observe that $G_{0,\alpha}(x) = \bar{c}x + \mathbb{E}h(x - D)$. This function is convex and, in view of Lemma 6.11, $G_{0,\alpha}(x) \to \infty$ as $|x| \to \infty$. The rest of the proof coincides with the proof of Theorem 6.10 with the functions $G_{t,v^0_\alpha,\alpha}$ replaced with the functions $G_{t,\alpha}$. □

By using the results of this section, Feinberg and Liang [18–20] obtained additional results for the inventory control problem. Feinberg and Liang [19] described the structure of optimal policies for all values of the discount factor $\alpha \geq 0$ for finite-horizon problems and for all values of $\alpha \in [0, 1)$ for infinite-horizon problems. In particular, the smallest possible values of the discount factor α^* mentioned in Theorem 6.10 are computed in Ref. 19. Though the general theory of MDPs implies that the value functions $v_{t,\alpha}(x)$, $G_{t,\alpha}(x)$, $v_\alpha(x)$, and $G_\alpha(x)$ are lower semi-continuous in x, it is proved in Ref. 19 that these functions are continuous. In particular, these continuity properties imply that, for total discounted cost criteria with finite and infinite horizons, the decisions to order up to the level S (S_t) are also optimal at the states s (s_t). Feinberg and Liang [18] proved that for the inventory control problem the average-cost optimality inequality in (6.20) holds in the stronger form of the optimality equation, the convergences $u_\alpha(\cdot) \to u(\cdot)$ and $G_\alpha(\cdot) \to H(\cdot)$ take place, as $\alpha \uparrow 1$, and the functions $u(x)$ and $G(x)$ are K-convex and continuous. Therefore, average-cost optimal (s, S) policies can be derived from the optimality equation, and the decision to place an order up to the level S at the state s is also optimal for the average-cost criterion. Feinberg and Liang [20] strengthened some of the results from Ref. 18 and extended them to the models with costs satisfying quasiconvexity conditions.

REMARK 6.13: This remark comments on the assumptions $\alpha \in [0, 1)$, $K \geq 0$, and $c > 0$. All the results of this article stated for the finite horizon hold with the same proofs for arbitrary $\alpha \geq 0$; see Feinberg and Liang [19] for detail. If $K = 0$, then it is well-known that it is possible to set s = S and $s_t = S_t$ for the corresponding optimal (s, S) policies, see e.g., Heyman and Sobel [23, Proposition 3–1], and such policies are called base stock or S-policies. Indeed, this follows from Lemma 6.8 and (6.18), (6.19) for discounted problems, and then from Theorem 6.10(iii) for problems with average costs per unit time. If $c = 0$, then Assumption **W*** holds. In particular, the function $c(x, a) = K1_{a>0} + \mathbb{E}h(x + a - D)$ is $\mathbb{K}$-inf-compact as a sum of a lower-semicontinuous function and a $\mathbb{K}$-inf-compact function; see Theorem 5.3(1). All the results formulated in the article for a fixed discount factor $\alpha \in [0, 1)$ remain correct for $\bar{c} = 0$. Furthermore, inequality (6.29) holds and therefore the conclusions of Theorem 6.12 hold. However, the function c is not inf-compact when $\bar{c} = 0$. For example, $c(-a, a) = K + \mathbb{E}h(-D) \nrightarrow +\infty$ as $a \to +\infty$. The proof of Assumption **B** in Proposition 6.3 is based on Theorem 4.2, which uses the assumption that the function c is inf-compact. So, for the long-term average-cost criterion, the results of this article do not cover the case $\bar{c} = 0$.

REMARK 6.14: For the inventory control problem, we have considered an MDP with $\mathbb{X} = \mathbb{R}$ and $A(x) = \mathbb{R}^+$ 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 $A(x) = \mathbb{Z}^+$, where $\mathbb{Z}$ is the set of integers and $\mathbb{Z}^+$ is the set of non-negative integers. 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 $\mathbb{X} = \mathbb{Z}$. All of the results of this article also 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.

ACKNOWLEDGMENT

Research of the first author was partially supported by NSF grants CMMI-1335296 and CMMI-1636193. Theorem 6.10, 6.12 of this article and some relevant preceding constructions and definitions are reprinted with some minor differences by permission from Ref. 11, where these theorems are presented without proofs. Copyright 2016, the Institute for Operations Research and the Management Sciences, 5521 Research Park Drive, Suite 200, Catonsville, MD 21228 USA.

REFERENCES

[1] A. Bensoussan, Dynamic programming and inventory control, IOS Press, Amsterdam, 2011.
[2] D.P. Bertsekas, Dynamic programming and optimal control, vol. 1, 2nd ed., Athena Scientific, Belmont, MA, 2000.
[3] D.P. Bertsekas, and S.E. Shreve, Stochastic optimal control: The discrete-time case, Athena Scientific, Belmont, MA, 1996.
[4] D. Beyer, F. Cheng, S.P. Sethi, and M. Taksar, Markovian demand inventory models, Springer, New York, 2010.
[5] D. Beyer, and S.P. Sethi, The classical average-cost inventory models of Iglehart and Veinott-Wagner revisited, J Optim Theory Appl 101 (1999), 523–555.
[6] P. Billingsley, Convergence of probability measures, Wiley, New York, 1968.
[7] D. Blackwell, Discounted dynamic programming, Ann Math Stat 36 (1965), 226–235.
[8] 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.
[9] 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.
[10] E.A. Feinberg, A generalization of 'expectation equals reciprocal of intensity' to non-stationary exponential distributions, J Appl Prob 31 (1994), 262–267.
[11] E.A. Feinberg, "Optimality conditions for inventory control," in: A. Gupta and A. Capponi (Editors) Tutorials in operations research. Optimization challenges in complex, networked, and risky systems, INFORMS, Cantonsville, MD, 2016, pp. 14–44.
[12] E.A. Feinberg, and P.O. Kasyanov, Continuity of minima: local results, Set-Valued Var Anal 23 (2015), 485–499.
[13] E.A. Feinberg, P.O. Kasyanov, and M. Voorneveld, Berge's maximum theorem for noncompact image sets, J Math Anal Appl 413 (2014), 1040–1046.
[14] E.A. Feinberg, P.O. Kasyanov, and N.V. Zadoianchuk, Average cost Markov decision processes with weakly continuous transition probabilities, Math Oper Res 37 (2012), 591–607.
[15] E.A. Feinberg, P.O. Kasyanov, and N.V. Zadoianchuk, Berge's theorem for noncompact image sets, J Math Anal Appl 397 (2013), 255–259.
[16] E.A. Feinberg, P.O. Kasyanov, and N.V. Zadoianchuk, Fatou's lemma for weakly converging probabilities, Theory Probab Appl 58 (2014), 683–689.
[17] E.A. Feinberg, and M.E. Lewis, Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem, Math Oper Res 32 (2007), 769–783.
[18] E.A. Feinberg, and Y. Liang, On the optimality equation for average cost Markov decision processes and its validity for inventory control, Ann Oper Res (2017). DOI 10.1007/s10479-017-2561-9
[19] E.A. Feinberg, and Y. Liang, Structure of optimal policies to periodic-review inventory models with convex costs and backorders for all values of discount factors, Ann Oper Res (2017). DOI 10.1007/s10479-017-2548-6
[20] E.A. Feinberg, and Y. Liang, Stochastic setup-cost inventory model with backorders and quasiconvex cost functions, Preprint ArXiv 1705.06814 (2017).
[21] A. Gut, Stopped random walks, Springer, New York, 1988.
[22] O. Hernández-Lerma, and J. B. Lasserre, Discrete-time Markov control processes: Basic optimality criteria, Springer, New York, 1996.
[23] D.P. Heyman, and M.J. Sobel, Stochastic models in operations research, vol. II, McGraw-Hill, New York, 1984.
[24] W.T. Huh, G. Janakiraman, and M. Nagarajan, Average cost single-stage inventory models: An analysis using a vanishing discount approach, Oper Res 59 (2011), 143–155.
[25] D. L. Iglehart, "Dynamic programming and stationary analysis of inventory problems," in H. Scarf, D. Gilford, and M. Shelly (Editors) Multistage inventory models and techniques, Stanford University Press, Stanford, CA, 1963, pp. 1–31.
[26] F. Luque-Vasques, and O. Hernández-Lerma, A counterexample on the semicontinuity lemma, Proc Am Math Soc 123 (1995), 3175–3176.
[27] E. Porteus. Foundations of stochastic inventory theory, Stanford University Press, Stanford, CA, 2002.
[28] S.I. Resnick, Adventures in stochastic processes, Birkhäuser, Boston, 1992.
[29] H. Scarf, "The optimality of (s, S) policies in the dynamic inventory problem," in: K. Arrow, S. Karlin, and P. Suppes (Editors) Mathematical methods in the social sciences, Stanford University Press, Stanford, CA, 1960, pp. 196–202.
[30] 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.
[31] M. Schäl, Average optimality in dynamic programming with general state space, Math Oper Res 18 (1993), 163–172.
[32] L.I. Sennott, Stochastic dynamic programming and the control of queueing systems, Wiley, New York, 1999.
[33] D. Simchi-Levi, X. Chen, and J. Bramel. The logic of logistics: Theory, algorithms, and applications for logistics management, 2nd ed., Springer, New York, 2005.
[34] R. Strauch, Negative dynamic programming, Ann Math Stat 37 (1966), 871–890.

[35] 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.
[36] A.F. Veinott, Jr., The status of mathematical inventory theory, Manag Sci 12 (1966), 745–777.
[37] A.F. Veinott, Jr., and H.M. Wagner, Computing optimal (s, S) policies, Manag Sci 11 (1965), 525–552.
[38] E. Zabel, A note on the optimality of (S, s) policies in inventory theory, Manag Sci 9 (1962), 123–125.
[39] Y. Zheng, A simple proof for the optimality of (s,S) policies in infinite horizon inventory systems, J Appl Prob 28 (1991), 802–810.
[40] P. H. Zipkin, Foundations of inventory management, McGraw-Hill, New York, 2000.