INFORMS Tutorials in Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Optimality Conditions for Inventory Control

Eugene A. Feinberg

TUTORIALS IN
OPERATIONS RESEARCH
INFORMS 2016

 | ISBN 978-0-9843378-9-7
http://dx.doi.org/10.1287/educ.2016.0145

Optimality Conditions for Inventory Control

Eugene A. Feinberg
Stony Brook University, Stony Brook, New York 11794, eugene.feinberg@stonybrook.edu

Abstract This tutorial describes recently developed general optimality conditions for Markov decision processes that have significant applications to inventory control. In particular, these conditions imply the validity of optimality equations and inequalities. They also imply the convergence of value iteration algorithms. For total discounted-cost problems, only two mild conditions on the continuity of transition probabilities and lower semicontinuity of one-step costs are needed. For average-cost problems, a single additional assumption on the finiteness of relative values is required. The general results are applied to periodic-review inventory control problems with discounted and average-cost criteria without any assumptions on demand distributions. The case of partially observable states is also discussed.

Keywords inventory control; Markov decision process; policy; optimality equation; sufficient conditions

1. Introduction

This tutorial describes recent progress in the theory of Markov decision processes (MDPs) with infinite-state and action sets that have significant applications to inventory control. Two groups of results are covered: (i) optimality conditions for MDPs with total-, discounted-, and average-cost criteria and (ii) optimality conditions for partially observable Markov decision processes (POMDPs) with total and discounted cost criteria.

Inventory control studies and applications are important motivating factors for studies of MDPs. The MDP studies provided important tools for the analysis of inventory control problems. The parallel development of these fields since the beginning of the second half of the 20th century is broadly recognized. For example, the abstract of the historical essay by Girlich and Chikán [37] on the history of inventory control studies states, "... we report how inventory problems have motivated the improvement of mathematical disciplines such as Markovian decision theory and optimal control of stochastic systems to provide a new basis of inventory theory in the second half of our century" (p. 351). However, over a long period of time, there was a gap between the modeling needs for inventory control, which require mathematical methods for the analysis of infinite-state controlled stochastic systems with unbounded action sets and weakly continuous transition probabilities, and available results for the corresponding models for MDPs. This gap was recently closed. Another topic covered in this tutorial is the recent progress in the development of optimality conditions for POMDPs. The literature on MDPs and inventory control is huge, and we do not attempt a comprehensive survey in this tutorial. For the most part, only directly relevant references are provided. The reader may find coverage of these topics in books on MDPs (Bäuerle and Rieder [3], Bertsekas and Shreve [12], Dynkin and Yushkevich [22], Feinberg and Shwartz [29], Heyman and Sobel [42], Hernández-Lerma [38], Hernández-Lerma and Lasserre [40, 41], Puterman [47], Sennott [54]) and on inventory management (Bensoussan [5], Heyman and Sobel [42], Porteus [46], Simchi-Levi et al. [59], Zipkin [68]).

Optimality results for MDPs provide sufficient conditions for the existence of stationary and Markov optimal policies satisfying optimality equations and inequalities, describe continuity properties of the value function, and guarantee the convergence of values and optimal actions when the horizon length tends to infinity or the discount factor tends to 1. These results provide useful tools to analyze specific inventory control problems and to prove the optimality of particular policies. In Section 4 this is illustrated with the classic periodic-review single-product stochastic inventory problem with nonnegative arbitrarily distributed i.i.d. demand. Most of the literature on inventory control is limited to discrete or continuous demand distributions.

Consider the classic periodic-review single-product stochastic inventory problem with backorders. For a finite horizon and continuous demand, Scarf [50] established under some conditions the optimality of (s, S) policies. Zabel [66] indicated some gaps in Scarf [50], corrected them, and mentioned in the last paragraph of his note that the proofs there can be adapted to arbitrary demand distributions. Iglehart [44] and Veinott and Wagner [64] established the optimality of (s, S) policies for the infinite horizon for continuous and discrete demand, respectively. Zheng [67] provided an alternative proof for discrete demand. Beyer and Sethi [13] described and corrected gaps in the proofs in Iglehart [44] and Veinott and Wagner [64]. As shown in Heyman and Sobel [42, Section 7.1], under appropriate conditions (s, S) policies are optimal for a finite-horizon problem with arbitrarily distributed demand. In general, (s, S) policies may not be optimal for finite horizons. For example, for a problem with convex holding costs, the appropriate condition is Assumption **GB** in Section 4. This assumption means that, as the amount of backordered inventory increases, the backordering cost per unit time becomes larger than the value of the backordered inventory. However, as shown in Veinott [63], for discrete demand, (s, S) policies are always optimal for the following three criteria: (i) infinite-horizon average costs per unit time, (ii) infinite-horizon discounted problems with a large discount factor, and (iii) finite-horizon problems with a large discount factor and appropriately selected terminal costs. Chen and Simchi-Levi [17, 18] described optimal policies for coordinating inventory control and pricing for finite and infinite-horizon problems with general demand under a technical assumption. If the price is fixed, the problem in Chen and Simchi-Levi [17, 18] becomes the periodic-review inventory control problem, the technical assumption becomes Assumption **GB**, and the results in Chen and Simchi-Levi [17, 18] imply the optimality of (s, S) policies. For coordinating inventory control and pricing, Huh et al. [43] provided a method for proving the optimality of stationary policies by adding specific assumptions that hold for inventory control to the MDP assumptions.

Using the results from Feinberg et al. [31] on the existence of stationary optimal policies and their properties for MDPs with general state and action sets and with possibly unbounded one-step cost functions, Feinberg and Lewis [26] proved the optimality of (s, S) policies for a general demand distribution for criteria (i)–(iii) mentioned in the previous paragraph. Feinberg and Liang [28] provided a complete description of optimal discounted policies for arbitrary demand. These results cover the results under Assumption **GB** as a special case. Feinberg and Liang [27] proved the validity of the optimality equation for average costs per unit time, while the general results for MDPs (Feinberg et al. [31]) imply only the validity the optimality inequality. The conclusions from Feinberg and Lewis [26], Feinberg and Liang [27, 28] are presented in Section 3.

Studies of MDPs started with investigations of models with finite-state and action sets. Problems with infinite-state and action sets were investigated later. The two classic objective criteria for infinite-horizon problems are (i) minimization of expected total discounted costs and (ii) minimization of long-run average costs per unit time. Problems with average-cost criteria are usually more difficult. In particular, optimality equations can be written for expected total costs under mild conditions, and for total expected discounted costs their analyses lead to the proof of optimality of stationary policies for infinite-horizon problems. For long-run average costs, stationary policies are optimal under stronger conditions than for discounted costs, and proofs of their optimality for average-cost criteria usually use the existence of stationary optimal policies for discounted criteria, when the discount factor increases to 1. This is the so-called vanishing discount factor approach. In particular, this approach can be used to establish the validity of optimality equations (sometimes called canonical equations) and inequalities for MDPs with long-run average costs. Average-cost optimality equations and inequalities imply the existence of optimal stationary policies for long-run average costs. In applications, average-cost optimality equations and inequalities can be written without an explicit use of the vanishing discount factor approach by using general results on the validity of average-cost optimality equations and inequalities for MDPs. However, as mentioned above, this approach is typically used in the theory of MDPs to establish the validity of such equations and inequalities.

Let us discuss optimality conditions for MDPs that are general enough to provide optimality conditions for broad classes of inventory control models. First, the state space should be a possibly unbounded subset of a Euclidean space. This level of generality is covered by Borel state spaces (more precisely, Borel subsets of complete separable metric spaces). Euclidean spaces are examples of Borel spaces, and the general theory of MDPs with Euclidean state spaces is not simpler than for Borel spaces. Similar to subsets of Euclidean spaces, Borel spaces either are finite, are countable, or have the cardinality of the continuum. A reader who is not familiar with the notion of Borel spaces may view all the state and action sets in this tutorial as subsets of Euclidean spaces. Second, the cost functions may be unbounded. More precisely, the cost functions should be inf-compact as a function of two variables: a state and action. For inventory control, inf-compact cost functions can be interpreted as lower semicontinuous functions tending to infinity if either the inventory/backorder or the order size tends to infinity. Cost functions may not be continuous. For example, they are not continuous in models with positive ordering costs. Third, transition probabilities should satisfy the property of continuity in distribution, also known under the name of weak continuity. In particular, transition probabilities are typically weakly continuous for periodic-review stochastic inventory control problems with arbitrary demand distributions; see Feinberg and Lewis [25, Section 4] for details. In particular, it is explained there that the case of setwise continuous transition probabilities, which is often considered in the MDP literature, typically covers only discrete and continuous demand distributions. Fourth, action sets may be unbounded. This corresponds to a potentially unlimited production/supply capacity. For example, if a production/supply capacity is limited, then (s, S) policies may not be optimal; see, e.g., Federgruen and Zipkin [23] and Shaoxiang [56].

For discounted costs, Shapley [57] introduced a zero-sum two-person stochastic game with finite state and action sets. If one of the players has only one action at each state, this model becomes an MDP. This publication is considered the first paper on MDPs. Blackwell [15] developed the theory for discounted costs and Borel state and action sets. In particular, Blackwell [15] studied problems with bounded costs and discovered that the objective functions may not be Borel measurable, and the dynamic programming approach to such problem should deal with more general policies than Borel measurable ones. The appropriate theory is developed in Bertsekas and Shreve [12]. Schäl [51] developed the theory for discounted costs, Borel state spaces, compact action sets, possibly unbounded above cost functions, and continuous transition probabilities. Results for two types of continuity for transition probabilities, setwise and weak continuity, are obtained in Schäl [51]. The results on weak continuity are more important for applications and more complicated. The theory for problems with setwise continuous transition probabilities and possibly noncompact action sets is described in Hernández-Lerma and Lasserre [40]. Feinberg and Lewis [25] provided results for discounted MDPs with weakly continuous transition probabilities, possibly uncountable action sets, and inf-compact cost functions. Feinberg et al. [31] introduced the notion of $\mathbb{K}$-inf-compact functions and obtained more general results than in Feinberg and Lewis [25]; see Theorem 5.1, which is a version of Feinberg et al. [31, Theorem 2] adapted in Feinberg and Lewis [26] to problems with possibly nonzero terminal costs.

For average costs per unit time, Blackwell [14] and Derman [19] established the existence of stationary optimal policies for the case of finite-state and action sets. Derman [20] and

Taylor [62] introduced optimality equations for infinite-state problems with bounded one-step costs. These equations and their version for multichain problems are called canonical in Dynkin and Yushkevich [22]. Sennott [53] introduced optimality conditions that lead to the validity of optimality inequalities whose solutions define stationary optimal policies; see also Sennott [54, 55] and the references therein. Cavazos-Cadena [16] provided an example when optimality inequalities do not hold in the form of equalities. Schäl [52] extended Sennott's results to Borel state spaces, compact action spaces, and weakly and setwise continuous transition probabilities. Hernández-Lerma [39] generalized Schäl's [52] results for setwise continuous transition probabilities to possibly noncompact action sets. Feinberg and Lewis [25] provided sufficient optimality conditions for weakly continuous transition probabilities and possibly noncompact action sets. Feinberg et al. [31] provided results for weakly continuous transition probabilities that generalize the corresponding results in Schäl [52] and Feinberg and Lewis [25]; see Section 5.2.

The second topic covered in this tutorial is optimality conditions for POMDPs and, in particular, for inventory control problems with incomplete information on inventory levels. Research on inventory management with incomplete information was pioneered by Bensoussan et al. [6, 7, 8, 9], where particular problems are studied and the existence of optimal policies and convergence of value iterations are established. In general, for POMDPs there is a well-known reduction, introduced by Aoki [1], Åström [2], Dynkin [21], and Shiryaev [58] of a POMDP to an MDP whose states are posterior probabilities of the states of the original process. This reduction holds for problems with Borel state, action, and observation sets, as well as with measurable transition probabilities (Bertsekas and Shreve [12], Hernández-Lerma [38], Rhenius [48], Yushkevich [65]). However, it provides little information about the existence of optimal policies and the validity of optimality equations.

This reduction is based on Bayes' formula, which has an explicit form only for problems with transition probabilities that either are discrete or have densities. As a result, except the case of finite-state, action, and observation sets, very little was known for a long time about the existence of optimal policies for POMDPs. Therefore, the common approach is to study applications by problem-specific methods. The general approach for verifying optimality conditions for POMDPs, applicable to a large variety of applications, is developed in Feinberg et al. [36], and one of the applications there deals with inventory control. The general optimality results on POMDPs are presented in Section 6, and an application to inventory control is presented in Section 7.

2. Markov Decision Processes: Definitions and Optimality Conditions

An MDP is defined by a tuple $\{\mathbb{X},\mathbb{A},P,c\}$, where $\mathbb{X}$ is the state space, $\mathbb{A}$ is the action space, P is the transition probability, and c is the one-step cost function. 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 $c\colon \mathbb{X}\times\mathbb{A}\to\bar{\mathbb{R}}=\mathbb{R}\cup\{+\infty\}$, and the system moves to the next state according to the probability distribution $P(\cdot\,|\,x,a)$ on $\mathbb{X}$. The function c is assumed to be bounded below and Borel measurable, and P is a transition probability; that is, $P(B\,|\,x,a)$ is a Borel function on $\mathbb{X}\times\mathbb{A}$ for each Borel subset B of $\mathbb{X}$, and $P(\cdot\,|\,x,a)$ is a probability measure on the Borel σ-field of $\mathbb{X}$ for each $(x,a)\in\mathbb{X}\times\mathbb{A}$.

The decision process proceeds as follows: At time $t=0,1,\ldots$, 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 $P(\cdot\,|\,x,a)$; and the process continues. Let $H_t=(\mathbb{X}\times\mathbb{A})^t\times\mathbb{X}$ be the set of histories for $t=0,1,\ldots.$ A (randomized) decision rule at epoch $t=0,1,\ldots$ 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 $h_t=(x_0,a_0,x_1,\ldots,a_{t-1},x_t)$, and

(ii) for any measurable subset $B \subseteq \mathbb{A}$, the function $\pi_t(B \mid \cdot)$ is measurable on H_t. A policy π is a sequence $(\pi_0, \pi_1, \ldots)$ of decision rules. Moreover, π is called nonrandomized if each probability measure $\pi_t(\cdot \mid h_t)$ is concentrated at one point. A nonrandomized policy is called Markov if all decisions depend only on the current state and time. A Markov policy is called stationary if all decisions depend only on the current state. Thus, a Markov policy ϕ is defined by a sequence $\phi_0, \phi_1, \ldots$ of measurable mappings $\phi_t \colon \mathbb{X} \to \mathbb{A}$. A stationary policy ϕ is defined by a measurable mapping $\phi \colon \mathbb{X} \to \mathbb{A}$.

The Ionescu Tulcea theorem (see Bertsekas and Shreve [12, pp. 140–141] or Hernández-Lerma and Lasserre [40, 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 the 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 $N = 0, 1, \ldots$ and a bounded below measurable function $\mathbf{F} \colon \mathbb{X} \to \bar{\mathbb{R}}$ called the terminal value, define the expected total discounted costs:

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

where $v_{0,\mathbf{F},\alpha}^\pi(x) = \mathbf{F}(x)$, $x \in \mathbb{X}$, $\alpha \geq 0$, and, if $N = \infty$, then $\alpha \in [0,1)$. 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 $x \in \mathbb{X}$, (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

$$w^\pi(x) := \limsup_{N \to \infty} \frac{1}{N} \mathbb{E}_x^\pi \sum_{t=0}^{N-1} c(x_t, a_t). \tag{2}$$

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

$$V(x) := \inf_{\pi \in \Pi} V^\pi(x), \tag{3}$$

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 $x \in \mathbb{X}$.

The defined model is too general for the existence of optimal policies. However, optimal policies exist under modest conditions, which typically hold for inventory control applications. The natural conditions for inventory control applications are that the transition probability P is weakly continuous and the cost function c is inf-compact.

The transition probability P is called weakly continuous if for every bounded continuous function $f \colon \mathbb{X} \to \mathbb{R}$ the function

$$\tilde{f}(x,a) := \int_{\mathbb{X}} f(y) P(dy \mid x, a), \quad x \in \mathbb{X}, a \in \mathbb{A},$$

is a continuous function on $\mathbb{X} \times \mathbb{A}$. For an $\bar{\mathbb{R}}$-valued function f, defined on a subset U of a metric space $\mathbb{U}$, consider the level sets

$$\mathcal{D}_f(\lambda; U) := \{ y \in U \colon f(y) \leq \lambda \}, \quad \lambda \in \mathbb{R}, \tag{4}$$

A function f is called *inf-compact* if all the level sets $\mathcal{D}_f(\lambda; U)$ are compact. In particular, the cost function c is defined on $\mathbb{U} := \mathbb{X} \times \mathbb{A}$ and the level sets for c are

$$\mathcal{D}_c(\lambda; \mathbb{X} \times \mathbb{A}) = \{ (x,a) \in \mathbb{X} \times \mathbb{A} \colon c(x,a) \leq \lambda \}, \quad \lambda \in \mathbb{R}. \tag{5}$$

As shown by Feinberg and Lewis [25], for the discounted costs, weak continuity of P and inf-compactness of c imply the existence of optimal policies. However, the condition that the function c is inf-compact can be relaxed by considering the class of $\mathbb{K}$-inf-compact functions.

For two sets U and V, where $U \subset V$, and for two functions f and g defined on V and U, respectively, function g defined on U is called *the restriction of f to U* if $g(x) = f(x)$ when $x \in U$.

Definition 2.1 (cf. Definition A.1 in the Appendix). Let $\tilde{\mathbb{S}}^i$ be metric spaces and $S^i \in \mathcal{B}(\tilde{\mathbb{S}}^i)$, where $S^i \neq \varnothing$, $i = 1, 2$. A function $f\colon S^1 \times S^2 \to \bar{\mathbb{R}}$ is called $\mathbb{K}$-inf-compact if, for every nonempty compact subset K of S^1, the restriction of this function to $K \times S^2$ is inf-compact.

For MDPs, Definition 2.1 corresponds to Definition A.1 of a $\mathbb{K}$-inf-compact function $f\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \bar{\mathbb{R}}$ on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$, where $\mathbb{S}^1$ and $\mathbb{S}^2$ are metric spaces, in the following way. Let $(\tilde{\mathbb{S}}^1, \rho^1)$ and $(\tilde{\mathbb{S}}^2, \rho^2)$, where ρ^1 and ρ^2 are metrics, be complete separable metric spaces in which the Borel sets $\mathbb{X}$ and $\mathbb{A}$ are defined. Let us consider the metric spaces $(\mathbb{X}, \rho^1)$, $(\mathbb{A}, \rho^2)$ and define $\mathbb{S}_1 := \mathbb{X}$, $\mathbb{S}_2 := \mathbb{A}$, and $\Phi(x) := \mathbb{A}$ for all $x \in \mathbb{X}$. The assumption that the function $f : \mathbb{X} \times \mathbb{A} \to \bar{\mathbb{R}}$ is $\mathbb{K}$-inf-compact in the sense of Definition 2.1 is equivalent to the assumption that the function $f\colon \mathrm{Gr}_{\mathbb{S}^1}(\Phi) \to \bar{\mathbb{R}}$ is $\mathbb{K}$-inf-compact in the sense of Definition A.1. In many inventory control applications, $\mathbb{X}$ and $\mathbb{A}$ are Polish spaces. In this case, it is natural to set $\tilde{\mathbb{S}}^1 = \mathbb{X}$ and $\tilde{\mathbb{S}}^2 = \mathbb{A}$. The examples include $\mathbb{X} = \mathbb{R}$, $\mathbb{X} = [0, \infty)$, $\mathbb{A} = \mathbb{R}$, and $\mathbb{A} = [0, \infty)$.

For a function $f\colon \mathbb{X} \times \mathbb{A} \to \bar{\mathbb{R}}$, $\mathbb{K}$-inf-compactness is a more general and natural property than inf-compactness. For example, for $\mathbb{X} = \mathbb{A} = \mathbb{R}$ the function $f(x, a) = |x - a|$ is $\mathbb{K}$-inf-compact, but it is not inf-compact. As shown in Feinberg et al. [31], the following assumption is sufficient for the existence of optimal policies for discounted MDPs.

Assumption W*. The following conditions hold:
(i) The transition probability P is weakly continuous.
(ii) The cost function c is $\mathbb{K}$-inf-compact.

We list some of the properties of MDPs that take place under Assumption W* (see Theorem 5.1 for details):

1. For a bounded below, lower semicontinuous terminal value function $\mathbf{F}$, the final-horizon optimality equation holds for all $\alpha \geq 0$:

$$v_{n+1,\mathbf{F},\alpha}(x) = \min_{a \in \mathbb{A}} \left\{ c(x,a) + \alpha \int_{\mathbb{X}} v_{n,\mathbf{F},\alpha}(y) P(dy \,|\, x, a) \right\}, \quad x \in \mathbb{X},\ n = 0, 1, \ldots, \tag{6}$$

where $v_{0,\mathbf{F},\alpha}(x) = \mathbf{F}(x)$ for all $x \in \mathbb{X}$. In particular, this is true for $\mathbf{F} \equiv 0$ and $v_{0,\alpha} \equiv 0$.

2. The function v_α is lower semicontinuous, where $\alpha \in [0, 1)$. If the function $\mathbf{F}$ is bounded below and lower semicontinuous, then the functions $v_{n,\mathbf{F},\alpha}$, for $n = 0, 1, \ldots$ and $\alpha \geq 0$, are lower semicontinuous. If, in addition, $\mathbf{F}(x) \leq v_\alpha(x)$ for all $x \in \mathbb{X}$, then $v_\alpha(x) = \lim_{n\to\infty} v_{n,\mathbf{F},\alpha}(x)$, where $\alpha \in [0, 1)$. In particular, this is true for $\mathbf{F} \equiv 0$; that is, $v_\alpha(x) = \lim_{n\to\infty} v_{n,\alpha}(x)$, where $\alpha \in [0, 1)$.

3. For $\alpha \in [0, 1)$, the infinite-horizon value function v_α satisfies the optimality equation

$$v_\alpha(x) = \min_{a \in \mathbb{A}} \left\{ c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) P(dy \,|\, x, a) \right\}, \quad x \in \mathbb{X}, \tag{7}$$

a stationary optimal policy exists, and a stationary policy ϕ is optimal if and only if

$$v_\alpha(x) = c(x, \phi(x)) + \alpha \int_{\mathbb{X}} v_\alpha(y) P(dy \,|\, x, \phi(x)), \quad x \in \mathbb{X}. \tag{8}$$

4. If the one-step cost function c is inf-compact, then the value function v_α is inf-compact, when $\alpha \in [0, 1)$. The same is true for the value functions $v_{n,\mathbf{F},\alpha}$, $n = 1, 2, \ldots$, when the terminal value $\mathbf{F}$ is a bounded below, lower semicontinuous function and $\alpha \geq 0$.

In particular, the fourth property is useful for proving the existence of stationary optimal policies for inventory control problems. It is well known that, for average costs per unit time, optimal policies may not exist under Assumption W*. For example, optimal policies may not exist for a countable state space and finite action sets (see, e.g., Ross [49, Section 5.1]) and for a finite-state set, compact action sets, and continuous transition probabilities and costs (see, e.g., Dynkin and Yushkevich [22, Section 7.8]). Next we formulate a general condition that typically holds for inventory control problems, which, together with Assumption W*, guarantees the existence of optimal policies for average cost MDPs. If $\inf_{x\in\mathbb{X}} w(x) < +\infty$, define for $\alpha \in [0,1)$:

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

Assumption B. The following conditions hold:
(i) $\inf_{x\in\mathbb{X}} w(x) < +\infty$; and
(ii) $\sup_{\alpha<1} u_\alpha(x) < \infty$ for all $x \in \mathbb{X}$.

We note that the function u_α is nonnegative, and Assumption **B** implies that m_α cannot take infinite values; see Schäl [52]. If Assumption **B**(i) does not hold, then the average-cost problem is trivial: all policies lead to infinite average losses per unit time. This assumption holds in all well-defined problems, and usually, it is easy to verify. The validity of Assumption **B**(ii) probably follows from various ergodicity and communicating conditions, but this relation has not been studied in the literature. As explained in the text following Theorem 5.5 below, Assumption **B**(ii) holds and can be easily verified for inventory control problems. As shown in Feinberg et al. [31], Assumptions W* and **B** imply the existence of stationary optimal policies for average-cost MDPs, which follows from the validity of optimality inequalities.

For $\alpha \in [0,1)$, consider

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

According to Schäl [52, Lemma 1.2], Assumption **B**(i) implies

$$0 \le \underline{w} \le \bar{w} \le w^* < +\infty. \tag{9}$$

According to Schäl [52, Proposition 1.3], if there exists a measurable function $u: \mathbb{X} \to [0,\infty)$ and a stationary policy ϕ satisfying the optimality inequality

$$\underline{w} + u(x) \ge c(x,\phi(x)) + \int u(y) P(dy \mid x, \phi(x)), \quad x \in \mathbb{X}, \tag{10}$$

then ϕ is average-cost optimal, and $w(x) = \underline{w} = \bar{w}$ for all $x \in \mathbb{X}$. Assumptions W* and **B** imply the existence of a stationary policy ϕ satisfying optimality inequality (10).

Another form of an optimality inequality was introduced in Feinberg et al. [31], where it was shown that, if there exists a measurable function $u: \mathbb{X} \to [0,+\infty)$ and a stationary policy ϕ such that

$$\bar{w} + u(x) \ge c(x,\phi(x)) + \int_{\mathbb{X}} u(y) P(dy \mid x, \phi(x)), \quad x \in \mathbb{X}, \tag{11}$$

then ϕ is average-cost optimal, and

$$w(x) = w^\phi(x) = \limsup_{\alpha\uparrow 1} (1-\alpha) v_\alpha(x) = \bar{w}, \quad x \in \mathbb{X}. \tag{12}$$

Observe that inequality (11) is weaker than (10) because (10) implies (11).

The existence of stationary optimal policies satisfying inequality (11) is proved in Feinberg et al. [31] under Assumptions W* and $\underline{\mathbf{B}}$ there, which consists of Assumption **B**(i) and the following assumption (Feinberg et al. [31]):

$$\liminf_{\alpha\uparrow 1} u_\alpha(x) < \infty, \quad \text{for all } x \in \mathbb{X}, \tag{13}$$

which is weaker than Assumption **B**(ii). However, an example of an MDP, satisfying Assumptions W* and $\underline{\mathbf{B}}$ but not satisfying Assumption **B**(ii), is currently unknown.

Remark 2.2. 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 a 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 $A_c(x)=\{a\in\mathbb{A}\colon c(x,a)<+\infty\}$, $x\in\mathbb{X}$. To transform an MDP with action sets $A(x)$ to an MDP with the action set $\mathbb{A}$, it is sufficient to set $c(x,a)=+\infty$ when $a\in\mathbb{A}\setminus A(x)$, $x\in\mathbb{X}$. Early works on MDPs by Blackwell [15] and Strauch [61] 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. MDPs Defined by Stochastic Equations

Inventory control problems are often defined by equations

$$x_{t+1}=F(x_t,a_t,D_{t+1}),\quad t=0,1,\dots, \tag{14}$$

where x_t is the amount of inventory available at the end of day t, a_t is the ordered quantity at the end of day t, and D_{t+1} is the demand on day $t+1$. For the classic periodic-review problem with backlogs, $F(x,a,D)=x+a-D$, and for a problem with lost sales, $F(x,a,D)=(x+a-D)^+$. The system can also incur losses of inventory, there could be lead times, and so on. So the function F can have a more complicated form, and interpretations of its parameters may be different for different problems. Also, in this paper we only consider independent and identically distributed demands—that is, $D_1,D_2,\dots$ are independent and identically distributed.

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, whose dynamics are defined by Equation (14), where $D_0,D_1,\dots$ are independent and identically distributed random variables with values in $\mathbb{S}$ whose distributions are defined by a probability measure μ and $F\colon\mathbb{X}\times\mathbb{A}\times\mathbb{S}\to\mathbb{X}$ is a measurable mapping.

Equation (14) defines the transition probability

$$P(B\mid x,a)=\int_{\mathbb{S}}1\{F(x,a,s)\in B\}\mu(ds),\quad B\in\mathcal{B}(\mathbb{S}), \tag{15}$$

from $\mathbb{X}\times\mathbb{A}\to\mathbb{X}$, and $P(\cdot\mid x_t,a_t)$ is the distribution of x_{t+1} given x_t and a_t, where 1 is the indicator function.

The following lemma relates Assumption W*(ii) to the problems defined by stochastic equations.

Lemma 3.1 (Hernández-Lerma [38, p. 92]). *If the function F is continuous, then the transition probability P is weakly continuous.*

Consider an MDP with the transition probability P defined by a continuous function F. If the one-step cost function c is inf-compact, then, for a random variable D with the same distribution as D_1, formulae (6)–(8) can be rewritten as

$$v_{n+1,\mathbf{F},\alpha}(x)=\min_{a\in\mathbb{A}}\{c(x,a)+\alpha\,\mathbb{E}\,v_{n,\mathbf{F},\alpha}(F(x,a,D))\},\quad x\in\mathbb{X},\ n=0,1,\dots, \tag{16}$$

$$v_\alpha(x)=\min_{a\in\mathbb{A}}\{c(x,a)+\alpha\,\mathbb{E}\,v_\alpha(F(x,a,D))\},\quad x\in\mathbb{X}, \tag{17}$$

and

$$v_\alpha(x)=c(x,\phi(x))+\alpha\int_{\mathbb{X}}v_\alpha(F(x,\phi(x),D)),\quad x\in\mathbb{X}. \tag{18}$$

Equation (10) becomes

$$\underline{w}+u(x)\ge c(x,\phi(x))+\mathbb{E}\,u(F(x,a,D)),\quad x\in\mathbb{X}, \tag{19}$$

and inequality (11) becomes the same as (19), with $\underline{w}$ replaced with $\bar{w}$.

4. The Classic Periodic-Review Problem with Backorders

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

$$x_{t+1} = x_t + a_t - D_{t+1}, \quad t = 0,1,2,\ldots, \tag{20}$$

where x_t is the inventory at the end of period t, a_t is the amount ordered at the end of period t, and D_{t+1} is the demand during period $(t+1)$. The demand is assumed to be i.i.d. In other words, the dynamics of the system is defined by Equation (14) with the function $F(x,a,D) = x + a - D$. Of course, this function is continuous. Here, we consider the case when there is a single commodity. In this case, x_t, a_t, and D_{t+1}, $t = 0,1,\ldots$, are real numbers.

A decision maker views the current inventory of a single commodity at the end of the day and makes an ordering decision. In the case of zero lead times considered here 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 nonnegative. The dynamics of the system are defined by (20). Let

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

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

Let us define the state space $\mathbb{X} = \mathbb{R}$; the action set $\mathbb{A} = \mathbb{R}^+$, where $\mathbb{R}^+ = [0,\infty)$; the transition probability P defined in (15) with $F(x,a,D) = x + a - D$; and the one-step cost function

$$c(x,a) = K 1_{\{a>0\}} + \bar{c}a + \mathbb{E}\, h(x + a - D).$$

The function c is inf-compact, and, of course, the function F is continuous. Therefore, Assumption W* holds. It is relatively easy to show that Assumption **B** holds. Thus, optimality equations hold for finite-horizon and infinite-horizon problems. In particular, they exist for problems with total discounted- and average-cost criteria.

Optimality equations and inequalities can be written as

$$v_{n+1,\mathbf{F},\alpha}(x) = \min\Big\{\min_{a>0}[K + G_{n,\mathbf{F},\alpha}(x+a)], G_{n,\mathbf{F},\alpha}(x)\Big\} - \bar{c}x, \tag{21}$$

$$v_\alpha(x) = \min\Big\{\min_{a>0}[K + G_\alpha(x+a)], G_\alpha(x)\Big\} - \bar{c}x, \tag{22}$$

$$w + u(x) \geq \min\Big\{\min_{a>0}[K + H(x+a)], H(x)\Big\} - \bar{c}x, \tag{23}$$

where $n = 0,1,\ldots$ and

$$G_{n,\mathbf{F},\alpha}(x) := \bar{c}x + \mathbb{E}\, h(x-D) + \alpha\, \mathbb{E}\, v_{n,\mathbf{F},\alpha}(x-D), \tag{24}$$

$$G_\alpha(x) := \bar{c}x + \mathbb{E}\, h(x-D) + \alpha\, \mathbb{E}\, v_\alpha(x-D), \tag{25}$$

$$H(x) := \bar{c}x + \mathbb{E}\, h(x-D) + \mathbb{E}\, u(x-D). \tag{26}$$

We also write $G_{n,\alpha}$ instead of $G_{n,\mathbf{F},\alpha}$ when $\mathbf{F} \equiv 0$.

Definition 4.1. Let s_t and S_t be real numbers such that $s_t \le S_t$, $t = 0, 1, \ldots$. 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 $x_t \ge s_t$. 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 is an (s, S) policy at all steps $t = 0, 1, \ldots$.

The standard method for proving the optimality of (s_t, S_t) and (s, S) policies for discounted costs was introduced by Scarf [50], and it is based on the notion of a K-convex function.

Definition 4.2. A function $f\colon \mathbb{R} \to \mathbb{R}$ is called K-convex, $K \ge 0$, if for each $x \le y$ and for each $\lambda \in (0, 1)$,

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

For an inf-compact function $g\colon \mathbb{R} \to \mathbb{R}$, let

$$S \in \operatorname*{arg\,min}_{x \in \mathbb{R}} \{g(x)\}, \tag{27}$$

$$s := \inf\{x \le S \mid g(x) \le K + g(S)\}. \tag{28}$$

These real numbers exist because the function g is inf-compact. In addition, s is defined uniquely and does not depend on the choice of S if there is more than one S satisfying (27).

The standard method for proving the optimality of (s_t, S_t) policies, $t = 0, 1, \ldots, N-1$, for N-horizon problems, $N = 1, 2, \ldots$, is to consider $g = G_{N-1,\,\alpha}$ and prove by induction that these functions are inf-compact and K-convex. As follows from the optimality equation (21), this implies the optimality of (s_t, S_t) policies with S_t and s_t defined by (27) and (28), respectively, with $g = G_{N-t=1,\,\alpha}$. The next step would be to consider $t \to \infty$ and prove the optimality of (s, S) policies for infinite horizon problems.

However, it is possible that the functions $G_{N-1,\,\alpha}$ are not inf-compact, and the described approach fails. Then the natural approach is to try to do the same steps for the function $G_{N-1,\,\mathbf{F},\,\alpha}$ for a specially selected terminal value function $\mathbf{F}$. The natural candidate is the function $\mathbf{F} = v_\alpha^0$, where v_α^0 is the infinite-horizon value for the problem with the ordering cost $K = 0$. It is possible to show that there exists $\alpha' \in [0, 1)$ such that the functions $G_{N-1,\,v_\alpha^0,\,\alpha}$ are inf-compact if $\alpha \in [\alpha', 1)$. This implies the optimality of (s_t, S_t) policies for all finite-horizon problems with the terminal value $\mathbf{F} = v_\alpha^0$ for all $\alpha \in [\alpha', 1)$, which implies optimality of (s, S) policies for the infinite horizon discounted criterion with the discount factor α. In addition, it is always true that $G_{N,\,v_\alpha^0,\,\alpha} \to G_\alpha$, and the following lemma holds.

Lemma 4.3 (Feinberg and Lewis [26], Veinott and Wagner [64]). *There exists $\alpha' \in [0, 1)$ such that $G_\alpha(x) \to \infty$ as $|x| \to \infty$ for all $\alpha \in [\alpha', 1)$ and for all setup costs $K \ge 0$.*

The optimality of (s, S)-optimal policies for large discount factors implies optimality of (s, S) policies for average costs per unit time. The following theorem takes place.

Theorem 4.4 (Feinberg and Lewis [26]). *Consider $\alpha' \in [0, 1)$, whose existence is stated in Lemma 4.3. The following statements hold for the inventory control problem:*

(i) *For $\alpha \ge \alpha'$ and $n = 0, 1, \ldots$, define $g(x) := G_{n,\,v_\alpha^0,\,\alpha}(x)$, $x \in \mathbb{R}$. Consider real numbers $S^*_{n,\,\alpha}$ satisfying (27) and $s^*_{n,\,\alpha}$ defined in (28) with $g = G_{n, v_\alpha^0, \alpha}$, $n = 0, 1, \ldots$. Then 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 $\mathbf{F}(x) = v_\alpha^0(x)$, $x \in \mathbb{R}$.*

(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 (27) and s_α defined in (28). Then the (s_α, S_α) policy is optimal for the discount factor α. Furthermore, the sequence of pairs $\{(s^*_{n,\,\alpha}, S^*_{n,\,\alpha})\}_{n=0,1,\ldots}$ is bounded, where $s^*_{n,\,\alpha}$ and $S^*_{n,\,\alpha}$ are described in statement* (i)*, $n = 0, 1, \ldots$. 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* $\{(s'_{\alpha_n}, S'_{\alpha_n})\}_{n=1,2,\ldots}$ *is bounded, and each limit point* (s', S') *defines an average-cost optimal* (s', S') *policy.*

As explained above, (s_t, S_t) policies may not be optimal for finite-horizon problems for all discount factors, and (s, S) may not be optimal for infinite-horizon discounted problems with a small discount factor. Let us consider the assumption on the growth of backordering costs that was probably introduced by Veinott and Wagner [64] for problems with discrete demand. This assumption ensures that the functions $G_{N-1,\alpha}$ and G_α are inf-compact, and, as explained above, this implies the optimality of (s_t, S_t) policies and (s, S) policies for finite-horizon and infinite-horizon discounted problems, respectively, for all $N = 1, 2, \ldots$ and for all $\alpha \in [0, 1)$.

Assumption GB. There exist $z, y \in \mathbb{R}$ such that $z < y$ and

$$\frac{h(y) - h(z)}{y - z} < -\bar{c}. \tag{29}$$

Lemma 4.5 (Chen and Simchi-Levi [17, 18], Feinberg and Lewis [26], Heyman and Sobel [42]). *Suppose that Assumption* **GB** *holds. Then the functions* $G_\alpha(x)$ *and* $G_{N,\alpha}(x)$, $N = 0, 1, \ldots$, *are inf-compact and* K*-convex.*

The following theorem describes the optimality of (s_t, S_t) policies and (s, S) policies for finite-horizon and infinite-horizon discounted problems under Assumption **GB**.

Theorem 4.6 (Chen and Simchi-Levi [17, 18], Feinberg and Lewis [26]). *Suppose that Assumption* **GB** *holds. Then,*

(i) *For* $\alpha \geq 0$ *and* $n = 0, 1, \ldots$, *consider real numbers* $S_{n,\alpha}$ *satisfying* (27) *and* $s_{n,\alpha}$ *defined in* (28), *with* $g(x) = G_{n,\alpha}(x)$, $x \in \mathbb{R}$. *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 an optimal policy for the* N*-horizon problem with the zero terminal values.*

(ii) *Let* $\alpha \in [0, 1)$. *Consider real numbers* S_α *satisfying* (27) *and* s_α *defined in* (28) *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* $\{(s_{n,\alpha}, S_{n,\alpha})\}_{n=0,1,\ldots}$ *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.

As stated in Theorem 4.4, (s, S) policies are optimal for average costs per unit time. However, Theorem 4.6 states the optimality of (s_t, S_t) policies and (s, S) policies for finite-horizon and infinite-horizon discounted problems for all discount factors only under Assumption **GB**. The structure of discount optimal policies for all discount factors is investigated in Feinberg and Liang [28], where the following parameters were introduced:

$$k_h := -\lim_{x \to -\infty} \frac{h(x)}{x} \tag{30}$$

and

$$\alpha^* := 1 - \frac{k_h}{\bar{c}}. \tag{31}$$

For example, $\alpha^* = 1 - h_-/\bar{c}$ for models with linear holding and backordering costs h considered in Bensoussan [5] and Bertsekas [11], when

$$h(x) = \begin{cases} h_+ x, & \text{if } x \geq 0, \\ -h_- x, & \text{otherwise,} \end{cases}$$

where h_- and h_+ are positive holding and backordering cost rates, and typically $h_- > h_+$.

The convexity and inf-compactness of h imply that $0 < k_h \le +\infty$. Therefore, $-\infty \le \alpha^* < 1$. Also, Assumption **GB** is equivalent to $\alpha^* < 0$. In addition, $\alpha' := \max\{\alpha^*, 0\}$ is the minimal possible value of the parameter α' whose existence is claimed in Lemma 4.3. These facts and their corollaries are summarized in the following theorem.

Theorem 4.7 (Feinberg and Liang [28]). *Assumption* **GB** *holds if and only if* $\alpha^* > 0$. *Therefore, if* $\alpha^* < 0$, *then the statements* (i) *and* (ii) *of Theorem* 4.6 *hold. In addition,* $\alpha' = \max\{\alpha^*, 0\}$ *is the minimal value of the parameter* α' *whose existence is stated in Lemma* 4.3. *Therefore, statements* (i) *and* (ii) *of Theorem* 4.4 *take place for* $\alpha' = \max\{\alpha^*, 0\}$.

Define $\mathbf{S}_0 := 0$ and

$$\mathbf{S}_t := \sum_{j=1}^{t} D_j, \quad t = 1, 2, \ldots. \tag{32}$$

Then, $\mathbb{E}[\mathbf{S}_t] = t\,\mathbb{E}[D] < +\infty$ for all $t = 0, 1, \ldots$.

Define the following function for all $t = 0, 1, \ldots$ and $\alpha \ge 0$:

$$f_{t,\alpha}(x) := \bar{c}x + \sum_{i=0}^{t} \alpha^i\, \mathbb{E}[h(x - \mathbf{S}_{i+1})], \quad x \in \mathbb{X}. \tag{33}$$

Observe that $f_{0,\alpha}(x) = \bar{c}x + \mathbb{E}[h(x - D)] = G_{0,\alpha}$. Since $h(x)$ is a convex function, then the function $f_{t,\alpha}(x)$ is convex for all $t = 0, 1, \ldots$ and $\alpha \ge 0$.

Let $F_{t,\alpha}(-\infty) := \lim_{x\to-\infty} f_{t,\alpha}(x)$ and

$$N_\alpha := \inf\{t = 0, 1, \ldots : F_{t,\alpha}(-\infty) = +\infty\}, \tag{34}$$

where the infimum of an empty set is $+\infty$. Since the function $h(x)$ is nonnegative, then the function $f_{t,\alpha}(x)$ is nondecreasing in t for all $x \in \mathbb{X}$ and $\alpha \ge 0$. Therefore, (i) N_α is nonincreasing in α, that is, $N_\alpha \le N_\beta$, if $\alpha > \beta$; and (ii) in view of the definition of N_α, for each $t \in \mathcal{N}_0$,

$$F_{t,\alpha}(-\infty) < +\infty, \quad \text{if } t < N_\alpha \qquad \text{and} \qquad F_{t,\alpha}(-\infty) = +\infty, \quad \text{if } t \ge N_\alpha. \tag{35}$$

The following theorem provides the complete description of optimal finite-horizon policies for all discount factors α.

Theorem 4.8 (Feinberg and Liang [28]). *Let* $\alpha > 0$. *Consider* α^* *defined in* (31). *If* $\alpha^* < 0$ *(that is, Assumption* **GB** *holds), then the statement of Theorem* 4.6(i) *holds. If* $0 \le \alpha^* < 1$, *then the following statements hold for the finite-horizon problem with the discount factor* α:

(i) *If* $\alpha \in [0, \alpha^*]$, *then a policy that never orders is optimal for every finite horizon* $N = 1, 2, \ldots$.

(ii) *If* $\alpha > \alpha^*$, *then* $N_\alpha < +\infty$, *and for a finite horizon* $N = 1, 2, \ldots$, *the following is true:*

(a) *if* $N \le N_\alpha$, *then a policy that never orders at steps* $t = 0, 1, \ldots, N-1$ *is optimal;*

(b) *if* $N > N_\alpha$, *then a policy that never orders at steps* $t = N - N_\alpha, \ldots, N-1$ *and follows the* $(s_{N-t-1,\alpha}, S_{N-t-1,\alpha})$ *policy at steps* $t = 0, \ldots, N - N_\alpha - 1$ *is optimal, where the real numbers* $S_{t,\alpha}$ *satisfy* (27), *and* $s_{t,\alpha}$ *are defined in* (28) *with* $g(x) := G_{t,\alpha}(x)$, $x \in \mathbb{X}$.

The conclusions of Theorem 4.8 are presented in Table 1 and Figure 1.

The following theorem provides the complete description of optimal infinite-horizon policies for all discount factors α.

Theorem 4.9 (Feinberg and Liang [28]). *Let* $\alpha \in [0, 1)$. *Consider* α^* *defined in* (31). *The following statements hold for the infinite-horizon problem with the discount factor* α:

TABLE 1. The structure of optimal policies for a discounted N-horizon problem with $N < +\infty$ and $\alpha \geq 0$.

$\alpha^* < 0$	$0 \leq \alpha^* < \alpha$	$\alpha^* \geq \alpha$
There is an optimal (s_t, S_t) policy.	For the natural number N_α defined in (34), if $N > N_\alpha$, then a policy, that never orders at steps $t = N - N_\alpha, \ldots, N-1$ and is an (s_t, S_t) policy at steps $t = 0, \ldots, N - N_\alpha - 1$, is optimal; if $N \leq N_\alpha$, then a policy that never orders is optimal.	The policy that never orders is optimal.

TABLE 2. The structure of optimal policies for a discounted infinite-horizon problem with $\alpha \in [0, 1)$.

$\alpha^* < \alpha$	$\alpha \leq \alpha^*$
There is an optimal (s, S) policy.	The policy that never orders is optimal.

(i) *if $\alpha^* < \alpha$, then an (s_α, S_α) policy is optimal, where the real numbers S_α satisfy* (27) *and s_a are defined in* (28) *with $g(x) := G_\alpha(x)$, $x \in \mathbb{X}$. Furthermore, a sequence of pairs $(s_{t,\alpha}, S_{t,\alpha})_{t=N_\alpha, N_\alpha+1,\ldots}$ considered in Theorem 4.8*(ii,b) *is bounded, and, for if (s^*_α, S^*_α) is a limit point of the sequence, then the (s^*_α, S^*_α) policy is optimal for the infinite-horizon problem with the discount factor α;*

(ii) *if $\alpha^* \geq \alpha$, then the policy that never orders is optimal.*

The conclusions of Theorem 4.9 are presented in Table 2 and Figure 2.

The above theorems describe optimal policies for all discount factors. However, it is possible that for a given discount factor, at some states there are multiple optimal actions. Therefore, there may exist multiple optimal policies. It is also possible to describe additional optimal policies; see Feinberg and Liang [28]. The results on MDPs imply that the functions v_α and $v_{N,\alpha}$ are lower semicontinuous. However, for this problem, they are continuous; see Feinberg and Liang [28]. In addition, optimality inequalities (10) and (23) hold in the form of equalities; see Feinberg and Liang [27].

5. MDPs with Infinite-State Spaces and Weakly Continuous Transition Probabilities

This section describes the theory of dynamic programming for infinite-state problems with weakly continuous transition probabilities. The main focus is on the existence of optimal policies and the validity of optimality equations for problems with discounted costs and optimality inequalities for average-cost problems. We also discuss the convergence of optimal values and actions when the horizon length tends to infinity for finite-horizon problems and when the discount factor increases to 1 for infinite-horizon problems.

FIGURE 1. The structure of optimal policies for a discounted N-horizon problem with $N < +\infty$ and $\alpha \geq 0$.

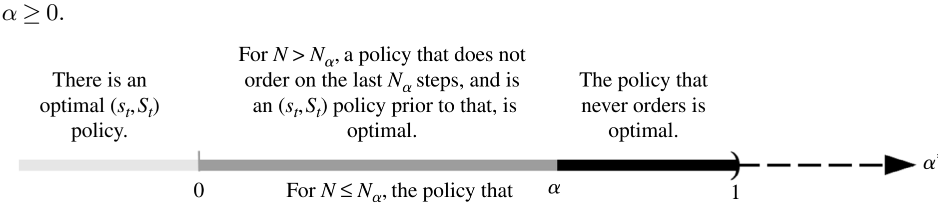
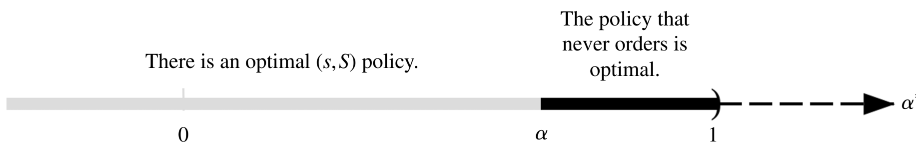


FIGURE 2. The structure of optimal policies for a discounted infinite-horizon problem with $\alpha \in [0,1)$.



5.1. Total Discounted Costs

The following theorem describes the validity of optimality equalities, the lower semicontinuity of value functions, and the convergence of value iterations. For zero terminal values, this theorem is presented in Feinberg et al. [31]. The case of nonzero terminal values is added in Feinberg and Lewis [26]. The case of inf-compact cost functions c, which leads to the inf-compactess of value functions, is studied in Feinberg and Lewis [25]. The inf-compactness of value functions is important for the analysis of average-cost problems. The proof of Theorem 5.1 uses the generalization of Berge's theorem, described in the appendix.

Theorem 5.1 (Feinberg and Lewis [26], Feinberg et al. [31]). *Let Assumption W* hold. Consider a bounded below, lower semicontinuous function* $\mathbf{F}\colon \mathbb{X} \to \bar{\mathbb{R}}$ *and* $\alpha \geq 0$. *Then,*

(i) *The functions* $v_{n,\mathbf{F},\alpha}$, $n = 0, 1, \ldots,$ *are lower semicontinuous.*

(ii) *The finite-horizon optimality equalities* (6) *hold with* $v_{0,\mathbf{F},\alpha}(x) = \mathbf{F}(x)$, *for all* $x \in \mathbb{X}$, *and the nonempty sets*

$$A_{n,\mathbf{F},\alpha}(x) := \left\{ a \in \mathbb{A}\colon v_{n+1,\mathbf{F},\alpha}(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_{n,\mathbf{F},\alpha}(y) P(dy \mid x,a) \right\}, \quad x \in \mathbb{X}, n = 0,1,\ldots,$$

satisfy the following properties:

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

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

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

(iv) *If the cost function* c *is inf-compact, the functions* $v_{n,\mathbf{F},\alpha}$, $n = 1, 2, \ldots,$ *are inf-compact.*

(v) *For* $\alpha \in [0,1)$, *if* $\mathbf{F}(x)$ *is constant or* $\mathbf{F}(x) \leq v_\alpha(x)$ *for all* $x \in \mathbb{X}$, *then* $v_{n,\mathbf{F},\alpha}(x) \to v_\alpha(x)$ *as* $n \to +\infty$ *for all* $x \in \mathbb{X}$.

(vi) *For* $\alpha \in [0,1)$, *the infinite-horizon optimality Equation* (7) *holds, and the nonempty sets*

$$A_\alpha(x) := \left\{ a \in \mathbb{A}\colon v_\alpha(x) = c(x,a) + \alpha \int_{\mathbb{X}} v_\alpha(y) P(dy \mid x,a) \right\}, \quad x \in \mathbb{X},$$

satisfy the following properties:

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

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

(vii) *For an infinite-horizon problem with* $\alpha \in [0,1)$, *there exists a stationary discount-optimal policy* ϕ_α, *and a stationary policy* ϕ_α *is optimal if and only if* $\phi_\alpha(x) \in A_\alpha(x)$ *for all* $x \in \mathbb{X}$.

(viii) *If the cost function* c *is inf-compact, then the infinite-horizon value function* v_α *is inf-compact,* $\alpha \in [0,1)$.

The following theorem describes convergence properties of optimal finite-horizon actions as the time horizon increases to infinity.

Theorem 5.2 (Feinberg and Lewis [26]). *Let Assumption* W* *hold and* $\alpha \in [0,1)$. *Let* $\mathbf{F}: \mathbb{X} \to \bar{\mathbb{R}}$ *be bounded below, lower semicontinuous, and such that for all* $x \in \mathbb{X}$,

$$\mathbf{F}(x) \le v_\alpha(x) \qquad \text{and} \qquad v_{1,\mathbf{F},\alpha}(x) \ge \mathbf{F}(x). \tag{36}$$

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

(i) *There is a compact subset* $D^*_\alpha(x)$ *of* $\mathbb{A}$ *such that* $A_{n,\mathbf{F},\alpha}(x) \subseteq D^*_\alpha(x)$ *for all* $t = 1, 2, \ldots,$ *where the sets* $A_{n,\mathbf{F},\alpha}(x)$ *are defined in Theorem* 5.1(ii).

(ii) *Each sequence* $\{a^{(n)} \in A_{n,\mathbf{F},\alpha}(x)\}_{n=1,2,\ldots}$ *is bounded, and all its limit points belong to* $A_\alpha(x)$.

Theorem 5.2 is useful for the analysis of the classic periodic-review inventory problem described in Section 4. As demonstrated in Table 1, (s_t, S_t) policies may not be optimal for finite-horizon problems, and the function $\mathbf{F} = v^0_\alpha$ is used to approximate optimal infinite-horizon thresholds, where v^0_α is the optimal value in the same problem with zero ordering costs.

5.2. Average Costs per Unit Time

We start with the formal introduction of Assumption **B**.

Assumption B. The following conditions hold:

(i) $\inf_{x\in\mathbb{X}} w(x) < +\infty$; and

(ii) $\liminf_{\alpha<1} u_\alpha(x) < \infty$ for all $x \in \mathbb{X}$.

Recall that the functions v_α and u_α are defined only for $\alpha \in [0,1)$. Let us set

$$u(x) := \liminf_{(y,\alpha)\to(x,1-)} u_\alpha(y), \quad x \in \mathbb{X}. \tag{37}$$

An equivalent definition is that $u(x)$ is the largest number such that $u(x) \le \liminf_{n\to\infty} u_{\alpha_n}(y_n)$ for all sequences $\{y_n \to x\}$ and $\{\alpha_n \to 1-\}$.

Theorem 5.3 (Feinberg et al. [31, Theorem 3]). *Suppose Assumptions* W* *and* **B** *hold. Then there exists a stationary policy* ϕ *satisfying* (11) *with* u *defined in* (37). *Thus, equalities* (12) *hold for this policy* ϕ. *Furthermore, the following statements hold:*

(i) *The function* $u: \mathbb{X} \to \mathbb{R}_+$ *is lower semicontinuous.*

(ii) *The nonempty sets*

$$A^*_u(x) := \left\{ a \in \mathbb{A}: \bar{w} + u(x) \ge c(x,a) + \int_{\mathbb{X}} u(y) P(dy \mid x, a) \right\}, \quad x \in \mathbb{X}, \tag{38}$$

satisfy the following properties:

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

(b) *for each* $x \in \mathbb{X}$, *the set* $A^*_u(x)$ *is compact.*

(iii) *A stationary policy* ϕ *is optimal for average costs and satisfies* (11) *with* u *defined in* (37), *if* $\phi(x) \in A^*_u(x)$ *for all* $x \in \mathbb{X}$.

(iv) *There exists a stationary policy* ϕ *with* $\phi(x) \in A_*(x) \subseteq A^*_u(x)$ *for all* $x \in \mathbb{X}$, *where*

$$A_*(x) := \left\{ a^* \in \mathbb{A}: u(x, a^*) = \inf_{a\in\mathbb{A}} \left\{ c(x,a) + \int_{\mathbb{X}} u(y) P(dy \mid x, a) \right\} \right\}, \quad x \in \mathbb{X}, \tag{39}$$

where

$$u(x,a^*) := c(x,a^*) + \int_{\mathbb{X}} u(y) P(dy|x,a^*), \qquad a^* \in \mathbb{A}.$$

(v) *If, in addition, the function* c *is inf-compact, then the function* u *is inf-compact.*

Stronger results hold under Assumption **B**.

Theorem 5.4 (Feinberg et al. [31, Theorem 4]). *Suppose Assumptions* W* *and* **B** *hold. Then there exists a nonnegative lower semicontinuous function u and a stationary policy ϕ satisfying* (10)*—that is, $\phi(x)\in A_u^*(x)$ for all $x\in\mathbb{X}$. Furthermore, every stationary policy ϕ for which* (10) *holds is optimal for the average costs per unit time criterion:*

$$w^{\phi}(x)=w(x)=w^*=\underline{w}=\bar{w}=\lim_{\alpha\uparrow 1}(1-\alpha)v_\alpha(x)=\lim_{N\to\infty}\frac{1}{N}v_{N,1}^{\phi}(x),\quad x\in\mathbb{X}. \tag{40}$$

Moreover, the following statements hold:

(i) *The nonempty sets $A_u^*(x), x\in\mathbb{X}$, satisfy the following properties:*
  (a) *the graph* $\mathrm{Gr}_{\mathbb{X}}(A_u^*)=\{(x,a): x\in\mathbb{X}, a\in A_u^*(x)\}$ *is a Borel subset of* $\mathbb{X}\times\mathbb{A}$*; and*
  (b) *for each $x\in\mathbb{X}$, the set $A_u^*(x)$ is compact.*

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

As an alternative to (37), as follows from Feinberg et al. [31, Theorems 3 and 4, p. 603], for each sequence $\alpha_n\to 1-$, the function u can be defined as

$$\tilde{u}(x):=\liminf_{(y,n)\to(x,\infty)}u_{\alpha_n}(y),\quad x\in\mathbb{X}. \tag{41}$$

An equivalent definition is that $\tilde{u}(x)$ is the largest number such that $\tilde{u}(x)\le\liminf_{n\to\infty}u_{\alpha_n}(y_n)$ for all sequences $\{y_n\to x\}$. It follows from these definitions that $u(x)\le\tilde{u}(x)$, $x\in\mathbb{X}$. 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 the cost function c is inf-compact, then the functions v_α, u, and $\tilde{u}$ are inf-compact as well; see Theorem 5.1 for the proof of this fact for v_α and Feinberg et al. [31, Theorem 4(e), Corollary 2] for u and $\tilde{u}$. We denote by $A_{\tilde{u}}^*(x)$ the sets defined in (37), when the function u is replaced with $\tilde{u}$.

In addition, if the one-step cost function c is inf-compact, the minima of the functions v_α possess additional properties. Set

$$X_\alpha:=\{x\in\mathbb{X}: v_\alpha(x)=m_\alpha\},\qquad \alpha\in[0,1). \tag{42}$$

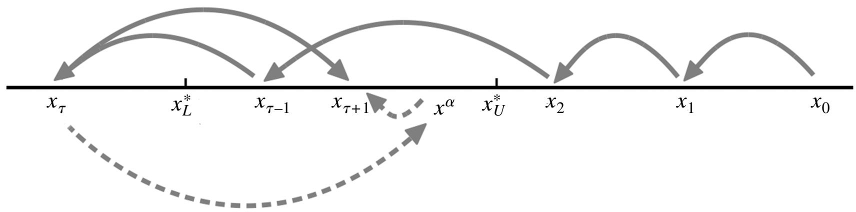
In view of Theorem 5.1(viii), the function v_α is inf-compact and $X_\alpha\neq\varnothing$. Since $X_\alpha=\{x\in\mathbb{X}: v_\alpha(x)\le m_\alpha\}$, this set is closed. The following theorem is useful for verifying the validity of Assumption **B**(ii) in inventory control applications; see Feinberg and Lewis [25, Lemma 5.1] and the references therein.

Theorem 5.5 (Feinberg et al. [31, Theorem 6]). *Let Assumptions* W* *and* **B**(*i*) *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 $\alpha\in[0,1)$.*

Theorem (5.5) implies that the minimum in $x\in\mathbb{X}$ of $v_\alpha(x)$ is achieved on a compact set $\mathcal{K}$, which does not depend on α. This typically means that to prove Assumption **B**(ii) it is sufficient to show that for each $x\in\mathbb{X}$ it is possible to reach every point in $\mathcal{K}$ in a way that the expected time and cost are finite. In inventory control applications, this can be shown by lowering the inventory levels below the levels in $\mathcal{K}$ and then by ordering up to a point in $\mathcal{K}$. Exact mathematical justifications are usually problem specific and use renewal theory. Here, we provide a short version of the proof from Feinberg and Lewis [26]. Choose $\mathcal{K}=[x_L^*,x_U^*]$; see Figure 3, where the existence of a set $\mathcal{K}$ is stated in Theorem 5.5, and this set can be chosen to be equal to a closed interval because each compact subset of $\mathbb{R}$ is contained in a closed finite interval. Let ϕ^α be a stationary optimal policy for a discount factor $\alpha\in[0,1)$, and let x^α be a state such that $v_\alpha(x^\alpha)=v_\alpha^{\phi^\alpha}(x^\alpha)=m_\alpha$. Since $x^\alpha\in X_\alpha$, then $x^\alpha\in[x_L^*,x_U^*]$. Consider a policy σ such that, if the initial point $x<x_L$, then σ orders up to the level that the policy ϕ^α would order at state x^α, and then σ makes the same decisions as ϕ^α. Since a move from state x_0 to x_1 can be presented as two instant moves—from x_0 to x^α and from x^α to x_1, as shown in Figure 3 (in this case, $\tau=0$, $x=x_\tau$, and only the move from x_τ to $x_{\tau+1}$ is relevant to (43)), then

$$v_\alpha(x)\le v_\alpha^\sigma(x)\le K+\bar{c}(x^\alpha-x)+v_\alpha(x^\alpha)\le K+\bar{c}(x_U^*-x)+m_\alpha,\quad x<x_L^*. \tag{43}$$

FIGURE 3. Verification of Assumption **B**(ii) for inventory control.



For the initial inventory level $x \geq x_L^*$, where $x = x_0$, the policy σ is defined in the following way. It does not order as long as the inventory level is greater than or equal to x_L^*. Then, as soon as the inventory level is less than x_L^*, the policy σ behaves in the same way it would if x_τ were the starting point, where $\tau := \inf\{t = 0, 1, \ldots : x_t < x_L^*\}$ is the first epoch when the inventory level is less than x_L^*. Standard arguments from renewal theory imply that $\mathbb{E}[x_\tau] > -\infty$ and $C(x,\tau) < +\infty$, where $C(x,\tau)$ is the expected total undiscounted holding (or backordering) cost paid until the system reaches the level x_τ. Then

$$v_\alpha(x) \leq v_\alpha^\sigma(x) \leq C(x,\tau) + K + \bar{c}\,\mathbb{E}[x_U^* - x_\tau] + m_\alpha, \quad x \geq x_L^*. \tag{44}$$

Inequalities (43) and (44) imply that Assumption **B**(ii) holds. Though the above proof was applied in Feinberg and Lewis [26] to the classic periodic-review system with backorders, it is generic and applicable to other systems. For problems with lost sales, the proof may be even simpler because it may be possible to define σ so that τ is the first time when there is no inventory. Then $x_\tau = 0$, and the expected cost of a lost sale will be added to the right-hand side of (43). This expected cost is typically finite.

Certain average-cost optimal policies can be approximated by discount optimal policies with a vanishing discount factor; see Feinberg et al. [31, Theorem 5]. The following theorem and its corollary follow from such approximations. In particular, the theorem and its corollary are useful for verifying that a limit point of optimal thresholds for vanishing discount factors is an optimal threshold for average costs per unit time.

Recall that, for the function $u(x)$ defined in (37), for each $x \in \mathbb{X}$ there exist sequences $\{\alpha_n \uparrow 1\}$ and $\{x^{(n)} \to x\}$, where $x^{(n)} \in \mathbb{X}$, $n = 1, 2, \ldots$, such that $u(x) = \lim_{n\to\infty} u_{\alpha_n}(x^{(n)})$. Similarly, for a sequence $\{\alpha_n \uparrow 1\}$, consider the function $\tilde{u}$ defined in (41). Then for each $x \in \mathbb{X}$ there exist a sequence $\{x^{(n)} \to x\}$ of points in $\mathbb{X}$ and a subsequence $\{\alpha_n^*\}_{n=1,2,\ldots}$ of the sequence $\{\alpha_n\}_{n=1,2,\ldots}$ such that $\tilde{u}(x) = \lim_{n\to\infty} u_{\alpha_n^*}(x^{(n)})$.

Theorem 5.6 (Feinberg and Lewis [26]). *Let Assumptions* $\mathbf{W}^*$ *and* **B** *hold. For* $x \in \mathbb{X}$ *and* $a^* \in \mathbb{A}$, *the following two statements hold:*

(i) *For a sequence* $\{(x^{(n)}, \alpha_n)\}_{n=1,2,\ldots}$ *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 a sequence of natural numbers* $\{n_k \to \infty\}_{k=1,2,\ldots}$ *and actions* $\{a^{(n_k)} \in A_{\alpha_{n_k}}(x^{(n_k)})\}_{k=1,2,\ldots}$, *such that* $a^{(n_k)} \to a^*$ *as* $k \to \infty$, *then* $a^* \in A_u^*(x)$, *where the function* u *is defined in* (37).

(ii) *Let* $\{\alpha_n \uparrow 1\}_{n=1,2,\ldots}$ *be a sequence of discount factors, let* $\{\alpha_n^*\}_{n=1,2,\ldots}$ *be its subsequence, and let* $\{x^{(n)} \to x\}_{n=1,2,\ldots}$ *be a sequence of states from* $\mathbb{X}$ *such that* $u_{\alpha_n^*}(x^{(n)}) \to \tilde{u}(x)$ *as* $n \to \infty$, *where the function* $\tilde{u}$ *is defined in* (41) *for the sequence* $\{\alpha_n\}_{n=1,2,\ldots}$. *If there are actions* $a^{(n)} \in A_{\alpha_n^*}(x^{(n)})$ *such that* $a^{(n)} \to a^*$ *as* $n \to \infty$, *then* $a^* \in A_{\tilde{u}}^*(x)$.

Corollary 5.7 (Feinberg and Lewis [26]). *Let Assumptions* $\mathbf{W}^*$ *and* **B** *hold. For* $x \in \mathbb{X}$ *and* $a^* \in \mathbb{A}$, *the following two statements hold:*

(i) *If each sequence* $\{(\alpha_n^*, x^{(n)})\}_{n=1,2,\ldots}$ *with* $0 \leq \alpha_n^* \uparrow 1$, $x^{(n)} \in \mathbb{X}$, *and* $x^{(n)} \to x$, $n = 1, 2, \ldots$, *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* (37).

(ii) *If there is a sequence* $\{\alpha_n \uparrow 1\}_{n=1,2,\dots}$, *such that for every sequence of states* $\{x_n \to x\}$ *from* $\mathbb{X}$ *there are actions* $a^n \in A_{\alpha_n}(x^{(n)})$, $n = 1,2,\dots$, *satisfying* $a_n \to a^*$ *as* $n \to \infty$, *then* $a^* \in A^*_{\tilde{u}}(x)$, *where the function* $\tilde{u}$ *is defined in* (41) *for the sequence* $\{\alpha_n\}_{n=1,2,\dots}$.

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

Theorem 5.8 (Feinberg and Lewis [26]). *Let Assumptions* W* *and* **B** *hold. For* $x \in \mathbb{X}$, *the following two statements 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* $\{\alpha_n\}_{n=1,2,\dots}$ *is a sequence of discount factors* $\alpha_n \in [0,1)$, *then every sequence of infinite-horizon* α_n*-optimal actions* $\{a^{(n)} \in A_{\alpha_n}(x)\}_{n=1,2,\dots}$ *is bounded and therefore has a limit point* $a^* \in \mathbb{A}$.

6. Partially Observable Markov Decision Processes

POMDPs model the situations when the current state of the system may be unknown and the decision maker uses indirect observations for decision making. A POMDP is defined by the same objects as an MDP but, in addition to the state space $\mathbb{X}$ and action space $\mathbb{A}$, there is an observation space $\mathbb{Y}$. The states and observations are linked by the transition probability $Q(dy_{t+1} \mid a_t, x_{t+1})$, from $\mathbb{A} \times \mathbb{X}$ to $\mathbb{Y}$, $t = 0,1,\dots$. Thus, a POMDP is defined as the tuple $\{\mathbb{X}, \mathbb{Y}, \mathbb{A}, P, Q, c\}$, where the Borel state and action spaces $\mathbb{X}$ and $\mathbb{A}$, the transition probability P, and the cost function c are the same objects as in an MDP. In addition, $\mathbb{Y}$ is the observation space, which is also assumed to be a Borel subset of a Polish space, and Q is the observation probability, which is a regular transition probability from $\mathbb{A} \times \mathbb{X}$ to $\mathbb{Y}$. Sometimes we say a transition kernel or a stochastic kernel instead of transition probability. Though the initial state of the system may be unknown, the decision maker knows the probability distribution of the initial state $p(dx_0)$, and there is an observation probability for the first observation $Q_0(dy_0 \mid x_0)$.

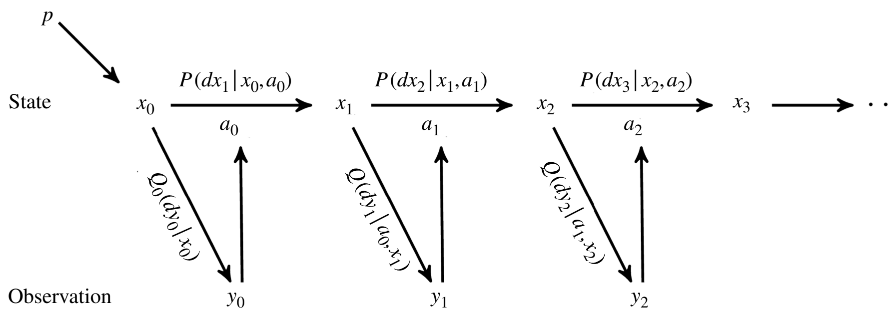
In various applications it is possible that there are continuous states and discrete observations or discrete states and continuous observations; both spaces can be discrete or continuous. So we consider a general situation by assuming that $\mathbb{X}$ and $\mathbb{Y}$ are Borel subsets of Polish spaces.

The following subsection describes the classic transformation of a POMDP to a completely observable MDP (COMDP), whose states are posterior probability distributions of states in the POMDP. This transformation was introduced by Aoki [1], Åström [2], Dynkin [21], and Shiryaev [58]. These ideas were advanced in the book by Striebel [61] and in the references provided in the following subsection. The main results of this section describe optimality conditions for POMDPs and COMDPs introduced in Feinberg et al. [36].

The POMDP evolves as follows. At time $t = 0$, the initial unobservable state x_0 has a given prior distribution p. The initial observation y_0 is generated according to the initial observation kernel $Q_0(\cdot \mid x_0)$. At each time epoch $t = 0,1,\dots$, if the state of the system is $x_t \in \mathbb{X}$ and the decision maker chooses an action $a_t \in \mathbb{A}$, then the cost $c(x_t, a_t)$ is incurred; the system moves to state x_{t+1} according to the transition law $P(\cdot \mid x_t, a_t)$. The observation $y_{t+1} \in \mathbb{Y}$ is generated by the observation kernels $Q(\cdot \mid a_t, x_{t+1})$, $t = 0,1,\dots$, and $Q_0(\cdot \mid x_0)$; see Figure 4. For the state space $\mathbb{X}$, denote by $\mathbb{P}(\mathbb{X})$ the set of probability measures on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$. We always consider a metric on $\mathbb{P}(\mathbb{X})$ consistent with the topology of weak convergence.

Define the *observable histories*: $h_0 := (p, y_0) \in H_0$ and $h_t := (p, y_0, a_0, \dots, y_{t-1}, a_{t-1}, y_t) \in H_t$ for all $t = 1,2,\dots$, where $H_0 := \mathbb{P}(\mathbb{X}) \times \mathbb{Y}$ and $H_t := H_{t-1} \times \mathbb{A} \times \mathbb{Y}$ if $t = 1,2,\dots$. Then a *policy* for the POMDP is defined as a sequence $\pi = \{\pi_t\}$ such that, for each $t = 0,1,\dots$, π_t is a transition kernel on $\mathbb{A}$ given H_t. Moreover, π is called *nonrandomized* if each probability measure $\pi_t(\cdot \mid h_t)$ is concentrated at one point. The *set of all policies* is denoted by Π. The Ionescu Tulcea theorem (Bertsekas and Shreve [12, pp. 140–141] or Hernández-Lerma and

FIGURE 4. POMDP Diagram.



Lassere [40, p. 178]) implies that, given a policy $\pi \in \Pi$, an initial distribution $p \in \mathbb{P}(\mathbb{X})$ and a sequence of transition probabilities $Q_0, \pi_0, P, Q, \pi_1, P, Q, \pi_2, \dots$ determine a unique probability measure P_p^π on the set of all trajectories $\mathbb{H}_\infty = (\mathbb{X} \times \mathbb{Y} \times \mathbb{A})^\infty$ endowed with the σ-field, which is the product of Borel σ-fields on $\mathbb{X}$, $\mathbb{Y}$, and $\mathbb{A}$, respectively. The expectation with respect to this probability measure is denoted by $\mathbb{E}_p^\pi$.

Let us specify a performance criterion. For a finite horizon $N = 0, 1, \dots$, and for a policy $\pi \in \Pi$, let the *expected total discounted costs* be

$$v_{N,\alpha}^\pi(p) := \mathbb{E}_p^\pi \sum_{t=0}^{N-1} \alpha^t c(x_t, a_t), \quad p \in \mathbb{P}(\mathbb{X}), \tag{45}$$

where $\alpha \geq 0$ is the discount factor, and $v_{0,\alpha}^\pi(p) = 0$. When $N = \infty$, we always assume $\alpha \in [0,1)$. We always assume that the function c is bounded below.

For any function $g^\pi(p)$, including $g^\pi(p) = v_{N,\alpha}^\pi(p)$ and $g^\pi(p) = v_\alpha^\pi(p)$, define the *optimal cost*

$$g(p) := \inf_{\pi \in \Pi} g^\pi(p), \quad p \in \mathbb{P}(\mathbb{X}),$$

where Π is *the set of all policies*. A policy π is called *optimal* for the respective criterion, if $g^\pi(p) = g(p)$ for all $p \in \mathbb{P}(\mathbb{X})$. For $g^\pi = v_{N,\alpha}^\pi$, the optimal policy is called *N-horizon discount optimal*; for $g^\pi = v_\alpha^\pi$, it is called *discount optimal*.

6.1. Reduction of POMDPs to MDPs

In this section, we formulate the well-known reduction of a POMDP to the corresponding COMDP (Bertsekas and Shreve [12], Dynkin and Yushkevich [22], Hernández-Lerma [38], Rhenius [48], Yushkevich [65]). This reduction constructs an MDP whose states are probability distributions on the original state space. These distributions are posterior distributions of states after the observations become known. In addition to posterior probabilities, they are also called belief probabilities and belief states in the literature. The reduction establishes the correspondence between certain classes of policies in MDPs and POMDPs and their performances. If an optimal policy is found for the COMDP, it defines in a natural way an optimal policy for the original POMDP. The reduction holds for measurable transition probabilities, observation probabilities, and one-step costs. Except for problems with discrete transition probabilities or with transition probabilities having densities (see Bäuerle and Rieder [3], Bensoussan [4]), almost nothing had been known until recently about the existence of optimal policies for POMDPs and how to find them.

To simplify notations, we sometimes drop the time parameter. Given a posterior distribution z of the state x at time epoch $t = 0, 1, \dots$, and given an action a selected at epoch t, denote by $R(B \times C \mid z, a)$ the joint probability that the state at time $(t+1)$ belongs to the set $B \in \mathcal{B}(\mathbb{X})$ and the observation at time $(t+1)$ belongs to the set $C \in \mathcal{B}(\mathbb{Y})$,

$$R(B \times C \mid z, a) := \int_{\mathbb{X}} \int_B Q(C \mid a, x') P(dx' \mid x, a) z(dx), \tag{46}$$

where R is a transition kernel on $\mathbb{X}\times\mathbb{Y}$ given $\mathbb{P}(\mathbb{X})\times\mathbb{A}$; see Bertsekas and Shreve [12], Dynkin and Yushkevich [22], Hernández-Lerma [38], or Yushkevich [65] for details. Therefore, the probability $R'(C\,|\,z,a)$ that the observation y at time n belongs to the set $C\in\mathcal{B}(\mathbb{Y})$ is

$$R'(C\,|\,z,a)=\int_{\mathbb{X}}\int_{\mathbb{X}}Q(C\,|\,a,x')P(dx'\,|\,x,a)z(dx), \tag{47}$$

where R' is a transition kernel on $\mathbb{Y}$ given $\mathbb{P}(\mathbb{X})\times\mathbb{A}$. By Bertsekas and Shreve [12, Proposition 7.27], there exists a transition kernel H on $\mathbb{X}$ given $\mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}$ such that

$$R(B\times C\,|\,z,a)=\int_C H(B\,|\,z,a,y)R'(dy\,|\,z,a). \tag{48}$$

The transition kernel $H(\cdot\,|\,z,a,y)$ defines a measurable mapping $H\colon\ \mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}\to\mathbb{P}(\mathbb{X})$, where $H(z,a,y)[\,\cdot\,]=H(\cdot\,|\,z,a,y)$. For each pair $(z,a)\in\mathbb{P}(\mathbb{X})\times\mathbb{A}$, the mapping $H(z,a,\cdot)\colon\ \mathbb{Y}\to\mathbb{P}(\mathbb{Y})$ is defined $R'(\cdot\,|\,z,a)$ a.s. uniquely in y; see Dynkin and Yushkevich [22, p. 309]. It is known that for a posterior distribution $z_t\in\mathbb{P}(\mathbb{X})$, action $a_t\in A(x)$, and an observation $y_{t+1}\in\mathbb{Y}$, the posterior distribution $z_{t+1}\in\mathbb{P}(\mathbb{X})$ is

$$z_{t+1}=H(z_t,a_t,y_{t+1}). \tag{49}$$

However, the observation y_{t+1} is not available in the COMDP model, and therefore y_{t+1} is a random variable with the distribution $R'(\cdot\,|\,z_t,a_t)$, and (49) is a stochastic equation that maps $(z_t,a_t)\in\mathbb{P}(\mathbb{X})\times\mathbb{A}$ to $\mathbb{P}(\mathbb{X})$. The stochastic kernel that defines the distribution of z_{t+1} on $\mathbb{P}(\mathbb{X})$ given $\mathbb{P}(\mathbb{X})\times\mathbb{X}$ is defined uniquely as

$$q(D\,|\,z,a):=\int_{\mathbb{Y}}1_D[H(z,a,y)]R'(dy\,|\,z,a), \tag{50}$$

where for $D\in\mathcal{B}(\mathbb{P}(\mathbb{X}))$,

$$1_D[u]=\begin{cases}1 & u\in D,\\ 0 & u\notin D;\end{cases}$$

see Hernández-Lerma [38, p. 87]. A particular choice of a stochastic kernel H in (48) does not change the measure $q(\cdot\,|\,z,a)$ since, for each pair $(z,a)\in\mathbb{P}(\mathbb{X})\times\mathbb{A}$, the mapping $H(z,a,\cdot)$: $\mathbb{Y}\to\mathbb{P}(\mathbb{Y})$ is defined $R'(\cdot\,|\,z,a)$ a.s. uniquely in y; see Dynkin and Yushkevich [22, p. 309].

The COMDP is defined as an MDP with parameters $(\mathbb{P}(\mathbb{X}),\mathbb{A},q,\bar{c})$, where we have the following:

(i) $\mathbb{P}(\mathbb{X})$ is the state space.
(ii) $\mathbb{A}$ is the action set available at all states $z\in\mathbb{P}(\mathbb{X})$.
(iii) The one-step cost function $\bar{c}$: $\mathbb{P}(\mathbb{X})\times\mathbb{A}\to\bar{\mathbb{R}}$ is defined as

$$\bar{c}(z,a):=\int_{\mathbb{X}}c(x,a)z(dx),\quad z\in\mathbb{P}(\mathbb{X}),\ a\in\mathbb{A}. \tag{51}$$

(iv) The transition probabilities q on $\mathbb{P}(\mathbb{X})$ given $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ is defined in (50).

If a stationary or Markov optimal policy for the COMDP exists and is found, it allows the decision maker to formulate an optimal policy for the POMDP. The details on how to do this can be found in Bertsekas and Shreve [12], Dynkin and Yushkevich [22], or Hernández-Lerma [38]. Therefore, a POMDP can be reduced to a COMDP. This reduction holds for measurable transition kernels P, Q, and Q_0. The measurability of these kernels and the cost function c lead to the measurability of transition probabilities and the cost function for the corresponding COMDP.

As follows from Theorem 5.1, if the COMDP satisfies Assumption W*, then optimal policies exist, they satisfy the optimality equation, and they can be found by value iterations. This is formulated in Theorem 6.2 below. The validity of Assumption W* for the COMDP is equivalent to the correctness of the following two hypotheses.

Hypothesis (i). The transition probability q from $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ to $\mathbb{P}(\mathbb{X})$ is weakly continuous.

Hypothesis (ii). The cost function $\bar{c}$: $\mathbb{P}(\mathbb{X})\times\mathbb{A}\to\bar{\mathbb{R}}$ is bounded below and $\mathbb{K}$-inf-compact on $\mathbb{P}(\mathbb{X})\times\mathbb{A}$.

The following theorem states the correctness of Hypothesis (ii). The question, whether Hypothesis (i) holds, is more difficult, and the following subsection is devoted to answering it.

Theorem 6.1 (Feinberg et al. [36]). *If the function c: $\mathbb{X}\times\mathbb{A}\to\bar{\mathbb{R}}$ is a bounded below, $\mathbb{K}$-inf-compact (inf-compact) function on $\mathbb{X}\times\mathbb{A}$, then the cost function $\bar{c}$: $\mathbb{P}(\mathbb{X})\times\mathbb{A}\to\bar{\mathbb{R}}$ defined for the COMDP in (51) is bounded below by the same constant and $\mathbb{K}$-inf-compact (inf-compact) on $\mathbb{P}(\mathbb{X})\times\mathbb{A}$.*

In addition to weak convergence, two types of convergence are mentioned in the next subsection: setwise convergence and convergence in total variation. Here, we recall their definitions.

Let $(P_n)_{n=1,2,\ldots}$ be a sequence of probability measures on a measurable space $(\mathbf{S},\mathcal{F})$. This sequence converges setwise to a probability measure P_0 on $(\mathbf{S},\mathcal{F})$ if $\lim_{n\to\infty}P_n(A)=P_0(A)$ for each $A\in\mathcal{F}$. This sequence converges in total variation if $\lim_{n\to\infty}\|P_n-P_0\|=0$, where $\|P_n(A)-P_0(A)\|=2\sup\{P_n(A)-P_0(A)\colon A\in\mathcal{F}\}$. Convergence in total variation implies setwise convergence. If $\mathbf{S}$ is a metric space and $\mathcal{F}$ is its Borel σ-field, then setwise convergence implies weak convergence. Recall that P^* is a regular transition probability from a metric space $\mathbf{S}_1$ to a metric space $\mathbf{S}_2$, if $P^*(\cdot\,|\,s)$ is a probability measure on $\mathbf{S}_1$ for each $s\in\mathbf{S}_2$ and $P^*(A\,|\,\cdot)$ is a Borel function on $\mathbf{S}_1$ for each Borel subset A of $\mathbf{S}_2$. A transition probability is weakly (setwise, in total variation) continuous, if, for every sequence $(s^n)_{n=1,2,\ldots}$ on S_1 converging to $s^0\in\mathbf{S}_1$, the sequence $(P^*(\cdot\,|\,s^n))_{n=1,2,\ldots}$ converges weakly (setwise, in total variation) to $P^*(\cdot\,|\,s^0)$. There are two mathematical tools that are useful for the analysis of convergence of probability measures and for the analysis of MDPs and POMDPs: Fatou's lemma for variable probabilities (see Feinberg et al. [33] and references therein) and uniform Fatou's lemma introduced in Feinberg et al. [35].

6.2. Optimality Conditions for Discounted POMDPs

For the COMDP, Assumption W* can be rewritten in the following form:

(i) $\bar{c}$ is $\mathbb{K}$-inf-compact on $\mathbb{P}(\mathbb{X})\times\mathbb{A}$.

(ii) The transition probability $q(\cdot\,|\,z,a)$ is weakly continuous in $(z,a)\in\mathbb{P}(\mathbb{X})\times\mathbb{A}$.

Theorem 5.1 has the following form for the COMDP $(\mathbb{P}(\mathbb{X}),\mathbb{A},q,\bar{c})$.

Theorem 6.2 (Feinberg et al. [36, Theorem 3.1]). *Let the* COMDP $(\mathbb{P}(\mathbb{X}),\mathbb{A},q,\bar{c})$ *satisfy Assumption* W*. *Then,*

(i) *The functions $v_{n,\alpha}$, $n=0,1,\ldots,$ and v_α are lower semicontinuous on $\mathbb{P}(\mathbb{X})$, and $v_{n,\alpha}(z)\to v_\alpha(z)$ as $n\to\infty$, for all $z\in\mathbb{P}(\mathbb{X})$.*

(ii) *For any $z\in\mathbb{P}(\mathbb{X})$ and $n=0,1,\ldots,$*

$$\begin{aligned} v_{n+1,\alpha}(z) &= \min_{a\in\mathbb{A}}\left\{\bar{c}(z,a)+\alpha\int_{\mathbb{P}(\mathbb{X})}v_{n,\alpha}(z')q(dz'\,|\,z,a)\right\}\\ &= \min_{a\in\mathbb{A}}\left\{\int_{\mathbb{X}}c(x,a)z(dx)\right.\\ &\qquad\left.+\alpha\int_{\mathbb{X}}\int_{\mathbb{X}}\int_{\mathbb{Y}}v_{t,\alpha}(H(z,a,y))Q(dy\,|\,a,x')P(dx'\,|\,x,a)z(dx)\right\}, \end{aligned} \tag{52}$$

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

$$A_{n,\alpha}(z) := \left\{a\in\mathbb{A}\colon v_{n+1,\alpha}(z)=c(z,a)+\alpha\int_{\mathbb{P}(\mathbb{X})} v_{n,\alpha}(z')q(dz'\mid z,a)\right\},$$

where $z\in\mathbb{P}(\mathbb{X})$, satisfy the following properties: (a) *the graph* $\mathrm{Gr}(A_{n,\alpha})=\{(z,a)\colon z\in\mathbb{P}(\mathbb{X}), a\in A_{n,\alpha}(z)\}$*, $n=0,1,\ldots,$ is a Borel subset of $\mathbb{P}(\mathbb{X})\times\mathbb{A}$; and* (b) *if $v_{n+1,\alpha}(z)=\infty$, then $A_{n,\alpha}(z)=\mathbb{A}$, and if $v_{n+1,\alpha}(z)<\infty$, then $A_{n,\alpha}(z)$ is compact.*

(iii) *For any $N=1,2,\ldots,$ there exists a Markov optimal N-horizon policy $(\phi_0,\ldots,\phi_{N-1})$ for the COMDP, and if for an N-horizon Markov policy $(\phi_0,\ldots,\phi_{N-1})$ the inclusions $\phi_{N-1-n}(z)\in A_{n,\alpha}(z)$, $z\in\mathbb{P}(\mathbb{X})$, $n=0,\ldots,N-1$, hold, then this policy is N-horizon optimal.*

(iv) *For $\alpha\in[0,1)$,*

$$\begin{aligned} v_\alpha(z) &= \min_{a\in\mathbb{A}}\left\{\bar{c}(z,a)+\alpha\int_{\mathbb{P}(\mathbb{X})} v_\alpha(z')q(dz'\mid z,a)\right\}\\ &= \min_{a\in\mathbb{A}}\left\{\int_{\mathbb{X}} c(x,a)z(dx)\right.\\ &\qquad\left.+\alpha\int_{\mathbb{X}}\int_{\mathbb{X}}\int_{\mathbb{Y}} v_\alpha(H(z,a,y))Q(dy\mid a,x')P(dx'\mid x,a)z(dx)\right\},\quad z\in\mathbb{P}(\mathbb{X}),\end{aligned}$$

and the nonempty sets

$$A_\alpha(z) := \left\{a\in\mathbb{A}\colon v_\alpha(z)=\bar{c}(z,a)+\alpha\int_{\mathbb{P}(\mathbb{X})} v_\alpha(z')q(dz'\mid z,a)\right\},\quad z\in\mathbb{P}(\mathbb{X}),$$

satisfy the following properties: (a) *the graph* $\mathrm{Gr}(A_\alpha)=\{(z,a)\colon z\in\mathbb{P}(\mathbb{X}), a\in\mathbb{A}_\alpha(z)\}$ *is a Borel subset of $\mathbb{P}(\mathbb{X})\times\mathbb{A}$; and* (b) *if $v_\alpha(z)=\infty$, then $A_\alpha(z)=\mathbb{A}$, and if $v_\alpha(z)<\infty$, then $A_\alpha(z)$ is compact.*

(v) *For an infinite horizon, there exists a stationary discount optimal policy ϕ_α for the COMDP, and a stationary policy ϕ is optimal if and only if $\phi(z)\in A_\alpha(z)$ for all $z\in\mathbb{P}(\mathbb{X})$.*

(vi) *If the function c is inf-compact, the functions $v_{n,\alpha}$, $n=1,2,\ldots,$ and v_α are inf-compact on $\mathbb{P}(\mathbb{X})$.*

Hernández-Lerma [38, Section 4.4] provided the following conditions for the existence of optimal policies for the COMDP: (a) $\mathbb{A}$ is compact, (b) the cost function c is bounded and continuous, (c) the transition probability $P(\cdot\mid x,a)$ and the observation kernel $Q(\cdot\mid a,x)$ are weakly continuous transition kernels, and (d) there exists a weakly continuous $H\colon\mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}\to\mathbb{P}(\mathbb{X})$ satisfying (48). Consider the following relaxed version of assumption (d).

Assumption H (Feinberg et al. [36]). There exists a transition kernel H on $\mathbb{X}$ given $\mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}$ satisfying (48) such that if a sequence $\{z^n\}\subseteq\mathbb{P}(\mathbb{X})$ converges weakly to $z\in\mathbb{P}(\mathbb{X})$, and $\{a^n\}\subseteq\mathbb{A}$ converges to $a\in\mathbb{A}$, $n\to\infty$, then there exists a subsequence $\{(z^{n_k},a^{n_k})\}_{k\geq1}\subseteq\{(z^n,a^n)\}_{n\geq1}$ such that

$$H(z^{n_k},a^{n_k},y)\text{ converges weakly to }H(z,a,y)\quad\text{as }k\to\infty,$$

and this convergence takes place $R'(\cdot\mid z,a)$ almost surely in $y\in\mathbb{Y}$.

The following theorem provides two sufficient conditions for weak continuity of q. Statement (ii) can be found in Hernández-Lerma [38, p. 90].

Theorem 6.3 (Feinberg et al. [36]). *If the transition probability $P(dx'\mid x,a)$ is weakly continuous, then each of the following two conditions implies weak continuity of the transition probability q from $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ to $\mathbb{P}(\mathbb{X})$:*

(i) *The transition probability $R'(dy\mid z,a)$ from $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ to $\mathbb{Y}$ is setwise continuous, and Assumption* H *holds.*

(ii) *The transition probability* $Q(dy\,|\,a,x)$ *from* $\mathbb{A}\times\mathbb{X}$ *to* $\mathbb{Y}$ *is weakly continuous, and there exists a weakly continuous* $H\colon\ \mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}\to\mathbb{P}(\mathbb{X})$ *satisfying* (48).

Weak continuity of the transition probability P and continuity of the transition probability Q in total variation imply that Assumption H holds, and this leads to the following theorem.

Theorem 6.4 (Feinberg et al. [36]). *Let the transition probability* $P(dx'\,|\,x,a)$ *from* $\mathbb{X}\times\mathbb{A}$ *to* $\mathbb{X}$ *be weakly continuous, and let the transition probability* $Q(dy\,|\,a,x)$ *from* $\mathbb{A}\times\mathbb{X}$ *to* $\mathbb{Y}$ *be continuous in total variation. Then the transition probability* $R'(dy\,|\,z,a)$ *from* $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ *to* $\mathbb{Y}$ *is setwise continuous, Assumption H holds, and the transition probability* q *from* $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ *to* $\mathbb{P}(\mathbb{X})$ *is weakly continuous.*

The following theorem, which follows from Theorems 6.1–6.3, relaxes assumptions (a), (b), and (d) in Hernández-Lerma [38, Section 4.4].

Theorem 6.5 (Feinberg et al. [36]). *Under the following conditions,*

(i) *The cost function* c *is* $\mathbb{K}$*-inf-compact.*

(ii) *Either*

- (a) *the transition probability* $R'(dy\,|\,z,a)$ *from* $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ *to* $\mathbb{Y}$ *is setwise continuous and Assumption H holds, or*
- (b) *the transition probability* $Q(dy\,|\,a,x)$ *from* $\mathbb{A}\times\mathbb{X}$ *to* $\mathbb{Y}$ *is weakly continuous and there exists a weakly continuous* $H\colon\ \mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}\to\mathbb{P}(\mathbb{X})$ *satisfying* (48);

the COMDP $(\mathbb{P}(\mathbb{X}),\mathbb{A},q,\bar{c})$ *satisfies Assumption* W*, *and therefore statements* (i)–(vi) *of Theorem* 6.2 *hold.*

Theorems 6.4 and 6.5 imply the following result.

Theorem 6.6 (Feinberg et al. [36]). *Let Assumption* W* *hold, and let the transition probability* $Q(dy\,|\,a,x)$ *from* $\mathbb{A}\times\mathbb{X}$ *to* $\mathbb{Y}$ *be continuous in total variation. Then statements* (i)–(vi) *of Theorem* 6.2 *hold.*

Theorem 6.5 assumes either the weak continuity of H or Assumption H together with the setwise continuity of R'. For some applications, including the inventory control applications described in Section 7, the filtering kernel H satisfies Assumption H for some observations, and it is weakly continuous for other observations. The following theorem is applicable to such situations.

Theorem 6.7 (Feinberg et al. [36]). *Let the observation space* $\mathbb{Y}$ *be partitioned into two disjoint subsets* $\mathbb{Y}_1$ *and* $\mathbb{Y}_2$ *such that* $\mathbb{Y}_1$ *is open in* $\mathbb{Y}$. *Suppose the following assumptions hold*:

(i) *The transition probabilities* P *to* $\mathbb{X}$ *from* $\mathbb{X}\times\mathbb{A}$ *to* $\mathbb{X}$ *and* Q *from* $\mathbb{A}\times\mathbb{X}$ *to* $\mathbb{Y}$ *are weakly continuous.*

(ii) *The measure* $R'(\cdot\,|\,z,a)$ *on* $(\mathbb{Y}_2,\mathcal{B}(\mathbb{Y}_2))$ *is setwise continuous in* $(z,a)\in\mathbb{P}(\mathbb{X})\times\mathbb{A}$; *that is, for every sequence* $\{(z^n,a^n)\}_{n=1,2,\ldots}$ *in* $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ *converging to* $(z,a)\in\mathbb{P}(\mathbb{X})\times\mathbb{A}$ *and for every* $C\in\mathcal{B}(\mathbb{Y}_2)$, *we have* $R'(C\,|\,z^n,a^n)\to R'(C\,|\,z,a)$.

(iii) *There exists a transition probability* H *from* $\mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}$ *to* $\mathbb{X}$ *satisfying* (48) *such that*

(a) *the transition probability* H *from* $\mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}_1$ *to* $\mathbb{X}$ *is weakly continuous; and*

(b) *Assumption* H *holds on* $\mathbb{Y}_2$; *that is, if a sequence* $\{z^{(n)}\}_{n=1,2,\ldots}\subseteq\mathbb{P}(\mathbb{X})$ *converges weakly to* $z\in\mathbb{P}(\mathbb{X})$ *and a sequence* $\{a^{(n)}\}_{n=1,2,\ldots}\subseteq\mathbb{A}$ *converges to* $a\in\mathbb{A}$, *then there exists a subsequence* $\{(z^{(n_k)},a^{(n_k)})\}_{k=1,2,\ldots}\subseteq\{(z^{(n)},a^{(n)})\}_{n=1,2,\ldots}$ *and a measurable subset* C *of* $\mathbb{Y}_2$ *such that* $R'(\mathbb{Y}_2\backslash C\,|\,z,a)=0$ *and* $H(z^{(n_k)},a^{(n_k)},y)$ *converges weakly to* $H(z,a,y)$, *for all* $y\in C$.

Then the transition probability q from $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ to $\mathbb{P}(\mathbb{X})$ is weakly continuous. If, in addition to the above conditions, the cost function c is $\mathbb{K}$-inf-compact, then the COMDP $(\mathbb{P}(\mathbb{X}),\mathbb{A},q,\bar{c})$ *satisfies Assumption* W*, *and therefore statements* (i)–(iv) *of Theorem* 6.2 *hold.*

The following corollary follows from Theorem 6.7.

Corollary 6.8 (Feinberg et al. [36]). *Let the observation space $\mathbb{Y}$ be partitioned into two disjoint subsets $\mathbb{Y}_1$ and $\mathbb{Y}_2$ such that $\mathbb{Y}_1$ is open in $\mathbb{Y}$ and $\mathbb{Y}_2$ is countable. Suppose the following assumptions hold*:

(i) *The transition probabilities P from $\mathbb{X}\times\mathbb{A}$ to $\mathbb{X}$ and Q from $\mathbb{A}\times\mathbb{X}$ to $\mathbb{Y}$ are weakly continuous.*

(ii) *$Q(y\,|\,a,x)$ is a continuous function on $\mathbb{A}\times\mathbb{X}$ for each $y\in\mathbb{Y}_2$.*

(iii) *There exists a stochastic kernel H on $\mathbb{X}$ given $\mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}$ satisfying* (48) *such that the stochastic kernel H on $\mathbb{X}$ given $\mathbb{P}(\mathbb{X})\times\mathbb{A}\times\mathbb{Y}_1$ is weakly continuous.*

Then assumptions (ii) *and* (iiib) *from Theorem* 6.7 *hold, and the transition probability q from $\mathbb{P}(\mathbb{X})\times\mathbb{A}$ to $\mathbb{P}(\mathbb{X})$ is weakly continuous. If, in addition to the above conditions, the cost function c is $\mathbb{K}$-inf-compact, then the COMDP $(\mathbb{P}(\mathbb{X}),\mathbb{A},q,\bar{c})$ satisfies Assumption* W*, *and therefore statements* (i)–(vi) *of Theorem* 6.2 *hold.*

In conclusion of this section, we would like to mention another model of a controlled Markov process with partial observations, in which the observation kernel Q is not defined explicitly, and a state of the system consists of two parts: one part of the state is observable and another one is not; see, e.g., Rhenius [48], Yushkevich [65], and Bäuerle and Rieder [3, Chap. 5]. In Feinberg et al. [34, 36], such models were called Markov decision models with incomplete information, and the most general known sufficient conditions for the existence of optimal policies for such models with the expected total costs are provided in Feinberg et al. [34, Theorem 6.2].

7. Inventory Control with Incomplete Information

Bensoussan et al. [8, 9] studied several inventory control problems for periodic-review systems, when the inventory manager (IM) may not have complete information about inventory levels. In Bensoussan et al. [8, 9], a problem with backorders is considered. In the model considered in Bensoussan et al. [8], the IM does not know the inventory level if it is nonnegative, and the IM knows the inventory level if it is negative. In the model considered in Bensoussan et al. [9], the IM only knows whether the inventory level is negative or nonnegative. In Bensoussan et al. [7], a problem with lost sales is studied where the IM only knows whether a lost sale happened or not. The underlying mathematical analysis is summarized in Bensoussan et al. [6], where additional references can be found. The analysis includes transformations of density functions of demand distributions.

This section describes periodic-review systems with backorders and lost sales, when some inventory levels are observable and some are not. The goal is to minimize the expected total costs. Demand distributions may not have densities. This model is introduced in Feinberg et al. [36, Section 8.2].

In the case of full observations, we model the problem as an MDP with the state space $\mathbb{X}=\mathbb{R}$ (the current inventory level), action space $\mathbb{A}=\mathbb{R}$ (the ordered amount of inventory), and action sets $\mathbb{A}(x)=\mathbb{A}$ available at states $x\in\mathbb{X}$. If in a state x the amount of inventory a is ordered, then the holding/backordering cost $h(x)$, ordering cost $C(a)$, and lost sale cost $G(x,a)$ are incurred, where it is assumed that h, C, and G are nonnegative lower semicontinuous functions with values in $\mathbb{R}$ and $C(a)\to+\infty$ as $|a|\to\infty$. Observe that the one-step cost function $c(x,a)=h(x)+C(a)+G(x,a)$ is $\mathbb{K}$-inf-compact on $\mathbb{X}\times\mathbb{A}$. For problems with backorders (no lost sales), usually $G(x,a)=0$ for all x and a.

Let D_t, $t=1,2,\ldots$, be i.i.d. random variables with the distribution function F_D, where D_t is the demand at epoch t. The dynamics of the system are defined by $x_{t+1}=F(x_t,a_t,D_{t+1})$, where x_t is the current inventory level and a_t is the ordered (or scrapped) inventory at epoch $t = 0, 1, \ldots$. For problems with backorders, $F(x_t, a_t, D_{t+1}) = x_t + a_t - D_{t+1}$, and for problems with lost sales, $F(x_t, a_t, D_{t+1}) = (x_t + a_t - D_{t+1})^+$. In both cases, F is a continuous function defined on $\mathbb{R}^3$. To simplify and unify the presentation, we do not assume $\mathbb{X} = [0, \infty)$ for models with lost sales. However, for problems with lost sales, it is assumed that the initial state distribution p is concentrated on $[0, \infty)$, and this implies that states $x < 0$ will never be visited. We assume that the distribution function F_D is atomless (an equivalent assumption is that the function F_D is continuous). The state transition law P on $\mathbb{X}$ given $\mathbb{X} \times \mathbb{A}$ is

$$P(B \,|\, x, a) = \int_{\mathbb{R}} 1\{F(x, a, s) \in B\}\, dF_D(s), \tag{53}$$

where $B \in \mathcal{B}(\mathbb{X})$, $x \in \mathbb{X}$, and $a \in \mathbb{A}$. Since we do not assume that demands are nonnegative, this model also covers cash balancing problems and problems with returns; see Feinberg and Lewis [25] and the references therein. In a particular case, when $C(a) = +\infty$ for $a < 0$, orders with negative sizes are infeasible, and, if an order is placed, the ordered amount of inventory should be positive.

As mentioned above, some states (inventory levels) $x \in \mathbb{X} = \mathbb{R}$ are observable and some are not. Let the inventory be stored in containers. From a mathematical perspective, containers are elements of a finite or countably infinite partition of $\mathbb{X} = \mathbb{R}$ into disjoint convex sets, and each of these sets is not a singleton. In other words, each container B_{i+1} is an interval (possibly open, closed, or semi-open) with ends d_i and d_{i+1} such that $-\infty \le d_i < d_{i+1} \le +\infty$, and the union of these disjoint intervals is $\mathbb{R}$. In addition, we assume that $d_{i+1} - d_i \ge \gamma$ for some constant $\gamma > 0$ for all containers; that is, the sizes of all the containers are uniformly bounded below by a positive number. We also follow the convention that the 0-inventory level belongs to a container with end points d_0 and d_1, and a container with end points d_i and d_{i+1} is labeled as the $(i+1)$th container B_{i+1}. Thus, container B_1 is the interval in the partition containing point 0. The containers' labels can be nonpositive. If there is a container with the smallest (or largest) finite label n, then $d_{n-1} = -\infty$ (or $d_n = +\infty$, respectively). If there are containers with labels i and j, then there are containers with all the labels between i and j. In addition, each container is either transparent or nontransparent. If the inventory level x_t belongs to a nontransparent container, the IM only knows which container the inventory level belongs to. If an inventory level x_t belongs to a transparent container, the IM knows that the amount of inventory is exactly x_t; see Figures 5–8.

For each nontransparent container with end points d_i and d_{i+1}, we fix an arbitrary point b_{i+1} satisfying $d_i < b_{i+1} < d_{i+1}$. For example, it is possible to set $b_{i+1} = 0.5d_i + 0.5d_{i+1}$, when

FIGURE 5. Example with known current inventory level.

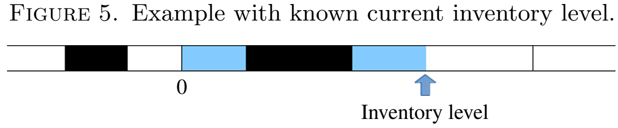


FIGURE 6. Example with unknown current inventory level.

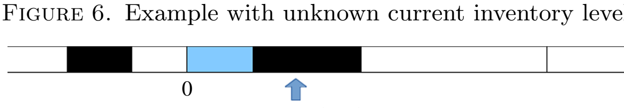
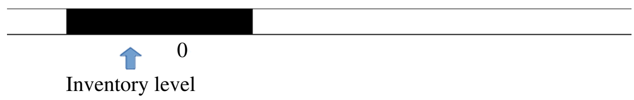


FIGURE 7. Example with known inventory levels and unknown backorder levels; the current inventory level is positive.

FIGURE 8. Example with inventory level 0 and current inventory level inside a nontransparent container.



$\max\{|d_i|, |d_{i+1}|\} < \infty$. If an inventory level belongs to a nontransparent container B_i, the IM observes $y_t = b_i$. Let L be the set of labels of the nontransparent containers. We set $Y_L = \{b_i: i \in L\}$ and define the observation set $\mathbb{Y} = \mathbb{T} \cup Y_L$, where $\mathbb{T}$ is the union of all transparent containers B_i (transparent elements of the partition). If the observation y_t belongs to a transparent container (in this case, $y_t \in \mathbb{T}$), then the IM knows that the inventory level $x_t = y_t$. If $y_t \in Y_L$ (in this case, $y_t = b_i$ for some i), then the IM knows that the inventory level belongs to the container B_i, and this container is nontransparent. Of course, the distribution of this level can be computed.

Let ρ be the Euclidean distance on $\mathbb{R}$: $\rho(a,b) = |a-b|$ for $a,b \in \mathbb{Y}$. On the state space $\mathbb{X} = \mathbb{R}$, we consider the metric $\rho_{\mathbb{X}}(a,b) = |a-b|$, if a and b belong to the same container, and $\rho_{\mathbb{X}}(a,b) = |a-b|+1$ otherwise, where $a,b \in \mathbb{X}$. The space $(\mathbb{X}, \rho_{\mathbb{X}})$ is a Borel subset of a Polish space (consisting of closed containers; that is, each finite point d_i is represented by two points: one belonging to the container B_i and another one to the container B_{i+1}). We notice that $\rho_{\mathbb{X}}(x^{(n)}, x) \to 0$ as $n \to \infty$ if and only if $|x^{(n)} - x| \to 0$ as $n \to \infty$ and the sequence $\{x^{(n)}\}_{n=N,N+1,\ldots}$ belongs to the same container as x for a sufficiently large N. Thus, convergence on $\mathbb{X}$ in the metric $\rho_{\mathbb{X}}$ implies convergence in the Euclidean metric. In addition, if $x \neq d_i$ for all containers i, then $\rho_{\mathbb{X}}(x^{(n)}, x) \to 0$ as $n \to \infty$ if and only if $|x^{(n)} - x| \to 0$ as $n \to \infty$. Therefore, for any open set B in $(\mathbb{X}, \rho_{\mathbb{X}})$, the set $B \setminus (\bigcup_i \{d_i\})$ is open in $(\mathbb{X}, \rho)$. We notice that each container B_i is an open and closed set in $(\mathbb{X}, \rho_{\mathbb{X}})$.

It is possible to show that the state transition law P given by (53) is weakly continuous in $(x,a) \in \mathbb{X} \times \mathbb{A}$. Set $\Psi(x) = x$ if the inventory level x belongs to a transparent container, and set $\Psi(x) = b_i$ if the inventory level belongs to a nontransparent container B_i with a label i. As follows from the definition of the metric $\rho_{\mathbb{X}}$, the function Ψ: $(\mathbb{X}, \rho_{\mathbb{X}}) \to (\mathbb{Y}, \rho)$ is continuous. Therefore, the observation transition probabilities Q_0 from $\mathbb{X}$ to $\mathbb{Y}$ and Q from $\mathbb{A} \times \mathbb{X}$ to $\mathbb{Y}$, $Q_0(C \,|\, x) := Q(C \,|\, a, x) := 1\{\Psi(x) \in C\}$, $C \in \mathcal{B}(\mathbb{Y})$, $a \in \mathbb{A}$, $x \in \mathbb{X}$, are weakly continuous.

If all the containers are nontransparent, the observation set $\mathbb{Y} = Y_L$ is countable, and conditions of Corollary 6.8 hold with $\mathbb{Y}_1 = \varnothing$ and $\mathbb{Y}_2 := \mathbb{Y}$. In particular, the function $Q(b_i \,|\, a, x) = 1\{x \in B_i\}$ is continuous if the metric $\rho_{\mathbb{X}}$ is considered on $\mathbb{X}$. If some containers are transparent and some are not, the conditions of Corollary 6.8 hold too. To verify this, we set $\mathbb{Y}_1 := \mathbb{T}$ and $\mathbb{Y}_2 := Y_L$, and note that $\mathbb{Y}_2$ is countable and the function $Q(b_i \,|\, x) = 1\{x \in B_i\}$ is continuous for each $b_i \in Y_L$ because B_i is open and closed in $(\mathbb{X}, \rho_{\mathbb{X}})$. Note that $H(B \,|\, z, a, y) = P(B \,|\, y, a)$ for any $B \in \mathcal{B}(\mathbb{X})$, $C \in \mathcal{B}(\mathbb{Y})$, $z \in \mathbb{P}(\mathbb{X})$, $a \in \mathbb{A}$, and $y \in \mathbb{T}$. The kernel H is weakly continuous on $\mathbb{P}(\mathbb{X}) \times \mathbb{A} \times \mathbb{Y}_1$. In addition, $\mathbb{T} = \bigcup_i B_i^{\mathrm{tr}}$, where B_i^{tr} are transparent containers, is an open set in $(\mathbb{X}, \rho_{\mathbb{X}})$. Thus the POMDP $(\mathbb{X}, \mathbb{Y}, \mathbb{A}, P, Q, c)$ satisfies the assumptions of Corollary 6.8. Thus, for the corresponding COMDP, there are stationary optimal policies for infinite-horizon problems with total discounted costs, optimal policies satisfy the optimality equations, and value iterations converge to the optimal value.

The models studied in Bensoussan et al. [7, 8, 9] correspond to the partition $B_1 = (-\infty, 0]$ and $B_2 = (0, +\infty)$, with the container B_2 being nontransparent and with the container B_1 being either nontransparent (backordered amounts are not known as in Bensoussan et al. [9]) or transparent (models with lost sales as in Bensoussan et al. [7], backorders are observable as in Bensoussan et al. [8]). Note that, since F_D is atomless, the probability that $x_t + a_t - D_{t+1} = 0$ is 0, $t = 0, 1, \ldots$.

The model provided in this subsection is applicable to other inventory control problems, and the conclusions of Corollary 6.8 hold for them too. For example, consider a periodic-review inventory system with backorders for which nonnegative inventory levels are known, and when the inventory level is negative, it is known that there is a backorder but its quantity is unknown. The partition consists of two containers: a nontransparent container $B_0 = (-\infty, 0)$ and a transparent container $B_1 = [0, +\infty)$.

8. Conclusions

The tutorial describes general sufficient conditions for the existence and characterization of optimal policies for Markov decision processes with possibly infinite-state spaces and unbounded action sets and costs. Expected total discounted-cost and average-cost criteria are considered. The described conditions imply the existence of optimal Markov policies in finite-horizon problems and the existence of optimal stationary policies for infinite-horizon problems. They imply the validity of optimality equations, convergence of value iterations, and continuity properties of value functions for discounted costs. They also imply the validity of optimality inequalities for average costs per unit time.

For discounted costs, these conditions consist of two assumptions: the transition probabilities are weakly continuous and the one-step cost function is $\mathbb{K}$-inf-compact. These two assumptions practically always hold for periodic-review stochastic inventory control problems. The $\mathbb{K}$-inf-compactness property of one-step costs is weaker than inf-compactness, which typically holds for cost functions for inventory control problems. One of the reasons for the generality of the results is that their derivation is linked to a new maximum theorem, which extends Berge's maximum theorem to possibly noncompact action sets.

For average-cost MDPs, the single additional assumption is that the relative value function is well defined. This assumption also holds for inventory control applications and can be verified easily.

The tutorial also describes optimality conditions for partially observable Markov decision processes with total discounted costs. These conditions imply the existence of optimal policies, validity of optimality equations, and convergence of value iterations. The results are illustrated with inventory control models for which some of the inventory levels are not observable.

The described results and methods are useful and insightful for investigating new and existing inventory control problems. As an illustration, a complete classification of possible solutions for the classic periodic-review stochastic single-product problem is described.

Acknowledgments

Some of the materials presented in this tutorial are based on results of work partially supported by the National Science Foundation [Grants CMMI-1335296 and CMMI-1636193]. The author thanks Jefferson Huang, Pavlo O. Kasyanov, Mark E. Lewis, Yan Liang, and Matthew J. Sobel for valuable comments.

Appendix. Berge's Maximum Theorem for Noncompact Action Sets and Some Properties of $\mathbb{K}$-Inf-Compact Functions

This appendix describes generalizations of Berge's maximum theorem and the relevant Berge theorem on semicontinuity of the value function to possibly noncompact action sets. These theorems are important for control theory, games, and mathematical economics. The major limitation of these theorems is that they require compact action sets. The generalizations provided in Feinberg et al. [30, 32] remove this limitation. Here, we present these results for metric spaces. With slight modifications, they hold for Hausdorff topological spaces (see Feinberg et al. [30]), but this level of generality is not needed for the results of this tutorial. Local versions of the results presented in this appendix can be found in Feinberg and Kasyanov [24].

Let $\mathbb{S}^1$ and $\mathbb{S}^2$ be metric spaces, $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \bar{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, and $\Phi\colon \mathbb{S}^1 \to 2^{\mathbb{S}^2} \setminus \{\varnothing\}$. Consider an optimization problem of the form

$$v(s^1) := \inf_{s^2 \in \Phi(s^1)} u(s^1, s^2), \quad \text{for each } s^1 \in \mathbb{S}^1, \tag{54}$$

which appears, for instance, in optimal control and game theory. Let $\mathbb{K}(\mathbb{S}^2)$ be the set of nonempty compact subsets of $\mathbb{S}^2$. Berge's theorem has the following formulation.

Berge's Theorem (Berge [10, p. 116]). *If $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \bar{\mathbb{R}}$ is a lower semicontinuous function and $\Phi\colon \mathbb{S}^1 \to \mathbb{K}(\mathbb{S}^2)$ is an upper semicontinuous, set-valued mapping, then the function $v\colon \mathbb{S}^1 \to \bar{\mathbb{R}}$ is lower semicontinuous.*

The well-known Berge's maximum theorem has the following formulation.

Berge's Maximum Theorem (Berge [10, p. 116]). *If $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \mathbb{R}$ is a continuous function and $\Phi\colon \mathbb{S}^1 \to \mathbb{K}(\mathbb{S}^2)$ is a continuous set-valued mapping, then the value function $v\colon \mathbb{S}^1 \to \mathbb{R}$ is continuous, and the solution multifunction $\Phi^*\colon \mathbb{S}^1 \to 2^{\mathbb{S}^2} \setminus \{\varnothing\}$, defined as*

$$\Phi^*(s^1) = \{s^2 \in \Phi(s^1)\colon v(s^1) = u(s^1, s^2)\}, \quad s^1 \in \mathbb{S}^1, \tag{55}$$

is upper semicontinuous and compact valued.

For an $\bar{\mathbb{R}}$-valued function f, defined on a nonempty subset U of a topological space $\mathbb{U}$, consider the level sets

$$\mathcal{D}_f(\lambda; U) = \{y \in U\colon f(y) \le \lambda\}, \quad \lambda \in \mathbb{R}.$$

We recall that a function f is called *inf-compact* (also sometimes called *lower semicompact*) on U if all the level sets $\mathcal{D}_f(\lambda; U)$ are compact. The following definition deals with the space $\mathbb{U} = \mathbb{S}^1 \times \mathbb{S}^2$ and its subsets $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$ and $\mathrm{Gr}_K(\Phi)$.

Definition A.1 (Feinberg et al. [32, Definition 1.1]). A function $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \bar{\mathbb{R}}$ is called $\mathbb{K}$-inf-compact on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$ if for every compact subset K of $\mathbb{S}^1$ this function is inf-compact on $\mathrm{Gr}_K(\Phi)$.

The following two theorems generalize Berge's theorem and Berge's maximum theorem, respectively, to possibly noncompact action sets.

Theorem A.2 (Feinberg et al. [32, Theorem 1.2]). *If the function $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \bar{\mathbb{R}}$ is $\mathbb{K}$-inf-compact on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$, then the function $v\colon \mathbb{S}^1 \to \bar{\mathbb{R}}$ is lower semicontinuous.*

Theorem A.2 (Feinberg et al. [30, Theorem 1.2]). *Assume that*
(i) $\Phi\colon \mathbb{S}^1 \to 2^{\mathbb{S}^2} \setminus \{\varnothing\}$ *is lower semicontinuous.*
(ii) $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \mathbb{R}$ *is $\mathbb{K}$-inf-compact and upper semicontinuous on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$.*
Then the value function $v : \mathbb{S}^1 \to \mathbb{R}$ is continuous, and the solution multifunction $\Phi^*\colon \mathbb{S}^1 \to \mathbb{K}(\mathbb{S}^2)$ is upper semicontinuous and compact valued.

The first statement of the following lemma implies that Theorems A.2 and A.3 are indeed generalizations of Berge's theorem and Berge's maximum theorem, respectively. The second statement indicates that the class of $\mathbb{K}$-inf-compact functions is broader than the class of inf-compact functions.

Lemma A.4 (Feinberg et al. [32, Lemma 2.1]). *The following statements hold:*
(i) *If $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \bar{\mathbb{R}}$ is lower semicontinuous on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$ and $\Phi\colon \mathbb{S}^1 \to \mathbb{K}(\mathbb{S}^2)$ is upper semicontinuous, then the function $u(\cdot,\cdot)$ is $\mathbb{K}$-inf-compact on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$.*
(ii) *If $u\colon \mathbb{S}^1 \times \mathbb{S}^2 \to \bar{\mathbb{R}}$ is inf-compact on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$, then the function $u(\cdot,\cdot)$ is $\mathbb{K}$-inf-compact on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$.*

Luque-Vásquez and Hernández-Lerma [45] provided an example with $\mathbb{S}^1 = \mathbb{R}$, $\mathbb{S}^2 = \Phi(s^1) = [0, \infty)$, continuous Φ, and continuous $u(s^1, s^2)$, which is inf-compact in s^2, where $v(s^1)$ is not lower semicontinuous. The following two lemmas indicate that $\mathbb{K}$-inf-compactness of u is stronger than its lower semicontinuity and inf-compactness in s^2.

Lemma A.5 (Feinberg et al. [32, Lemma 2.2]). *If $u(\cdot,\cdot)$ is a $\mathbb{K}$-inf-compact function on $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$, then for every $s^1 \in \mathbb{S}^1$, the function $u(s^1, \cdot)$ is inf-compact on $\Phi(s^1)$.*

Lemma A.6 (Feinberg et al. [32, Lemma 2.3]). *A* $\mathbb{K}$*-inf-compact function* $u(\cdot,\cdot)$ *on* $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$ *is lower semicontinuous on* $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$.

The following lemma provides the necessary and sufficient condition for $\mathbb{K}$-inf-compactness. This condition is used in Assumption W* in Feinberg et al. [31] instead of equivalent Definition A.1.

Lemma A.7 (Feinberg et al. [32, Lemma 2.5]). *The function* $u(\cdot,\cdot)$ *is* $\mathbb{K}$*-inf-compact on* $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$ *if and only if the following two conditions hold:*

(i) $u(\cdot,\cdot)$ *is lower semicontinuous on* $\mathrm{Gr}_{\mathbb{S}^1}(\Phi)$.

(ii) *If a sequence* $\{s_n^1\}_{n=1,2,\ldots}$ *with values in* $\mathbb{S}^1$ *converges and its limit* s^1 *belongs to* $\mathbb{S}^1$, *then any sequence* $\{s_n^2\}_{n=1,2,\ldots}$ *with* $s_n^2 \in \Phi(s_n^1)$, $n = 1,2,\ldots$, *satisfying the condition that the sequence* $\{u(s_n^1, s_n^2)\}_{n=1,2,\ldots}$ *is bounded above has a limit point* $s^2 \in \Phi(s^1)$.

References

[1] M. Aoki. Optimal control of partially observable Markovian systems. *Journal of the Franklin Institute* 280(5):367–386, 1965.

[2] K. J. Åström. Optimal control of Markov decision processes with incomplete state estimation. *Journal of Mathematical Analysis and Applications* 10(1):174–205, 1965.

[3] N. Bäuerle and U. Rieder. *Markov Decision Processes with Applications to Finance*. Springer, New York, 2011.

[4] A. Bensoussan. *Stochastic Control of Partially Observable Systems*. Cambridge University Press, Cambridge, UK, 1992.

[5] A. Bensoussan. *Dynamic Programming and Inventory Control*. IOS Press, Amsterdam, 2011.

[6] A. Bensoussan, M. Çakanyıldırım, and S. P. Sethi. Filtering for discrete-time Markov processes and applications to inventory control with incomplete information. D. Crisan and B. Rozovskii, eds. *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, New York, 500–525, 2011.

[7] A. Bensoussan, M. Çakanyıldırım, and S. P. Sethi. Partially observed inventory systems: The case of zero-balance walk. *SIAM Journal on Control and Optimization* 46(1):176–209, 2007.

[8] A. Bensoussan, M. Çakanyıldırım, J. A. Minjárez-Sosa, and S. P. Sethi. Partially observed inventory systems: The case of rain checks. *SIAM Journal on Control and Optimization* 47(5):2490–2519, 2008.

[9] A. Bensoussan, M. Çakanyıldırım, S. P. Sethi, and R. Shi. An incomplete information inventory model with presence of inventories or backorders as only observations. *Journal of Optimization Theory and Applications* 146(3):544–580, 2010.

[10] C. Berge. *Topological Spaces*. Macmillan, New York, 1963.

[11] D. P. Bertsekas. *Dynamic Programming and Optimal Control*, 2nd ed. Vol. 1. Athena Scientific, Belmont, MA, 2000.

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

[13] D. Beyer and S. P. Sethi. The classical average-cost inventory models of Iglehart and Veinott-Wagner revisited. *Journal of Optimization Theory and Applications* 101(3):523–555, 1999.

[14] D. Blackwell. Discrete dynamic programming. *Annals of Mathematical Statistics* 33(2): 719–726, 1962.

[15] D. Blackwell. Discounted dynamic programming. *Annals of Mathematical Statistics* 36(1): 226–235, 1965.

[16] R. Cavazos-Cadena. A counterexample on the optimality equation in Markov decision chains with the average cost criterion. *Systems and Control Letters* 16(5):387–392, 1991.

[17] X. Chen and D. Simchi-Levi. Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The finite horizon case. *Operations Research* 52(6): 387–392, 2004.

[18] X. Chen and D. Simchi-Levi. Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The infinite horizon case. *Mathematics of Operations Research* 29(3):698–723, 2004.

[19] C. Derman. On sequential decisions and Markov chains. *Management Science* 9(1):16–24, 1962.

[20] C. Derman. Denumerable state Markovian decision processes. *Annals of Mathematical Statistics* 37(6):1545–1553, 1966.

[21] E. B. Dynkin. Controlled random sequences. *Theory of Probability and Its Applications* 10(1): 1–14, 1965.

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

[23] A. Federgruen and P. Zipkin. An inventory model with limited production capacity and uncertain demands I. The average-cost criterion. *Mathematics of Operations Research* 11(2):193–207, 1986.

[24] E. A. Feinberg and P. O. Kasyanov. Continuity of minima: Local results. *Set-Valued and Variational Analysis* 23(3):485–499, 2015.

[25] E. A. Feinberg and M. E. Lewis. Optimality inequalities for average cost Markov decision processes and the stochastic cash balance problem. *Mathematics of Operations Research* 32(4): 769–783, 2007.

[26] E. A. Feinberg and M. E. Lewis. On the convergence of optimal actions for Markov decision processes and the optimality of (s, S) policies for inventory control. Preprint arXiv:1507.05125, http://arxiv.org/pdf/1507.05125.pdf, 2015.

[27] E. A. Feinberg and Y. Liang. On the optimality equation for average cost Markov decision processes and its validity for inventory control. Preprint arXiv:1609.03984, http://arxiv.org/pdf/1609.08252.pdf, 2016.

[28] E. A. Feinberg and Y. Liang. Structure of optimal solutions to periodic-review total-cost inventory control problems. Preprint arXiv:1609.03984, http://arxiv.org/pdf/1609.03984.pdf, 2016.

[29] E. A. Feinberg and A. Shwartz, eds. *Handbook of Markov Decision Processes: Methods and Applications*. Kluwer, Boston, 2002.

[30] E. A. Feinberg, P. O. Kasyanov, and M. Voorneveld. Berge's maximum theorem for noncompact image sets. *Journal of Mathematical Analysis and Applications* 413(2):1040–1046, 2014.

[31] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk. Average cost Markov decision processes with weakly continuous transition probabilities. *Mathematics of Operations Research* 37(4):591–607, 2012.

[32] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk. Berge's theorem for noncompact image sets. *Journal of Mathematical Analysis and Applications* 37(1):255–259, 2013.

[33] E. A. Feinberg, P. O. Kasyanov, and N. V. Zadoianchuk. Fatou's lemma for weakly converging probabilities. *Theory of Probability and Its Applications* 58(4):683–689, 2014.

[34] E. A. Feinberg, P. O. Kasyanov, and M. Z. Zgurovsky. Convergence of probability measures and Markov decision models with incomplete information. *Proceedings of the Steklov Institute of Mathematics* 287(1):96–117, 2014.

[35] E. A. Feinberg, P. O. Kasyanov, and M. Z. Zgurovsky. Uniform Fatou's lemma. *Journal of Mathematical Analysis and Applications* 444(1):550–567, 2016.

[36] E. A. Feinberg, P. O. Kasyanov, and M. Z. Zgurovsky. Partially observable total-cost Markov decision processes with weakly continuous transition probabilities. *Mathematics of Operations Research* 41(2):656–681, 2016.

[37] H.-J. Girlich and A. Chikán. The origins of dynamic inventory modelling under uncertainty: (The men, their work and connection with the Stanford Studies). *International Journal of Production Economics* 71(1–3):351–363, 2001.

[38] O. Hernández-Lerma. *Adaptive Markov Control Processes*. Springer, New York, 1989.

[39] O. Hernández-Lerma. Average optimality in dynamic programming on Borel spaces—Unbounded costs and controls. *Systems and Control Letters* 17(5):237–242, 1991.

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

[41] O. Hernández-Lerma and J. B. Lasserre. *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York, 1999.

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

[43] W. T. Huh, G. Janakiraman, and M. Nagarajan. Average cost single-stage inventory models: An analysis using a vanishing discount approach. *Operations Research* 59(1):143–155, 2011.

[44] D. L. Iglehart. Dynamic programming and stationary analysis of inventory problems. H. Scarf, D. Gilford, and M. Shelly, eds. *Multistage Inventory Models and Techniques*. Stanford University Press, Stanford, CA, 1-31, 1963.

[45] F. Luque-Vásques and O. Hernández-Lerma. A counterexample on the semicontinuity of minima. *Proceedings of the American Mathematical Society* 123(10):3175–3176, 1995.

[46] E. Porteus. *Foundations of Stochastic Inventory Theory*. Stanford University Press, Stanford, CA, 2002.

[47] M. L. Puterman. *Markov Decision Processes*. John Wiley & Sons, Hoboken, NJ, 2005.

[48] D. Rhenius. Incomplete information in Markovian decision models. *Annals of Statistics* 2(6): 1327–1334, 1974.

[49] S. M. Ross. *Introduction to Stochastic Dynamic Programming*. Academic Press, New York, 1983.

[50] H. Scarf. The optimality of (s, S) policies in the dynamic inventory problem. K. Arrow, S. Karlin, and P. Suppes, eds. *Mathematical Methods in the Social Sciences*. Stanford University Press, Stanford, CA, 196–202, 1959.

[51] M. Schäl. Conditions for optimality in dynamic programming and for the limit of n-stage optimal policies to be optimal. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete* 32(3):179–196, 1975.

[52] M. Schäl. Average optimality in dynamic programming with general state space. *Mathematics of Operations Research* 18(1):163–172, 1993.

[53] L. I. Sennott. A new condition for the existence of optimal stationary policies in average cost Markov decision processes. *Operations Research Letters* 5(1):17–23, 1986.

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

[55] L. I. Sennott. Average reward optimization theory for denumerable state spaces. E. A. Feinberg and A. Shwartz, eds. *Handbook of Markov Decision Processes: Methods and Applications*. Kluwer, Boston, 153–172, 2002.

[56] C. Shaoxiang. The infinite horizon periodic review problem with setup costs and capacity constraints: A partial characterization of the optimal policy. *Operations Research* 52(3): 409–421, 2004.

[57] L. S. Shapley. Stochastic games. *Proceedings of the National Academy of Sciences of the United States of America* 39(10):1095–1100, 1953.

[58] A. N. Shiryaev. Some new results in the theory of controlled random processes. *Selected Translations in Mathematical Statistics and Probability* 8(6):49–130, 1969.

[59] D. Simchi-Levi, X. Chen, and J. Bramel. *The Logic of Logistics: Theory, Algorithms, and Applications for Logistics and Supply Chain Management*. Springer, New York, 2005.

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

[61] C. Striebel. *Optimal Control for Discrete Time Stochastic Systems*. Springer-Verlag, Berlin, 1975.

[62] H. M. Taylor III. Markovian sequential replacement processes. *Annals of Mathematical Statistics* 36(4):1677–1694, 1965.

[63] A. F. Veinott. On the optimality of (s, S) inventory policies: New conditions and new proof. *SIAM Journal on Applied Mathematics* 14(5):1067–1083, 1966.

[64] A. F. Veinott and H. M. Wagner. Computing optimal (s, S) policies. *Management Science* 11(5):525–552, 1965.

[65] A. A. Yushkevich. Reduction of a controlled Markov model with incomplete data to a problem with complete information in the case of Borel state and control spaces. *Theory of Probability and Its Applications* 21(1):153–158, 1976.

[66] E. Zabel. A note on the optimality of (S, s) policies in inventory theory. *Management Science* 9(1):123–125, 1962.

[67] Y. Zheng. A simple proof for the optimality of (s, S) policies in infinite horizon inventory systems. *Journal of Applied Probability*, 28(4):802–810, 1991.

[68] P. H. Zipkin. *Foundations of Inventory Management*. McGraw-Hill, New York, 2000.