

Average Cost Optimality Inequality for Markov Decision Processes with Borel Spaces and Universally Measurable Policies^{*†}

Huizhen Yu[‡]

Abstract

We consider average-cost Markov decision processes (MDPs) with Borel state and action spaces and universally measurable policies. For the nonnegative cost model and an unbounded cost model with a Lyapunov-type stability character, we introduce a set of new conditions under which we prove the average cost optimality inequality (ACOI) via the vanishing discount factor approach. Unlike most existing results on the ACOI, our result does not require any compactness and continuity conditions on the MDPs. Instead, the main idea is to use the almost-uniform-convergence property of a pointwise convergent sequence of measurable functions as asserted in Egoroff’s theorem. Our conditions are formulated in order to exploit this property. Among others, we require that for each state, on selected subsets of actions at that state, the state transition stochastic kernel is majorized by finite measures. We combine this majorization property of the transition kernel with Egoroff’s theorem to prove the ACOI.

Keywords:

Markov decision processes; Borel spaces; universally measurable policies;
average cost; optimality inequality; majorization conditions

^{*}This research was funded by DeepMind, AMII and Alberta Innovates—Technology Futures (AITF).

[†]This paper consists of the main results originally given in the author’s arXiv eprint [54, Section 3] and newly added sections for an extended discussion and illustrative examples.

[‡]RLAI Lab, Department of Computing Science, University of Alberta, Canada (janey.hzyu@gmail.com)

Contents

1	Introduction	3
2	Preliminaries	5
2.1	Definitions of some Sets and Functions	5
2.2	Borel-space MDPs	5
2.3	Some Basic Optimality Properties of Two Models	7
3	Main Results: The ACOI	9
3.1	The (UC) Case	9
3.1.1	Majorization Condition	9
3.1.2	Optimality Results	11
3.2	The (PC) Case	13
3.2.1	Majorization Condition	14
3.2.2	Optimality Result	15
4	Further Discussion and Illustrative Examples	16
4.1	About Assumption 3.3(i), Epi-limits of Functions, and Their Relation with ACOIs .	17
4.2	About Assumption 3.3(ii), Weak Sequential Compactness, and Alternative Proof of Lemmas 3.8, 3.15	18
4.3	About the Assumption $\sup_{\alpha \in (0,1)} \ h_\alpha\ _w < \infty$ in the (UC) Case	19
4.4	About Geometric Ergodicity Conditions, ACOI, and ACOE in the (UC) Case	21
4.5	An Illustrative Example for the (PC) Case	21
4.5.1	Two Helpful Lemmas	22
4.5.2	Single-Product Inventory System	22
Appendix A	Supplementary Materials for Section 4	26
A.1	Proofs of Proposition 4.1 and Lemma 4.5	26
A.2	Derivations of (4.19)-(4.20) for the Proof of Proposition 4.7	27
A.3	An Illustrative Example for the (UC) Case	28
A.3.1	Definitions of Action Sets	29
A.3.2	Bounding $\{h_\alpha\}$	30
A.3.3	Satisfying the Majorization Condition	31
	References	32

1 Introduction

We consider discrete-time Markov decision processes (MDPs) with Borel state and action spaces, under the average cost criterion where the objective is to minimize the (limsup) expected long-run average cost per unit time. Specifically, we consider the universal measurability framework, which involves lower semi-analytic (l.s.a.) one-stage cost functions and universally measurable (u.m.) policies. It is a mathematical formulation of MDPs developed to resolve measurability difficulties in dynamic programming on Borel spaces [4, 5, 43, 44, 45, 47]. An in-depth study of this theoretical framework is given in the monograph [3, Part II], and optimality properties of finite- and infinite-horizon problems with discounted and undiscounted total cost criteria have been analyzed (see e.g., [3, Part II] and [31, 53, 56]). The average cost problem has not been thoroughly studied in this framework, however; there are only a few prior results (which we will discuss below). The primary purpose of this paper is to investigate the subject further, and our central focus will be on the average cost optimality inequality (ACOI).

The study of ACOI was initiated by Sennott [42], who proved it for countable-space MDPs; prior to [42], the ACOE (average cost optimality equation) was the research focus. Cavazos-Cadena's counterexample [6] showed that the ACOI is more general: in his example, the ACOI has a solution and yet the ACOE does not. For Borel-space MDPs, the ACOI was first established by Schäl [41] for one-stage costs that are bounded below and under, alternatively, two types of compactness and continuity conditions. Specifically, the first (resp. second) case requires that the state transition stochastic kernel is continuous with respect to (w.r.t.) setwise convergence (resp. weak convergence) and the one-stage cost function lower semicontinuous (l.s.c.) in the action variable (resp. the state and action variables). The compactness conditions on the action sets in [41] were weakened by Hernández-Lerma [20] (to inf-compactness for the first case) and by Feinberg, Kasyanov, and Zadoianchuk [14] (to \mathbb{K} -inf-compactness for the second case; a similar, slightly stronger condition was proposed by Costa and Dufour [8] independently). Extensions of these results to unbounded one-stage costs and to ACOEs have also been studied; see the textbook accounts given in Hernández-Lerma and Lasserre [22, Chap. 10] and for more recent advances, see Vega-Amaya [49, 50, 51], Jaśkiewicz and Nowak [28], Feinberg et al. [13, 16] and the references therein. (For related earlier researches, see also the survey by Arapostathis et al. [1].)

In this paper, we prove the ACOI for two types of MDPs in the universal measurability framework: the nonnegative cost model and an unbounded cost model with a Lyapunov-type stability property. These models have been studied in the references just mentioned. As in some of those studies, to prove the ACOI, we use the vanishing discount factor approach, which treats the average cost problem as the limiting case of the discounted problems, and we adopt some boundedness conditions formulated in [14, 22, 41] regarding the optimal value functions of the discounted problems.

Different from those studies, however, our results require *no compactness and continuity conditions* on the one-stage cost function and state transition stochastic kernel. Instead, we introduce a set of new conditions of, what we call, the majorization type: among others, we require that for each state, on selected subsets of actions at that state, the state transition stochastic kernel is majorized by finite measures (see Assumptions 3.3, 3.12). Our main idea is to combine these majorization properties with Egoroff's theorem, which asserts that pointwise convergence of functions is "almost" uniform convergence as measured by a given finite measure (cf. footnote 8), and which allows us to extract arbitrarily large sets (large as measured by the majorizing finite measures) on which certain functions involved in our analyses have desired uniform convergence properties. With this technique, we obtain the ACOI for the two MDP models mentioned above (see Theorems 3.5, 3.13). These results can be applied to a class of MDPs with discontinuous dynamics and one-stage costs.

For comparison, let us discuss several prior researches relevant to average-cost MDPs with u.m. policies. Dynkin and Yushkevich [11, Chap. 7.9] and Piunovskiy [36] studied the characteristic properties of canonical systems—a general form of the ACOE together with stationary policies that solve or almost solve the ACOE. ([11, Chap. 7.9] considers an abstract model with desired measurable

structures; [36] considers the universal measurability framework.)

Gubenko and Shtatland used a contraction-based fixed point approach to prove the ACOE under, alternatively, a minorization condition and a majorization on the state transition stochastic kernel [19, Thms. 2, 2'] (measurability issues are assumed away in these theorems). Their majorization condition not only differs in essential ways from ours but is also too stringent to be practical (see Remark 3.4 for details). But their approach with the minorization condition is a fruitful one and has close connections with ergodic theory for Markov chains; in particular, their minorization condition implies that the MDP is uniformly geometrically ergodic (cf. Section 4.3). Their contraction argument was extended by Kurano [30] to a multistep-contraction argument for average-cost MDPs with u.m. policies. The main results of [19, 30] on the ACOE concern the case of bounded one-stage costs.

For unbounded one-stage costs, a contraction-based fixed point approach to proving the ACOE was proposed more recently. It was originally formulated by Vega-Amaya [49], building on the result of Hernández-Lerma and Lasserre on positive Harris recurrent Markov chains, as well as on the prior research of Gordienko and Hernández-Lerma [18], which relates a class of MDPs with unbounded one-stage costs to w -geometrically ergodic Markov chains. In [49] (also the recent work [51]), this fixed point approach is applied to MDPs with continuity/compactness properties. The approach was taken by Jaśkiewicz [27] to establish the ACOE for semi-Markovian decision processes (SMDPs) in the universal measurability framework. The conditions of [27, 49] generalize the minorization condition of [19], and the ACOE results of [27, 49] are applicable to a class of MDPs that are uniformly w -geometrically ergodic w.r.t. certain weight functions w on the state spaces (cf. the discussion in Section 4.4).

Meyn [33] studied the convergence of policy iteration for average-cost MDPs. He obtained the ACOE and the existence of a stationary average-cost optimal policy, under a set of conditions on the Markov chains induced by stationary policies that could be generated by the policy iteration algorithm, together with a continuity condition on the differential cost functions of those policies and the one-stage cost function. A large part of his analysis is based on general state space Markov chain theory and requires no compactness/continuity conditions. However, it does not directly apply to Borel-space MDPs in the universal measurability framework due to known measurability issues in policy iteration. It is still an open problem to extend the method of [33] to average-cost MDPs with u.m. policies.

We remark that these prior studies differ significantly from our work in both the approaches taken and the results obtained. When compactness/continuity conditions are absent, a major tool used in many of those studies is ergodic theory for Markov chains, whereas, by exploiting Egoroff's theorem, our analyses and results can be applied to non-ergodic MDPs (cf. the examples in Section 4 and Appendix A.3).

Let us also mention two related results from our separate recent work. In [55], we formulated another condition of the majorization type to make use of Lusin's theorem when applying a direct method, the minimum pair approach, to study the average cost problems; this result is for countable discrete action spaces and strictly unbounded one-stage costs. In [54, sect. 2.2] (see also Theorem 2.2), for the two MDP models considered in this paper, we gave a characterization of the structures of the optimal cost functions and optimal/ ϵ -optimal policies in the average cost problems, without any extra conditions.

The rest of this paper is organized as follows. In Section 2, we give background materials about Borel-space MDPs. In Section 3, we propose new majorization type conditions and prove the ACOI for two MDP models. Further discussion and illustrative examples are given in Section 4 and Appendix A.

2 Preliminaries

In this section, we first introduce Borel-space MDPs in the universal measurability framework. To prepare the stage for subsequent analyses, we then review several basic optimality properties for the nonnegative cost and unbounded cost models we consider, under the average and discounted cost criteria. We start with certain sets and functions that lie at the foundation of Borel-space MDPs.

2.1 Definitions of some Sets and Functions

We consider separable metrizable spaces. A *Borel space* (a.k.a. standard Borel space) is a separable metrizable space that is homeomorphic to a Borel subset of some Polish space [3, Def. 7.7]. For a Borel space X , let $\mathcal{B}(X)$ denote the Borel σ -algebra and $\mathcal{P}(X)$ the set of probability measures on $\mathcal{B}(X)$ (we will call them Borel probability measures). With the topology of weak convergence, $\mathcal{P}(X)$ is also a Borel space [3, Chap. 7.4]. Each $p \in \mathcal{P}(X)$ has a unique extension on a larger σ -algebra $\mathcal{B}_p(X)$ generated by $\mathcal{B}(X)$ and all the subsets of X with p -outer measure 0, and this extension is called the *completion of p* (cf. [9, Chap. 3.3]). The *universal σ -algebra* on X is defined as $\mathcal{U}(X) := \bigcap_{p \in \mathcal{P}(X)} \mathcal{B}_p(X)$. $\mathcal{U}(X)$ -measurable functions on X are thus measurable w.r.t. the completion of any Borel probability measure; these functions are called *u.m. (universally measurable)*.

If X and Y are Borel spaces, a *Borel or u.m. stochastic kernel* on Y given X is a function $q : X \rightarrow \mathcal{P}(Y)$ that is measurable from the space $(X, \mathcal{B}(X))$ or $(X, \mathcal{U}(X))$, respectively, to the space $(\mathcal{P}(Y), \mathcal{B}(\mathcal{P}(Y)))$; see [3, Def. 7.12, Prop. 7.26, Lem. 7.28]. We use the notation $q(dy|x)$ for the stochastic kernel. When q is a continuous function, we say the stochastic kernel is *continuous* (a.k.a. *weakly continuous* or *weak Feller* in the literature).

An *analytic set* in a Polish space is the image of a Borel subset of some Polish space under a Borel measurable function (cf. [3, Prop. 7.41], [9, sect. 13.2]). A function $f : D \rightarrow [-\infty, \infty]$ is called *l.s.a. (lower semi-analytic)*, if D is an analytic set and for every $a \in \mathbb{R}$, the level set $\{x \in D \mid f(x) \leq a\}$ of f is analytic [3, Def. 7.21]. An equivalent definition is that the epigraph of f , $\{(x, a) \mid x \in D, f(x) \leq a, a \in \mathbb{R}\}$, is analytic (cf. [3, p. 186]). For comparison, f is *l.s.c. (lower semicontinuous)* if its epigraph is closed. A Borel measurable extended-real-valued function on a Borel space is l.s.a. and an l.s.a. function is u.m., since in a Polish space every Borel set is analytic and every analytic set is u.m. ([3, Cor. 7.42.1], [9, Thm. 13.2.6]).

The properties of analytic sets give rise to many properties of l.s.a. functions that are important for dynamic programming in Borel-space MDPs. The most critical is the Jankov-von Neumann measurable selection theorem [3, Prop. 7.49], which asserts that for an analytic set D in the product space $X \times Y$ of two Borel spaces, with $\text{proj}_X(D)$ being the projection of D on X , there exists an analytically measurable¹ function $\phi : \text{proj}_X(D) \rightarrow Y$ such that the graph of ϕ lies in D . This theorem gives rise to a measurable selection theorem for partial minimization of l.s.a. functions on product spaces [3, Prop. 7.50]. For Borel-space MDPs, these properties of analytic sets and l.s.a. functions are closely related to the validity of value iteration, the structure of the optimal value functions, and the existence of optimal or nearly optimal policies and their structures, some of which we will discuss in Sections 2.2-2.3. Due to space limit, however, we do not list these properties and will provide references where we use them in this paper.²

2.2 Borel-space MDPs

In the universal measurability framework, a Borel-space MDP has the following elements and model assumptions (cf. [3, Chap. 8.1]):

- The state space \mathbb{X} and the action space \mathbb{A} are *Borel spaces*.

¹I.e., measurable w.r.t. the σ -algebra generated by the analytic sets.

²For l.s.a. functions, we refer the reader to the papers [5, 31, 44] and the monograph [3, Chap. 7]; for general properties of analytic sets, see also the books [11, Appendix 2] and [35, 46].

- The control constraint is specified by a set-valued map $A : x \mapsto A(x)$, where for each state $x \in \mathbb{X}$, $A(x) \subset \mathbb{A}$ is a nonempty set of admissible actions at that state, and the graph of $A(\cdot)$, $\Gamma = \{(x, a) \mid x \in \mathbb{X}, a \in A(x)\} \subset \mathbb{X} \times \mathbb{A}$, is *analytic*.
- The one-stage cost function $c : \Gamma \rightarrow [-\infty, +\infty]$ is *l.s.a.*
- State transitions are governed by $q(dy \mid x, a)$, a *Borel measurable* stochastic kernel on \mathbb{X} given $\mathbb{X} \times \mathbb{A}$.

We consider infinite-horizon control problems. A policy consists of a sequence of stochastic kernels on \mathbb{A} that specify for each stage, which admissible actions to apply, given the history up to that stage. In particular, a *u.m. policy* is a sequence $\pi = (\mu_0, \mu_1, \dots)$, where for each $k \geq 0$, $\mu_k(da_k \mid x_0, a_0, \dots, a_{k-1}, x_k)$ is a u.m. stochastic kernel on \mathbb{A} given $(\mathbb{X} \times \mathbb{A})^k \times \mathbb{X}$ and obeys the control constraint of the MDP:

$$\mu_k(A(x_k) \mid x_0, a_0, \dots, a_{k-1}, x_k) = 1 \quad \forall (x_0, a_0, \dots, a_{k-1}, x_k) \in (\mathbb{X} \times \mathbb{A})^k \times \mathbb{X}. \quad (2.1)$$

(As Γ is analytic, the sets $A(x)$ are u.m. [3, Lem. 7.29]; the probability of $A(x_k)$ here is measured w.r.t. the completion of $\mu_k(da_k \mid x_0, a_0, \dots, a_{k-1}, x_k)$.) A policy π is *Borel measurable* if each component μ_k is a Borel measurable stochastic kernel; π is then also u.m. by definition. (A Borel measurable policy, however, may not exist [4].) We define the policy space Π of the MDP to be the set of u.m. policies. We shall simply refer to these policies as policies, dropping the term “u.m.,” if there is no confusion or no need to emphasize their measurability.

We define several subclasses of policies in the standard way. A policy π is *nonrandomized* if $\mu_k(da_k \mid x_0, a_0, \dots, a_{k-1}, x_k)$ is a Dirac measure that assigns probability one to a single action in $A(x_k)$, for every $(x_0, a_0, \dots, a_{k-1}, x_k)$ and $k \geq 0$. A policy π is *semi-Markov* if for every $k \geq 0$, the function $(x_0, a_0, \dots, a_{k-1}, x_k) \mapsto \mu_k(da_k \mid x_0, a_0, \dots, a_{k-1}, x_k)$ depends only on (x_0, x_k) ; *Markov* if for every $k \geq 0$, that function depends only on x_k ; *stationary* if π is Markov and $\mu_k = \mu$ for all $k \geq 0$. For the stationary case, we simply write μ for $\pi = (\mu, \mu, \dots)$. A nonrandomized stationary policy μ can also be viewed as a function that maps each $x \in \mathbb{X}$ to an action in $A(x)$, so for such μ , we will use both notations $\mu(x)$, $\mu(da \mid x)$ in the paper.

Because the graph Γ of the control constraint $A(\cdot)$ is analytic, by the Jankov-von Neumann selection theorem [3, Prop. 7.49], there exists at least one u.m. nonrandomized stationary policy. Thus the policy space Π is non-empty.

We consider the average cost criterion and the discounted cost criterion. By [3, Prop. 7.45], given a policy $\pi \in \Pi$ and an initial state distribution $p_0 \in \mathcal{P}(\mathbb{X})$, the collection of stochastic kernels $\mu_0(da_0 \mid x_0)$, $q(dx_1 \mid x_0, a_0)$, $\mu_1(da_1 \mid x_0, a_0, x_1)$, $q(dx_2 \mid x_1, a_1)$, \dots determines uniquely a probability measure on the universal σ -algebra on $(\mathbb{X} \times \mathbb{A})^\infty$. The *n-stage value function* and the *average cost function* of π are defined, respectively, by³

$$J_n(\pi, x) := \mathbb{E}_x^\pi \left[\sum_{k=0}^{n-1} c(x_k, a_k) \right], \quad J(\pi, x) := \limsup_{n \rightarrow \infty} J_n(\pi, x)/n, \quad x \in \mathbb{X},$$

where \mathbb{E}_x^π denotes expectation w.r.t. the probability measure induced by π and the initial state $x_0 = x$. Define the *optimal average cost function* by

$$g^*(x) := \inf_{\pi \in \Pi} J(\pi, x) = \inf_{\pi \in \Pi} \limsup_{n \rightarrow \infty} J_n(\pi, x)/n, \quad x \in \mathbb{X}.$$

For $0 < \alpha < 1$, define the α -discounted value function of a policy π by

$$v_\alpha^\pi(x) := \limsup_{n \rightarrow \infty} \mathbb{E}_x^\pi \left[\sum_{k=0}^{n-1} \alpha^k c(x_k, a_k) \right], \quad x \in \mathbb{X},$$

³In general, for a u.m. function $f : (\mathbb{X} \times \mathbb{A})^\infty \rightarrow [-\infty, +\infty]$, define $\mathbb{E}f := \mathbb{E}f^+ - \mathbb{E}f^-$ where $f^+ = \max\{0, f\}$ and $f^- = -\min\{0, f\}$; if $\mathbb{E}f^+ = \mathbb{E}f^- = +\infty$, we adopt the convention $\infty - \infty = -\infty + \infty = \infty$. In the MDPs of our interest, however, we will not encounter such summations.

and the *optimal α -discounted value function* by

$$v_\alpha(x) := \inf_{\pi \in \Pi} v_\alpha^\pi(x), \quad x \in \mathbb{X}.$$

The functions $J_n(\pi, \cdot)$, $J(\pi, \cdot)$ and $v_\alpha^\pi(\cdot)$ are u.m. by [3, Prop. 7.46, Lem. 7.30(2)]. As will be discussed in the next subsection, for the two MDP models we consider, the optimal cost functions g^* and v_α are l.s.a.

Let $\mathcal{M}(\mathbb{X})$ (resp. $\mathcal{A}(\mathbb{X})$) denote the space of extended-real-valued u.m. (resp. l.s.a.) functions on \mathbb{X} . For $0 < \alpha \leq 1$, define *dynamic programming operators* T_α that map $v \in \mathcal{M}(\mathbb{X})$ to a function on \mathbb{X} according to⁴

$$(T_\alpha v)(x) := \inf_{a \in A(x)} \left\{ c(x, a) + \alpha \int_{\mathbb{X}} v(y) q(dy \mid x, a) \right\}, \quad x \in \mathbb{X}.$$

For $\alpha = 1$, we simply write T for T_α . By the properties of analytic sets and l.s.a. functions (cf. [3, Chap. 7]), T_α and T map $\mathcal{A}(\mathbb{X})$ into $\mathcal{A}(\mathbb{X})$.

The following subclasses of u.m. and l.s.a. functions will be needed shortly. For a u.m. function $w : \mathbb{X} \rightarrow (0, +\infty)$, which we shall refer to as a *weight function*, let

$$\mathcal{M}_w(\mathbb{X}) := \{f \mid \|f\|_w < \infty, f \in \mathcal{M}(\mathbb{X})\}, \quad \text{where } \|f\|_w := \sup_{x \in \mathbb{X}} |f(x)|/w(x).$$

Note that $(\mathcal{M}_w(\mathbb{X}), \|\cdot\|_w)$ is a Banach space, and $\mathcal{A}(\mathbb{X}) \cap \mathcal{M}_w(\mathbb{X})$ is a closed subset of this space.⁵

2.3 Some Basic Optimality Properties of Two Models

We consider two classes of MDPs. The first is the nonnegative cost model where the one-stage cost function $c \geq 0$. We shall refer to this model as (PC) in what follows. For the average cost or discounted problem, it is equivalent to the case where c is bounded from below.

The second model, designated as (UC), involves unbounded costs: the function c can be unbounded from below or above, but it needs to satisfy a growth condition and moreover, there is a Lyapunov-type condition on the dynamics of the MDP. The precise definition is as follows.

Definition 2.1 (the model (UC)). *There exist a u.m. weight function $w(\cdot) \geq 1$ and constants $b, \hat{c} \geq 0$ and $\lambda \in [0, 1)$ such that for all $x \in \mathbb{X}$,*

- (a) $\sup_{a \in A(x)} |c(x, a)| \leq \hat{c} w(x);$
- (b) $\sup_{a \in A(x)} \int_{\mathbb{X}} w(y) q(dy \mid x, a) \leq \lambda w(x) + b.$

For (UC), its definition ensures that the average cost function of any policy π satisfies $\|J(\pi, \cdot)\|_w \leq \ell$ for the constant $\ell = \hat{c}b/(1 - \lambda)$. Hence the optimal average cost function also satisfies $\|g^*\|_w \leq \ell$ and in particular, g^* is finite everywhere. For (PC), $g^* \geq 0$ and it is possible that at some state x , $g^*(x) = +\infty$ (this possibility will be eliminated in Section 3 under further assumptions on the MDP model).

For both (PC) and (UC), the average-cost MDPs have the general optimality properties given in the next theorem, which is proved by the author [54, sects. 2.2, A.2]. In what follows, by an ϵ -optimal or optimal policy (with no mention of an initial state), we mean a policy that is ϵ -optimal or optimal for *all initial states*.

⁴Here and throughout the paper, the integration of a u.m. function w.r.t. $p \in \mathcal{P}(\mathbb{X})$ is defined w.r.t. the completion of p (cf. Section 2.1).

⁵This is because convergence in the $\|\cdot\|_w$ norm implies pointwise convergence, and pointwise limits of a sequence of l.s.a. functions are l.s.a. [3, Lem. 7.30(2)].

Theorem 2.2 (average-cost optimality results; [54, Thm. 2.1]). (PC)(UC)

- (i) *The optimal average cost function g^* is l.s.a.*
- (ii) *For each $\epsilon > 0$, there exists a (u.m.) randomized semi-Markov ϵ -optimal policy. If there exists an optimal policy for each state $x \in \mathbb{X}$, then there exists a (u.m.) randomized semi-Markov optimal policy.*

In Section 3, we will use the vanishing discount factor approach to prove the ACOI for (PC) and (UC) under additional conditions. That analysis starts with the optimality equations for the α -discounted cost criteria (α -DCOE) given below:

Theorem 2.3 (the α -DCOE). (PC)(UC) *For $\alpha \in (0, 1)$, the optimal value function v_α is l.s.a. and satisfies the α -DCOE $v_\alpha = T_\alpha v_\alpha$, i.e.,*

$$v_\alpha(x) = \inf_{a \in A(x)} \left\{ c(x, a) + \alpha \int_{\mathbb{X}} v_\alpha(y) q(dy \mid x, a) \right\}, \quad x \in \mathbb{X}.$$

For (PC), v_α is the smallest nonnegative solution of the α -DCOE in $\mathcal{A}(\mathbb{X})$; for (UC), v_α is the unique solution of the α -DCOE in the space $\mathcal{A}(\mathbb{X}) \cap \mathcal{M}_w(\mathbb{X})$. Furthermore, in both cases, for each $\epsilon > 0$, there exists a nonrandomized stationary ϵ -optimal policy.

Remark 2.4 (about the proof of Theorem 2.3). Under certain compactness and continuity conditions, proofs of the α -DCOE for (PC) and (UC) can be found in e.g., the papers [14, 40] and the books [21, Chap. 5], [22, Chap. 8]. In our case, Theorem 2.3 is proved by using the results of [3, Part II] for general Borel-space MDPs. Specifically, for (PC), this theorem is implied by the optimality results for the nonnegative model [3, Props. 9.8, 9.10, 9.19]. For (UC), it can be shown⁶ that for some u.m. weight function $\tilde{w} \geq w$, the operator T_α is a contraction on the closed subset $\mathcal{A}(\mathbb{X}) \cap \mathcal{M}_{\tilde{w}}(\mathbb{X})$ of the Banach space $(\mathcal{M}_{\tilde{w}}(\mathbb{X}), \|\cdot\|_{\tilde{w}})$, with contraction modulus $\beta \in (\alpha, 1)$. We use this contraction property of T_α together with the correspondence between the MDP and the so-called deterministic control model (DM) defined in [3, Chap. 9] to prove Theorem 2.3 for (UC). The proof can be found in [54, Appendix A.3].⁷ \square

As another preparation for Section 3, the next lemma states an implication of the ACOI on the existence and structure of average-cost optimal or nearly optimal policies. For comparison, note that in the general case where g^* need not be constant, one can only assert the existence of a randomized semi-Markov ϵ -optimal policy (cf. Theorem 2.2). The proof of this lemma uses mostly standard arguments and can be found in [54, Appendix A.4].

Lemma 2.5 (a consequence of ACOI). *Consider the models (PC) and (UC) with the average cost criterion. Suppose that the optimal average cost function g^* is constant and finite. Suppose also that for some real-valued $h \in \mathcal{A}(\mathbb{X})$, with $h \geq 0$ for (PC) and $\|h\|_w < \infty$ for (UC), the ACOI holds: $g^* + h \geq Th$, i.e.,*

$$g^* + h(x) \geq \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} h(y) q(dy \mid x, a) \right\}, \quad x \in \mathbb{X}. \quad (2.2)$$

Then there exist a nonrandomized Markov optimal policy and, for each $\epsilon > 0$, a nonrandomized stationary ϵ -optimal policy. If, in addition, the infimum in the right-hand side of the ACOI is attained for every $x \in \mathbb{X}$, then there exists a nonrandomized stationary optimal policy.

⁶See [54, Lem. A.2, Appendix A.4] for the proof. The construction of the weight function $\tilde{w} \geq w$ and the proof of this contraction property are similar to the analysis given in [22, pp. 45–46].

⁷The proof is similar to, but does not follow exactly the one given in [3, Chap. 9] for bounded one-stage costs; see [54, Remark A.1, Appendix A.3] for further explanations, including when one can reduce the case to that of bounded costs by using Veinott's similarity transformation [48, 52].

3 Main Results: The ACOI

In this section, we introduce our majorization condition for the (PC) and (UC) models and study the ACOI via the vanishing discount factor approach. The arguments for (PC) and (UC) are similar but differ in details, so we will discuss the two models separately. We consider (UC) first.

3.1 The (UC) Case

Let \bar{x} be some fixed state and consider the relative value functions of the α -discounted problems:

$$h_\alpha(x) := v_\alpha(x) - v_\alpha(\bar{x}), \quad x \in \mathbb{X}. \quad (3.1)$$

Take a sequence $\alpha_n \uparrow 1$ such that for some $\rho^* \in \mathbb{R}$,

$$(1 - \alpha_n) v_{\alpha_n}(\bar{x}) \rightarrow \rho^* \quad \text{as } n \rightarrow \infty. \quad (3.2)$$

(This is possible because the model conditions of (UC) imply that for each $x \in \mathbb{X}$, $(1 - \alpha) v_\alpha(x)$ is bounded in α over $(0, 1)$; cf. [22, sect. 10.4.A].) Consider the corresponding sequence of functions $h_n := h_{\alpha_n}$. Define

$$\underline{h} := \liminf_{n \rightarrow \infty} h_n, \quad \underline{h}_n := \inf_{m \geq n} h_m, \quad n \geq 0, \quad (3.3)$$

$$\bar{h} := \limsup_{n \rightarrow \infty} h_n, \quad \bar{h}_n := \sup_{m \geq n} h_m, \quad n \geq 0. \quad (3.4)$$

Note that as $n \rightarrow \infty$, $\underline{h}_n \uparrow \underline{h}$ and $\bar{h}_n \downarrow \bar{h}$.

Assumption 3.1. *For the model (UC), the set of functions $\{h_\alpha \mid \alpha \in (0, 1)\}$ as defined above is bounded in $\mathcal{M}_w(\mathbb{X})$, i.e., $\sup_{\alpha \in (0, 1)} \|h_\alpha\|_w < \infty$.*

This assumption is extracted from the analysis of the ACOI given in the book [22, Chap. 10.4], where the MDP is assumed to be uniformly w -geometrically ergodic, which ensures $\sup_{\alpha \in (0, 1)} \|h_\alpha\|_w < \infty$. In Section 4.3 we shall discuss this w -geometric ergodicity condition as well as other types of sufficient conditions for Assumption 3.1.

The next lemma follows directly from Theorem 2.3 and [3, Lem. 7.30(2)]:

Lemma 3.2. (UC) *Under Assumption 3.1, all the functions \underline{h} , \bar{h} , \underline{h}_n , \bar{h}_n , $n \geq 0$, are l.s.a. and lie in a bounded subset of $\mathcal{M}_w(\mathbb{X})$.*

3.1.1 Majorization Condition

We now introduce our majorization condition for the (UC) model. The first two parts of this condition will also be used for the (PC) model later. Let $\mathbb{1}(\cdot)$ denote the indicator function.

Assumption 3.3. *In (UC), for each $x \in \mathbb{X}$ and $\epsilon > 0$, the following hold:*

- (i) *There exist a subset $K_\epsilon(x) \subset A(x)$ and $0 < \bar{\alpha} < 1$ such that for all $\alpha \in [\bar{\alpha}, 1)$,*

$$\inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha \int_{\mathbb{X}} v_\alpha(y) q(dy \mid x, a) \right\} \leq v_\alpha(x) + \epsilon. \quad (3.5)$$

- (ii) *There exists a finite measure ν on $\mathcal{B}(\mathbb{X})$ such that*

$$\sup_{a \in K_\epsilon(x)} q(B \mid x, a) \leq \nu(B) \quad \forall B \in \mathcal{B}(\mathbb{X}). \quad (3.6)$$

(iii) The weight function $w(\cdot)$ is uniformly integrable w.r.t. $\{q(dy | x, a) \mid a \in K_\epsilon(x)\}$ in the sense that

$$\lim_{\ell \rightarrow \infty} \sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} w(y) \mathbb{1}[w(y) \geq \ell] q(dy | x, a) = 0. \quad (3.7)$$

Note that in the above, $\bar{\alpha}$ and ν can depend on x and ϵ . Note also that the one-stage cost function $c(\cdot)$ is not required to have any special properties.

Assumption 3.3(i) is motivated by the theory on epi-convergence of functions, in particular, by the relation between the infima of epi-limits and pointwise limits of a sequence of functions and the implication of this relation on the validity of the ACOI in MDPs. In some circumstances, the existence of a compact subset $K_\epsilon(x)$ with the property (3.5) is close to being necessary for the ACOI. We shall elaborate on this in Section 4.1 (cf. Proposition 4.1, Cor. 4.2) after we prove the ACOI, since the roles of the above conditions can be better seen then.

The choice of the subset $K_\epsilon(x)$ of actions in Assumption 3.3(i) is important for the subsequent parts (ii)-(iii) of this assumption. Although (i) is trivially satisfied if we let $K_\epsilon(x) = A(x)$, this makes it harder or sometimes impossible to satisfy (ii)-(iii). For instance, if $A(x) = \mathbb{R}$ and $q(dy | x, a)$ is the normal distribution $\mathcal{N}(a, 1)$ on \mathbb{R} , no finite measure can majorize these distributions for all $a \in \mathbb{R}$. On the other hand, identifying a proper subset of actions that satisfies (i) is in general a difficult problem, because it requires the knowledge of the family of optimal value functions $\{v_\alpha\}$. An additional difficulty in the (UC) case is that, unlike in the (PC) case, the cost function $c(x, \cdot)$ here must be bounded for each state x , preventing us from using “coercivity” of $c(x, \cdot)$ to eliminate actions. (For this reason, in the illustrative example given in Appendix A.3 for the (UC) model, we will set $K_\epsilon(x) = A(x)$.)

Assumption 3.3(ii) is the key condition that allows us to apply Egoroff’s theorem in proving the ACOI. The class of MDPs for which Assumption 3.3(ii) can hold naturally are those where for all $a \in K_\epsilon(x)$, $q(dy | x, a)$ has a density $f_{x,a}$ w.r.t. a common (σ -finite) reference measure φ . The pointwise supremum of the density functions, $f_x := \sup_{a \in K_\epsilon(x)} f_{x,a}$ (or a measurable function that upper-bounds it), when it is integrable w.r.t. φ , defines a finite measure ν , with $d\nu = f_x d\varphi$, that has the desired majorization property (3.6).

Assumption 3.3(ii) need not be satisfied by continuous state transition stochastic kernels. For example, if $A(x) = [0, 1]$ and $q(dy | x, a) = \delta_a$ (the Dirac measure at a), there is no finite measure with the desired majorization property (there is also no nontrivial measure ν that can satisfy the minorization condition $q(dy | x, a) \geq \nu(dy)$ for all a here). This is not a “defect” in the majorization condition but a reflection of the difference in nature of topological and measure-theoretic properties; a deeper discussion of Assumption 3.3(ii) will be given in Section 4.2 to show what it entails. Due to this difference, however, there are some inevitable limitations in the analysis technique we present in this paper. We shall discuss this subject further in the example in Section 4.5 (cf. Remark 4.11).

Regarding Assumption 3.3(iii), roughly speaking, in the proof of the ACOI, we will use it to handle what is leftover after the application of Egoroff’s theorem. Verifying this condition can be straightforward when the weight function $w(\cdot)$ has a simple analytical expression (e.g., when $x \in \mathbb{R}$ and $w(x) = e^x$ or x^2) and the properties of the integrals involved are known. For example, if $w(\cdot)$ is bounded from above on the union of the supports of the probability measures $q(dy | x, a)$, $a \in K_\epsilon(x)$, such as in the case where the union is contained in a compact set and $w(\cdot)$ is continuous, then Assumption 3.3(iii) is clearly satisfied. If $c(\cdot)$ is bounded, $w(\cdot)$ can be chosen to be constant and Assumption 3.3(iii) then holds trivially. Assumption 3.3(iii) is also implied by the slightly stronger condition $\int w d\nu < \infty$, where ν is the majorizing finite measure in Assumption 3.3(ii).

Let us end this discussion by clarifying the relation between our majorization condition and the one from Gubenko and Shtatland’s early work. Later we will discuss further the conditions involved in our proof of the ACOI and give illustrative examples in Section 4 and Appendix A.3.

Remark 3.4 (about the majorization condition in [19]). In their contraction-based fixed point approach, Gubenko and Shtatland’s majorization condition [19, sect. 3, Condition (II)] is like a sym-

metric counterpart of their minorization condition, and it requires that there exists a finite measure ν on $\mathcal{B}(\mathbb{X})$ such that

$$q(B \mid x, a) \leq \nu(B) \quad \forall B \in \mathcal{B}(\mathbb{X}), (x, a) \in \Gamma, \quad \text{and} \quad \nu(\mathbb{X}) < 2. \quad (3.8)$$

Note that here the same measure ν needs to majorize $q(dy \mid x, a)$ for all states and admissible actions, whereas in our Assumption 3.3, ν can be different for each state. The requirement $\nu(\mathbb{X}) < 2$ (needed for converting T into a contraction) is too stringent and renders their condition (3.8) impractical. \square

3.1.2 Optimality Results

We now prove the ACOI for (UC) under the assumptions introduced earlier. Recall that the relative value functions $\{h_n\}$, the functions \underline{h} , $\{\underline{h}_n\}$, \bar{h} , $\{\bar{h}_n\}$, and also the scalar $\rho^* = \lim_{n \rightarrow \infty} (1 - \alpha_n) v_{\alpha_n}(\bar{x})$ are defined in (3.1)-(3.4).

Theorem 3.5 (the ACOI for (UC)). *Under Assumptions 3.1, 3.3 for the (UC) model, the optimal average cost function $g^*(\cdot) = \rho^*$ and with $\underline{h} \in \mathcal{A}(\mathbb{X}) \cap \mathcal{M}_w(\mathbb{X})$ as given in (3.3), the pair (ρ^*, \underline{h}) satisfies the ACOI:*

$$\rho^* + \underline{h}(x) \geq \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy \mid x, a) \right\}, \quad x \in \mathbb{X}. \quad (3.9)$$

Hence there exist a nonrandomized Markov optimal policy and for each $\epsilon > 0$, a nonrandomized stationary ϵ -optimal policy.

Note that this theorem implies that ρ^* does not depend on our choice of the sequence $\{\alpha_n\}$ or the state \bar{x} , and $\lim_{\alpha \rightarrow 1} (1 - \alpha) v_{\alpha}(x) = g^*$ (the constant optimal average cost) for all $x \in \mathbb{X}$.

This theorem can be compared with the prior ACOI result for (UC) with compactness/continuity conditions given in the book [22, Thm. 10.3.1].

In the case of bounded one-stage costs, another existing result due to Ross [38, Thm. 2] says that when \mathbb{A} is finite, if $\{h_{\alpha}\}$ is uniformly bounded and equicontinuous, then the ACOE holds for a continuous function h and constant g^* . For comparison, treating the case of bounded one-stage costs as a special case of (UC) with $w(\cdot) \equiv 1$, we have the following corollary from Theorem 3.5:

Corollary 3.6. *In an MDP with bounded $c(\cdot)$, suppose that for each $x \in \mathbb{X}$, there is a finite measure ν on $\mathcal{B}(\mathbb{X})$ such that $\sup_{a \in A(x)} q(B \mid x, a) \leq \nu(B)$ for all $B \in \mathcal{B}(\mathbb{X})$. Then, if $\{h_{\alpha}\}$ is uniformly bounded, the ACOI (3.9) holds.*

Remark 3.7. It will be obvious from the proof of Theorem 3.5 that if we have $\underline{h} = \lim_{n \rightarrow \infty} h_n$ in addition, then the ACOI (3.9) holds with equality. The extra condition on the convergence of $\{h_n\}$ can be stringent and hard to check in practice, however. A discussion about existing results on the ACOE in the (UC) case and whether Theorem 3.5 can be strengthened to ACOE will be given in Section 4.3. \square

We now proceed to prove Theorem 3.5. First, we prove an important lemma, in which we combine the majorization condition with Egoroff's theorem. (An alternative proof of this lemma will be given in Section 4.2.)

Lemma 3.8. (UC) *Let Assumptions 3.1, 3.3(ii)-(iii) hold, and let $K_{\epsilon}(x) \subset A(x)$ be the set in the latter condition for a given $x \in \mathbb{X}$ and $\epsilon > 0$. Then*

$$\lim_{n \rightarrow \infty} \inf_{a \in K_{\epsilon}(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} \underline{h}_n(y) q(dy \mid x, a) \right\} = \inf_{a \in K_{\epsilon}(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy \mid x, a) \right\}.$$

Proof. For the state x , $\epsilon > 0$, and the set $K_\epsilon(x)$ given in the lemma, let ν be the corresponding finite measure on $\mathcal{B}(\mathbb{X})$ in Assumption 3.3(ii). Recall that $\underline{h}_n \uparrow \underline{h}$ and these functions are real-valued and u.m. (Lemma 3.2). Therefore, by Egoroff's Theorem [9, Thm. 7.5.1],⁸ for any $\delta > 0$, there exists a u.m. set $D_\delta \subset \mathbb{X}$ with $\nu(\mathbb{X} \setminus D_\delta) < \delta$ such that on the set D_δ , \underline{h}_n converges to \underline{h} uniformly as $n \rightarrow \infty$. Consequently, for any $\eta > 0$, it holds for all n sufficiently large that

$$\int_{D_\delta} (\underline{h}(y) - \underline{h}_n(y)) q(dy \mid x, a) \leq \eta \quad \forall a \in A(x). \quad (3.10)$$

We now bound the integral of $\underline{h} - \underline{h}_n$ on the complement set $\mathbb{X} \setminus D_\delta$. By Lemma 3.2, for all $n \geq 0$, $\|\underline{h} - \underline{h}_n\|_w \leq \ell$ for some constant ℓ . So for all $a \in A(x)$,

$$\int_{\mathbb{X} \setminus D_\delta} (\underline{h}(y) - \underline{h}_n(y)) q(dy \mid x, a) \leq \ell \int_{\mathbb{X} \setminus D_\delta} w(y) q(dy \mid x, a).$$

By the choice of D_δ and the majorization property of ν in Assumption 3.3(ii),

$$\sup_{a \in K_\epsilon(x)} q(\mathbb{X} \setminus D_\delta \mid x, a) \leq \nu(\mathbb{X} \setminus D_\delta) < \delta. \quad (3.11)$$

By an alternative characterization of uniform integrability [9, Thm. 10.3.5], (3.11) together with the uniform integrability condition in Assumption 3.3(iii) implies that for any given $\eta > 0$, it holds for all δ sufficiently small that $\int_{\mathbb{X} \setminus D_\delta} w(y) q(dy \mid x, a) \leq \eta$ for all $a \in K_\epsilon(x)$. Therefore, given $\eta > 0$, by choosing a small enough δ , we can make

$$\sup_{a \in K_\epsilon(x)} \int_{\mathbb{X} \setminus D_\delta} (\underline{h}(y) - \underline{h}_n(y)) q(dy \mid x, a) \leq \eta. \quad (3.12)$$

Combining this with (3.10), we obtain that for all n sufficiently large,

$$\sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} (\underline{h}(y) - \underline{h}_n(y)) q(dy \mid x, a) \leq 2\eta$$

and hence

$$\inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} \underline{h}_n(y) q(dy \mid x, a) \right\} \geq \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} \underline{h}(y) q(dy \mid x, a) \right\} - 2\eta.$$

Since η is arbitrary and $\underline{h}_n \leq \underline{h}$, the lemma follows by letting $n \rightarrow \infty$ on both sides of the preceding inequality and using also the fact that

$$(1 - \alpha_n) \sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} |\underline{h}(y)| q(dy \mid x, a) \rightarrow 0$$

since $\sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} |\underline{h}(y)| q(dy \mid x, a) < \infty$ by Assumption 3.1 and the model condition of (UC). \square

Proof of Theorem 3.5. For each $x \in \mathbb{X}$ and $\epsilon > 0$, by the α -DCOE (Theorem 2.3) and Assumption 3.3(i), for all n sufficiently large

$$(1 - \alpha_n) v_{\alpha_n}(\bar{x}) + h_n(x) = \inf_{a \in A(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} h_n(y) q(dy \mid x, a) \right\} \quad (3.13)$$

$$\begin{aligned} &\geq \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} h_n(y) q(dy \mid x, a) \right\} - \epsilon \\ &\geq \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} \underline{h}_n(y) q(dy \mid x, a) \right\} - \epsilon, \end{aligned} \quad (3.14)$$

⁸Egoroff's Theorem [9, Thm. 7.5.1]: Let (X, \mathcal{S}, ν) be finite measure space. Let f_n and f be measurable functions from X into a separable metric space. Suppose $f_n(x) \rightarrow f(x)$ for ν -almost all x . Then for any $\delta > 0$, there is a set D with $\nu(X \setminus D) < \delta$ such that $f_n \rightarrow f$ uniformly on D .

where the last inequality used the fact $\underline{h}_n \leq h_n$ and that \underline{h}_n is u.m. (Lemma 3.2). Letting $n \rightarrow \infty$ in both sides of (3.14), we have

$$\begin{aligned} \rho^* + \underline{h}(x) + \epsilon &\geq \liminf_{n \rightarrow \infty} \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} \underline{h}_n(y) q(dy | x, a) \right\} \\ &= \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) \right\} \\ &\geq \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) \right\}, \end{aligned}$$

where the equality follows from Lemma 3.8. Since this holds for every $x \in \mathbb{X}$ and ϵ is arbitrary, the desired inequality (3.9) is proved.

To show $g^*(\cdot) = \rho^*$, as in the analysis in [28], it suffices to show that for the pair (ρ^*, \bar{h}) , where $\bar{h} = \limsup_{n \rightarrow \infty} h_n$ as we recall, the opposite inequality holds:

$$\rho^* + \bar{h}(x) \leq \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \bar{h}(y) q(dy | x, a) \right\}, \quad x \in \mathbb{X}. \quad (3.15)$$

This will yield $g^*(\cdot) = \rho^*$ because the inequality (3.15) implies that for every state x , every policy π has average cost $J(\pi, x) \geq \rho^*$, whereas the inequality (3.9) just proved implies the existence of a policy π with $J(\pi, x) \leq \rho^*$ (the proofs of these two facts are standard). From the α -DCOE (3.13), we have that for all $a \in A(x)$,

$$(1 - \alpha_n) v_{\alpha_n}(\bar{x}) + h_n(x) \leq c(x, a) + \alpha_n \int_{\mathbb{X}} h_n(y) q(dy | x, a), \quad (3.16)$$

so by taking limit supremum of both sides as $n \rightarrow \infty$ and using also the fact $\bar{h}_n \geq h_n$ by definition, we have

$$\rho^* + \bar{h}(x) \leq c(x, a) + \limsup_{n \rightarrow \infty} \alpha_n \int_{\mathbb{X}} \bar{h}_n(y) q(dy | x, a) = c(x, a) + \int_{\mathbb{X}} \bar{h}(y) q(dy | x, a),$$

where the last equality follows from the dominated convergence theorem, in view of Lemma 3.2 and the model condition of (UC). This proves (3.15) and hence $g^*(\cdot) = \rho^*$ as discussed earlier. Finally, that $\underline{h} \in \mathcal{A}(\mathbb{X}) \cap \mathcal{M}_w(\mathbb{X})$ follows from Lemma 3.2, and the existence of a Markov optimal policy and stationary ϵ -optimal policy follows from the ACOI proved above and Lemma 2.5. \square

The simple ACOE result mentioned in Remark 3.7 follows from the ACOI just proved and the inequality (3.15) in the proof above since $\underline{h} = \bar{h}$ in that case.

3.2 The (PC) Case

We now consider the nonnegative cost model (PC). Let

$$m_\alpha := \inf_{x \in \mathbb{X}} v_\alpha(x), \quad \alpha \in (0, 1).$$

We shall work under the following assumption taken from the ACOI literature:

Assumption 3.9. *For the model (PC):*

- (G) *For some policy π and state x , the average cost $J(\pi, x) < \infty$.*
- (B) *For every $x \in \mathbb{X}$, $\liminf_{\alpha \uparrow 1} (v_\alpha(x) - m_\alpha) < \infty$.*

Remark 3.10. Precursors to these conditions were introduced in the early works of Sennott [42] and Schäl [41] and then evolved as the research progressed. The condition (G) is the same as that in [41] and by [41, Lem. 1.2(b)], it implies that

$$\limsup_{\alpha \rightarrow 1} (1 - \alpha) m_\alpha \leq \inf_{x \in \mathbb{X}} g^*(x) < \infty.$$

The condition (B) was introduced by Feinberg, Kasyanov, and Zadoianchuk [14] to weaken the condition (B) in [41], which is

$$\sup_{\alpha \in (0,1)} (v_\alpha(x) - m_\alpha) = \sup_{\alpha \in (0,1)} h_\alpha(x) < \infty \quad \forall x \in \mathbb{X}. \quad (3.17)$$

Under (G), (3.17) is equivalent to $\limsup_{\alpha \uparrow 1} h_\alpha(x) < \infty$ for $x \in \mathbb{X}$ [14, Lem. 5]. Sennott first showed in her work on countable-space MDPs [42] that in proving the ACOI, it suffices to require the family of relative value functions to be bounded from above pointwise instead of uniformly. \square

Let

$$h_\alpha := v_\alpha - m_\alpha, \quad \underline{h} := \liminf_{\alpha \rightarrow 1} h_\alpha. \quad (3.18)$$

For each $\alpha \in (0, 1)$, define a function \underline{h}_α as

$$\underline{h}_\alpha(x) := \inf_{\beta \in [\alpha, 1)} h_\beta(x), \quad x \in \mathbb{X}. \quad (3.19)$$

Note that

$$\underline{h}_\alpha \leq \underline{h}_\beta \quad \forall \alpha \leq \beta < 1, \quad \text{and} \quad \underline{h}_\alpha \uparrow \underline{h} \quad \text{as } \alpha \uparrow 1. \quad (3.20)$$

Before proceeding, let us prove that these functions are l.s.a., so that we can take their integrals in the subsequent analysis.

Lemma 3.11. (PC) *Under Assumption 3.9, the functions \underline{h} and \underline{h}_α , $\alpha \in (0, 1)$, are real-valued and l.s.a.*

Proof. Clearly, the functions \underline{h} and \underline{h}_α are real-valued under Assumption 3.9(B). In order to prove that they are l.s.a., let us treat the relative value functions h_α , $\alpha \in (0, 1)$, as a function of (α, x) . We have proved in [54, Lem. A.1, Appendix A.4] that $v_\alpha(x)$ is an l.s.a. function of (α, x) on $(0, 1) \times \mathbb{X}$ (this proof uses the deterministic control model corresponding to the MDP mentioned in Remark 2.4 in Section 2.3). Next, consider m_α as a function of α . Since the one-stage cost $c \geq 0$ in (PC), for each policy π and initial state x , the α -discounted value $v_\alpha^\pi(x)$ is non-decreasing as α increases. Therefore, $m_\alpha = \inf_{x \in \mathbb{X}} v_\alpha(x)$ is also monotonically non-decreasing as α increases. It follows that m_α is a Borel measurable function of α on $(0, 1)$ and so is $-m_\alpha$. Then, by [3, Lem. 7.30(4)], $h_\alpha(x) = v_\alpha(x) - m_\alpha$ is l.s.a. in (α, x) . Now, since for each α , \underline{h}_α is the partial minimization of $h_\beta(x)$ over $\beta \in [\alpha, 1)$, \underline{h}_α is an l.s.a. function of x by [3, Prop. 7.47]. Finally, since $\underline{h}_\alpha \uparrow \underline{h}$ as $\alpha \uparrow 1$, we can write \underline{h} as the pointwise limit of \underline{h}_{α_n} for a sequence $\alpha_n \uparrow 1$, and therefore, \underline{h} is l.s.a. by [3, Lem. 7.30(2)]. \square

3.2.1 Majorization Condition

We now introduce a majorization condition for (PC). Its first two parts are the same as those of Assumption 3.3 for (UC).

Assumption 3.12. *In the model (PC), for each $x \in \mathbb{X}$ and $\epsilon > 0$,*

- (i) *Assumption 3.3(i) holds;*
- (ii) *Assumption 3.3(ii) holds;*
- (iii) *Assumption 3.3(iii) holds with \underline{h} in place of the weight function w .⁹*

$$\lim_{\ell \rightarrow \infty} \sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} \underline{h}(y) \mathbb{1}[\underline{h}(y) \geq \ell] q(dy \mid x, a) = 0. \quad (3.21)$$

⁹The integrals of \underline{h} appearing here are well-defined since \underline{h} is u.m. by Lemma 3.11.

Recall that in Assumption 3.3(i), for each state x and $\epsilon > 0$, we need to choose a subset $K_\epsilon(x) \subset A(x)$ that contains ϵ -optimal actions of the minimization problems associated with the right-hand sides of the α -DCOE for all α sufficiently large. For the (PC) model, if $c(x, \cdot)$ is unbounded above but the family $\{h_\alpha\}$ is pointwise bounded (i.e., (3.17) holds), a strict subset of $A(x)$ can be found as a natural candidate for $K_\epsilon(x)$, without the need for detailed knowledge about $c(\cdot)$ or $\{h_\alpha\}$. For an illustrative example, see the example and the proof of Proposition 4.9 in Section 4.5.

About Assumption 3.12(iii), in practice, to verify it without knowing \underline{h} , one could first find an upper bound $\hat{h} \geq \underline{h}$ and then verify the condition (3.21) for \hat{h} instead:

$$\lim_{\ell \rightarrow \infty} \sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} \hat{h}(y) \mathbb{1}[\hat{h}(y) \geq \ell] q(dy | x, a) = 0.$$

An upper bound \hat{h} can be available as a byproduct of verifying the pointwise boundedness condition on the family $\{h_\alpha\}$ (either Assumption 3.9(B) or the stronger condition (3.17)). An illustrative example will be given in Section 4.5.

3.2.2 Optimality Result

The following theorem can be compared with the existing ACOI results that require continuity/compactness conditions [14, 41, 50].

Theorem 3.13 (the ACOI for (PC)). *For the (PC) model, suppose Assumptions 3.9 and 3.12 hold. Let $\rho^* = \limsup_{\alpha \uparrow 1} (1 - \alpha) m_\alpha$ and let the real-valued function $\underline{h} \in \mathcal{A}(\mathbb{X})$ be as given in (3.18). Then the optimal average cost function $g^*(\cdot) = \rho^*$, and the pair (ρ^*, \underline{h}) satisfies the ACOI:*

$$\rho^* + \underline{h}(x) \geq \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) \right\}, \quad x \in \mathbb{X}. \quad (3.22)$$

Hence there exist a nonrandomized Markov optimal policy and for each $\epsilon > 0$, a nonrandomized stationary ϵ -optimal policy.

Remark 3.14. The countable-state finite-action MDP in Cavazos-Cadena's counterexample for ACOE [6] satisfies both assumptions of Theorem 3.13. In particular, regarding the majorization condition, in the finite-action case, Assumption 3.12(i)-(ii) hold trivially for $K_\epsilon(x) = A(x)$; in that example $\underline{h} = 0$, so Assumption 3.12(iii) also holds. This shows that without further conditions, we cannot improve the ACOI result in Theorem 3.13 to the ACOE. \square

We now prove Theorem 3.13. The proof is similar to that of the ACOI in the (UC) case, but there are some subtle differences, one of which being that in this case, for different states, we choose possibly different sequences $\alpha_n \uparrow 1$ and work with the corresponding relative value functions h_{α_n} . We start with a lemma similar to Lemma 3.8 for (UC); it follows from Egoroff's theorem and our majorization condition.

Lemma 3.15. (PC) *Let Assumptions 3.9, 3.12(ii)-(iii) hold, and let $K_\epsilon(x) \subset A(x)$ be the set in the latter condition for a given $x \in \mathbb{X}$ and $\epsilon > 0$. Then for any sequence $\alpha_n \uparrow 1$, with $\underline{h}_n = \underline{h}_{\alpha_n}$, we have*

$$\lim_{n \rightarrow \infty} \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} \underline{h}_n(y) q(dy | x, a) \right\} = \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) \right\}.$$

Proof. Note first that under Assumption 3.9, $\underline{h}_n \uparrow \underline{h} < \infty$ as $n \rightarrow \infty$, and the integrals in the lemma are well-defined since $\underline{h}_n, \underline{h}$ are u.m. by Lemma 3.11. We follow the proof for Lemma 3.8 except for two small changes. The first change is that here, when dealing with the complement set $\mathbb{X} \setminus D_\delta$ after applying Egoroff's theorem, in order to obtain the bound (3.12):

$$\sup_{a \in K_\epsilon(x)} \int_{\mathbb{X} \setminus D_\delta} (\underline{h}(y) - \underline{h}_n(y)) q(dy | x, a) \leq \eta \quad \text{for all } n \text{ sufficiently large,}$$

for a given $\eta > 0$ and sufficiently small $\delta > 0$, we use the fact that $\underline{h} - \underline{h}_n \leq \underline{h}$ and we also use the uniform integrability condition in Assumption 3.12(iii) for the nonnegative function \underline{h} , instead of the weight function w . The second change is that near the end of the proof, to show that

$$(1 - \alpha_n) \sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

we now use the fact that $\sup_{a \in K_\epsilon(x)} \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) < \infty$, which is again implied by the uniform integrability condition on \underline{h} in Assumption 3.12(iii). \square

Proof of Theorem 3.13. Consider an arbitrary $x \in \mathbb{X}$ and fix it in the proof below. Choose a sequence $\{\alpha_n\}$ such that

$$\lim_{n \rightarrow \infty} h_{\alpha_n}(x) = \liminf_{\alpha \rightarrow 1} h_\alpha(x) = \underline{h}(x).$$

Note that $\rho^* \geq \limsup_{n \rightarrow \infty} (1 - \alpha_n) m_{\alpha_n}$. Recall also that $\rho^* \leq \inf_{y \in \mathbb{X}} g^*(y)$ by [41, Lem. 1.2(b)]. Define $h_n := h_{\alpha_n}$, $\underline{h}_n := \underline{h}_{\alpha_n}$. Then $\underline{h}_n \leq h_n$ and $\underline{h}_n \uparrow \underline{h}$ as $n \rightarrow \infty$.

Let $\epsilon > 0$. For the fixed state x , by subtracting αm_α from both sides of the α -DCOE (Theorem 2.3), we have that for all $n \geq 0$,

$$(1 - \alpha_n) m_{\alpha_n} + h_n(x) = \inf_{a \in A(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} h_n(y) q(dy | x, a) \right\}. \quad (3.23)$$

Then similarly to the derivation of (3.14) in the proof for Theorem 3.5, we apply Assumption 3.12(i) and replace h_n by \underline{h}_n in the right-hand side above before letting $n \rightarrow \infty$ in both sides of (3.23). This results in the inequality

$$\rho^* + \underline{h}(x) \geq \liminf_{n \rightarrow \infty} \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \alpha_n \int_{\mathbb{X}} \underline{h}_n(y) q(dy | x, a) \right\} - \epsilon \quad (3.24)$$

where $K_\epsilon(x)$ is the subset of actions in Assumption 3.12(i) for the fixed state x and $\epsilon > 0$. Next, by (3.24) and Lemma 3.15, we have

$$\begin{aligned} \rho^* + \underline{h}(x) + \epsilon &\geq \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) \right\} \\ &\geq \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a) \right\}. \end{aligned}$$

Since this holds for an arbitrary $x \in \mathbb{X}$ and an arbitrary $\epsilon > 0$, we obtain the desired ACOI. Since $\rho^* \leq g^*(\cdot)$ [41, Lem. 1.2(b)], the ACOI just established implies that we must have $g^*(\cdot) = \rho^*$ [41, Prop. 1.3]. Finally, the existence of a Markov average-cost optimal policy and stationary ϵ -optimal policy follows from Lemma 3.11, the ACOI proved above, and Lemma 2.5. This completes the proof. \square

4 Further Discussion and Illustrative Examples

In this section, we first explain the origin of the first part of our majorization condition and discuss an important implication of the second part of that condition (cf. Sections 4.1-4.2). We will then discuss the boundedness condition (Assumption 3.1) on the family of relative value functions in the (UC) case, a related w -geometric ergodicity condition, as well as stronger conditions for ensuring the ACOE (cf. Sections 4.3-4.4). Finally, we will illustrate the application of our ACOI result for the (PC) model using an inventory control example (cf. Section 4.5); a similar example for (UC) is given in Appendix A.3.

4.1 About Assumption 3.3(i), Epi-limits of Functions, and Their Relation with ACOIs

Assumption 3.3(i) is the first part of our majorization condition for both (UC) and (PC) models. In introducing this assumption, we have been influenced by the theory on epi-convergence of functions [37, Chap. 7] and more specifically, by the relation between the epi-limits and pointwise limits of sequences of functions and certain inequalities for the infima of those functions and their limits.

If $\{f_n\}$ is a sequence of extended-real-valued functions on a metric space Y , the *outer limit* of their epigraphs,¹⁰ $\limsup_{n \rightarrow \infty} \text{epi}(f_n)$, corresponds to the epigraph of some (extended-real-valued) function on Y . This function, denoted by $\underline{\text{e-lim}}_n f_n$, is called the *lower epi-limit* of $\{f_n\}$. By definition, $\underline{\text{e-lim}}_n f_n$ is always l.s.c. and lies below the pointwise limit $\liminf_{n \rightarrow \infty} f_n$, and moreover,

$$\liminf_{n \rightarrow \infty} \inf_{y \in Y} f_n(y) \leq \inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y) \leq \inf_{y \in Y} \liminf_{n \rightarrow \infty} f_n(y). \quad (4.1)$$

Thus, if we want the equality $\liminf_{n \rightarrow \infty} \inf_{y \in Y} f_n(y) = \inf_{y \in Y} \liminf_{n \rightarrow \infty} f_n(y)$, we will need $\liminf_{n \rightarrow \infty} \inf_{y \in Y} f_n(y) = \inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y)$ to hold at least.

The proposition below gives a necessary, sometimes sufficient, condition for the latter equality. It is similar to [37, Thm. 7.31(a)]; because [37] deals with finite-dimensional spaces only, for clarity, we include a proof of this result in Appendix A.1.

Proposition 4.1. *Let $\{f_n\}$ be a sequence of extended-real-valued functions on a metric space Y such that $\liminf_{n \rightarrow \infty} \inf_{y \in Y} f_n(y)$ is finite. If*

$$\liminf_{n \rightarrow \infty} \inf_{y \in Y} f_n(y) = \inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y), \quad (4.2)$$

then the following condition holds: for every $\epsilon > 0$, there exist a compact set $K \subset Y$ and a subsequence $\{f_{n_j}\}$ such that

$$\inf_{y \in K} f_{n_j}(y) \leq \inf_{y \in Y} f_{n_j}(y) + \epsilon \quad \text{for all } j \text{ sufficiently large.} \quad (4.3)$$

If $\lim_{n \rightarrow \infty} \inf_{y \in Y} f_n(y)$ exists and is finite, this condition implies (4.2).

Let us discuss now what these results imply for the validity of the ACOI in MDPs. Consider some fixed state $\hat{x} \in \mathbb{X}$. Let Assumption 3.1 or 3.9 hold for (UC) or (PC), respectively. By redefining, if necessary, the sequence of discount factors α_n in the proofs of the ACOI in Section 3, we can suppose that the corresponding relative value functions $h_n := h_{\alpha_n}$ are such that $\lim_{n \rightarrow \infty} h_n(\hat{x}) = \underline{h}(\hat{x})$ and also that $\lim_{n \rightarrow \infty} (1 - \alpha_n) m_{\alpha_n}$ exists if the model is (PC). Define functions f_n on $A(\hat{x})$ by

$$f_n(a) = c(\hat{x}, a) + \alpha_n \int_{\mathbb{X}} h_n(y) q(dy \mid \hat{x}, a), \quad a \in A(\hat{x}).$$

Then the following limit exists and is finite:

$$\lim_{n \rightarrow \infty} \inf_{a \in A(\hat{x})} f_n(a) = \rho(\hat{x}) + \underline{h}(\hat{x}),$$

where $\rho(\hat{x}) = \rho^*$ for (UC) and $\rho(\hat{x}) = \lim_{n \rightarrow \infty} (1 - \alpha_n) m_{\alpha_n} \leq \rho^*$ for (PC). Applying Proposition 4.1 and (4.1) to this sequence $\{f_n\}$ then gives the next corollary:

¹⁰The outer limit of a sequence $\{E_n\}$ of sets in a metric space is defined as $\limsup_{n \rightarrow \infty} E_n := \bigcap_{m \geq 1} \text{cl}(\bigcup_{n \geq m} E_n)$, where cl denotes the closure of a set. In our case, E_n is the epigraph of f_n , $\text{epi}(f_n) := \{(y, r) \mid f_n(y) \leq r, y \in Y, r \in \mathbb{R}\}$, and the set $\limsup_{n \rightarrow \infty} \text{epi}(f_n)$ consists of all the points $(y, r) \in Y \times \mathbb{R}$ such that $r \geq \lim_{k \rightarrow \infty} f_{n_k}(y_k)$ for some subsequence $\{f_{n_k}\}$ of $\{f_n\}$ and some sequence of points $y_k \rightarrow y$.

Corollary 4.2. *Let Assumption 3.1 (resp. 3.9) hold for the (UC) (resp. (PC)) model, and consider the preceding $\{\alpha_n\}$, $\{h_n\}$, \hat{x} and $\rho(\hat{x})$. Suppose that for some $\epsilon > 0$, there does not exist a compact subset $K \subset A(\hat{x})$ with the property that*

$$\inf_{a \in K} \left\{ c(\hat{x}, a) + \alpha_{n_j} \int_{\mathbb{X}} v_{\alpha_{n_j}}(y) q(dy \mid \hat{x}, a) \right\} \leq v_{\alpha_{n_j}}(\hat{x}) + \epsilon \quad (4.4)$$

for some subsequence $\{\alpha_{n_j}\}$ and all j sufficiently large. Then

$$\rho(\hat{x}) + \underline{h}(\hat{x}) < \inf_{a \in A(\hat{x})} \left\{ c(\hat{x}, a) + \liminf_{n \rightarrow \infty} \int_{\mathbb{X}} h_n(y) q(dy \mid \hat{x}, a) \right\}. \quad (4.5)$$

While the strict inequality (4.5) does not contradict the ACOI in general, it rules out the ACOI in the following situations, which could happen in a given problem: $h_n \uparrow \underline{h}$ and, if the model is (PC), $\rho(\hat{x}) = \rho^*$ in addition. In these cases, on the right-hand side of (4.5), we have $\liminf_{n \rightarrow \infty} \int_{\mathbb{X}} h_n(y) q(dy \mid \hat{x}, a) = \int_{\mathbb{X}} \underline{h}(y) q(dy \mid \hat{x}, a)$ for each a by the monotone convergence theorem, so (4.5) becomes

$$\rho^* + \underline{h}(\hat{x}) < \inf_{a \in A(\hat{x})} \left\{ c(\hat{x}, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy \mid \hat{x}, a) \right\},$$

violating the ACOI at the state \hat{x} . It is this observation that led us to introduce a condition like the inequality (4.4) as part of our majorization condition for the ACOI. The inequality (4.4) itself is unwieldy to verify, since, among others, it depends on the particular choice of the sequence $\{\alpha_n\}$. So we required a similar inequality to hold for all sufficiently large α instead, for each state x , and this gave rise to Assumption 3.3(i). (In the latter, the subset of actions is not required to be compact, for that is not needed in our proof of the ACOI.)

Finally, note that Proposition 4.1 only deals with the first relation in (4.1) and does not guarantee equality to hold throughout (4.1), for the second inequality in (4.1) can still be strict. Likewise, Assumption 3.3(i) alone (even with a nontrivial, compact strict subset $K_\epsilon(x)$ of $A(x)$) is in general not sufficient for the ACOI to hold. It is the two other parts of Assumption 3.3 that gave us the rest of the help needed.

4.2 About Assumption 3.3(ii), Weak Sequential Compactness, and Alternative Proof of Lemmas 3.8, 3.15

The existence of majorizing finite measures required by (3.6) in Assumption 3.3(ii) has the following important implication. For each state x , the set of probability measures, $\mathcal{Q}_\epsilon(x) := \{q(dy \mid x, a) \mid a \in K_\epsilon(x)\}$, viewed as a subset in the Banach space of finite signed Borel measures on \mathbb{X} (endowed with the total variation norm), must be *weakly sequentially compact*, i.e., sequentially compact¹¹ w.r.t. the weak topology on the latter space induced by its topological dual [24, Prop. 1.4.4, Cor. 1.4.5]. Hence, any sequence in $\mathcal{Q}_\epsilon(x)$ has a subsequence that converges *setwise* to some probability measure in $\mathcal{P}(\mathbb{X})$ (cf. [24, Chaps. 1.4.1, 1.5.2]).

This gives us an alternative proof of Lemmas 3.8, 3.15. Let us sketch it for Lemma 3.8 here (the arguments for these two lemmas are similar). As in the previous proof of Lemma 3.8, by Assumptions 3.1, 3.3(iii), and the fact $\underline{h}_n \leq \underline{h}$, to prove the lemma, it suffices to prove the inequality

$$\ell_1 := \liminf_{n \rightarrow \infty} \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}_n(y) q(dy \mid x, a) \right\} \geq \inf_{a \in K_\epsilon(x)} \left\{ c(x, a) + \int_{\mathbb{X}} \underline{h}(y) q(dy \mid x, a) \right\} =: \ell_2.$$

Now, take a sequence $\{a_n\} \subset K_\epsilon(x)$ and choose a subsequence $\{a_{n_k}\}$ from it such that (i) ℓ_1 equals

$$\liminf_{n \rightarrow \infty} \left\{ c(x, a_n) + \int_{\mathbb{X}} \underline{h}_n(y) q(dy \mid x, a_n) \right\} = \lim_{k \rightarrow \infty} \left\{ c(x, a_{n_k}) + \int_{\mathbb{X}} \underline{h}_{n_k}(y) q(dy \mid x, a_{n_k}) \right\},$$

¹¹Here the notion of sequential compactness is as defined by Dunford and Schwartz [10, Def. I.6.10].

and (ii) $q(dy | x, a_{n_k})$ converges setwise to some probability measure $\bar{p} \in \mathcal{P}(\mathbb{X})$ as $k \rightarrow \infty$. Using the setwise convergence property (ii), the fact $\underline{h}_n \uparrow \underline{h}$, and the uniform integrability condition in Assumption 3.3(iii), it can be shown¹² that as $k \rightarrow \infty$,

$$\int_{\mathbb{X}} \underline{h}_{n_k}(y) q(dy | x, a_{n_k}) \rightarrow \int_{\mathbb{X}} \underline{h} d\bar{p}, \quad \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a_{n_k}) \rightarrow \int_{\mathbb{X}} \underline{h} d\bar{p}.$$

This implies $\ell_1 = \lim_{k \rightarrow \infty} \{c(x, a_{n_k}) + \int_{\mathbb{X}} \underline{h}(y) q(dy | x, a_{n_k})\} \geq \ell_2$, the desired inequality to establish Lemma 3.8.

Note that the limit measure \bar{p} in the above need not lie in $\{q(dy | x, a) \mid a \in A(x)\}$; even if the subsequence $a_{n_k} \rightarrow \bar{a} \in A(x)$, it does not imply $q(dy | x, \bar{a}) = \bar{p}$. In other words, the weak sequential compactness of $\mathcal{Q}_\epsilon(x)$ entailed by the majorization condition in Assumption 3.3(ii) is different from the condition of $q(dy | x, a)$ being setwise continuous in a (for each x) considered by [20, 41].

The above discussion also suggests two possible directions to improve the results of this paper. One is to exploit the full potential of Egoroff's theorem, and the other is to use a measure-theoretic approach instead of Egoroff's theorem, under weaker conditions than Assumption 3.3(ii).

Finally, we note that previously, [17, 25] applied majorization type conditions and the weak sequential compactness property to completely or partially observable, discounted-cost MDPs. However, their focuses are on the existence of nonrandomized stationary optimal policies (with the initial distribution being fixed in [17]) for l.s.c. MDP models.

4.3 About the Assumption $\sup_{\alpha \in (0,1)} \|h_\alpha\|_w < \infty$ in the (UC) Case

This is Assumption 3.1. It holds if the MDP is *uniformly w -geometrically ergodic*,¹³ as shown by [22, Lem. 10.4.2]—this result applies in our case as well, in view of Theorem 2.3 on the existence of stationary ϵ -optimal policies in the α -discounted problems. Geometrically or w -geometrically ergodic Markov chains and their application to MDPs are well-studied; see the book [22] for an in-depth discussion and a large body of the related literature.

Assumption 3.1 can also hold in MDPs that do not possess ergodicity properties. A simple class of such MDPs is the “invariant” model considered by Assaf [2]: $A(x) = \mathbb{A}$ for all $x \in \mathbb{X}$ and the state transition stochastic kernel is $q(dy | a)$, independent of x . In this case, if $|c(x, a)| \leq \hat{c}w(x)$ for all $(x, a) \in \Gamma$ and a weight function $w(\cdot)$, then

$$|v_\alpha(x) - v_\alpha(\bar{x})| \leq \sup_{a \in \mathbb{A}} |c(x, a) - c(\bar{x}, a)| \leq \hat{c}(w(x) + w(\bar{x})) \quad \forall \alpha \in (0, 1), \quad (4.6)$$

and hence $\sup_{\alpha \in (0,1)} \|h_\alpha\|_w < \infty$. In the special case of bounded one-stage costs, the family $\{h_\alpha\}$ is uniformly bounded.

For (UC) with an unbounded weight function $w(\cdot)$, Assumption 3.1 also holds for “partially invariant” models given in the next example.

Example 4.1 (partially invariant models). Consider a slight extension of Assaf's invariant models [2] in the case of (UC). Let $b, \hat{c} \geq 0$ and $\lambda \in [0, 1)$ be the constants in the model conditions of (UC),

¹²To prove this, one can use arguments from [13, 23] or [39, p. 231] about limit theorems for integrals involving setwise convergent measures.

¹³This means that every stationary nonrandomized policy μ induces a Markov chain on \mathbb{X} with a unique invariant probability measure φ_μ , and the n -step transition kernel $q_\mu^n(dy | x)$ of this Markov chain satisfies that $\|q_\mu^n - \varphi_\mu\|_w \leq Rr^n$ for all $n \geq 1$, where

$$\|q_\mu^n - \varphi_\mu\|_w := \sup_{x \in \mathbb{X}} \sup_{|f| \leq w, f \in \mathcal{M}(\mathbb{X})} w(x)^{-1} \cdot \left| \int_{\mathbb{X}} f(y) q_\mu^n(dy | x) - \int_{\mathbb{X}} f(y) \varphi_\mu(dy) \right|,$$

and $R > 0, 0 < r < 1$ are constants independent of the policy μ (cf. [22, Assumption 10.2.2]). If $w(\cdot) \equiv 1$, the MDP is called *uniformly geometrically ergodic*.

and let the weight function $w(\cdot)$ be l.s.a. Suppose that for some $\lambda' \in (\lambda, 1)$, the MDP model is “invariant” on the subset of states

$$\hat{\mathbb{X}} := \{x \in \mathbb{X} \mid w(x) \leq b/(\lambda' - \lambda)\},$$

in the sense that for all $x \in \hat{\mathbb{X}}$, $\hat{A} := A(x)$ is the same and $q(dy \mid x, a)$ depends only on a . Then (4.6) holds for $x, \bar{x} \in \hat{\mathbb{X}}$ with \hat{A} in place of \mathbb{A} , so

$$|v_\alpha(x) - v_\alpha(\bar{x})| \leq \hat{c}(w(x) + w(\bar{x})) \leq 2b\hat{c}/(\lambda' - \lambda) \quad \forall x, \bar{x} \in \hat{\mathbb{X}}. \quad (4.7)$$

Fix $\bar{x} \in \hat{\mathbb{X}}$. Let us bound $v_\alpha(x) - v_\alpha(\bar{x})$ for $x \notin \hat{\mathbb{X}}$. Let $\tau = \inf\{n \geq 0 \mid x_n \in \hat{\mathbb{X}}\}$, the first entrance time to $\hat{\mathbb{X}}$. By the (UC) model condition (b) (cf. Def. 2.1) and the definition of the set $\hat{\mathbb{X}}$, we have that

$$\int_{\mathbb{X}} w(y) q(dy \mid x, a) \leq \lambda' w(x) \quad \forall x \notin \hat{\mathbb{X}}, a \in A(x). \quad (4.8)$$

It implies that under any policy π ,

$$\mathbb{E}_x^\pi[\sum_{n=0}^{\tau-1} w(x_n)] \leq \ell w(x) \quad \forall x \notin \hat{\mathbb{X}}, \quad (4.9)$$

where ℓ is some constant independent of π and x . This inequality follows from a general comparison theorem [32, Thm. 15.2.5] and can also be verified directly (in fact, $\ell = (1 - \lambda')^{-1}$ in this case). For each $\alpha \in (0, 1)$, let μ_α be a stationary ϵ -optimal policy for the α -discounted problem (cf. Theorem 2.3). Using the inequality (4.9) together with the (UC) model condition (a) and the strong Markov property, we have

$$\begin{aligned} v_\alpha(x) &\leq \mathbb{E}_x^{\mu_\alpha}[\sum_{n=0}^{\tau-1} \alpha^n c(x_n, a_n) + \alpha^\tau v_\alpha(x_\tau)] \\ &\leq \hat{c}\ell w(x) + v_\alpha(\bar{x}) + \mathbb{E}_x^{\mu_\alpha}[\alpha^\tau v_\alpha(x_\tau) - v_\alpha(\bar{x})], \end{aligned} \quad (4.10)$$

$$\begin{aligned} v_\alpha(x) &\geq \mathbb{E}_x^{\mu_\alpha}[\sum_{n=0}^{\tau-1} \alpha^n c(x_n, a_n) + \alpha^\tau v_\alpha(x_\tau)] - \epsilon \\ &\geq -\hat{c}\ell w(x) + v_\alpha(\bar{x}) + \mathbb{E}_x^{\mu_\alpha}[\alpha^\tau v_\alpha(x_\tau) - v_\alpha(\bar{x})] - \epsilon. \end{aligned} \quad (4.11)$$

Now

$$\begin{aligned} |\mathbb{E}_x^{\mu_\alpha}[\alpha^\tau v_\alpha(x_\tau) - v_\alpha(\bar{x})]| &= |\mathbb{E}_x^{\mu_\alpha}[\alpha^\tau (v_\alpha(x_\tau) - v_\alpha(\bar{x})) + (\alpha^\tau - 1)v_\alpha(\bar{x})]| \\ &\leq \sup_{x' \in \hat{\mathbb{X}}} |v_\alpha(x') - v_\alpha(\bar{x})| + \mathbb{E}_x^{\mu_\alpha}[1 - \alpha^\tau] |v_\alpha(\bar{x})|. \end{aligned} \quad (4.12)$$

The first term in (4.12) is bounded by a constant by (4.7). To bound the second term, note that $\mathbb{E}_x^{\mu_\alpha}[1 - \alpha^\tau] \leq (1 - \alpha) \mathbb{E}_x^{\mu_\alpha}[\tau]$, and $\mathbb{E}_x^{\mu_\alpha}[\tau] \leq \ell w(x)$ by (4.9), whereas due to the (UC) model condition, $(1 - \alpha)|v_\alpha(\bar{x})|$ is bounded by some constant $\ell_{\bar{x}}$ independent of α . Therefore, the second term in (4.12) is bounded by $\ell\ell_{\bar{x}}w(x)$. We thus have

$$|\mathbb{E}_x^{\mu_\alpha}[\alpha^\tau v_\alpha(x_\tau) - v_\alpha(\bar{x})]| \leq 2b\hat{c}/(\lambda' - \lambda) + \ell\ell_{\bar{x}}w(x). \quad (4.13)$$

Combining this relation with the upper and lower bounds on $v_\alpha(x)$ given in (4.10)-(4.11) and taking $\epsilon \rightarrow 0$, we obtain that for all $\alpha \in (0, 1)$,

$$-\tilde{\ell}_{\bar{x}}w(x) - \tilde{\ell} \leq v_\alpha(x) - v_\alpha(\bar{x}) \leq \tilde{\ell}_{\bar{x}}w(x) + \tilde{\ell}$$

where $\tilde{\ell}_{\bar{x}} = \hat{c}\ell + \ell\ell_{\bar{x}}$ and $\tilde{\ell} = 2b\hat{c}/(\lambda' - \lambda)$ are constants independent of α and x . This shows that $\sup_{\alpha \in (0, 1)} \|h_\alpha\|_w < \infty$. \square

Remark 4.3 (ACOI for invariant and partially invariant models). For invariant models, Assaf [2] proved the ACOE when the one-stage costs are bounded and the action space is finite. Schäl [41, sect. 4, p. 169] showed that, for a Borel action space, if c is bounded below with $\sup_{a \in \mathbb{A}} c(x, a) < \infty$ for all $x \in \mathbb{X}$, then $\sup_{\alpha \in (0, 1)} h_\alpha(x) < \infty$ for all $x \in \mathbb{X}$. Thus, in that case, the ACOI holds under our majorization condition for (PC), by Theorem 3.13. In the (UC) case, by Theorem 3.5, the ACOI also holds under our majorization condition, for both the invariant models and the partially invariant models described in the preceding example. \square

We will give an illustrative example in Appendix A.3 to demonstrate another way to ensure that Assumption 3.1 holds in an MDP that is not necessarily ergodic. There we will use some arguments from Example 4.1. In general, however, verifying Assumption 3.1 is a hard problem since it involves the optimal value functions $\{v_\alpha\}$.

4.4 About Geometric Ergodicity Conditions, ACOI, and ACOE in the (UC) Case

For a uniformly w -geometrically ergodic MDP (cf. footnote 13 in Section 4.3), under compactness/continuity and additional conditions, the ACOI has been strengthened to the ACOE; see e.g., Gordienko and Hernández-Lerma [18, Thm. 2.8], Hernández-Lerma and Vega-Amaya [26, Thm. 3.5(a)], Jaśkiewicz and Nowak [28, Thm. 3], and also the book [22, Thm. 10.3.6]. As recounted in Section 1, the ACOE was also directly obtained through a contraction-based, fixed point approach in Vega-Amaya [49, 51] for MDPs under certain compactness and continuity conditions, and in Jaśkiewicz [27] for SMDPs with u.m. policies.¹⁴ (In the special case $w(\cdot) \equiv 1$, the ACOE was obtained earlier, through the fixed point approach, by Gubenko and Shtatland [19], as mentioned in the introduction.)

For these ACOE results to hold in an MDP, among others, the MDP must be uniformly w -geometrically ergodic,¹⁵ and moreover, there must exist a nontrivial σ -finite measure φ on \mathbb{X} such that under any stationary nonrandomized policy μ , the induced Markov chain $\{x_n\}$ is φ -irreducible (see e.g., [22, 32, 34] for definition). The existence of such a common irreducibility measure appears either in the conditions of the aforementioned results (e.g., [22, Thm. 10.3.6]) or follows as an implication of their assumptions (e.g., [27, 49]).

The uniform w -geometric ergodicity condition implies that for any *two* stationary nonrandomized policies, there must be a common irreducibility measure, but a common irreducibility measure for *all* stationary nonrandomized policies need not exist. This is illustrated by the following counterexample.

Example 4.2 (non-existence of common irreducibility measure). Let \mathbb{X} be the unit circle and $\mathbb{A} = \{0, 1, 2\} = A(x)$ for all $x \in \mathbb{X}$. Represent states by their angular coordinates. Let $q(dy | x, a)$ be the uniform distribution on $[\frac{2a}{3}\pi, \frac{2a}{3}\pi + \pi] =: I_a$ for $a \in \mathbb{A}$ and every x . Then the MDP is uniformly geometrically ergodic, and the *maximal irreducibility measures* (see [32, 34] for definition) that are possible under any nonrandomized stationary policy, are (up to equivalence) the Lebesgue measure restricted on the intervals I_a , $a = 0, 1, 2$, on the unions of any two such intervals, and on $[0, 2\pi]$. Since the intersection of these intervals is empty, there is no common irreducibility measure for all stationary nonrandomized policies in this MDP. \square

This example shows that the uniform w -geometric ergodicity condition is more general than the conditions required by the fixed point approach studied in [27, 49, 51], whereas Assumption 3.1 is more general than the former condition, as discussed in Section 4.3. This suggests that most likely, the ACOI in our Theorem 3.5 cannot be strengthened to ACOE without additional assumptions. However, we do not know a counterexample for ACOE under the conditions of Theorem 3.5.

4.5 An Illustrative Example for the (PC) Case

In this subsection, we use an inventory control example to illustrate the application of our ACOI result for the (PC) case. This example is adapted from an example in [41, p. 170]. The original example has compact action sets $A(x)$ and i.i.d. random demands. We make $A(x)$ non-compact and

¹⁴One difference between the minorization conditions in [27] and [49, 51] is that the former is stated in terms of “small sets,” whereas the latter “small functions” (cf. [34, Chap. 2.3]).

¹⁵In [49, 51], this follows as an implication of the conditions involved, by [49, Thm. 3.3] together with [29, Thm. 6] (see also [22, Thm. 7.3.11]).

also allow action-dependent random demands, with which the state transition stochastic kernel need not be continuous. We will verify two major conditions for applying Theorem 3.13: our majorization condition and the condition (B), $\sup_{\alpha \in (0,1)} h_\alpha(x) < \infty$, $x \in \mathbb{X}$, which implies Assumption 3.9(B).

4.5.1 Two Helpful Lemmas

The condition (B) is not easy to check because it involves the optimal value functions v_α , whose structures are problem-dependent and hard to characterize in general. Most of our efforts will be on verifying this condition. We will need two helpful lemmas given below. They are from [41, Lems. 4.4, 4.6], with slight modifications in the first one. Their proof arguments are outlined in [41], which utilize various stopping times. For completeness, we include a proof for the second lemma (Lemma 4.5 below) in Appendix A.1, as it is less obvious.

Let $\underline{c}(x) := \inf_{a \in A(x)} c(x, a)$ and $\underline{g}^* := \inf_{x \in \mathbb{X}} g^*(x)$. Note that $\underline{c}(\cdot)$ is l.s.a. [3, Prop. 7.47] and $\underline{g}^* < \infty$ under Assumption 3.9(G). Define also

$$\underline{v}_\alpha(x) := \inf_{a \in A(x)} \int_{\mathbb{X}} v_\alpha(y) q(dy | x, a), \quad \underline{m}_\alpha := \inf_{x \in \mathbb{X}} \underline{v}_\alpha(x).$$

Note that $\alpha \underline{m}_\alpha \leq m_\alpha \leq \underline{m}_\alpha$, the function \underline{v}_α is l.s.a. [3, Props. 7.47, 7.48], and by [3, Prop. 7.50], for any $\epsilon > 0$, there is a (u.m.) nonrandomized stationary policy μ_α^ϵ with

$$\int_{\mathbb{X}} v_\alpha(y) q(dy | x, \mu_\alpha^\epsilon(x)) \leq \underline{v}_\alpha(x) + \epsilon \quad \forall x \in \mathbb{X}. \quad (4.14)$$

Denote by \mathcal{F}_n the σ -algebra generated by $(x_0, a_0, \dots, x_n, a_n)$.

Lemma 4.4 (cf. [41, Lem. 4.4]). *Let $\eta, \epsilon > 0$. Let π be any policy and τ be any stopping time w.r.t. $\{\mathcal{F}_n\}$ such that $\alpha \underline{v}_\alpha(x_\tau) \leq m_\alpha + \eta$ on $\{\tau < \infty\}$. Then*

$$h_\alpha(x) \leq \eta + \epsilon + \mathbb{E}_x^\pi \left[\sum_{n=0}^{\tau-1} c(x_n, a_n) + c(x_\tau, \mu_\alpha^\epsilon(x_\tau)) \right] \quad \forall x \in \mathbb{X}.$$

Lemma 4.5 (cf. [41, Lem. 4.6]). *Let Assumption 3.9(G) hold. Let $\epsilon > 0$ and $G \subset \Gamma$ be a u.m. set such that $\int_{\mathbb{X}} \underline{c}(y) q(dy | x, a) \geq \underline{g}^* + \epsilon$ for all $(x, a) \in G$. Then, for some $\alpha_\epsilon < 1$, it holds for all $\alpha \geq \alpha_\epsilon$ that*

$$\int_{\mathbb{X}} v_\alpha(y) q(dy | x, a) \geq \underline{m}_\alpha + \epsilon/2 \quad \forall (x, a) \in G.$$

4.5.2 Single-Product Inventory System

Let $\mathbb{X} = \mathbb{A} = \mathbb{R}$, $A(x) = [x, +\infty)$, and $c(x, a) = \kappa(a - x) + \psi(a)$ where $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $\psi : \mathbb{R} \rightarrow \mathbb{R}_+$. Let the states evolve according to $x_{n+1} = a_n - \xi_n$, where, for the n th stage:

- x_n and a_n are the stock levels before and after placing the order, respectively;
- $\xi_n \geq 0$ is the demand or consumption of the product, which is a random variable that does not depend on the history $(x_0, a_0, \xi_0, \dots, x_n)$ given a_n ;
- $\kappa(a_n - x_n)$ and $\psi(a_n)$ correspond to the ordering cost and the holding or shortage/backordering cost, respectively.

We assume that for all $n \geq 0$, the conditional distributions of ξ_n given a_n are parametrized by the values of a_n (and Borel measurable in a_n), independent of n . These probability distributions will be denoted by F_a , $a \in \mathbb{R}$. In what follows, we shall write $\xi_n(a_n)$ for ξ_n to emphasize the dependence on a_n , and we use $\xi(a)$ to denote a generic random variable with probability distribution F_a and $\mathbb{E}[\xi(a)]$ to denote its expectation w.r.t. F_a .

We impose two sets of conditions on the functions $\kappa(\cdot), \psi(\cdot)$ and the demand distributions F_a . The first set is given below and will be used shortly to prove that the family $\{h_\alpha\}$ is pointwise

bounded under those conditions. When we consider the majorization condition later, we will add a second set of conditions.

Throughout this example, let Assumption 3.9(G) hold¹⁶ (hence $\underline{g}^* < \infty$).

Assumption 4.6 (conditions on $\kappa(\cdot)$, $\psi(\cdot)$, and $\{\xi(a)\}$).

- (i) The ordering cost $\kappa(\cdot)$ and the holding or shortage cost $\psi(\cdot)$ are bounded above on bounded intervals; moreover, $\kappa(\cdot)$ is non-decreasing and $\kappa(0) = 0$.
- (ii) $\liminf_{z \rightarrow \infty} \kappa(z) > \underline{g}^*$ and $\liminf_{|a| \rightarrow \infty} \psi(a) > \underline{g}^*$.
- (iii) For any $\bar{a} \in \mathbb{R}$, $\{\xi(a) \mid a \geq \bar{a}\}$ is uniformly integrable.
- (iv) For any bounded interval $I \subset \mathbb{R}$ and any given $z \geq 0$,

$$\inf_{a \in I} \mathbb{E}[\xi(a)] > 0 \quad \text{and} \quad \sup_{a \in I} \mathbb{E}[\kappa(z + \xi(a))] < \infty.$$

Note that here $\kappa(\cdot)$ and $\psi(\cdot)$ need not be l.s.c., and $q(dy \mid x, a)$ need not be continuous. As a simple example, these functions can be piecewise continuous; even in this case, the compactness/continuity conditions required in [8, 14, 41] are violated.

Proposition 4.7. Under Assumption 4.6, for all $x \in \mathbb{R}$, $\sup_{\alpha \in (0,1)} h_\alpha(x) < \infty$.

Proof. We follow the line of reasoning in [41, sect. 4]. When computing upper bounds on $h_\alpha(x)$, however, we will need to use more complex arguments because the random demands in our case are not i.i.d. as in [41].

Let us write $q(dy \mid a)$ for $q(dy \mid x, a)$, since state transitions in this example depend on a only. First, we use Lemma 4.5 to assert the existence of a bounded interval $[L, M]$ such that for some $\bar{\epsilon} > 0$ and $\alpha_{\bar{\epsilon}} < 1$, it holds for all $\alpha \geq \alpha_{\bar{\epsilon}}$ that

$$\exists y_\alpha \in [L, M] \quad \text{with} \quad \int_{\mathbb{X}} v_\alpha(y) q(dy \mid y_\alpha) \leq \underline{m}_\alpha + \bar{\epsilon}/2. \quad (4.15)$$

(This is proved by a direct calculation of a set G in the condition of Lemma 4.5 using the definition of the function \underline{c} and Assumption 4.6(ii)-(iii).) The existence of such y_α allows us to define a stopping time τ needed in applying Lemma 4.4 to bound h_α : since for all $x' \leq y_\alpha$, we have $y_\alpha \in A(x')$ and

$$\alpha \underline{v}_\alpha(x') \leq \alpha \int_{\mathbb{X}} v_\alpha(y) q(dy \mid y_\alpha) \leq \alpha \underline{m}_\alpha + \bar{\epsilon}/2 \leq m_\alpha + \bar{\epsilon}/2, \quad (4.16)$$

stopping at any state $x' \leq y_\alpha$ will satisfy the condition of Lemma 4.4 on τ . This is the basic idea to bound $\sup_{\alpha \in (0,1)} h_\alpha(x)$. We now proceed to calculate the bounds on $h_\alpha(x)$ for each $x \in \mathbb{R}$ and an arbitrary $\alpha \geq \alpha_{\bar{\epsilon}}$.

Define a stationary policy μ by

$$\mu(x) = x \quad \text{if } x \geq L, \quad \mu(x) = y_\alpha \quad \text{if } x < L.$$

Let $\tau := \inf\{n \geq 0 \mid x_n < L\}$. We will apply Lemma 4.4 with this τ and $\pi = \mu$. To this end, let us bound $\mathbb{E}_x^\mu \left[\sum_{n=0}^{\tau-1} c(x_n, a_n) + c(x_\tau, y_\alpha) \right]$ independently of α . The case $x < L$ is simple: we have $\tau = 0$ and under Assumption 4.6(i),

$$c(x, y_\alpha) = \kappa(y_\alpha - x) + \psi(y_\alpha) \leq \kappa(M - x) + \sup_{y \in [L, M]} \psi(y) < \infty. \quad (4.17)$$

¹⁶In fact, (G) holds under Assumption 4.6(i) and (iv): it can be easily verified that $J(\pi, 0) < \infty$ for the nonrandomized stationary policy π that has $\pi(x) = 0$ for $x < 0$ and $\pi(x) = x$ for $x \geq 0$.

For $x \geq L$, since $\kappa(0) = 0$ by Assumption 4.6(i), we have

$$\begin{aligned} \mathbb{E}_x^\mu \left[\sum_{n=0}^{\tau-1} c(x_n, a_n) + c(x_\tau, y_\alpha) \right] &= \mathbb{E}_x^\mu \left[\sum_{n=0}^{\tau-1} \psi(x_n) + \psi(y_\alpha) + \kappa(y_\alpha - x_\tau) \right] \\ &\leq \mathbb{E}_x^\mu[\tau] \cdot \sup_{y \in [L, x \vee M]} \psi(y) + \mathbb{E}_x^\mu[\kappa(y_\alpha - x_\tau)] \end{aligned} \quad (4.18)$$

where $x \vee M := \max\{x, M\}$. So we need to bound $\mathbb{E}_x^\mu[\tau]$ and $\mathbb{E}_x^\mu[\kappa(y_\alpha - x_\tau)]$.

We apply a comparison theorem [32, Prop. 11.3.2] together with other arguments to bound $\mathbb{E}_x^\mu[\tau]$, which in turn yields a bound on $\mathbb{E}_x^\mu[\kappa(y_\alpha - x_\tau)]$. The derivations are given in Appendix A.2. The obtained upper bounds are as follows: for $x \geq L$,

$$\mathbb{E}_x^\mu[\kappa(y_\alpha - x_\tau)] \leq \mathbb{E}_x^\mu[\tau] \cdot \sup_{y \in [L, x \vee M]} \mathbb{E}[\kappa(M - L + \xi(y))], \quad (4.19)$$

$$\mathbb{E}_x^\mu[\tau] \leq 2(x - L + D_x)/\Delta_x, \quad (4.20)$$

where $\Delta_x = \inf_{y \in [L, x \vee M]} \mathbb{E}[\xi(y)] > 0$ (cf. Assumption 4.6(iv)), and $0 < D_x < \infty$ is the smallest real number such that

$$\sup_{y \in [L, x \vee M]} \mathbb{E}[\xi(y) \mathbb{1}(\xi(y) > D_x)] \leq \Delta_x/2. \quad (4.21)$$

Such D_x exists by Assumption 4.6(iii) and the fact that viewed as a function of z ,

$$\sup_{y \in [L, x \vee M]} \mathbb{E}[\xi(y) \mathbb{1}(\xi(y) > z)]$$

is continuous from the right. The upper bounds in (4.20) and (4.19) are thus finite (cf. Assumption 4.6(iv)).

Combining (4.17)-(4.20), we have

$$\mathbb{E}_x^\mu \left[\sum_{n=0}^{\tau-1} c(x_n, a_n) + c(x_\tau, y_\alpha) \right] \leq H(x) \quad \forall x \in \mathbb{R}, \alpha \geq \alpha_{\bar{\epsilon}},$$

where the function H is finite everywhere, independent of α , and given by

$$H(x) = \kappa(M - x) + \sup_{y \in [L, M]} \psi(y) \quad \text{for } x < L, \quad (4.22)$$

and for $x \geq L$,

$$H(x) = \frac{2(x - L + D_x)}{\Delta_x} \cdot \left\{ \sup_{y \in [L, x \vee M]} \psi(y) + \sup_{y \in [L, x \vee M]} \mathbb{E}[\kappa(M - L + \xi(y))] \right\}. \quad (4.23)$$

We can now apply Lemma 4.4 to prove $\sup_{\alpha \in (0,1)} h_\alpha(x) < \infty$ for all $x \in \mathbb{R}$.

Let $\alpha \geq \alpha_{\bar{\epsilon}}$. For all $x \leq L$, by (4.15), $\int_{\mathcal{X}} v_\alpha(y) q(dy | y_\alpha) \leq \underline{v}_\alpha(x) + \bar{\epsilon}/2$, so for applying Lemma 4.4, we can define a policy $\mu_\alpha^{\bar{\epsilon}/2}$ that satisfies (4.14) for $\epsilon = \bar{\epsilon}/2$, in such a way that $\mu_\alpha^{\bar{\epsilon}/2}(x) = y_\alpha$ for $x \leq L$. We have $\alpha \underline{v}_\alpha(x_\tau) \leq m_\alpha + \bar{\epsilon}/2$ by (4.16), so by Lemma 4.4 and the preceding proof, for all $x \in \mathbb{R}$ and $\alpha \geq \alpha_{\bar{\epsilon}}$,

$$h_\alpha(x) \leq \bar{\epsilon} + \mathbb{E}_x^\mu \left[\sum_{n=0}^{\tau-1} c(x_n, a_n) + c(x_\tau, y_\alpha) \right] \leq \bar{\epsilon} + H(x) < \infty.$$

As $H(x)$ is independent of α , this inequality implies $\limsup_{\alpha \uparrow 1} h_\alpha(x) < \infty$, which, under Assumption 3.9(G), is equivalent to $\sup_{\alpha \in (0,1)} h_\alpha(x) < \infty$ by [14, Lem. 5]. \square

We now tackle Assumption 3.12 (the majorization condition). Let us place two more conditions on the one-stage costs and the demand distributions:

Assumption 4.8 (additional conditions on $c(\cdot)$ and $\{\xi(a)\}$).

- (i) $\lim_{|a| \rightarrow \infty} c(x, a) = +\infty$ for each $x \in \mathbb{R}$.
- (ii) For any bounded interval $I \subset \mathbb{R}$, $\{F_a \mid a \in I\}$ have densities w.r.t. the Lebesgue measure that are uniformly bounded above, and $\cup_{a \in I} \text{supp}(F_a)$, the union of their supports, is bounded.

Assumption 4.8(i) will help us eliminate actions and find proper subsets of actions to use as $K_\epsilon(x)$ in the majorization condition. Assumption 4.8(ii) and the function H derived in the proof of Proposition 4.7 will be useful in verifying the uniform integrability condition on \underline{h} in Assumption 3.12(iii), since $\underline{h} \leq H + \bar{\epsilon}$. In Assumption 4.8(ii), the requirement on $\{F_a\}$ to have densities is a limitation of our technique and will be discussed further in Remark 4.11. The bounded-support condition is much stronger than Assumption 4.6(iii); it can be relaxed if one knows more precisely how the tails of $F_a, a \in I$, decay and how $H(x)$ varies with x as $x \rightarrow -\infty$.

Proposition 4.9. *Assumptions 4.6 and 4.8 imply Assumption 3.12.*

Proof. Let $\bar{\epsilon} > 0, \alpha_{\bar{\epsilon}} \in (0, 1)$, and the function H be as in the proof of Proposition 4.7. That proof showed that under Assumption 4.6, $h_\alpha(x) \leq \bar{\epsilon} + H(x)$ for all $x \in \mathbb{R}$ and $\alpha \geq \alpha_{\bar{\epsilon}}$. Consider now an arbitrary state x . By Assumption 4.8(i), there exists $\bar{a} \geq x$ such that $c(x, a) \geq \underline{g}^* + 2\bar{\epsilon} + H(x)$ for all $a \geq \bar{a}$. Then, from the α -DCOE

$$(1 - \alpha) m_\alpha + h_\alpha(x) = \inf_{a \in A(x)} \{c(x, a) + \alpha \int_{\mathcal{X}} h_\alpha(y) q(dy \mid x, a)\}$$

and the fact $\limsup_{\alpha \uparrow 1} (1 - \alpha) m_\alpha \leq \underline{g}^*$, we have that for all α sufficiently large,

$$\inf_{a \in [x, \bar{a}]} \{c(x, a) + \alpha \int_{\mathcal{X}} h_\alpha(y) q(dy \mid x, a)\} = (1 - \alpha) m_\alpha + h_\alpha(x).$$

Thus, for any $\epsilon > 0$, Assumption 3.12(i) holds with $K_\epsilon(x) = K := [x, \bar{a}]$. Next, for the majorizing finite measure ν required in Assumption 3.12(ii), in view of Assumption 4.8(ii), we can simply let ν be a multiple of the Lebesgue measure on the bounded set $E := \cup_{a \in K} \{a - \text{supp}(F_a)\}$ and the trivial measure on the complement set E^c .

Finally, given Assumption 4.8(ii) and the fact $\underline{h} \leq \bar{\epsilon} + H$, for Assumption 3.12(iii) to hold, it suffices that H is bounded on bounded intervals. The expression (4.22)-(4.23) of H shows that this is the case. In particular, by Assumption 4.6(i) and (iv), if x lies in a bounded interval I , all those terms in (4.22)-(4.23) that involve the functions $\kappa(\cdot)$ and $\psi(\cdot)$ must be bounded above by some constants depending on I . The remaining term is $2(x - L + D_x)/\Delta_x$ (which bounds $\mathbb{E}_x^\mu[\tau]$) for $x \geq L$. Let I be a bounded interval in $[L, \infty)$. By Assumption 4.6(iv), we have

$$\Delta := \inf_{x \in I} \Delta_x = \inf_{x \in I} \inf_{y \in [L, x \vee M]} \mathbb{E}[\xi(y)] > 0.$$

We also have $\sup_{x \in I} D_x < \infty$, because by Assumption 4.6(iii), there exists some $0 < D < \infty$ such that

$$\sup_{y \geq L} \mathbb{E}[\xi(y) \mathbb{1}(\xi(y) > D)] \leq \Delta/2 \leq \Delta_x/2 \quad \forall x \in I,$$

and this implies that $D \geq D_x$ for all $x \in I$, since D_x is by definition the smallest number satisfying (4.21). Hence $\sup_{x \in I} 2(x - L + D_x)/\Delta_x < \infty$. Thus H is bounded on any bounded interval, and consequently, Assumption 3.12(iii) holds. \square

We have shown that the conditions in Theorem 3.13 are met and the ACOI holds. In connection with this example, let us make two final remarks regarding the structure of stationary optimal/ ϵ -optimal policies and limitation of our majorization technique.

Remark 4.10. For l.s.c. models studied in e.g., [14, 15, 22, 28, 41], besides the ACOI or ACOE, one can also relate the optimal actions for discounted problems to those for the average cost problem in the limit as the discount factor $\alpha \uparrow 1$. Moreover, for inventory control, if certain convexity

properties are present, there exist optimal policies with particularly simple structures like the policy μ in this example, i.e., the so-called (s, S) policies and their generalizations (see e.g., [7, 12, 15]). For discontinuous MDP models, in general, one cannot guarantee such structural properties for the stationary optimal or ϵ -optimal policies. The proofs of Theorems 3.5 and 3.13 show that if for $\epsilon > 0$, the set-valued map $x \mapsto K_\epsilon(x)$ has an analytic graph, then in the average cost problem, there exists a nonrandomized stationary ϵ -optimal policy μ such that $\mu(x) \in K_\epsilon(x)$ for all $x \in \mathbb{X}$. This, however, provides only a “loose” connection between some good actions in the discounted problems and those in the average cost problem, for the sets $K_\epsilon(x)$ can be quite large, as shown by this example for (PC) and the example for (UC) in Appendix A.3. \square

Remark 4.11. Typically, the random demands in inventory control problems are allowed to be zero, as long as they are positive with positive probabilities. We do not allow this in Assumption 4.8(ii), because no finite measure can majorize the measures $r_a \delta_a$ with $r_a > 0$ for all $a \in I$. This limitation is due to the nature of the measure-theoretic properties underlying our majorization condition, as discussed in Sections 3.1.1 and 4.2. For similar problems in which $q(dy | x, a)$ is not atomless, one possible way to proceed is to approximate the original MDP by one in which each action corresponds to some probability distribution over the feasible actions in the original problem. Then the majorization idea and the analysis technique presented in this paper may still be potentially useful for analyzing the approximating MDP, thereby shedding light on the original average cost problem. \square

Appendix A Supplementary Materials for Section 4

A.1 Proofs of Proposition 4.1 and Lemma 4.5

Proof of Proposition 4.1. We prove the second statement first. By the first inequality in (4.1), we need to prove $\lim_{n \rightarrow \infty} \inf_{y \in Y} f_n(y) \geq \inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y)$, and in turn, by the assumption (4.3), it is sufficient to prove that for each $\epsilon > 0$, $\ell := \liminf_{j \rightarrow \infty} \inf_{y \in K} f_{n_j}(y) \geq \inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y)$. To this end, consider a sequence $\{y_j\}$ in K with $f_{n_j}(y_j) - \inf_{y \in K} f_{n_j}(y) \rightarrow 0$ as $j \rightarrow \infty$. Since K is compact, there exists a subsequence $y_{j_i} \rightarrow \bar{y} \in K$ such that, with $n'_i := n_{j_i}$ and $y'_i := y_{j_i}$, $f_{n'_i}(y'_i) \rightarrow \ell$ as $i \rightarrow \infty$. Since $(y'_i, f_{n'_i}(y'_i)) \in \text{epi}(f_{n'_i})$ for all i and $y'_i \rightarrow \bar{y}$, by the definition of the lower epi-limit, we have $(\underline{\text{e-lim}}_n f_n)(\bar{y}) \leq \lim_{i \rightarrow \infty} f_{n'_i}(y'_i) = \ell$ and hence $\inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y) \leq \ell$ as desired.

To prove the first statement, for every $\epsilon > 0$, let \bar{y} be any point that satisfies

$$(\underline{\text{e-lim}}_n f_n)(\bar{y}) \leq \inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y) + \epsilon/3.$$

By the definition of the lower epi-limit, there exist a subsequence $\{n_j\}$ and points $\{y_j\}$ with $y_j \rightarrow \bar{y}$ and $f_{n_j}(y_j) \rightarrow (\underline{\text{e-lim}}_n f_n)(\bar{y})$ as $j \rightarrow \infty$. Let $K = \{y_j\}_{j \geq 1} \cup \{\bar{y}\}$, which is a compact subset of Y . Since $\inf_{y \in K} f_{n_j}(y) \leq f_{n_j}(y_j)$, it follows that for all j sufficiently large,

$$\inf_{y \in K} f_{n_j}(y) \leq (\underline{\text{e-lim}}_n f_n)(\bar{y}) + \epsilon/3 \leq \inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y) + 2\epsilon/3,$$

whereas (4.2) implies that $\inf_{y \in Y} (\underline{\text{e-lim}}_n f_n)(y) \leq \inf_{y \in Y} f_{n_j}(y) + \epsilon/3$ for all j sufficiently large. Hence (4.3) holds, as desired. \square

Proof of Lemma 4.5. First, note that for all α larger than some $\alpha_\epsilon < 1$, we have $(1 - \alpha)\underline{m}_\alpha \leq \underline{g}^* + \epsilon/4$. This follows from the fact that $\alpha(\underline{m}_\alpha - m_\alpha) \leq (1 - \alpha)m_\alpha$ (since $\alpha \underline{m}_\alpha \leq m_\alpha \leq \underline{m}_\alpha$) and $\limsup_{\alpha \rightarrow \infty} (1 - \alpha)m_\alpha \leq \underline{g}^*$.

Consider an arbitrary $\alpha \geq \alpha_\epsilon$ and $(x, a) \in G$. Let π be a policy that applies action a at $x_0 = x$ and then applies a stationary policy that is $\epsilon/4$ -optimal for the α -discounted problem. Let

$\tau = \inf \{n \geq 0 \mid (x_n, a_n) \notin G\}$ ($\tau = \infty$ if this set is empty). We have

$$\begin{aligned} \int_{\mathbb{X}} v_\alpha(y) q(dy \mid x, a) + \frac{\epsilon}{4} &\geq \mathbb{E}_x^\pi \left[\sum_{n=1}^{\infty} \alpha^{n-1} c(x_n, a_n) \right] \\ &\geq \mathbb{E}_x^\pi \left[\sum_{n=1}^{\tau} \alpha^{n-1} c(x_n, a_n) + \alpha^\tau v_\alpha(x_{\tau+1}) \right] \\ &= \sum_{n=1}^{\infty} \mathbb{E}_x^\pi [\mathbb{1}(\tau > n-1) \alpha^{n-1} c(x_n, a_n) + \mathbb{1}(\tau = n) \alpha^n v_\alpha(x_{n+1})]. \end{aligned}$$

In view of the property of G , for $n \geq 1$, on $\{\tau > n-1\}$,

$$\mathbb{E}_x^\pi [c(x_n, a_n) \mid \mathcal{F}_{n-1}] \geq \int_{\mathbb{X}} \underline{c}(x_n) q(dx_n \mid x_{n-1}, a_{n-1}) \geq \underline{g}^* + \epsilon \geq (1-\alpha) \underline{m}_\alpha + 3\epsilon/4,$$

and in particular, $\mathbb{E}_x^\pi [c(x_1, a_1)] \geq (1-\alpha) \underline{m}_\alpha + 3\epsilon/4$. We also have $\mathbb{E}_x^\pi [v_\alpha(x_{n+1}) \mid \mathcal{F}_n] \geq \underline{m}_\alpha$. Combining these relations, we obtain that

$$\int_{\mathbb{X}} v_\alpha(y) q(dy \mid x, a) + \epsilon/4 \geq 3\epsilon/4 + \mathbb{E}_x^\pi \left[\sum_{n=1}^{\tau} \alpha^{n-1} \cdot (1-\alpha) \underline{m}_\alpha + \alpha^\tau \underline{m}_\alpha \right] = 3\epsilon/4 + \underline{m}_\alpha,$$

which gives the desired inequality for all $\alpha \geq \alpha_\epsilon$ and $(x, a) \in G$. \square

A.2 Derivations of (4.19)-(4.20) for the Proof of Proposition 4.7

We first prove the bound (4.20) on $\mathbb{E}_x^\mu[\tau]$ for $x \geq L$. Recall $\Delta_x = \inf_{y \in [L, x \vee M]} \mathbb{E}[\xi(y)] > 0$ and $0 < D_x < \infty$ is such that

$$\sup_{y \in [L, x \vee M]} \mathbb{E}[\xi(y) \mathbb{1}(\xi(y) > D_x)] \leq \Delta_x/2.$$

Consider nonnegative random variables $Z_n := (x_n - L + D_x)^+$ for $n \geq 0$, where $(\cdot)^+ := \max\{\cdot, 0\}$. Denote by \mathcal{F}_n the σ -algebra generated by $(x_0, a_0, \dots, x_n, a_n)$. Using the definition of the policy μ , a direct calculation shows that when $x_n \geq L$,

$$\mathbb{E}_x^\mu [Z_{n+1} \mid \mathcal{F}_n] = \mathbb{E}_x^\mu [(x_n - \xi_n(x_n) - L + D_x)^+ \mid \mathcal{F}_n] \leq Z_n - \frac{1}{2} \Delta_x,$$

and when $x_n < L$,

$$\mathbb{E}_x^\mu [Z_{n+1} \mid \mathcal{F}_n] \leq Z_n + y_\alpha - L + D_x.$$

By a comparison theorem [32, Prop. 11.3.2] (which is an implication of Dynkin's formula [32, Thm. 11.3.1]), this implies that for the stopping time $\tau = \inf\{n \geq 0 \mid x_n < L\}$,

$$\mathbb{E}_x^\mu \left[\sum_{n=0}^{\tau-1} \frac{1}{2} \Delta_x \right] \leq Z_0 = x - L + D_x.$$

Therefore,

$$\mathbb{E}_x^\mu[\tau] \leq 2(x - L + D_x)/\Delta_x,$$

which is the desired inequality (4.20).

We now bound $\mathbb{E}_x^\mu[\kappa(y_\alpha - x_\tau)]$ for $x \geq L$ and derive the inequality (4.19). Since $\kappa(\cdot)$ is non-decreasing by Assumption 4.6(i) and

$$y_\alpha \leq M, \quad x_{\tau-1} \geq L, \quad x_\tau = x_{\tau-1} - \xi_{\tau-1}(x_{\tau-1})$$

by the definition of the stopping time τ and policy μ , we have

$$\begin{aligned}
\mathbb{E}_x^\mu[\kappa(y_\alpha - x_\tau)] &\leq \mathbb{E}_x^\mu[\kappa(M - L + \xi_{\tau-1}(x_{\tau-1}))] \\
&\leq \mathbb{E}_x^\mu\left[\sum_{n=0}^{\tau-1} \kappa(M - L + \xi_n(x_n))\right] \\
&= \mathbb{E}_x^\mu\left[\sum_{n=0}^{\infty} \mathbb{1}(\tau > n) \kappa(M - L + \xi_n(x_n))\right] \\
&= \mathbb{E}_x^\mu\left[\sum_{n=0}^{\infty} \mathbb{1}(\tau > n) \mathbb{E}_x^\mu[\kappa(M - L + \xi_n(x_n)) \mid \mathcal{F}_n]\right] \\
&\leq \mathbb{E}_x^\mu\left[\sum_{n=0}^{\infty} \mathbb{1}(\tau > n) \sup_{y \in [L, x \vee M]} \mathbb{E}[\kappa(M - L + \xi(y))]\right].
\end{aligned}$$

Therefore,

$$\mathbb{E}_x^\mu[\kappa(y_\alpha - x_\tau)] \leq \mathbb{E}_x^\mu[\tau] \cdot \sup_{y \in [L, x \vee M]} \mathbb{E}[\kappa(M - L + \xi(y))].$$

This is the desired inequality (4.19).

A.3 An Illustrative Example for the (UC) Case

This example is adapted from an inventory-production system example in the book [22, Examples 8.6.2, 10.9.3]. In the original example, the random demands at each stage are i.i.d. and for the average cost problem, the production level must always be below the expected demand so as to ensure that the w -geometric ergodicity condition of [22, Thm. 10.3.1] holds. We relax these requirements in our example so that the state transition stochastic kernel need not be continuous and the MDP need not be uniformly w -geometrically ergodic. To show that the ACOI holds in this example, most of our efforts will be on making sure that $\sup_{\alpha \in (0,1)} \|h_\alpha\|_w < \infty$.

Regarding notation, for $x, y \in \mathbb{R}$, denote $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. Recall also that $(x)^+ = \max\{x, 0\}$.

Let the state and action spaces be $\mathcal{X} = \mathcal{A} = \mathbb{R}_+$. The states evolve according to

$$x_{n+1} = (x_n + z_n - \xi_n)^+,$$

where x_n is the stock level and z_n the amount of production at the beginning of the n th stage, and ξ_n is the random demand during that stage. As in Section 4.5.2, we let actions correspond to the stock levels after production:

$$a_n = x_n + z_n.$$

We assume that given a_n , the demand ξ_n depends on the value of a_n and is conditionally independent of the history $(x_0, a_0, \xi_0, x_1, \dots, \xi_{n-1}, x_n)$. As before, we denote the demand distributions by F_a , $a \in \mathbb{R}_+$, we write $\xi_n(a_n)$ for ξ_n to emphasize the dependence on a_n , and we use $\xi(a)$ to denote a generic random variable with probability distribution F_a and $\mathbb{E}[\xi(a)]$ to denote its expectation w.r.t. F_a .

Let the one-stage cost function be given by

$$c(x, a) = \kappa(z) + \psi(a) - s \mathbb{E}[a \wedge \xi(a)] \quad \text{with } z = a - x.$$

The functions $\kappa : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ specify the production cost and the maintenance/holding cost, respectively, and $s \mathbb{E}[a \wedge \xi(a)]$ is the sales revenue with s being the unit sale price. This definition of $c(\cdot)$ is similar to the one in [22, Example 8.6.2], except that we do not require $\kappa(\cdot)$ and $\psi(\cdot)$ to be continuous or l.s.c.

Assumption A.1 (conditions on one-stage costs). *On \mathbb{R}_+ , the production cost $\kappa(\cdot)$ is bounded on bounded intervals, and the maintenance or holding cost $\psi(\cdot)$ is bounded above by some polynomial function.*

A.3.1 Definitions of Action Sets

We assume that beyond certain stock level L , the demand $\xi(a)$ becomes “saturated” and follows a fixed probability distribution, no longer affected by a . Accordingly, we separate the states into three groups,

$$\{0\}, \quad (0, L), \quad [L, +\infty).$$

The state 0 will be special in this example. The feasible action sets $A(x)$ are bounded intervals of the form $[x, a_x]$, and we define them for each group of states differently, by placing different constraints on the admissible production levels z :

- (i) For all $x \geq L$, $z \in [0, \theta]$ for some $\theta > 0$ (so $A(x) = [x, x + \theta]$). We will let the parameter θ be smaller than the mean demand, in order to satisfy the (UC) model conditions. The precise definition of θ will be given below.
- (ii) For $0 < x < L$, $z \in [0, \theta_x]$ for some $\theta_x > 0$, and $\sup_{x \in (0, L)} \theta_x < \infty$. There are no other restrictions on θ_x ; in particular, the production level z can exceed the expected demand $\mathbb{E}[\xi(x + z)]$. (Regarding measurability, it suffices that θ_x is a Borel measurable function of x .)
- (iii) For the state 0, $z \in [0, \bar{a}]$ where \bar{a} will be made large enough so that $A(0) = [0, \bar{a}]$ contains $A(x)$ for a large subset of states x . The precise definition of $A(0)$ will be given below. The purpose of such a choice is to ensure $\sup_{\alpha \in (0, 1)} \|h_\alpha\|_w < \infty$ in this example; more specifically, it will be used to ensure that $\{h_\alpha\}$ can be bounded below by a multiple of the weight function w of our choice.

We now proceed to define the weight function $w(\cdot)$, the action set $A(0)$, and the maximum production level θ allowed at states $x \geq L$. Some conditions on the demand distributions will be needed:

Assumption A.2 (conditions on $\{\xi(a)\}$).

- (i) For $a \geq L$, F_a is the same, independent of a .
- (ii) $\inf_{a > 0} \mathbb{E}[\xi(a)] > 0$.
- (iii) $\{\xi(a) \mid a \geq 0\}$ is uniformly integrable.

These conditions will be used in the subsequent analysis to ensure that all the conditions required by Theorem 3.5 are met, except for the majorization condition. (For the latter, another condition will be added; cf. Assumption A.3.)

Choice of θ and Parameters in the (UC) Model: Similarly to [22, Example 10.9.3], for states $x \geq L$, we set the upper limit θ on the production levels to be such that for $a \geq L$,

$$0 < \theta < \mathbb{E}[\xi(a)]$$

(cf. Assumption A.2(i)-(ii)), and we then have that for some $r > 0$ (and all $a \geq L$),

$$\lambda := \mathbb{E}\left[e^{r(\theta - \xi(a))}\right] < 1.$$

Let the weight function w be

$$w(x) = e^{rx}.$$

By a direct calculation similar to that given in [22, pp. 72-73],

$$\sup_{z \in [0, \theta]} \int_{\mathcal{X}} w(y) q(dy \mid x, x + z) \leq \lambda w(x) + 1 \quad \forall x \geq L. \quad (\text{A.1})$$

For states $x < L$, since $\cup_{x < L} A(x)$ is bounded, we have

$$b := \sup_{x < L, a \in A(x)} \int_{\mathbb{X}} w(y) q(dy \mid x, a) < \infty.$$

The (UC) model condition (b) is then satisfied, for the preceding constants λ, b and weight function $w(\cdot)$. In view of Assumption A.1 and the fact that the admissible production levels are uniformly bounded for all states, the (UC) model condition (a) on the one-stage cost function is also satisfied for some constant \hat{c} .

Choice of $A(0)$: For the state 0, we choose an especially large set $A(0)$ as follows. For some $\lambda' \in (\lambda, 1)$, let $\tilde{L} = \max\{L, -\ln(\lambda' - \lambda)/r\}$. Then $(\lambda' - \lambda)e^{r\tilde{L}} \geq 1$ and by (A.1),

$$\sup_{z \in [0, \theta]} \int_{\mathbb{X}} w(y) q(dy \mid x, x+z) \leq \lambda' w(x) \quad \forall x \geq \tilde{L} \geq L. \quad (\text{A.2})$$

Let $A(0) = [0, \bar{a}]$ be such that

$$A(0) \supset A(x) \quad \forall 0 < x \leq \tilde{L} \quad (\text{A.3})$$

(e.g., let $\bar{a} = (\tilde{L} + \theta) \vee \sup_{x \in (0, L)} (x + \theta_x) < \infty$). As mentioned and will be shown shortly, the purpose of this choice of $A(0)$ is to make sure that in our subsequent analysis, $h_\alpha(x)$ can be bounded from below independently of α .

A.3.2 Bounding $\{h_\alpha\}$

Set $\bar{x} = 0$ in the definition of the relative value functions $h_\alpha(x) = v_\alpha(x) - v_\alpha(\bar{x})$, $\alpha \in (0, 1)$. We prove below that under the preceding assumptions, $\sup_{\alpha \in (0, 1)} \|h_\alpha\|_w < \infty$.

Let $\hat{\mathbb{X}} := [0, \tilde{L}]$. We derive upper/lower bounds on $h_\alpha(x)$, first for $x \in \hat{\mathbb{X}}$ and then for $x \notin \hat{\mathbb{X}}$. To calculate the upper bounds, we consider the policy π that makes no production so that $z_n = 0$ and $a_n = x_n$ always, and we bound first the expected hitting time to the state $\bar{x} = 0$ under π . Let $\bar{\tau} := \inf\{n \geq 0 \mid x_n = 0\}$.

Bounding $\mathbb{E}_x^\pi[\bar{\tau}]$ for $x \in \hat{\mathbb{X}}$: The derivation is similar to that of (4.20) in the example in Section 4.5.2 (see Appendix A.2). Let $\Delta := \inf_{y \in [0, \tilde{L}]} \mathbb{E}[\xi(y)] > 0$ (cf. Assumption A.2(ii)). By Assumption A.2(iii), there exists $0 < D < \infty$ such that

$$\sup_{y \in [0, \tilde{L}]} \mathbb{E}[\xi(y) \mathbb{1}(\xi(y) > D)] \leq \Delta/2.$$

For $n \geq 0$, define $\tilde{x}_0 := x_0 = x$, $\tilde{x}_{n+1} := x_n - \xi_n(x_n)$; then $\tilde{x}_{n+1} = x_{n+1}$ if $\xi_n(x_n) \leq x_n$. Let $Z_n := (\tilde{x}_n + D)^+$ and let \mathcal{F}_n be the σ -algebra generated by (x_0, x_1, \dots, x_n) . A direct calculation shows that

$$\mathbb{E}_x^\pi[Z_{n+1} \mid \mathcal{F}_n] \leq \begin{cases} Z_n - \Delta/2 & \text{when } x_n > 0; \\ Z_n + D & \text{when } x_n = 0. \end{cases}$$

Applying a comparison theorem [32, Prop. 11.3.2] to the nonnegative random variables $\{Z_n\}$, we obtain

$$\mathbb{E}_x^\pi\left[\sum_{n=0}^{\bar{\tau}-1} \frac{1}{2} \Delta\right] \leq Z_0 = x + D,$$

so

$$\mathbb{E}_x^\pi[\bar{\tau}] \leq 2(x + D)/\Delta \leq 2(\tilde{L} + D)/\Delta. \quad (\text{A.4})$$

Bounding $h_\alpha(x)$ from above for $x \in \hat{\mathbb{X}}$: The calculations and reasoning are similar to those given in Example 4.1. We have

$$\begin{aligned} v_\alpha(x) &\leq \mathbb{E}_x^\pi \left[\sum_{n=0}^{\bar{\tau}-1} \alpha^n c(x_n, a_n) + \alpha^{\bar{\tau}} v_\alpha(\bar{x}) \right] \\ &\leq \mathbb{E}_x^\pi \left[\sum_{n=0}^{\bar{\tau}-1} \alpha^n c(x_n, a_n) \right] + v_\alpha(\bar{x}) + \mathbb{E}_x^\pi [1 - \alpha^{\bar{\tau}}] \cdot |v_\alpha(\bar{x})| \\ &\leq \hat{c} \mathbb{E}_x^\pi [\bar{\tau}] \cdot \sup_{x' \in \hat{\mathbb{X}}} w(x') + v_\alpha(\bar{x}) + \ell_{\bar{x}} \mathbb{E}_x^\pi [\bar{\tau}], \end{aligned} \quad (\text{A.5})$$

where, to derive the last inequality, for the first term, we used the (UC) model condition (a) and the fact that the states $x_n \in \hat{\mathbb{X}}$ for all n under π , and for the last term, we used the fact that $\mathbb{E}_x^\pi [1 - \alpha^{\bar{\tau}}] \leq (1 - \alpha) \mathbb{E}_x^\pi [\bar{\tau}]$ and in (UC), $|(1 - \alpha)v_\alpha(\bar{x})|$ is bounded by some constant $\ell_{\bar{x}}$ for all $\alpha \in (0, 1)$.

Bounding $h_\alpha(x)$ from below for $x \in \hat{\mathbb{X}}$: By the construction of this example, $A(\bar{x}) = A(0) \supset A(x)$ for all $x \in \hat{\mathbb{X}} = [0, \tilde{L}]$ and $q(dy | x, a)$ depends on a only. Therefore,

$$v_\alpha(x) - v_\alpha(\bar{x}) \geq -\sup_{a \in A(x)} |c(x, a) - c(\bar{x}, a)| \geq -2\hat{c} \sup_{x' \in \hat{\mathbb{X}}} w(x'). \quad (\text{A.6})$$

Bounding $h_\alpha(x)$ for $x \notin \hat{\mathbb{X}}$: Let $\tau = \inf\{n \geq 0 \mid x_n \in \hat{\mathbb{X}}\}$. In view of (A.2), the same derivations (4.9)-(4.11) and reasoning given in Example 4.1 apply here and yield the inequalities:

$$\begin{aligned} v_\alpha(x) &\leq \hat{c} \ell w(x) + v_\alpha(\bar{x}) + \mathbb{E}_x^{\mu_\alpha} [\alpha^\tau v_\alpha(x_\tau) - v_\alpha(\bar{x})], \\ v_\alpha(x) &\geq -\hat{c} \ell w(x) + v_\alpha(\bar{x}) + \mathbb{E}_x^{\mu_\alpha} [\alpha^\tau v_\alpha(x_\tau) - v_\alpha(\bar{x})] - \epsilon, \end{aligned}$$

where $\ell > 0$ is some constant independent of α and x ; $\epsilon > 0$ is some arbitrary small number; and μ_α is a stationary ϵ -optimal policy for the α -discounted problem. Similarly to the derivations of (4.12)-(4.13) given in Example 4.1, we can bound the third term in the above inequalities as

$$|\mathbb{E}_x^{\mu_\alpha} [\alpha^\tau v_\alpha(x_\tau) - v_\alpha(\bar{x})]| \leq \sup_{x' \in \hat{\mathbb{X}}} |v_\alpha(x') - v_\alpha(\bar{x})| + \ell \ell_{\bar{x}} w(x),$$

where, by (A.4)-(A.6),

$$\sup_{x' \in \hat{\mathbb{X}}} |v_\alpha(x') - v_\alpha(\bar{x})| \leq 2\hat{c} e^{r\tilde{L}} + (\hat{c} e^{r\tilde{L}} + \ell_{\bar{x}}) \cdot 2(\tilde{L} + D)/\Delta,$$

which is a constant independent of α .

Combining the preceding relations shows that for some constants $\tilde{\ell}_1, \tilde{\ell}_2$ independent of α and x ,

$$|v_\alpha(x) - v_\alpha(\bar{x})| \leq \tilde{\ell}_1 w(x) + \tilde{\ell}_2, \quad x \in \mathbb{X}.$$

It now follows that $\sup_{\alpha \in (0,1)} \|h_\alpha\|_w < \infty$, so Assumption 3.1 is satisfied in this example.

A.3.3 Satisfying the Majorization Condition

We impose an additional condition on the demand distributions:

Assumption A.3. *The demand distributions $\{F_a \mid a > 0\}$ have densities w.r.t. the Lebesgue measure that are uniformly bounded above.*

The majorization condition (Assumption 3.3) holds easily under the preceding assumptions. In particular, for each state $x \geq 0$ and $\epsilon > 0$, with $A(x) = [x, a_x]$, we observe the following:

- Let $K_\epsilon(x) = A(x)$; then Assumption 3.3(i) holds trivially.
- For Assumption 3.3(ii), the required majorizing finite measure ν can be defined as

$$\nu := f^{\max} \varphi + \delta_0,$$

where f^{\max} is an upper bound on the densities of $\{F_a \mid a > 0\}$ (cf. Assumption A.3); φ equals the Lebesgue measure on $[0, a_x]$ and the trivial measure on (a_x, ∞) ; and δ_0 is the Dirac measure at 0. This finite measure ν majorizes $q(dy | x, a)$, $a \in [x, a_x]$ and thus meets the requirement in Assumption 3.3(ii).

- Finally, notice that $x_1 \in [0, a_x]$ if $x_0 = x$, whereas $w(y) = e^{ry}$ is bounded on $[0, a_x]$. Therefore, the uniform integrability condition required by Assumption 3.3(iii) is also trivially satisfied.

We have now verified all the conditions in Theorem 3.5 and can therefore conclude that the ACOI holds in this example.

Acknowledgments

The author would like to thank Professor Eugene Feinberg and two anonymous reviewers for important critical comments on the previous version of this manuscript, and for pointing her to several related early and recent works on Borel-space MDPs. The author is also grateful to Dr. Martha Steenstrup, who read parts of her preliminary draft and gave her advice on improving the presentation.

References

- [1] A. ARAPOSTATHIS, V. S. BORKAR, E. FERNÁNDEZ-GAUCHERAND, M. K. GHOSH, AND S. I. MARCUS, *Discrete-time controlled Markov processes with average cost criterion: A survey*, SIAM J. Control Optim., 31 (1993), pp. 282–344.
- [2] D. ASSAF, *Invariant problems in dynamic programming—average reward criterion*, Stoch. Proc. Appl., 10 (1980), pp. 313–322.
- [3] D. P. BERTSEKAS AND S. E. SHREVE, *Stochastic Optimal Control: The Discrete Time Case*, Academic Press, New York, 1978.
- [4] D. BLACKWELL, *A Borel set not containing a graph*, Ann. Math. Statist., 39 (1968), pp. 1345–1347.
- [5] D. BLACKWELL, D. FREEDMAN, AND M. ORKIN, *The optimal reward operator in dynamic programming*, Ann. Probability, 2 (1974), pp. 926–941.
- [6] R. CAVAZOS-CADENA, *A counterexample on the optimality equation in Markov decision chains with the average cost criterion*, System and Control Lett., 16 (1991), pp. 387–392.
- [7] X. CHEN AND D. SIMCHI-LEVI, *Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: The infinite horizon case*, Math. Oper. Res., 29 (2004), pp. 698–723.
- [8] O. L. V. COSTA AND F. DUFOUR, *Average control of Markov decision processes with Feller transition probabilities and general action spaces*, J. Math. Anal. Appl., 396 (2012), pp. 58–69.
- [9] R. M. DUDLEY, *Real Analysis and Probability*, Cambridge University Press, Cambridge, 2002.
- [10] N. DUNFORD AND J. T. SCHWARTZ, *Linear Operators, Part I*, Wiley, New York, 1957.
- [11] E. B. DYNKIN AND A. A. YUSHKEVICH, *Controlled Markov Processes*, Springer, New York, 1979.
- [12] E. A. FEINBERG, *Optimality conditions for inventory control*, in INFORMS Tutorials in Operations Research, INFORMS, 2016, pp. 14–44.
- [13] E. A. FEINBERG, P. O. KASYANOV, AND Y. LIANG, *Fatou’s lemma in its classic form and Lebesgue’s convergence theorems for varying measures with applications to MDPs*, 2019, <https://arxiv.org/abs/1902.01525>. To appear in Theory Probab. Appl.
- [14] E. A. FEINBERG, P. O. KASYANOV, AND N. V. ZADOIANCHUK, *Average cost Markov decision processes with weakly continuous transition probabilities*, Math. Oper. Res., 37 (2012), pp. 591–607.
- [15] E. A. FEINBERG AND M. E. LEWIS, *On the convergence of optimal actions for Markov decision processes and the optimality of (s, S) inventory policies*, Naval Res. Logist., 65 (2018), pp. 619–637.
- [16] E. A. FEINBERG AND Y. LIANG, *On the optimality equation for average cost Markov decision processes and its validity for inventory control*, Ann. Oper. Res., (2017). <https://doi.org/10.1007/s10479-017-2561-9>.
- [17] J. GONZÁLEZ-HERNÁNDEZ AND O. HERNÁNDEZ-LERMA, *Envelopes of sets of measures, tightness, and Markov control processes*, Appl. Math. Optim., 40 (1999), pp. 377–392.
- [18] E. GORDIENKO AND O. HERNÁNDEZ-LERMA, *Average cost Markov control processes with weighted norms: existence of canonical policies*, Appl. Math. (Warsaw), 23 (1995), pp. 199–218.
- [19] L. G. GUBENKO AND E. S. SHTATLAND, *On controlled, discrete-time Markov decision processes*, Theory Probab. Math. Statist., 7 (1975), pp. 47–61.

- [20] O. HERNÁNDEZ-LERMA, *Average optimality in dynamic programming on Borel spaces—unbounded costs and controls*, System and Control Lett., 17 (1991), pp. 337–242.
- [21] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, Springer, New York, 1996.
- [22] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999.
- [23] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Fatou’s and Lebesgue’s convergence theorems for measures*, J. Appl. Math. Stoch. Anal., 13 (2000), pp. 137–146.
- [24] O. HERNÁNDEZ-LERMA AND J. B. LASSERRE, *Markov Chains and Invariant Probabilities*, Birkhäuser Verlag, Basel, 2003.
- [25] O. HERNÁNDEZ-LERMA AND R. ROMERA, *Limiting discounted-cost control of partially observable stochastic systems*, SIAM J. Control Optim., 40 (2001), pp. 348–369.
- [26] O. HERNÁNDEZ-LERMA AND O. VEGA-AMAYA, *Infinite-horizon Markov control processes with undiscounted cost criteria: From average to overtaking optimality*, Appl. Math. (Warsaw), 25 (1998), pp. 153–178.
- [27] A. JAŚKIEWICZ, *Semi-Markov control processes with non-compact action spaces and discontinuous costs*, Appl. Math. (Warsaw), 36 (2009), pp. 29–42.
- [28] A. JAŚKIEWICZ AND A. S. NOWAK, *On the optimality equation for average cost Markov control processes with Feller transition probabilities*, J. Math. Anal. Appl., 316 (2006), pp. 495–509.
- [29] N. V. KARTASHOV, *Inequalities in theorems of ergodicity and stability for Markov chains with common phase space. II*, Theory Probab. Appl., 30 (1985), pp. 507–515.
- [30] M. KURANO, *Markov decision processes with a Borel measurable cost function—the average case*, Math. Oper. Res., 11 (1986), pp. 309–320.
- [31] A. MAITRA AND W. SUDDERTH, *The optimal reward operator in negative dynamic programming*, Math. Oper. Res., 17 (1992), pp. 921–931.
- [32] S. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, Cambridge University Press, Cambridge, 2nd ed., 2009.
- [33] S. P. MEYN, *The policy iteration algorithm for average reward Markov decision processes with general state space*, IEEE Trans. Automat. Contr., 42 (1997), pp. 1663–1680.
- [34] E. NUMMELIN, *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge University Press, Cambridge, 1984.
- [35] K. R. PARTHASARATHY, *Probability Measures on Metric Spaces*, Academic Press, New York, 1967.
- [36] A. B. PIUNOVSKIY, *General Markov models with the infinite horizon*, Problems of Control and Information Theory, 18 (1989), pp. 169–182.
- [37] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer, Berlin, 1st ed., 1998.
- [38] S. M. ROSS, *Arbitrary state Markovian decision processes*, Ann. Math. Statist., 39 (1968), pp. 2118–2122.
- [39] H. L. ROYDEN, *Real Analysis*, Macmillan, New York, 1968.
- [40] M. SCHÄL, *Conditions for optimality in dynamic programming and for the limit of n -stage optimal policies to be optimal*, Z. Wahrscheinlichkeitstheorie verw. Gebiete, 32 (1975), pp. 179–196.
- [41] M. SCHÄL, *Average optimality in dynamic programming with general state space*, Math. Oper. Res., 18 (1993), pp. 163–172.
- [42] L. I. SENNOTT, *Average cost optimal stationary policies in infinite state Markov decision processes with unbounded costs*, Oper. Res., 37 (1989), pp. 626–633.
- [43] S. E. SHREVE, *Resolution of measurability problems in discrete-time stochastic control*, in Stochastic Control Theory and Stochastic Differential Systems, Springer, Berlin, 1979, pp. 580–587.
- [44] S. E. SHREVE AND D. P. BERTSEKAS, *Alternative theoretical frameworks for finite horizon discrete-time stochastic optimal control*, SIAM J. Control Optim., 16 (1978), pp. 953–978.
- [45] S. E. SHREVE AND D. P. BERTSEKAS, *Universally measurable policies in dynamic programming*, Math. Oper. Res., 4 (1979), pp. 15–30.
- [46] S. M. SRIVASTAVA, *A Course on Borel Sets*, Springer, New York, 1998.
- [47] R. E. STRAUCH, *Negative dynamic programming*, Ann. Math. Statist., 37 (1966), pp. 871–890.
- [48] J. VAN DER WAL, *Stochastic Dynamic Programming*, The Mathematical Centre, Amsterdam, 2nd ed., 1984.
- [49] O. VEGA-AMAYA, *The average cost optimality equation: A fixed point approach*, Bol. Soc. Mat. Mexi-

- cana, 9 (2003), pp. 185–195.
- [50] O. VEGA-AMAYA, *On the vanishing discount factor approach for Markov decision processes with weakly continuous transition probabilities*, J. Math. Anal. Appl., 426 (2015), pp. 978–985.
 - [51] O. VEGA-AMAYA, *Solutions of the average cost optimality equation for Markov decision processes with weakly continuous kernel: The fixed-point approach revisited*, J. Math. Anal. Appl., 464 (2018), pp. 152–163.
 - [52] A. F. VEINOTT, *On discrete dynamic programming with sensitive discount optimality criteria*, Ann. Math. Statist., 40 (1969), pp. 1635–1660.
 - [53] H. YU, *On convergence of value iteration for a class of total cost Markov decision processes*, SIAM J. Control Optim., 53 (2015), pp. 1982–2016.
 - [54] H. YU, *On Markov decision processes with Borel spaces and an average cost criterion*, 2019, <https://arxiv.org/abs/1901.03374>.
 - [55] H. YU, *On the minimum pair approach for average-cost Markov decision processes with countable discrete action spaces and strictly unbounded costs*, SIAM J. Control Optim., 58 (2020), pp. 660–685.
 - [56] H. YU AND D. P. BERTSEKAS, *A mixed value and policy iteration method for stochastic control with universally measurable policies*, Math. Oper. Res., 40 (2015), pp. 926–968.