On Linear Programming for Constrained and Unconstrained Average-Cost Markov Decision Processes with Countable Action Space and Strictly Unbounded Costs

Huizhen Yu†

Abstract

We consider the linear programming approach for constrained and unconstrained Markov decision processes (MDPs) with the long-run average cost criterion, where the class of MDPs in our study have a Borel state space and a discrete, countable action space. Under a strict unboundedness condition on the one-stage costs and a recently introduced majorization condition on the state transition stochastic kernel, we study infinite-dimensional linear programs for the average-cost MDPs and prove the absence of duality gap and other optimality results. A characteristic of these results is that they do not require a lower semicontinuous MDP model and as such, they can be applied to countable action space MDPs where the dynamics and one-stage costs are discontinuous in the state variable.

Keywords:
Markov decision processes; Borel state space; countable action space; average cost; constraints minimum pair; majorization condition; infinite-dimensional linear programs; duality

∗This research was supported by a grant from Alberta Innovates—Technology Futures.
†RLAI Lab, Department of Computing Science, University of Alberta, Canada (janey.hzyu@gmail.com)
# Linear Programming for Average-Cost MDPs

## Contents

1. **Introduction** ................................................................. 3

2. **Preliminaries** ............................................................... 4
   2.1  MDP Model, Average Cost Criterion, and Minimum Pair Approach ....... 5
       2.1.1 Average Cost Criterion and Minimum Pair ............................ 5
       2.1.2 Model Assumptions and Existence of Stationary Minimum Pair ....... 6
   2.2  Linear Programs in Topological Vector Spaces .......................... 7

3. **Linear Programming for Average-Cost MDPs** ............................ 8
   3.1  Primal and Dual Linear Programs ........................................ 9
   3.2  Optimality Results and Discussion ....................................... 11
   3.3  Proofs ............................................................................. 14
       3.3.1 Proof of Theorem 3.1 ..................................................... 14
       3.3.2 Proof of Prop. 3.2 ....................................................... 19

4. **Extension to Constrained Average-Cost MDPs** ............................ 20
   4.1  Model Assumptions and Existence of Stationary Optimal Pairs ......... 21
   4.2  Linear Programming Formulation and Optimality Results ............... 22
   4.3  Proofs ............................................................................. 25
       4.3.1 Proof of Theorem 4.2 (Outline) ....................................... 25
       4.3.2 Proof of Theorem 4.3 (Outline) ....................................... 27
       4.3.3 Proofs of Props. 4.4 and 4.5 .......................................... 28

References ................................................................................. 30


1 Introduction

We consider discrete-time Markov decision processes (MDPs) with the long-run average cost criterion. Our focus will be on the linear programming (LP) approach, for a class of unconstrained and constrained MDPs that have a Borel state space, a discrete, countable action space, and strictly unbounded one-stage costs.

LP methods for average-cost MDPs have a long history and an extensive literature (see e.g., [11, 23, 26, 22] on some early work on finite-space MDPs, [27, 28, 32] on countable state space and countable or compact action space MDPs, and [14, 20, 23, 31, 38] on Borel space MDPs; see also the books [1, 15, 21, 22, 35] and their references). For the special case of strictly unbounded costs we consider, where the one-stage costs are nonnegative and unbounded off compact sets, there is a line of research that connects the LP approach with the minimum pair approach for average-cost MDPs (see e.g., [22, Chaps. 11-12]). The latter approach is also related to the convex analytic approach [8]. Its idea is to consider average costs of all policies for all initial distributions, with the goal of finding a stationary policy and an associated initial distribution that together attain the minimal average cost, where the associated initial distribution is not an arbitrary one but an invariant probability measure induced by that stationary policy. In this paper, we shall refer to such a policy and initial distribution pair as a stationary minimum pair. When the one-stage costs are strictly unbounded, under various conditions on the MDP model (to be explained shortly), a stationary minimum pair can be proved to exist. Finding such a pair can be formulated as a linear program in the space of stationary policies and their induced invariant probability measures. This provides a method to solve the average-cost MDP or to gain further insights about its dynamic programming-related properties, such as optimality equations, through the duality relationships in LP. This method can be applied to general multichain MDPs, which is an advantage since the chain structure of an MDP can be complicated, and hard to analyze when the state space is uncountably infinite.

To our knowledge, Denardo [10] was the first to propose this LP method for solving finite-space multichain MDPs (although he focused more on algorithms than on the minimum pair idea as a general approach). For infinite Borel space MDPs, the minimum pair approach was introduced by Kurano [30], motivated by the ideas of occupancy measures from Borkar [6, 7], and it was further developed by Hernández-Lerma [16] and Vega-Amaya [37] (see also [22, Chap. 11]). Kurano considered compact spaces. Hernández-Lerma and Vega-Amaya considered non-compact spaces with strictly unbounded costs, and obtained further results, including strong and pathwise average-cost optimality results, besides the existence of a stationary minimum pair. Hernández-Lerma and Lasserre [20] (see also the book chapters [22, Chaps. 12] and [23]) formulated an LP framework for Borel space average-cost MDPs by using the theory of infinite-dimensional LP (Anderson and Nash [2]). They characterized the relation between the values of the primal/dual linear programs and the minimal average cost of an MDP, and proved the absence of duality gap under tightness conditions closely related to the minimum pair method. Before [20], a much earlier duality result was proved by Yamada [38] for compact Euclidean state and action spaces and bounded costs, under geometric ergodicity conditions. Additional results and generalizations of some of the results of [20] were given by Hernández-Lerma and González-Hernández [17]. Extensions of the LP framework to constrained MDPs were subsequently studied by Kurano et al. [31] for compact spaces and by Hernández-Lerma et al. [19] for non-compact spaces.

Our work builds upon the earlier researches on Borel space constrained and unconstrained MDPs just mentioned. The action space in those prior results is more general than the countable action space we deal with in this paper. However, except for [38], they all require a lower semicontinuous MDP model assumption; namely, they require the one-stage cost function to be lower semicontinuous and the state transition stochastic kernel to be (weakly) continuous ([38] involves different continuity conditions; see Remark 3.2 for details). This is a restriction.

Recently, we introduced in [40] a majorization condition on the state transition stochastic kernel, to replace the lower semicontinuous model assumption, for the case of countable action spaces
Linear Programming for Average-Cost MDPs

(with the discrete topology), and we obtained the existence of a stationary minimum pair and other average-cost optimality results analogous to those for lower semicontinuous MDPs given by [16, 30, 37]. The idea of the majorization condition, roughly speaking, is as follows. We require the existence of finite Borel measures on the state space that can majorize certain sub-stochastic kernels created from the state transition stochastic kernel, at all admissible state-action pairs (see Assumption 2.1(M)). We then use those majorizing finite measures in combination with Lusin’s theorem [13, Theorem 7.5.2], and this allows us to extract arbitrarily large sets (large as measured by a given finite measure) on which certain Borel measurable functions involved in our analysis have desired continuity properties. With this technique—although its application range is currently limited to the case of countable action spaces, we are able to avoid the lower semicontinuous model assumption and obtain results in [40] that can be applied to MDPs with discontinuous dynamics and one-stage costs.

The purpose of the present paper is to study further the implications of the majorization condition in the LP context, for both unconstrained and constrained MDPs. The main contributions can be summarized as follows:

(i) For an unconstrained average-cost MDP, under the strictly unbounded cost condition and the majorization condition, we prove there is no duality gap between the primal and dual linear programs for an LP formulation (see Theorem 3.1).

(ii) For a constrained average-cost MDP, under similar conditions, we prove the existence of stationary optimal pair and stationary lexicographically optimal pair (which are analogous to stationary minimum pairs for an unconstrained MDP), and we then prove the absence of duality gap for an LP formulation (see Theorem 4.2 and Theorem 4.3, respectively).

In addition, we also discuss the maximizing sequences of dual linear programs and their relation with certain versions of average cost optimality equations (ACOE) (see Props. 3.2 for unconstrained MDPs and Props. 4.4, 4.5 for constrained MDPs). Our results for unconstrained (respectively, constrained) MDPs given in this paper can be compared with some of the prior results in [22, Chap. 12] and [23] (respectively, [19] and [31]) for lower semicontinuous models.

We comment that although we focus exclusively on the average cost criterion in this paper, with minor changes in the proof arguments, the majorization condition can also be applied to constrained or multi-objective discounted-cost MDPs similar to those studied in [14, 18, 24], for finding constrained optimal or Pareto optimal policies (for a given initial distribution) using the LP approach, in the case of countable action spaces. We also remark that a different majorization condition has been introduced in our recent work [39] to replace the lower semicontinuous model assumption in the vanishing discount factor approach for average-cost MDPs with both Borel state and action spaces.

The rest of this paper is organized as follows. Section 2 gives background materials about the average-cost MDP model, some prior optimality results for the minimum pair approach, and an overview of linear programs in topological vector spaces. Section 3 presents our LP formulation and duality results for unconstrained MDPs, and Section 4 the extension to constrained MDPs.

2 Preliminaries

We first introduce some notations and basic definitions that will be needed throughout the paper. For a topological space $X$, $\mathcal{B}(X)$ denotes the Borel $\sigma$-algebra on $X$, and $\mathcal{P}(X)$ denotes the set of probability measures on $\mathcal{B}(X)$. In general, measures on $\mathcal{B}(X)$ will be referred to as Borel measures.

A Borel space is a separable metrizable space that is homeomorphic to a Borel subset of some Polish space (i.e., a separable and completely metrizable space) [3, Chap. 7]. Let $X$ and $Y$ be Borel spaces. A Borel measurable stochastic kernel on $Y$ given $X$, denoted $q(\text{dy} \mid x)$, is a Borel measurable function from $X$ into $\mathcal{P}(Y)$, where the space $\mathcal{P}(Y)$ is endowed with the topology of weak convergence. If this function $q(\text{dy} \mid x)$ from $X$ into $\mathcal{P}(Y)$ is continuous, we call it a continuous stochastic kernel.
(in the literature it is also called weak Feller or weakly continuous).

For the space $\mathcal{P}(X)$ or more generally, the space of finite Borel measures on $X$, besides the topology of weak convergence just mentioned, other topologies will also be considered later in this paper, when these spaces appear in infinite-dimensional linear programs.

In what follows, we present background materials about average-cost MDPs (Section 2.1) and infinite-dimensional linear programs in topological vector spaces (Section 2.2).

### 2.1 MDP Model, Average Cost Criterion, and Minimum Pair Approach

We consider an MDP with state space $\mathcal{X}$ and action space $\mathcal{A}$, where $\mathcal{X}$ is a Borel space and $\mathcal{A}$ is a countable space endowed with the discrete topology. The control constraint is specified by a set-valued map $A : \mathcal{X} \to 2^\mathcal{A}$; in particular, at a state $x \in \mathcal{X}$, the set of admissible actions is given by a nonempty set $A(x) \subseteq \mathcal{A}$, and the graph of the map $A(\cdot)$,

$$\Gamma := \{(x, a) \mid x \in \mathcal{X}, a \in A(x)\},$$

is assumed to be a Borel subset of $\mathcal{X} \times \mathcal{A}$. If an action $a \in A(x)$ is taken at state $x$, a one-stage cost $c(x, a)$ is incurred, followed by a probabilistic state transition. We assume that the one-stage cost function $c : \mathcal{X} \times \mathcal{A} \to [0, +\infty]$ is nonnegative and Borel measurable, finite-valued on $\Gamma$ and taking the value $+\infty$ outside $\Gamma$. For state transition, we assume that it is governed by a Borel measurable stochastic kernel $q(dy \mid x, a)$ on $\mathcal{X}$ given $\mathcal{X} \times \mathcal{A}$.

A policy is a sequence of stochastic kernels on $\mathcal{A}$ that specify how to take actions at each stage, given the history up to that stage. More precisely, for infinite-horizon average cost problems that we consider, a Borel measurable policy is an infinite sequence $\pi := (\mu_0, \mu_1, \ldots)$ where for each $n \geq 0$, $\mu_n(da_n \mid x_0, a_0, \ldots, a_{n-1}, x_n)$ is a Borel measurable stochastic kernel on $\mathcal{A}$ given $(\mathcal{X} \times \mathcal{A})^n \times \mathcal{X}$ and obeys the control constraint of the MDP:

$$\mu_n(A(x_n) \mid x_0, a_0, \ldots, a_{n-1}, x_n) = 1, \quad \forall (x_0, a_0, \ldots, a_{n-1}, x_n) \in (\mathcal{X} \times \mathcal{A})^n \times \mathcal{X}.$$

Such a policy is called stationary if the function $(x_0, a_0, \ldots, a_{n-1}, x_n) \mapsto \mu_n(da_n \mid x_0, a_0, \ldots, a_{n-1}, x_n)$ depends only on the state $x_n$, in the same way for every $n \geq 0$. In this case, we can write the policy as $\pi = (\mu, \mu, \ldots)$ for a Borel measurable stochastic kernel $\mu(da \mid x)$ on $\mathcal{A}$ given $\mathcal{X}$ that obeys the control constraint of the MDP, and we will simply designate this policy by $\mu$.

Let $\Pi$ denote the space of Borel measurable policies and $\Pi_s$, the subset of all stationary policies in $\Pi$. Given that the action space $\mathcal{A}$ is countable, $\Pi$ and $\Pi_s$ are nonempty (see e.g., [40, Sec. 2]), and these Borel measurable policies will be adequate for our purpose—henceforth, we shall simply call them policies. We also note that in the above and throughout the paper, for notational simplicity, although $\mathcal{A}$ is countable, we write probability measures on $\mathcal{A}$ using the general notation for probability measures on a possibly uncountably infinite space.

#### 2.1.1 Average Cost Criterion and Minimum Pair

In an MDP, a policy $\pi \in \Pi$ and an initial (state) distribution $\zeta \in \mathcal{P}(\mathcal{X})$ induce a stochastic process $\{(x_n, a_n)\}_{n \geq 0}$ on the infinite product of state and action spaces, $(\mathcal{X} \times \mathcal{A})^\infty$. The probability measure for this process is uniquely determined by the initial distribution $\zeta$, the sequence of stochastic kernels in $\pi$, and the state transition stochastic kernel $q(dy \mid x, a)$ [Prop. 7.28]. We denote this probability measure by $\mathbb{P}^\pi_\zeta$ and the corresponding expectation operator by $\mathbb{E}^\pi_\zeta$.

The long-run expected average cost of the policy $\pi$ for the initial distribution $\zeta$ is defined by

$$J(\pi, \zeta) := \limsup_{n \to \infty} n^{-1} \mathbb{E}^\pi_\zeta\left[\sum_{k=0}^{n-1} c(x_k, a_k)\right].$$

We shall also refer to $J(\pi, \zeta)$ as the average cost of the pair $(\pi, \zeta)$. With the minimum pair approach, we consider the average costs of all policy and initial distribution pairs, and among these pairs, of special interest are the types of pairs given in the following definitions.
Let $\rho^*$ be the minimal average cost over all policies and initial distributions:

$$\rho^* := \inf_{\zeta \in \mathcal{P}(\mathcal{X})} \inf_{\pi \in \Pi} J(\pi, \zeta).$$

**Definition 2.1** (minimum pair). A pair $(\pi^*, \zeta^*) \in \Pi \times \mathcal{P}(\mathcal{X})$ is called a minimum pair if and only if $(\text{iff})$ $J(\pi^*, \zeta^*) = \rho^*$.

**Definition 2.2** (stationary pair and stationary minimum pair).

(a) For a stationary policy $\mu \in \Pi_s$ and an initial distribution $p \in \mathcal{P}(\mathcal{X})$, if $p$ is an invariant probability measure of the Markov chain induced by $\mu$ on $\mathcal{X}$, we call $(\mu, p)$ a stationary pair. The set of all stationary pairs is denoted by $\Delta_s$.

(b) If $(\mu^*, p^*) \in \Delta_s$ is a minimum pair, we call it a stationary minimum pair.

**Remark 2.1.** In the references [20, 22], the stationary policy in what we call a stationary pair $(\mu, p)$ is referred to as a "stable policy" if the average cost $J(\mu, p)$ is finite. In the reference [9], the probability measure $\gamma(d(x, a)) = \mu(da|x)p(dx)$ associated with a stationary pair $(\mu, p)$ is called an "ergodic occupation measure"; see Section 3.1 for a further discussion on such probability measures.

### 2.1.2 Model Assumptions and Existence of Stationary Minimum Pair

We now impose additional conditions on the MDP model. Recall that $\Gamma = \{(x, a) \mid x \in \mathcal{X}, a \in A(x)\}$ and it is the set of state and admissible action pairs. For a set $B$ in some space, let $B^c$ denote its complement; for a set $B \subset \mathcal{X} \times \mathcal{A}$, let $\text{proj}_\mathcal{X}(B)$ denote the projection of $B$ on $\mathcal{X}$.

**Assumption 2.1.**

**G** For some $\pi \in \Pi$ and $\zeta \in \mathcal{P}(\mathcal{X})$, the average cost $J(\pi, \zeta) < \infty$.

**SU** There exists a nondecreasing sequence of compact sets $\Gamma_n \uparrow \Gamma$ such that

$$\lim_{n \to \infty} \inf_{(x, a) \in \Gamma_n} c(x, a) = +\infty.$$

**M** For each compact set $K \in \{\text{proj}_\mathcal{X}(\Gamma_n)\}$, there exist an open set $O \supset K$ and a finite measure $\nu$ on $\mathcal{B}(\mathcal{X})$ such that

$$q((O \setminus D) \cap B \mid x, a) \leq \nu(B), \quad \forall B \in \mathcal{B}(\mathcal{X}), \ (x, a) \in \Gamma,$$  

where $D \subset \mathcal{X}$ is some closed set (possibly empty) such that restricted to $D \times \mathcal{A}$, the state transition stochastic kernel $q(dy \mid x, a)$ is continuous and the one-stage cost function $c$ is lower semicontinuous.

The first two conditions in this assumption are standard. Condition (G) excludes vacuous problems. Condition (SU) defines the case of strictly unbounded one-stage costs. As mentioned earlier, in the literature they have been used on lower semicontinuous MDP models, to derive optimality and LP duality results for those MDPs [16, 22, 30, 37].

Condition (M) was introduced in our recent work [40] to replace the lower semicontinuity model conditions, and we use the set $D$ to separate a "continuous part" of the model from the rest, in order to sharpen the condition (M), although this condition can also be used with $D$ being the empty set. The condition (M) seems natural for problems where the probability measures $\{q(\cdot \mid x, a) \mid (x, a) \in \Gamma\}$ have densities on $\mathcal{X} \setminus D$ with respect to (w.r.t.) a common $\sigma$-finite reference measure and those density functions are bounded uniformly from above. See [40] Example 3.2 and Remark 3.1 for some specific examples of situations where (M) is naturally satisfied or cannot be satisfied.

Under the preceding assumption, the following results are proved in [40] (see [40, Theorem 3.5] for additional optimality properties of a stationary minimum pair). They are analogous to the prior results for lower semicontinuous MDPs [16, 22, 30, 37], and they will serve as the starting point for the analyses we present in this paper.
Theorem 2.2 (optimality of stationary pairs [40, Prop. 3.2, Theorem 3.3]). Under Assumption 2.1, the following hold:

(i) For any pair \((\pi, \zeta) \in \Pi \times \mathcal{P}(\mathbb{K})\) with \(J(\pi, \zeta) < \infty\), there exists a stationary pair \((\bar{\mu}, \bar{p}) \in \Delta_s\) with \(J(\bar{\mu}, \bar{p}) \leq J(\pi, \zeta)\).

(ii) There exists a stationary minimum pair \((\mu^*, p^*) \in \Delta_s\).

2.2 Linear Programs in Topological Vector Spaces

In this subsection, we give a brief overview of infinite-dimensional linear programs in topological vector spaces. We refer the reader to the books [2, 36] for in-depth studies of these subjects, and to the book [22, Chap. 12.2] for a more detailed introduction than ours. Here we shall focus on a few basic concepts and results that we will need.

We consider topological vector spaces over the real field. A topological vector space is a vector space with a topology that is compatible with its algebraic structure (namely, with that topology, the addition and multiplication operations are continuous; see [36, Chap. I, Section 3]). Let \(X, Y\) be two (real) vector spaces, and let \(0\) denote the element zero for both spaces. The pair \((X, Y)\) is called a dual pair if there is a bilinear form \(\langle \cdot, \cdot \rangle : X \times Y \to \mathbb{R}\) such that

- for each \(x \neq 0\) in \(X\), there exists some \(y \in Y\) with \(\langle x, y \rangle \neq 0\),
- for each \(y \neq 0\) in \(Y\), there exists some \(x \in X\) with \(\langle x, y \rangle \neq 0\).

For a dual pair \((X, Y)\), the coarsest topology on \(X\) under which the function \(|\langle \cdot, y \rangle|\) is continuous for every \(y \in Y\) is called the weak topology on \(X\) determined by \(Y\), and denoted by \(\sigma(X, Y)\). By symmetry, \((Y, X)\) is also a dual pair and \(\sigma(Y, X)\), the weak topology on \(Y\) determined by \(X\), is likewise defined. With the respective weak topology, the space \(X\) or \(Y\) is a topological vector space that is separated (i.e., a Hausdorff space) and locally convex (i.e., every point in the space has a base of convex neighborhoods) [36, Chap. II, Section 3]. Convergence in \(X\) under the weak topology can be characterized as follows: a net \(\{x_i\}_{i \in I}\) in \(X\) converges to \(\bar{x} \in X\) iff

\[ \langle x_i, y \rangle \to \langle \bar{x}, y \rangle, \quad \forall y \in Y. \]

We consider equality-constrained linear programs and their dual linear programs in topological vector spaces. The definitions of these programs involve the following objects:

- two dual pairs of vector spaces \((X, Y)\) and \((Z, W)\), with each space endowed with its respective weak topology;
- a linear mapping \(L : X \to Z\) that is required to be weakly continuous (i.e., \(L\) is continuous under the topology \(\sigma(X, Y)\) for \(X\) and the topology \(\sigma(Z, W)\) for \(Z\));
- a convex cone \(\Lambda\) in \(X\) and its dual cone \(\Lambda^*\) in \(Y\) defined as

\[ \Lambda^* := \{ y \in Y \mid \langle x, y \rangle \geq 0, \forall x \in \Lambda \}. \]

The convex cones \(\Lambda\) and \(\Lambda^*\) induce a partial ordering "\(\leq\)" on \(X\) and \(Y\), respectively:

\[ x_1 \leq x_2 \text{ iff } x_2 - x_1 \in \Lambda; \quad y_1 \leq y_2 \text{ iff } y_2 - y_1 \in \Lambda^*. \]

The linear mapping \(L\) appears in the constraints of a linear program designated as the primal program \((P)\). Associated with \(L\) is another linear mapping \(L^*\) on the space \(W\), called the adjoint (or transpose) of \(L\), that maps each \(w \in W\) to a linear form on \(X\) and is defined by the identity relation (where \(\langle x, L^* w \rangle\) stands for \((L^* w)(x)):\n
\[ \langle x, L^* w \rangle := \langle Lx, w \rangle, \quad \forall x \in X, \ w \in W. \]

An important property of \(L\) and \(L^*\) is given by the following proposition:
Proposition 2.3 (Chap. II, Prop. 12 and its corollary). A linear mapping $L : X \to Z$ is weakly continuous if and only if $L^*(W) \subset Y$. If $L$ is weakly continuous, so is $L^*$.

This proposition also gives a convenient way to verify whether a linear mapping is weakly continuous or not. When $L$ is weakly continuous, with the weakly continuous mapping $L^* : W \to Y$, one can define the dual of the primal linear program.

Let $c \in Y$ and $b \in Z$. Consider an equality-constrained primal linear program (P) and its dual linear program (P$^*$) defined as follows:

$$(P) \quad \text{minimize} \quad \langle x, c \rangle \quad \text{subject to} \quad Lx = b, \quad x \in \Lambda.$$  \hspace{1cm} (2.2)

$$(P^*) \quad \text{maximize} \quad \langle b, w \rangle \quad \text{subject to} \quad -L^*w + c \in \Lambda^*, \quad w \in W.$$  \hspace{1cm} (2.3)

Similarities between these programs and standard finite-dimensional linear programs can be seen by writing the constraints $x \in \Lambda$ and $-L^*w + c \in \Lambda^*$ equivalently as $x \geq 0$ and $L^*w \leq c$, respectively.

If the program (P) or (P$^*$) has a feasible solution, it is said to be consistent; if it admits an optimal solution, it is said to be solvable. Let $\inf(P)$ and $\sup(P^*)$ denote the values of (P) and (P$^*$), respectively. The elementary duality theory (see [2, Chap. 3]) asserts that if (P) and (P$^*$) are both consistent, then

$$\sup (P^*) \leq \inf (P).$$

If equality holds, we say there is no duality gap.

There are several sufficient conditions for the absence of duality gap. For our purpose, however, one duality theorem (Theorem 2.4 below) will be the most important. It characterizes the relation between the value of (P$^*$) and the subvalue of (P), which is defined as follows.

Consider the set $H \subset Z \times \mathbb{R}$ defined by

$$H := \{(Lx, \langle x, c \rangle + r) \mid x \in \Lambda, \ r \geq 0\}.$$  \hspace{1cm} (2.4)

Let $\overline{H}$ denote the closure of $H$ in the weak topology $\sigma(\mathbb{R} \times \mathbb{R})$ (corresponding to the dual pair $(\mathbb{R} \times \mathbb{R}, \mathbb{R} \times \mathbb{R})$). We call (P) subconsistent iff there exists some $r \in \mathbb{R}$ with $(b, r) \in \overline{H}$. When (P) is subconsistent, the subvalue of (P) is defined by

$$\text{subvalue} (P) := \inf \{r \mid (b, r) \in \overline{H}\}.$$  

For comparison, note that $\inf(P) = \inf \{r \mid (b, r) \in H\}$. Note also that if $\rho$ is the subvalue of (P), then by the definition of the closure $\overline{H}$, there exists some net $\{x_i\}_{i \in I}$ with $x_i \in \Lambda$ for all $i$ and $Lx_i \to b$ and $\langle x_i, c \rangle \to \rho$, where $x_i$ need not be feasible for (P).

Theorem 2.4 (subconsistency and duality [2, Theorem 3.3]). (P) is subconsistent with a finite subvalue $\underline{\rho}$ if and only if $(P^*)$ is consistent with a finite value $\rho$.

We will apply this theorem in analyzing the duality relationship between the primal and dual linear programs for average-cost MDPs.

3 Linear Programming for Average-Cost MDPs

In this section we study the LP approach for the average-cost MDP under Assumption 2.1. Roughly speaking, the primal linear program (P) is formulated to find a stationary minimum pair among the stationary pairs of the average cost MDP—this is viable since under Assumption 2.1 the set
of stationary pairs is nonempty and a stationary minimum pair exists (cf. Theorem \(\text{(2.2)}\)). The dual linear program (\(P^*\)) is then determined by the primal program and the two dual pairs of vector spaces involved in the formulation (cf. Section \(\text{(2.2)}\)). We present the LP formulation and our main duality results in Sections \(\text{3.1} \) \text{and} \(\text{3.2} \) respectively, and we then give the proofs in Section \(\text{3.3} \).

Our formulation of the primal linear program is the same as that given by the prior work \(\text{(22, Chap. 12.3)}\), and it avoids a condition on the state transition stochastic kernel used in \(\text{(22, Chap. 12.3)}\), without affecting the desired duality result (cf. Remark \(\text{(3.3)}\)). This LP formulation we present is one instance of a general class of formulations discussed in the prior work \(\text{(20, Sec. 4)}\); however, for the sake of completeness, we will give a detailed account of it using the terminologies introduced in Section \(\text{2.2} \).

Regarding notations, in what follows, \(\mathbb{R}_+\) denotes the set of nonnegative numbers. For \(X = \mathbb{X} \) or \(\Gamma\), \(\mathcal{M}(X)\) denotes the space of finite (signed) Borel measures on \(X\), and \(\mathcal{P}(X)\) the set of real-valued Borel measurable functions on \(X\). We write \(\mathcal{M}^+(X)\) or \(\mathcal{P}^+(X)\) for the subset of those nonnegative elements in \(\mathcal{M}(X)\) or \(\mathcal{P}(X)\), and we will use similar notations for the subspaces of \(\mathcal{M}(X)\) or \(\mathcal{P}(X)\).

For the one-stage cost function \(c(\cdot)\), we will also need to work with its restriction to the set \(\Gamma\) of state and admissible action pairs (on which \(c(\cdot)\) is finite as we recall). For notational simplicity, we shall use the same notation \(c\) or \(c(\cdot)\) for the restriction of \(c(\cdot)\) to \(\Gamma\), and the context will make it clear which function is involved in the discussion.

Likewise, for a Borel measure \(\gamma\) on \(\Gamma\), sometimes we will also need to work with its extension to the whole state-action space \(\mathbb{X} \times \mathbb{A}\), which is simply a Borel measure concentrated on \(\Gamma\), and conversely, if \(\gamma\) is a Borel measure on \(\mathbb{X} \times \mathbb{A}\) concentrated on \(\Gamma\), sometimes we will need to consider its restriction to \(\Gamma\). In such cases, for notational simplicity, we will use the same notation \(\gamma\) for both measures.

### 3.1 Primal and Dual Linear Programs

To define a minimization problem on stationary pairs, let us first explain a well-known, many-to-one correspondence between a stationary pair \((\mu, p) \in \Delta_s\) and a Borel probability measure \(\gamma\) on \(\Gamma\) that satisfies

\[
\hat{\gamma}(B) = \int_{\mathbb{X}} q(B \mid x, a) \gamma(d(x, a)), \quad \forall B \in \mathcal{B}(\mathbb{X}),
\]

\[(3.1)\]

where \(\hat{\gamma}\) denotes the marginal of \(\gamma\) on \(\mathbb{X}\). The correspondence is essentially given by

\[
\gamma(d(x, a)) = \mu(da \mid x)p(dx),
\]

\[(3.2)\]

and has the property that

\[
J(\mu, p) = \int c d\gamma.
\]

\[(3.3)\]

Indeed, for \((\mu, p) \in \Delta_s\), as \(p\) is an invariant probability measure on \(\mathbb{X}\) induced by \(\mu\), we have

\[
p(B) = \int_{\mathbb{X}} \int_{\mathbb{A}} q(B \mid x, a) \mu(da \mid x)p(dx), \quad \forall B \in \mathcal{B}(\mathbb{X}).
\]

\[(3.4)\]

This is the same as \(\text{(3.1)}\) for the probability measure \(\gamma\) given by \(\text{(3.2)}\), since the marginal of \(\gamma\) is \(\hat{\gamma} = p\) and \(\mu\) obeys the control constraint of the MDP. The equality \(\text{(3.3)}\) follows from the definition of the average cost and the stationarity of the Markov chain under \(\mu\) when the initial distribution is \(p\). Conversely, given a probability measure \(\gamma\) satisfying \(\text{(3.1)}\), by \(\text{(2.3, Cor. 7.27.2)}\), we can decompose \(\gamma\) as in \(\text{(3.2)}\) with \(p = \hat{\gamma}\) and \(\mu(da \mid x)\) being a Borel measurable stochastic kernel on \(\mathbb{A}\) given \(\mathbb{X}\) that obeys the control constraint of the MDP. Then, since \(\gamma\) satisfies \(\text{(3.1)}\), the pair \((\mu, p)\) with \(p = \hat{\gamma}\) satisfies \(\text{(3.4)}\), which means that \(p\) is invariant for the Markov chain induced by \(\mu\) and hence \((\mu, p)\) is a stationary pair. The policy \(\mu\) here is in general not unique; however, by stationarity, every \((\mu, p)\) from this decomposition of \(\gamma\) has the same average cost \(\text{(3.3)}\).
Due to this correspondence between \((\mu, p)\) and \(\gamma\), finding a stationary minimum pair can be expressed as a minimization problem, \(\inf \int c \, d\gamma\), over the set of probability measures \(\gamma\) that satisfy (3.1) (in the reference [9], this set is called the set of “ergodic occupation measures”).

Before proceeding to write down the optimization problem for the primal program, we also need to restrict attention to those stationary pairs that have finite average costs, so that \(\infty\) does not appear in the objective or constraints of the primal program. The following definitions are introduced for this purpose. Consider a positive weight function \(w : \Gamma \to \mathbb{R}_+\),

\[
  w(x, a) := 1 + c(x, a), \quad (x, a) \in \Gamma.
\]

Let \(\mathcal{M}_w(\Gamma)\) be the set of finite, signed Borel measures on \(\Gamma\) w.r.t. which the function \(w\) is integrable:

\[
  \mathcal{M}_w(\Gamma) := \{ \gamma \in \mathcal{M}(\Gamma) \mid \int w \, d|\gamma| < \infty \}.
\]

Let \(\mathcal{F}_w(\Gamma)\) be the set of Borel measurable functions \(\phi\) on \(\Gamma\) such that

\[
  |\phi| \leq kw \quad \text{for some } k > 0.
\]

Then every \(\phi \in \mathcal{F}_w(\Gamma)\) is integrable w.r.t. all \(\gamma \in \mathcal{M}_w(\Gamma)\). By (3.3) and the definition of \(w(\cdot)\), if a stationary pair \((\mu, p)\) has finite average cost, then the corresponding probability measure \(\gamma \in \mathcal{M}_w(\Gamma)\).

We are now ready to define the linear programs for the average-cost MDP. Let us specialize the programs \((P)\) and \((P^*)\) given in Section 2.2 by identifying the objects involved in these programs with the spaces of measures or functions and constraints involved in the MDP:

- The dual pair \((X, Y) = (\mathcal{M}_w(\Gamma), \mathcal{F}_w(\Gamma))\), with the bilinear form

  \[
  \langle \gamma, \phi \rangle := \int_{\Gamma} \phi \, d\gamma, \quad \gamma \in \mathcal{M}_w(\Gamma), \phi \in \mathcal{F}_w(\Gamma).
  \]

- The dual pair \((Z, W) = (\mathbb{R} \times \mathcal{M}(\mathcal{X}), \mathbb{R} \times \mathcal{F}_b(\mathcal{X}))\), where \(\mathcal{M}(\mathcal{X})\) is the set of finite, signed Borel measures on \(\mathcal{X}\) as defined earlier, \(\mathcal{F}_b(\mathcal{X})\) is the set of bounded Borel measurable functions on \(\mathcal{X}\), and the bilinear form on \((\mathbb{R} \times \mathcal{M}(\mathcal{X})) \times (\mathbb{R} \times \mathcal{F}_b(\mathcal{X}))\) is defined as

  \[
  \langle (r, \xi), (\rho, h) \rangle := r\rho + \int_{\mathcal{X}} h \, d\xi, \quad (r, \xi) \in \mathbb{R} \times \mathcal{M}(\mathcal{X}), (\rho, h) \in \mathbb{R} \times \mathcal{F}_b(\mathcal{X}).
  \]

- The convex cone \(\Lambda = \mathcal{M}_w^+(\Gamma)\), the subset of nonnegative measures in \(\mathcal{M}_w(\Gamma)\). The dual cone of \(\Lambda\) is \(\Lambda^* = \mathcal{F}_w^+(\Gamma)\), the subset of nonnegative functions in \(\mathcal{F}_w(\Gamma)\).

- The objective function of the primal program \((P)\) is \(\langle \gamma, c \rangle\), and the feasible set of \((P)\) is defined by the following constraints:

  \[
  \gamma \in \mathcal{M}_w^+(\Gamma), \quad \gamma(\Gamma) = 1, \quad \hat{\gamma}(B) = \int_{\Gamma} q(B \mid x, a) \gamma(d(x, a)) \quad \forall B \in \mathcal{B}(\mathcal{X}),
  \]

where \(\hat{\gamma}\) is the marginal of \(\gamma\) on \(\mathcal{X}\). Thus, the feasible solutions of \((P)\) correspond to those stationary pairs with finite average costs, and the objective is to minimize the average cost over them. In the form of \((P)\) discussed in Section 2.2, the two equality constraints in (3.5) can be written as

\[
  L\gamma = b := (1, 0),
\]

where \(\theta\) is the trivial measure on \(\mathcal{X}\) (i.e., \(\theta(B) \equiv 0\) for all \(B \in \mathcal{B}(\mathcal{X})\)), and the linear mapping \(L\) is defined as \(L : \mathcal{M}_w(\Gamma) \to \mathbb{R} \times \mathcal{M}(\mathcal{X})\) with \(L = (L_0, L_1)\) where, for \(\gamma \in \mathcal{M}_w(\Gamma)\),

\[
  L_0\gamma := \gamma(\Gamma),
\]

\[
  (L_1\gamma)(B) := \hat{\gamma}(B) - \int_{\Gamma} q(B \mid x, a) \gamma(d(x, a)) \quad \forall B \in \mathcal{B}(\mathcal{X}).
\]
• The adjoint of $L$ is the linear mapping $L^* : \mathbb{R} \times \mathbb{F}_b(X) \to \mathbb{F}_w(\Gamma)$ that maps each $(\rho, h) \in \mathbb{R} \times \mathbb{F}_b(X)$ to the function

$$L^*(\rho, h)(x, a) := \rho + h(x) - \int_X h(y) q(dy | x, a), \quad (x, a) \in \Gamma. \quad (3.8)$$

Since $L^*(\mathbb{R} \times \mathbb{F}_b(X)) \subset \mathbb{F}_w(\Gamma)$, both $L$ and $L^*$ are weakly continuous (36, Chap. II, Prop. 12 and its corollary); see also Prop. 2.3. The inequality constraint in the program (P*) is

$$-L^*(\rho, h) + c \in \mathbb{F}_w^+(\Gamma).$$

We can write this constraint as $L^*(\rho, h) \leq c$ or more explicitly, as

$$\rho + h(x) - \int_X h(y) q(dy | x, a) \leq c(x, a), \quad \forall (x, a) \in \Gamma. \quad (3.9)$$

The objective function of the dual program (P*) is $\langle b, (\rho, h) \rangle = \langle (1, 0), (\rho, h) \rangle = \rho$.

Expressed in the form introduced in Section 2.2, the primal and dual linear programs for the average-cost MDP are:

(P) \begin{align*}
{\text{minimize}} & \quad \langle \gamma, c \rangle \\
{\text{subject to}} & \quad L\gamma = (1, 0), \quad \gamma \in \mathbb{M}_w^+(\Gamma). \quad (3.10)
\end{align*}

(P*) \begin{align*}
{\text{maximize}} & \quad \rho \\
{\text{subject to}} & \quad L^*(\rho, h) \leq c, \quad \rho \in \mathbb{R}, \ h \in \mathbb{F}_b(X). \quad (3.11)
\end{align*}

A few properties of these programs are easy to see. From the discussion at the beginning of this subsection about the relation between stationary pairs and feasible solutions $\gamma$ of the primal program (P), it is clear that under Assumption 2.1 the existence of a stationary minimum pair (Theorem 2.2) ensures that (P) is both consistent and solvable. The consistency of the dual program (P*) is trivial: since $c \geq 0$, $\rho = 0$ and $h(\cdot) \equiv 0$ give a feasible solution. We then have $0 \leq \sup(P^*) \leq \inf(P) = \rho^*$ under Assumption 2.1.

In the next subsection, we will address the duality between (P) and (P*). There we will also examine a connection between (P*) and the ACOE (average cost optimality equation) for the MDP, through a maximizing sequence of (P*). Such a sequence is defined as a sequence $\{ (\rho_n, h_n) \}$ of feasible solutions of (P*) with the property that $\rho_n \uparrow \sup(P^*)$.

3.2 Optimality Results and Discussion

Our main result of this section is the following duality theorem. It can be compared with the prior result of [22, Chap. 12.3, Theorem 12.3.4] for average-cost semicontinuous MDPs.

**Theorem 3.1** (consistency and absence of duality gap). Under Assumption 2.1, the linear programs (P) and (P*) in (3.10)-(3.11) satisfy the following:

(i) (P) is consistent and solvable, and (P*) is consistent.

(ii) There is no duality gap, and the value of (P) and (P*) is $\rho^*$.

**Remark 3.1** (about the proof of Theorem 3.1). Besides the differences in assumptions, one difference between our proof of the absence of duality gap and the proof given in the prior work [22, Chap. 12.3C] is the following. The approach of the latter proof is to show that the set $H$ in (2.4) is weakly closed (i.e., $\overline{H} = H$), which is a sufficient condition for the absence of duality gap, and which requires one to show that every point of $\overline{H}$ is in $H$. Our proof uses the duality between the
subvalue ρ of (P) and the value of (P∗) [2, Theorem 3.3] (cf. Theorem 2.4), with which it suffices to show that a single point of Π, namely, ((1, 0), ρ), is in H. Thus the proof is simpler in this respect.

We can also prove that H is weakly closed under our assumptions. This requires some minor changes in the proof arguments used in [10], which we will also use to prove Theorem 3.1 in particular, we only need to change slightly the finite measures used when applying Lusin’s theorem. Nonetheless, it will take some space to explain the details of those changes, and this is another reason that we choose to use the duality theorem [2, Theorem 3.3] instead in our proof. □

Remark 3.2 (comparison with a duality result in [38, Theorem 3]). Yamada proved an absence of duality gap result [38] for compact Euclidean state and action spaces, under continuity conditions on the MDP model different from the lower semicontinuous model assumption we mentioned. His continuity conditions can be related to our model assumptions, so let us explain in more detail how our assumptions and duality result compare with his:

Among others, Yamada assumed that c(x, a) is continuous in a for each fixed x, and q(dy | x, a) has a density p(y | x, a) w.r.t. the Lebesgue measure, where p(y | x, a) is continuous in (y, a) for each fixed x [38, (A2) and (A3)]. In our case, since the action space has the discrete topology, trivially, c(x, a) and q(dy | x, a) are continuous in a for each fixed x, so there are similarities to Yamada’s conditions. Our majorization condition (M) is, however, entirely different from Yamada’s geometric ergodicity condition [38, (A1) and (A4)], in which he required the density function p(y | x, a) to be bounded away from zero uniformly for all (x, a) ∈ Γ. Using this condition together with the continuity and other assumptions, he proved the absence of duality gap result [38, Theorem 3]. Both his conditions and his proof arguments are very different from ours. □

Remark 3.3 (about the formulation of (P∗) and its solvability). In defining (P∗), we have chosen the space M_b(Γ) of bounded Borel measurable functions to form the dual pair with the space M(Γ) of finite Borel measures. With this choice, (P∗) is in general not solvable (i.e., an optimal solution may not exist), since the inequality

$$\rho^* + h(x) \leq c(x, a) + \int_{\mathcal{X}} h(y) q(dy | x, a), \quad \forall (x, a) \in \Gamma,$$

need not admit a bounded solution h.

As mentioned earlier, our LP formulation is only an instance of the class of formulations discussed in [20, Sec. 4]. A different dual program (P∗) is studied in [22, Chap. 12.3]. It involves, instead of (M(Γ), F_b(Γ)), the dual pair (M_{w_0}(Γ), F_{w_0}(Γ)), where the two spaces are defined similarly to M_w(Γ) and F_w(Γ), respectively: with w_0(x) := 1 + \inf_{a \in A(x)} c(x, a), x ∈ X,

$$M_{w_0}(\mathcal{X}) := \left\{ p \in M(\mathcal{X}) \mid \int_{\mathcal{X}} w \, d|p| < \infty \right\}, \quad F_{w_0}(\mathcal{X}) := \left\{ h \in F(\mathcal{X}) \mid |h| < k w_0 \text{ for some } k > 0 \right\}.$$

This choice leaves more room for (P∗) to admit an optimal solution. However, a disadvantage is that to ensure the weak continuity of the linear mapping L, an additional condition on the state transition stochastic kernel is required (cf. [22, Chap. 12.3A, Assumption 12.3.1]): for some constant k > 0,

$$\int_{\mathcal{X}} \inf_{a' \in A(y)} c(y, a') q(dy | x, a) \leq k (1 + c(x, a)), \quad \forall (x, a) \in \Gamma. \quad (3.12)$$

Yet, since the costs are strictly unbounded, this condition (3.12) is neither needed for the existence of a minimum pair, nor needed for the absence of duality gap between (P) and (P∗).

Also, the use of the dual pair (M_{w_0}(Γ), F_{w_0}(Γ)) alone cannot guarantee that (P∗) has an optimal solution, for which one would still need to make additional assumptions about the functions h_n in a maximizing sequence \{ (\rho_n, h_n) \} for (P∗) (cf. [22, Chap. 12.4B, Theorem 12.4.2]). This makes it less appealing to us to have the dual pair (M_{w_0}(Γ), F_{w_0}(Γ)) with its extra condition [38, 24] in the LP formulation.
For these reasons, we have formulated (P*) differently. Accordingly, we treat the result on ACOE given in the next proposition not as the property of a dual optimal solution, which may not exist, but as a potential consequence of the results from the LP approach.

As just noted, the dual program (P*) in our formulation need not admit an optimal solution. However, because there is no duality gap, one can still obtain a version of ACOE for the MDP from a maximizing sequence \( \{(\rho_n, h_n)\} \) of (P*), under certain conditions on \( h_n \). The arguments are essentially the same as those for [22, Chap. 12.4B, Theorem 12.4.2(c)] (although that theorem requires the functions \( h_n \) to be uniformly bounded by some multiple of the function \( w_0 \) mentioned in Remark 3.3). We include the result in the proposition below. Here we consider nonnegative or nonpositive functions \( h_n \) (such maximizing sequences exist since adding a constant to \( h_n \) does not affect the feasibility and the value of the objective function of the solution \( (\rho_n, h_n) \)). The limiting function \( h^* = \limsup_{n \to \infty} h_n \) is then bounded below or above, respectively.

One can also consider more general cases of \( h_n \) and require \( h^* \) to be in \( \mathcal{M}_{w_0}(\mathcal{X}) \) for some weight function \( w_0 \), similarly to [22, Chap. 12.4B, Theorem 12.4.2(c)].

**Proposition 3.2** (ACOE for \( p^* \)-almost all states). Suppose Assumption [22] hold. Let \( (\mu^*, p^*) \) be a stationary minimum pair, and let \( \{(\rho_n, h_n)\} \) be a maximizing sequence of the dual program (P*). Then, with \( h^* = \limsup_{n \to \infty} h_n \), in either of the two cases:

(i) \( h_n \geq 0 \), \( \int h^* \, dp^* < +\infty \), and \( \int \sup_n h_n(y) \, q(dy \mid x, a) < +\infty \) for all \( (x, a) \in \Gamma \);

(ii) \( h_n \leq 0 \) and \( \int h^* \, dp^* > -\infty \),

the function \( h^* \) is finite \( p^* \)-almost everywhere and satisfies that

\[
\rho^* + h^*(x) \leq c(x, a) + \int_{\mathcal{X}} h^*(y) \, q(dy \mid x, a), \quad \forall (x, a) \in \Gamma,
\]

and that for \( p^* \)-almost all \( x \in \mathcal{X} \),

\[
\rho^* + h^*(x) = \inf_{a \in \mathcal{A}(x)} \left\{ c(x, a) + \int_{\mathcal{X}} h^*(y) \, q(dy \mid x, a) \right\} = \int_{\mathcal{A}(x)} \left\{ c(x, a) + \int_{\mathcal{X}} h^*(y) \, q(dy \mid x, a) \right\} \mu^*(da \mid x).
\]

**Remark 3.4** (about nonrandomized stationary optimal policies). From (3.14)-(3.15) and the fact that \((\mu^*, p^*)\) is a stationary pair, one can deduce that there exists a subset \( \hat{\mathcal{X}} \subset \mathcal{X} \) with \( p^*(\hat{\mathcal{X}}) = 1 \) and a nonrandomized, Borel measurable stationary policy, i.e., a Borel measurable function \( f : \hat{\mathcal{X}} \to \mathcal{A} \) with \( f(x) \in \mathcal{A}(x) \) for all \( x \in \hat{\mathcal{X}} \), such that (i) on \( \hat{\mathcal{X}} \), \( f \) attains the minimum in the ACOE (3.14):

\[
\rho^* + h^*(x) = c(x, f(x)) + \int_{\mathcal{X}} h^*(y) \, q(dy \mid x, f(x)), \quad \forall x \in \hat{\mathcal{X}},
\]

and \( \mu^* \) the set \( \hat{\mathcal{X}} \) is absorbing under \( f \), namely, \( q(\hat{\mathcal{X}} \mid x, f(x)) = 1 \) for all \( x \in \hat{\mathcal{X}} \).

More specifically, let \( \mathcal{X}' \) be the set on which (3.14) and (3.15) hold. First, one can construct a set \( \check{\mathcal{X}} \subset \mathcal{X}' \) with \( p^*(\check{\mathcal{X}}) = 1 \) that is absorbing under the policy \( \mu^* \), by applying the same proof of [33, Prop. 4.2.3(ii)] to the stationary Markov chain on \( \mathcal{X} \) induced by \( \mu^* \) with the initial distribution \( p^* \). One can then use (3.14)-(3.15) to construct the desired nonrandomized Borel measurable policy \( f \) on \( \check{\mathcal{X}} \), either directly by using the fact that the action space is countable in our case, or by using the Blackwell and Ryll-Nardzewski theorem [3, Theorem 2] as discussed in [20, Remark 4.6].

This gives the desired nonrandomized policy \( f \) and absorbing set \( \check{\mathcal{X}} \). Then, consider the states in \( \check{\mathcal{X}} \). From (3.16), under further conditions such as \( \lim_{n \to \infty} n^{-1} \mathbb{E}_x [h^*(x_n)] \geq 0 \) for all \( x \in \check{\mathcal{X}} \), one can use standard arguments to obtain that the policy \( f \) is average-cost optimal for all initial states \( x \in \check{\mathcal{X}} \) (see e.g., the discussion in [20, Sec. 3] on canonical triplets).
3.3 Proofs

Let us first recall a few definitions and facts about probability measures on a metrizable space $X$. Let $\mathcal{C}_b(X)$ denote the set of real-valued, bounded continuous functions on $X$. By definition, a sequence of probability measures $p_n \in \mathcal{P}(X)$ converges weakly to some $p \in \mathcal{P}(X)$, denote $p_n \to p$, iff $\int f dp_n \to \int f dp$ for all $f \in \mathcal{C}_b(X)$. If $E$ is a family of probability measures in $\mathcal{P}(X)$ such that for any $\epsilon > 0$, there is a compact set $K \subset X$ with $p(K) > 1 - \epsilon$ for all $p \in E$, we say that $E$ is tight.

By Prohorov’s theorem [4, Theorem 6.1], any sequence in a tight family $E$ has a further subsequence that converges weakly to a probability measure in $\mathcal{P}(X)$. We will use this fact many times in the present section as well as in Section 4 for some family $E \subset \mathcal{P}(\Gamma)$ that satisfies $\sup_{r \in E} (\gamma, c) < \infty$.

By the strict unboundedness condition on $c$ given in Assumption 2.1 (SU), such a family $E$ must be tight (as can be seen easily from the condition (SU) and the definition of tightness).

3.3.1 Proof of Theorem 3.1

The consistency of (P) and (P*) and the solvability of (P) were already discussed in Section 3.1 where we also showed that under Assumption 2.1 0 $\leq$ $\sup(P^*) \leq \inf(P) = \rho^*$.

We now prove that there is no duality gap between (P) and (P*)). Our approach is to use [2, Theorem 3.3] (cf. Theorem 2.4) in Section 2.2, which asserts the equality between the subvalue of (P) and the value of (P*) when they are finite. Specifically, recall from Section 2.2 that the subvalue $(\rho = (\rho_i)_{i \in \mathcal{I}})$ for (P) is defined as

$$\rho := \inf \{ r \mid ((1,0), r) \in \overline{H} \}$$

where the set $H \subset \mathbb{R} \times \mathcal{M}(\mathcal{X}) \times \mathbb{R}$ is given by

$$H := \{(L_{\gamma_i}, (\gamma_i, c) + r) \mid \gamma_i \in \mathcal{M}_b^+(\Gamma), \ r \geq 0\}, \tag{3.17}$$

and $\overline{H}$ is the closure of $H$ in the weak topology $\sigma(\mathbb{R} \times \mathcal{M}(\mathcal{X}) \times \mathbb{R}, \mathbb{R} \times \mathcal{M}(\mathcal{X}) \times \mathbb{R})$. Since (P) and (P*) are consistent, $\sup(P^*)$ is finite and equals the subvalue $\rho$ by [2, Theorem 3.3] (cf. Theorem 2.4).

So, to show $\inf(P) = \sup(P^*)$, we need to prove $\rho^* = \rho$. In what follows, we will prove that

$$(\rho = (\rho_i)_{i \in \mathcal{I}}) \in H,$$

by constructing a stationary pair whose average cost is no greater than $\rho$. This will give us $\rho^* = \rho$ (since it implies $\rho \geq \rho^*$, whereas $\rho^* \geq \rho$). The proof will proceed in four steps, with the first three steps making preparations for the last one.

**Step (i):** From the definition of $\rho$, it follows that $((1,0), \rho) \in \overline{H}$ and moreover, there exist a direct set $\mathcal{I}$ and a net $\{\gamma_i\}_{i \in \mathcal{I}} \in \mathcal{M}_b^+(\Gamma)$ with

$$(L_{\gamma_i}, (\gamma_i, c)) \to ((1,0), \rho)$$

in the $\sigma(\mathbb{R} \times \mathcal{M}(\mathcal{X}) \times \mathbb{R}, \mathbb{R} \times \mathcal{M}(\mathcal{X}) \times \mathbb{R})$ topology. This means that

$$\gamma_i(\Gamma) \to 1, \tag{3.18}$$

$$\int \chi h(x) \gamma_i(dx) - \int \chi \int \chi h(y) q(dy \mid x, a) \gamma_i(dx, a) \to 0, \quad \forall h \in \mathcal{M}_b(\mathcal{X}), \tag{3.19}$$

$$\gamma_i \to \rho. \tag{3.20}$$

In view of (3.18), there exists $\bar{i} \in \mathcal{I}$ such that for all $i \geq \bar{i}$, $\gamma_i(\Gamma) > 0$. Then, since all $\gamma_i$ are nonnegative measures and $\gamma_i(\Gamma) \to 1$, by restricting attention to $\gamma_i, i \geq \bar{i}$, and considering the normalized measures $\gamma_i(\cdot)/\gamma_i(\Gamma)$ instead of $\gamma_i$, we can redefine the net $\{\gamma_i\}_{i \in \mathcal{I}}$ in the above so that every $\gamma_i$ is a probability measure on $\mathcal{B}(\Gamma)$:

$$\gamma_i(\Gamma) = 1, \quad \forall i \in \mathcal{I}. \tag{3.19}$$
**Step (ii):** Next, from the net \(\{\gamma_n\}_{n \in \mathbb{Z}}\), we will extract a sequence of probability measures with the property that the convergence in (3.19) holds for a countable subset of the functions in \(\mathcal{F}_b(\mathcal{X})\). We start by defining this subset. It consists of two countable families of functions, \(\hat{\mathcal{C}}_b(\mathcal{X})\) and \(\mathcal{F}_b(\mathcal{X})\). The set \(\hat{\mathcal{C}}_b(\mathcal{X})\) involves continuous bounded functions that will be used to determine if two probability measures on \(\mathcal{X}\) are equal. The set \(\mathcal{F}_b(\mathcal{X})\) involves indicator functions of certain sets in \(\mathcal{X}\) that will be important in the subsequent proof to handle the discontinuities in the MDP model by using Lusin's theorem. Their precise definitions are as follows.

Let \(\mathcal{C}_b(\mathcal{X})\) denote the set of (real-valued) bounded continuous functions on \(\mathcal{X}\). Since \(\mathcal{X}\) is metrizable, by [34, Theorem 6.6], there exists a countable set

\[
\hat{\mathcal{C}}_b(\mathcal{X}) := \{h_1, h_2, \ldots\} \subset \mathcal{C}_b(\mathcal{X})
\]

such that in \(\mathcal{P}(\mathcal{X})\), a sequence of probability measures \(p_n \xrightarrow{\text{w}} p \in \mathcal{P}(\mathcal{X})\) if and only if

\[
\int h \, dp_n \to \int h \, dp, \quad \forall h \in \hat{\mathcal{C}}_b(\mathcal{X}).
\]

Then by [13, Prop. 11.3.2], for any \(p, p' \in \mathcal{P}(\mathcal{X})\),

\[
p = p' \iff \int h \, dp = \int h \, dp', \quad \forall h \in \hat{\mathcal{C}}_b(\mathcal{X}).
\]

The countable set \(\hat{\mathcal{C}}_b(\mathcal{X})\) is the first family of functions we will need.

We now define the other countable family \(\mathcal{F}_b(\mathcal{X})\) of indicator functions mentioned earlier. The definition of this set involves some new notations and Lusin's theorem.

Let \(\mathbb{Z}_+\) denote the set of all positive integers. For \(m \in \mathbb{Z}_+\), define the truncated one-stage cost function \(c^m(\cdot) := \min\{c(\cdot), m\}\) on \(\mathcal{X} \times \mathcal{A}\) (later, a technical argument in Step (iv) of our proof will involve these \(c^m\) functions). For each \(j \in \mathbb{Z}_+\), corresponding to the compact set \(\Gamma_j\) in Assumption 2.1, let \((O_j, D_j, \nu_j)\) be the open set, the closed set, and the finite measure, respectively, in Assumption 2.1(M) for \(K = \text{proj}_j(\Gamma_j)\). Let \(F_j := \text{proj}_j(\Gamma_j)\), the projection of \(\Gamma_j\) on \(\mathcal{A}\). Then the set \(F_j\) is compact, and since \(\mathcal{A}\) is countable and discrete, this means that the set \(F_j\) is finite.

**Lemma 3.3.** For each \(j, m \in \mathbb{Z}_+\) and \(\ell \in \mathbb{Z}_+\), there exist closed subsets \(B_{j,m,\ell}^1\) and \(B_{j,\ell}^2\) of \(\mathcal{X}\) such that the following hold:

(i) \(\nu_j(\mathcal{X} \setminus B_{j,m,\ell}^1) \leq \ell^{-1}\) and \(\nu_j(\mathcal{X} \setminus B_{j,\ell}^2) \leq \ell^{-1}\);

(ii) restricted to the set \(B_{j,m,\ell}^1 \times F_j\), the function \(c^m(\cdot)\) is continuous, and restricted to the set \(B_{j,\ell}^2 \times F_j\), the state transition stochastic kernel \(q(dy \mid \cdot, \cdot)\) is continuous.

**Proof.** This lemma is a consequence of Lusin's theorem (see [13, Theorem 7.5.2]), which asserts that if \(f\) is a Borel measurable function from a topological space \(\mathcal{X}\) into a separable metric space \(S\) and \(\nu\) is a finite, closed regular Borel measure on \(\mathcal{X}\), then for any \(\delta > 0\), there is a closed set \(B\) such that \(\nu(\mathcal{X} \setminus B) < \delta\) and the restriction of \(f\) to \(B\) is continuous.

We apply this theorem with \(X = \mathcal{X}\) and \(\nu = \nu_j\) for each \(j\) in the lemma. Since \(\mathcal{X}\) is a metrizable topological space, every finite Borel measure is closed regular by [13, Theorem 7.1.3], and therefore, the finite measure \(\nu_j\) in the lemma meets the condition in Lusin’s theorem.

For each \(j, m, \ell \in \mathbb{Z}_+\), to find the desired closed set \(B_{j,m,\ell}^1\), we apply Lusin's theorem with \(X = \mathcal{X}, S = \mathbb{R}, \nu = \nu_j\) and \(\delta = \ell^{-1}/|F_j|\), and with the function \(f(\cdot) = c^m(\cdot, a)\) for each action \(a \in F_j\). This gives us, for each \(a \in F_j\), a closed set \(E_a\) such that \(\nu_j(\mathcal{X} \setminus E_a) < \delta\) and restricted to \(E_a\), \(c^m(\cdot, a)\) is continuous. Then the closed set \(B_{j,m,\ell}^1 := \bigcap_{a \in F_j} E_a\) has the desired property that \(\nu_j(\mathcal{X} \setminus B_{j,m,\ell}^1) \leq \ell^{-1}\) and restricted to \(B_{j,m,\ell}^1 \times F_j\), \(c^m(\cdot, \cdot)\) is continuous.

For each \(j, \ell \in \mathbb{Z}_+\), the desired closed set \(B_{j,\ell}^2\) is constructed similarly, by applying Lusin’s theorem to the state transition stochastic kernel \(q(dy \mid x, a)\), which is a \(\mathcal{P}(\mathcal{X})\)-valued Borel measurable function on \(\mathcal{X} \times \mathcal{A}\). Specifically, let \(X = \mathcal{X}, S = \mathcal{P}(\mathcal{X}), \nu = \nu_j\), and \(\delta = \ell^{-1}/|F_j|\). (Since \(\mathcal{X}\) is separable and metrizable, by [3, Prop. 7.20], \(\mathcal{P}(\mathcal{X})\) is also a separable metrizable space and hence meets the
condition for the space $S$ in Lusin’s theorem.) We apply Lusin’s theorem to $f(\cdot) = q(dy | \cdot, a)$ for each $a \in F_j$ to obtain a closed set $E_a$ such that $\nu_j(X \setminus E_a) < \delta$ and restricted to $E_a$, $q(dy | \cdot, a)$ is continuous. We then let the desired set $B^2_{j,\ell} = \cap_{a \in F_j} E_a$. \hfill \Box

We group $(O_j, D_j, \nu_j, B^1_{j,m,\ell})$, $(O_j, D_j, \nu_j, B^2_{j,\ell})$ in the preceding proof into two countable collections $W_1$ and $W_2$:

$$W_1 := \{(O_j, D_j, \nu_j, B^1_{j,m,\ell}) \mid j, m, \ell \in \mathbb{Z}_+\}, \quad W_2 := \{(O_j, D_j, \nu_j, B^2_{j,\ell}) \mid j, \ell \in \mathbb{Z}_+\}.$$ Let $\mathbb{1}_E$ denote the indicator function for a set $E$. Finally, define a countable set $\hat{\mathbb{1}}_b(X)$ of indicator functions on $X$ by

$$\hat{\mathbb{1}}_b(X) := \{\mathbb{1}_E(\cdot) \mid E = (O \setminus D) \cap B^c \text{ for some } (O, D, \nu, B) \in W_1 \cup W_2\}. \quad (3.22)$$

Note that the sets $E$ in $(3.22)$ are open sets (since $O$ is open and $D, B$ are closed); this fact will be useful later.

We now extract a desirable sequence from the net $\{\gamma_i\}_{i \in \mathcal{I}}$:

**Lemma 3.4.** There exists a sequence $\{\gamma_n\}_{n \geq 0} \subset \{\gamma_i\}_{i \in \mathcal{I}}$ such that

$$\int_X h(x) \gamma_n(dx) - \int_X \int_X h(y) q(dy \mid x, a) \gamma_n(d(x, a)) \to 0, \quad \forall h \in \hat{\mathbb{C}}_b(X) \cup \hat{\mathbb{1}}_b(X), \quad (3.23)$$

$$\langle \gamma_n, c \rangle \to \bar{\rho}, \quad (3.24)$$

**Proof.** Let us order the functions in the countable set $\hat{\mathbb{C}}_b(X) \cup \hat{\mathbb{1}}_b(X)$ as $h_1, h_2, \ldots$. Choose any $\tilde{i}_0 \in \mathcal{I}$ and let $\gamma_n = \gamma_{i_0}$ for $n = 0$. For each $n \geq 1$, by $(3.19)-(3.20)$, there exists $i_n \in \mathcal{I}, \tilde{i}_n \geq \tilde{i}_{n-1}$ such that for all $i \geq \tilde{i}_n,$

$$\left| \int_X h(x) \gamma_i(dx) - \int_X \int_X h(y) q(dy \mid x, a) \gamma_i(d(x, a)) \right| \leq n^{-1}, \quad \forall h \in \{h_1, h_2, \ldots, h_n\},$$

$$\bar{\rho} - n^{-1} \leq \langle \gamma_i, c \rangle \leq \bar{\rho} + n^{-1}.$$ Let $\gamma_n = \gamma_{i_n}$. The resulting sequence $\{\gamma_n\}_{n \geq 0}$ satisfies $(3.23)-(3.24)$. \hfill \Box

**Step (iii):** Henceforth, we work with the sequence $\{\gamma_n\}$ of probability measures given by Lemma 3.5. The relation $(3.24)$ together with Assumption 2.1 (SU) implies that $\{\gamma_n\}$ is a tight family of probability measures on $\mathcal{B}(\Gamma)$. So by Prohorov’s theorem [4, Theorem 6.1], it has a subsequence that converges weakly to some probability measure $\bar{\gamma}$ on $\mathcal{B}(\Gamma)$. To simplify notation, let us use the same notation $\{\gamma_n\}$ to denote the convergent subsequence. Thus $\gamma_n \overset{w}{\to} \bar{\gamma}$.

By [3, Cor. 7.27.2], the probability measure $\bar{\gamma}$ can be decomposed into its marginal $\bar{\mu}$ on $X$ and a stochastic kernel $\bar{\mu}$ on $\mathcal{X}$ given $X$ that obeys the control constraint of the MDP; i.e.,

$$\bar{\gamma}(d(x, a)) = \bar{\mu}(da \mid x) \bar{\mu}(dx).$$

This gives us a stationary policy $\bar{\mu}$. Before we investigate the property of the pair $(\bar{\mu}, \bar{\rho})$ in the next step, we need the following majorization property:

**Lemma 3.5.** For every $(O, D, \nu, B) \in W_1 \cup W_2$,

$$\limsup_{n \to \infty} \gamma_n((O \setminus D) \cap B^c) \leq \nu(B^c), \quad \bar{\rho}((O \setminus D) \cap B^c) \leq \nu(B^c).$$
Proof. For \((O, D, \nu, B) \in W_1 \cup W_2\), let \(E = (O \setminus D) \cap B^c\) and since the indicator function \(\mathbb{1}_E \in \hat{F}_h(\mathcal{X})\), we have, by \((3.23)\) in Lemma \(3.4\) that
\[
\epsilon_n := \left| \hat{\gamma}_n(E) - \int_{\mathcal{X}} q(E \mid x, a) \gamma_n(dx, a) \right| \to 0.
\]
We also have, by Assumption \((2.1) M)\),
\[
\int_{\mathcal{X}} q(E \mid x, a) \gamma_n(dx, a) \leq \int_{\mathcal{X}} \nu(B^c) \gamma_n(dx, a) = \nu(B^c).
\]
Hence \(\hat{\gamma}_n(E) \leq \nu(B^c) + \epsilon_n\) for all \(n \geq 0\); consequently, \(\limsup_{n \to \infty} \hat{\gamma}_n(E) \leq \nu(B^c)\).

Now \(\hat{\gamma}_n \to \bar{p}\) (since \(\gamma_n \to \bar{\gamma}\)) and \(E\) is an open set (since \(O\) is open and \(D, B\) are closed). Therefore, by \([13\) Theorem 11.1.1] and the first part of the proof, \(\bar{p}(E) \leq \liminf_{n \to \infty} \hat{\gamma}_n(E) \leq \nu(B^c)\).

Step (iv): We are now ready to prove that \(((1, \emptyset), \bar{\rho}) \in H\).

Lemma 3.6. The pair \((\bar{\mu}, \bar{p})\) is a stationary pair with \(J(\bar{\mu}, \bar{p}) = \langle \bar{\gamma}, c \rangle \leq \bar{\rho}\).

Proof outline. We will only outline the proof, because the arguments for this lemma are essentially the same as those given in our prior work on the existence and optimality properties of stationary pairs \([40\) Sec. 4.1 and Sec. 4.3.1\]. By Lemma \(3.4\) it suffices to prove the inequality
\[
\langle \bar{\gamma}, c \rangle \leq \lim_{n \to \infty} \langle \gamma_n, c \rangle = \bar{\rho}
\]
and to prove that for all \(h \in \hat{C}_b(\mathcal{X}),\)
\[
\lim_{n \to \infty} \int_{\mathcal{X}} h(y) q(dy \mid x, a) \gamma_n(dx, a) = \int_{\mathcal{X}} h(y) q(dy \mid x, a) \bar{\gamma}(dx, a). \tag{3.26}
\]
To see the sufficiency of \((3.26)\) and \((3.20)\), note that \((3.20)\), together with \((3.23)\) in Lemma \(3.4\) and the fact \(\lim_{n \to \infty} \int_{\mathcal{X}} h d\gamma_n = \int_{\mathcal{X}} h d\bar{p}\) for all \(h \in \hat{C}_b(\mathcal{X})\) (since \(\gamma_n \to \bar{p}\)), will imply that
\[
\int_{\mathcal{X}} h(y) q(dy \mid x, a) \bar{\gamma}(dx, a) = \int_{\mathcal{X}} h \bar{p}(dx), \quad \forall h \in \hat{C}_b(\mathcal{X}).
\]
In turn, this will imply that \(\bar{p}\) is identical to the probability measure \(\int_{\mathcal{X}} q(\cdot \mid x, a) \bar{\gamma}(dx, a)\) (cf. \((3.24)\)), thus proving that \(\bar{p}\) is an invariant probability measure for the Markov chain induced by the policy \(\bar{\mu}\) and hence \((\bar{\mu}, \bar{p})\) is a stationary pair. Then the first relation \((3.25)\) will give us the desired inequality \(J(\bar{\mu}, \bar{p}) = \langle \bar{\gamma}, c \rangle \leq \bar{\rho}\).

Proving \((3.25)\): The proof of \((3.26)\) is essentially the same as that given in \([40\) Sec. 4.1, proofs of Lemmas 4.3 and 4.9\]. Below, we sketch the main proof arguments (see the proofs in \([40\) for the details of each step):

1. To show \((3.25)\), it suffices to show that for each \(m \in \mathbb{Z}_+\),
\[
\int c^m d\bar{\gamma} \leq \liminf_{n \to \infty} \int c^m d\gamma_n. \tag{3.27}
\]
(In the above, the probability measures \(\bar{\gamma}\) and \(\gamma_n\) are extended from \(\Gamma\) to \(\mathcal{X} \times \mathcal{A}\), and \(c^m\) is the truncated one-stage cost function \(\min\{c(\cdot), m\}\), as we recall.)

2. Fix \(m\). To prove \((3.27)\), consider arbitrarily small \(\epsilon = \delta = \ell^{-1}\), for some arbitrarily large \(\ell \in \mathbb{Z}_+\). The tightness of \(\{\gamma_n\}\) and the fact that \(\gamma_n \to \bar{\gamma}\), together with Assumption \((2.1) SU\), allow us to choose \(\ell \in \mathbb{Z}_+\) large enough so that for the compact set \(\Gamma_j\) in Assumption \((2.1) SU\), we have \(\gamma_n(\Gamma_j) \leq \epsilon\) for all \(n\) and \(\bar{\gamma}(\Gamma_j) \leq \epsilon\). This in turn allows us to bound \(\int_{\Gamma_j} c^m d\gamma_n\) and \(\int_{\Gamma_j} c^m d\bar{\gamma}\) by \(mc\), an negligible term when we take \(\epsilon \to 0\). Consequently, to prove \((3.27)\), we can focus on the integrals of \(c^m\) on the compact set \(\Gamma_j\) and on bounding the difference
\[
\int_{\Gamma_j} c^m d\gamma_n - \int_{\Gamma_j} c^m d\bar{\gamma}. \tag{3.28}
\]
3. We now handle the term (3.28) — this is where we apply Lusin’s theorem and the majorization property given in Assumption 2.1(M). Corresponding to \( \Gamma_j \), let us choose the element \((O, D, \nu, B) := (O_j, D_j, \nu_j, B^1_{j,m,\ell}) \in \mathcal{W}_1 \) (cf. the definition of the set \( \mathcal{W}_1 \) given in Step (ii)). By the definition of the set \( B^1_{j,m,\ell} \) (cf. Lemma 3.3 in Step (ii)), the function \( c^m \) is continuous on the closed set \( B \times F_j \) and \( \nu(B^c) \leq \delta = \ell^{-1} \). We handle the continuous part of \( c^m \) separately from the rest of \( c^m \). For the continuous part, we apply the Tietze-Urysohn extension theorem [13, Theorem 2.6.4] to extend the restriction of \( c^m \) to the closed set \( (D \cup B) \times F_j \) (on which \( c^m \) is lower semicontinuous in view of the property of \( D \) given in Assumption 2.1(M)) to a nonnegative lower semicontinuous function \( \tilde{c}^m \) on the entire space \( \mathcal{X} \times \mathcal{A} \). Since \( \gamma_n \to \tilde{\gamma} \), by [21, Prop. E.2],

\[
\liminf_{n \to \infty} \int \tilde{c}^m \, d\gamma_n \geq \int \tilde{c}^m \, d\tilde{\gamma}.
\]

We then handle the difference between \( c^m \) and \( \tilde{c}^m \). These two functions differ only outside the set \( (D \cup B) \times F_j \). By using the fact \( \nu(B^c) \leq \delta, \int \nu(B^c) \gamma \leq \delta \), the majorization property given in Lemma 3.5 and the bounds \( \int_{\Gamma_j} c^m \, d\gamma_n \leq m\epsilon, \int_{\Gamma_j} c^m \, d\tilde{\gamma} \leq m\epsilon \) from Step 2, we can calculate that

\[
\limsup_{n \to \infty} \left| \int_{\mathcal{X} \times \mathcal{A}} (c^m - \tilde{c}^m) \, d\gamma_n \right| \leq m(\delta + \epsilon), \quad \left| \int_{\mathcal{X} \times \mathcal{A}} (c^m - \tilde{c}^m) \, d\tilde{\gamma} \right| \leq m(\delta + \epsilon).
\]

4. Finally, putting all the pieces together gives us the inequality

\[
\liminf_{n \to \infty} \int c^m \, d\gamma_n \geq \int c^m \, d\tilde{\gamma} - 2m(\delta + \epsilon).
\]

By letting \( \ell \to \infty \) so that \( \delta, \epsilon \to 0 \), the desired relation (3.27) follows.

**Proving (3.26):** The proof of (3.26) is similar to the above and essentially the same as that given in [40, Sec. 4, proofs of Lemmas 4.4 and 4.10]. We outline the main arguments below (see [40] for detailed derivations):

1. Consider an arbitrary \( h \in \hat{C}_b(\mathcal{X}) \). Let \( \epsilon = \delta = \ell^{-1} \), for some arbitrarily large \( \ell \in \mathbb{Z}_+ \). Proceed as in Step 2 of the proof of (3.25) to choose \( j \in \mathbb{Z}_+ \), large enough so that for the compact set \( \Gamma_j \) in Assumption 2.1(SU), we have \( \gamma_n(\Gamma_j) \leq \epsilon \) for all \( n \) and \( \tilde{\gamma}(\Gamma_j) \leq \epsilon \).

2. Define a function \( \phi(x, \cdot) := \int_{\mathcal{X}} h(y) q(dy \mid x) \) on \( \mathcal{X} \times \mathcal{A} \). Corresponding to the chosen \( j \) and \( \ell \), choose the element \((O, D, \nu, B) := (O_j, D_j, \nu_j, B^2_{j,\ell}) \in \mathcal{W}_2 \). By the definition of the set \( B^2_{j,\ell} \) (cf. Lemma 3.3 in Step (ii)), \( \nu(B^c) \leq \delta = \ell^{-1} \) and on the closed set \( B \times F_j \), \( q(dy \mid \cdot) \) is continuous. Then, since \( q(dy \mid \cdot) \) is also continuous on the closed set \( D \times \mathcal{A} \) (cf. Assumption 2.1(M)) and \( h \) is a continuous function, we have, by [3, Prop. 7.30], that the function \( \phi \) is continuous on the closed set \( (D \cup B) \times F_j \). We now treat the continuous part of \( \phi \) separately: by the Tietze-Urysohn extension theorem [13, Theorem 2.6.4], the restriction of \( \phi \) to \( (D \cup B) \times F_j \) can be extended to a bounded continuous function \( \tilde{\phi} \) on the entire space \( \mathcal{X} \times \mathcal{A} \), with \( \|\tilde{\phi}\|_{\infty} \leq \|\phi\|_{\infty} \leq \|h\|_{\infty} \). Since \( \gamma_n \to \tilde{\gamma} \), we have

\[
\lim_{n \to \infty} \int \tilde{\phi} \, d\gamma_n = \int \tilde{\phi} \, d\tilde{\gamma}.
\]

We then handle the difference between \( \phi \) and \( \tilde{\phi} \). These two functions differ only outside the set \( (D \cup B) \times F_j \). By using the fact \( \nu(B^c) \leq \delta \), the majorization property given in Lemma 3.5 and the bounds \( \gamma_n(\Gamma_j) \leq \epsilon, \tilde{\gamma}(\Gamma_j) \leq \epsilon \) from Step 1, we can calculate that

\[
\limsup_{n \to \infty} \left| \int_{\mathcal{X} \times \mathcal{A}} (\phi - \tilde{\phi}) \, d\gamma_n \right| \leq 2\|h\|_{\infty} \cdot (\delta + \epsilon), \quad \left| \int_{\mathcal{X} \times \mathcal{A}} (\phi - \tilde{\phi}) \, d\tilde{\gamma} \right| \leq 2\|h\|_{\infty} \cdot (\delta + \epsilon).
\]
3. Finally, putting all the pieces together gives us the bound

\[
\limsup_{n \to \infty} \left| \int \phi \, d\gamma_n - \int \phi \, d\tilde{\gamma} \right| \leq 4\|h\|_{\infty} \cdot (\delta + \epsilon).
\]

By letting \( \ell \to \infty \) so that \( \delta, \epsilon \to 0 \), the desired relation (3.26) follows.

The lemma now follows from (3.25)-(3.26), as discussed earlier.

By Lemma 3.6, \((1, 0, \rho) = (L\tilde{\gamma}, \langle \tilde{\gamma}, c \rangle + \tilde{r})\) for \( \tilde{r} = \rho - \langle \tilde{\gamma}, c \rangle \geq 0 \). Thus \((1, 0, \rho) \in H\) and consequently, \( \rho = \rho^* \). This completes the proof of Theorem 3.1.

### 3.3.2 Proof of Prop. 3.2

The proof is similar to that of [22, Chap. 12.4B, Theorem 12.4.2(c)]. Suppose we have shown that

\[
\rho^* + h^*(x) \leq c(x, a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x, a), \quad \forall (x, a) \in \Gamma,
\]

and suppose also that

\[
\int |h^*| \, dp^* < \infty.
\]

Then, since \((\mu^*, p^*)\) is a stationary minimum pair, we have

\[
\rho^* = \int_{\mathcal{X}} \int_{\mathcal{A}} c(x, a) \mu^*(da \mid x) \, p^*(dx),
\]

\[-\infty < \int_{\mathcal{X}} h^*(x) \, dp^* = \int_{\mathcal{X}} \int_{\mathcal{A}} \int_{\mathcal{X}} h^*(y) q(dy \mid x, a) \mu^*(da \mid x) \, p^*(dx) < +\infty,
\]

and hence

\[
\int_{\mathcal{X}} \int_{\mathcal{A}} \left\{ \rho^* + h^*(x) - c(x, a) - \int_{\mathcal{X}} h^*(y) q(dy \mid x, a) \right\} \mu^*(da \mid x) \, p^*(dx) = 0.
\]

This together with (3.29) implies that for \( p^* \)-almost all \( x \in \mathcal{X} \),

\[
\rho^* + h^*(x) - \int_{\mathcal{A}} \left\{ c(x, a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x, a) \right\} \mu^*(da \mid x) = 0,
\]

which in turn implies that for \( p^* \)-almost all \( x \in \mathcal{X} \),

\[
\rho^* + h^*(x) = \int_{\mathcal{A}} \left\{ c(x, a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x, a) \right\} \mu^*(da \mid x)
\]

\[
\geq \inf_{a \in \mathcal{A}(x)} \left\{ c(x, a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x, a) \right\}.
\]

Then, by (3.29), equality must hold in (3.31), and this gives the desired ACOE (3.14) and (3.15).

We now verify that in the two cases given in the proposition, (3.29) and (3.30) hold. In both cases, (3.30) holds by assumption. Regarding (3.29), in both cases, since \( \{(\rho_n, h_n)\} \) is a maximizing sequence of \((P^*)\), \( \rho_n \uparrow \rho^* \) by Theorem 3.1 and for all \( n \geq 0 \),

\[
\rho_n + h_n(x) \leq c(x, a) + \int_{\mathcal{X}} h_n(y) q(dy \mid x, a), \quad \forall (x, a) \in \Gamma.
\]
In the case (i), for each \((x, a) \in \Gamma\), the assumption \(\int_{x} \sup_n h_n(x) q(dy \mid x, a) < +\infty\) implies that
\[
\limsup_{n \to \infty} \int_{x} h_n(x) q(dy \mid x, a) \leq \int_{x} \limsup_{n \to \infty} h_n(x) q(dy \mid x, a) \tag{3.33}
\]
by Fatou’s lemma. Letting \(n \to \infty\) and taking limit superior on both sides of (3.29), we obtain (3.32). Likewise, in the case (ii), since the functions \(h_n \leq 0\), we also have (3.33) by Fatou’s lemma. The relation (3.29) then follows by letting \(n \to \infty\) in (3.32), as in the case (i). This completes the proof of Prop. 3.2.

4 Extension to Constrained Average-Cost MDPs

In this section, we extend our results for an unconstrained average-cost MDP to a constrained one. Let the state and action spaces and the state transition stochastic kernel of the MDP be the same as before. Consider multiple one-stage cost functions on \(X \times A\): \(c_0, c_1, \ldots, c_d\). We assume that these functions are nonnegative and Borel measurable, finite on \(\Gamma\) and taking the value +\(\infty\) outside \(\Gamma\). The goal is to minimize the average cost w.r.t. \(c_0\), while keeping the average costs w.r.t. \(c_1, \ldots, c_d\) within given limits.

More specifically, let \(\kappa := (\kappa_1, \ldots, \kappa_d) \geq 0\) be prescribed upper limits on the average costs in the constraints. For a policy \(\pi\) and initial distribution \(\zeta\), let \(J_i(\pi, \zeta)\) denote the average cost of this pair w.r.t. \(c_i, i = 0, 1, \ldots, d\). Define the feasible set of policy and initial distribution pairs by
\[
\mathcal{S} := \{(\pi, \zeta) \in \Pi \times \mathcal{P}(X) \mid J_0(\pi, \zeta) < \infty, J_i(\pi, \zeta) \leq \kappa_i, i = 1, \ldots, d\}. \tag{4.1}
\]
Define the optimal average cost of this constrained problem to be
\[
\rho^*_c := \inf_{(\pi, \zeta) \in \mathcal{S}} J_0(\pi, \zeta).
\]

As before, within the feasible set \(\mathcal{S}\), we are especially interested in those stationary pairs, where stationary pairs are as defined in Def. 2.2(a) (recall that the set of all stationary pairs is denoted by \(\Delta_s\)). Analogous to the minimum pairs and stationary minimum pairs for an unconstrained MDP, let us define optimal pairs and stationary optimal pairs for the constrained MDP. (In the reference [31], what we call optimal pairs are referred to as constrained optimal pairs.)

**Definition 4.1** (optimal pairs).
(a) We call \((\pi^*, \zeta^*) \in \Pi \times \mathcal{P}(X)\) an optimal pair for the constrained MDP iff
\[
(\pi^*, \zeta^*) \in \mathcal{S} \quad \text{and} \quad J_0(\pi^*, \zeta^*) = \rho^*_c.
\]
(b) We call an optimal pair \((\pi^*, \zeta^*)\) lexicographically optimal iff for each \((\pi, \zeta) \in \mathcal{S}\), either \(J_i(\pi^*, \zeta^*) = J_i(\pi, \zeta)\) for all \(0 \leq i \leq d\), or for some \(\tilde{d} \leq d\),
\[
J_i(\pi^*, \zeta^*) = J_i(\pi, \zeta) \quad \forall i \leq \tilde{d} - 1, \quad J_{\tilde{d}}(\pi^*, \zeta^*) < J_{\tilde{d}}(\pi, \zeta).
\]

**Definition 4.2** (stationary optimal pairs). If a stationary pair \((\mu^*, p^*) \in \Delta_s\) is (lexicographically) optimal for the constrained MDP, we call it a stationary (lexicographically) optimal pair.

In the rest of this section, we first adapt the strict unboundedness (SU) condition and the majorization (M) condition to accommodate multiple one-stage cost functions in the constrained MDP, and under those modified conditions we show that stationary optimal pairs exist (Section 4.1). We then formulate primal/dual linear programs for the constrained MDP and present duality results that are analogous to the ones for unconstrained problems (Section 4.2). The proofs of the results of this section are given in Section 4.3.
4.1 Model Assumptions and Existence of Stationary Optimal Pairs

We impose the following conditions on the constrained MDP model:

**Assumption 4.1.**

(G) The feasible set \( S \neq \emptyset \).

(SU) There exists a nondecreasing sequence of compact sets \( \Gamma_n \uparrow \Gamma \) such that for some \( 0 \leq i \leq d \),

\[
\lim_{n \to \infty} \inf_{(x,a) \in \Gamma_n^c} c_i(x,a) = +\infty.
\]

(M) For each compact set \( K \in \{\text{proj}_X(\Gamma_n)\} \), there exist an open set \( O \supset K \) and a finite measure \( \nu \) on \( \mathcal{B}(X) \) such that

\[
q((O \setminus D) \cap B \mid x,a) \leq \nu(B), \quad \forall B \in \mathcal{B}(X), \ (x,a) \in \Gamma,
\]

where \( D \subset X \) is some closed set (possibly empty) such that restricted to \( D \times A \), the state transition stochastic kernel \( q(dy \mid x,a) \) is continuous and all the one-stage cost functions \( c_i \), \( 0 \leq i \leq d \), are lower semicontinuous.

This assumption is similar to Assumption 2.1. The condition (G) is to exclude vacuous problems. The condition (SU) is the same as that considered in [19] for the constrained MDP, and it differs from the (SU) condition in Assumption 2.1 in that here we require some one-stage cost function in the constrained problem to be strictly unbounded. The condition (M) is almost identical to that in Assumption 2.1 except that here \( D \subset X \) is required to be a closed set on which every one-stage cost function in the constrained problem is lower semicontinuous in the state variable. (As before, the set \( D \) is used here to sharpen the condition (M) by treating a "continuous" part of the model separately, and the condition (M) can be used with \( D = \emptyset \).)

Theorem 4.2 below extends our early results [40, Prop. 3.2 and Theorem 3.3] (cf. Theorem 2.2) to the constrained MDP. In particular, its part (i) can be compared with Theorem 2.2(i), and its parts (ii)-(iii) with Theorem 2.2(ii). The proof will be outlined in Section 4.3, and it is mostly based on the arguments given in [40]—roughly speaking, the present majorization condition allows us to apply the same reasoning used in [40] to every one-stage cost function \( c_i \) in the constrained MDP.

Parts (i)-(ii) of this theorem are also comparable with the results of [19, Theorem 3.2] and [31, Lemma 1.1 and the solvability part of Lemma 2.3] for constrained, lower semicontinuous MDPs. Part (iii) concerns lexicographically optimal solutions of the constrained MDP, which can be related to solutions for multi-objective MDPs similar to those discussed in [24].

**Theorem 4.2 (optimality of stationary pairs).** Under Assumption 4.1, the following hold:

(i) For any pair \((\pi, \zeta) \in S\), there exists a stationary pair \((\bar{\mu}, \bar{p}) \in \Delta_s \cap S\) with

\[
J_i(\bar{\mu}, \bar{p}) \leq J_i(\pi, \zeta), \quad \forall i = 0, \ldots, d.
\]

(ii) There exists a stationary optimal pair \((\mu^*, p^*) \in \Delta_s \cap S\).

(iii) There exists a stationary lexicographically optimal pair \((\mu^*, p^*) \in \Delta_s \cap S\).

**Remark 4.1.** It is known that even in a finite state and action MDP, for a given initial state or distribution, there need not exist a stationary optimal policy for the constrained average-cost problem (see [26, Sec. 4] for an interesting counterexample that is due to Derman [12]). The difference between this known fact and the existence of a stationary optimal pair in Theorem 4.2 is that in the constrained MDP here, the initial distribution is not given and there is freedom of choosing it to optimize the average-costs.
Remark 4.2 (pathwise average costs of $\mu^*$). Suppose that in the part (ii) or (iii) of Theorem 4.2 the policy $\mu^*$ induces on $\mathcal{X}$ a positive Harris recurrent Markov chain (see e.g., [33, Chap. 10.1] for the definition). Then, by the ergodic properties of such Markov chains and by the same proof of [40, Theorem 3.5(b)], we have that for all initial distributions $\zeta$, $P^{\mu^*_\zeta}$-almost surely,

$$\lim_{n \to \infty} n^{-1}\sum_{k=0}^{n-1} c_i(x_k, a_k) = J_i(\mu^*, p^*), \quad i = 0, 1, \ldots, d.$$ 

In other words, almost surely, on each sample path, the pathwise average costs of the policy $\mu^*$ w.r.t. $c_i$, $i = 1, 2, \ldots, d$, are also within the prescribed limits $\kappa_i$, while its pathwise average cost w.r.t. $c_0$ equals $\rho^{\ast}_c$ as well.

4.2 Linear Programming Formulation and Optimality Results

Similarly to the unconstrained case, for the constrained MDP, the primal linear program (P) is formulated to minimize the average cost over feasible stationary pairs, by utilizing the correspondence between a stationary pair and a probability measure that satisfies (3.1) discussed at the beginning of Section 3.2. Under Assumption 4.1, the existence of a stationary optimal pair given by Theorem 4.2 ensures that such a pair can be obtained by solving the primal program (P). The dual linear program $(P^*)$ is, as before, determined by (P) and two dual pairs of vector spaces we choose.

We now define precisely (P) and $(P^*)$ for the constrained MDP, by identifying the spaces and linear mappings involved in the general LP formulation given in Section 2.2. To define the primal linear program (P), we consider the dual pair of vector spaces

$$(\mathcal{M}_w(\Gamma) \times \mathbb{R}^d, F_w(\Gamma) \times \mathbb{R}^d)$$

where the weight function $w : \Gamma \to \mathbb{R}_+$ is given by

$$w(x, a) = 1 + \sup_{0 \leq i \leq d} c_i(x, a), \quad (x, a) \in \Gamma.$$

The bilinear form associated with this dual pair is defined as the sum of the bilinear forms associated with the two dual pairs, $(\mathcal{M}_w(\Gamma), F_w(\Gamma))$ and $(\mathbb{R}^d, \mathbb{R}^d)$; i.e.,

$$\langle (\gamma, \alpha), (\phi, \alpha') \rangle := \langle \gamma, \phi \rangle + \langle \alpha, \alpha' \rangle = \int_{\Gamma} \phi \, d\gamma + \sum_{i=1}^{d} \alpha_i \alpha'_i$$

for $(\gamma, \alpha) \in \mathcal{M}_w(\Gamma) \times \mathbb{R}^d$ and $(\phi, \alpha') \in F_w(\Gamma) \times \mathbb{R}^d$.

The feasible set of (P) corresponds to the subset of stationary pairs that are feasible for the constrained MDP, and it is defined by the following constraints:

$$\gamma \in \mathcal{M}_w^+(\Gamma), \quad \gamma(\Gamma) = 1, \quad \gamma(B) = \int_{\Gamma} q(B \mid x, a) \gamma(d(x, a)), \quad \forall B \in \mathcal{B}(\mathcal{X}),$$

and

$$\alpha \geq 0, \quad (\langle \gamma, c_1 \rangle, \ldots, \langle \gamma, c_d \rangle) + \alpha = \kappa.$$

The objective of (P) is to minimize the average cost $\langle \gamma, c_0 \rangle$. We can state the primal program (P) in the form introduced in Section 2.2 as follows:

$$(P) \quad \text{minimize} \quad \langle \gamma, c_0 \rangle$$

subject to

$L(\gamma, \alpha) = (1, 0, \kappa), \quad \gamma \in \mathcal{M}_w^+(\Gamma), \quad \alpha \geq 0$
The objective function of \( (P^*) \) is given by \( L = (L_0, L_1, L_2) \) with

\[
L_0(\gamma, \alpha) := \gamma(\Gamma),
\]

\[
L_1(\gamma, \alpha)(B) := \hat{\gamma}(B) - \int_M q(B \mid x, a) \gamma(d(x, a)), \quad \forall B \in \mathcal{B}(X),
\]

\[
L_2(\gamma, \alpha) := (\langle \gamma, c_1 \rangle, \ldots, \langle \gamma, c_d \rangle) + \alpha,
\]

for \( \gamma \in \mathbb{M}_w(\Gamma) \) and \( \alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d \).

To define the dual linear program \( (P^*) \), we consider the dual pair of vector spaces

\[
(\mathbb{R} \times \mathbb{M}(X) \times \mathbb{R}^d, \mathbb{R} \times \mathcal{F}_b(X) \times \mathbb{R}^d),
\]

with the bilinear form defined as the sum of the bilinear forms for the three dual pairs, \( (\mathbb{R}, \mathbb{R}), (\mathbb{M}(X), \mathcal{F}_b(X)), \) and \( (\mathbb{R}^d, \mathbb{R}^d) \), similar to (4.2). From the definition of \( L \), the adjoint mapping \( L^* \) can be identified: it is the linear mapping \( L^* = (L_1^*, L_2^*) : \mathbb{R} \times \mathcal{F}_b(X) \times \mathbb{R}^d \to \mathbb{R}_w(\Gamma) \times \mathbb{R}^d \) where

\[
L_1^*(\rho, h, \beta) := \rho + h(x) - \int_X h(y) q(dy \mid x, a) + \sum_{i=1}^d \beta_i c_i(x, a),
\]

\[
L_2^*(\rho, h, \beta) := \beta,
\]

for \( (\rho, h, \beta) \in \mathbb{R} \times \mathcal{F}_b(X) \times \mathbb{R}^d \). Clearly, \( L^*(\mathbb{R} \times \mathcal{F}_b(X) \times \mathbb{R}^d) \subseteq \mathbb{R}_w(\Gamma) \times \mathbb{R}^d \) as claimed, so both linear mappings \( L \) and \( L^* \) are weakly continuous (36, Chap. II, Prop. 12 and its corollary); cf. Prop. 2.3. The objective function of \( (P^*) \) is

\[
\langle (1, 0, \kappa), (\rho, h, \beta) \rangle = \rho + \sum_{i=1}^d \beta_i \kappa_i.
\]

Let us state the dual program \( (P^*) \) in the form introduced in Section 2.2

\[
(P^*) \quad \text{maximize} \quad \rho + \sum_{i=1}^d \beta_i \kappa_i
\]

subject to \( L^*(\rho, h, \beta) \leq (c_0, 0), \quad \rho \in \mathbb{R}, \ h \in \mathcal{F}_b(X), \ \beta \in \mathbb{R}^d. \) (4.9)

Note that the inequality constraint in (4.9) is the same as the cone constraint \(-L^*(\rho, h, \beta) + (c_0, 0) \in \mathbb{F}_w(\Gamma) \times \mathbb{R}^d_+ \) (cf. Section 2.2), and it can be expressed more explicitly as

\[
\rho + h(x) - \int_X h(y) q(dy \mid x, a) + \sum_{i=1}^d \beta_i c_i(x, a) \leq c_0(x, a), \quad \forall (x, a) \in \Gamma,
\]

\[
\beta \leq 0.
\]

The next theorem about the primal/dual programs \( (P) \) and \( (P^*) \) is an extension of Theorem 3.1 to the constrained MDP. The solvability of \( (P) \) is a consequence of the existence of a stationary optimal pair given in Theorem 4.2(ii). The absence of duality gap is the main result of this section, and its proof uses essentially the same proof arguments for Theorem 3.1(ii).

**Theorem 4.3** (consistency and absence of duality gap). Under Assumption 4.1 the following hold for the linear programs \( (P) \) and \( (P^*) \) given in (4.3) and (4.9):

(i) \( (P) \) is consistent and solvable, and \( (P^*) \) is consistent.

(ii) There is no duality gap, and the value of \( (P) \) and \( (P^*) \) is \( \rho^*_c \).
This theorem is comparable with the prior results \[ \text{[18, Theorem 4.4] and [31, Lemma 2.3]} \] on the LP approach for constrained lower semicontinuous MDPs \( [31] \) considers compact spaces and \[ \text{[19, non-compact spaces].} \] Besides the differences in model assumptions, our formulation of the dual program \( (P^* ) \) also differs from that in \[ \text{[19].} \] The main difference lies in the choice of the spaces \( \mathcal{M}(\mathcal{X}) \) and \( \mathcal{F}_b(\mathcal{X}) \) for \( (P^* ) \). As in the unconstrained case, our motivation for this choice is to avoid an extra condition on the state transition stochastic kernel used in \[ \text{[19], which is the same condition [3.12] from [22, Chap. 12.3]} \] that we discussed earlier in Remark \( \text{[3.3]} \). For the same reason as explained in Remark \( \text{[3.3]} \) the dual program \( (P^* ) \) as we formulated above need not admit an optimal solution.

The next two propositions are about the solution properties of the dual program \( (P^* ) \). Recall that a sequence of feasible solutions \( (\rho_n, h_n, \beta_n) \) of \( (P^* ) \) is called a maximizing sequence if \( \rho_n \uparrow \text{sup}(P^* ) \). We first examine the boundedness property of \( \{ \beta_n \} \). Denote by \( \beta_{n,j} \), the \( j \)th component of \( \beta_n \). Let us separate the constraints of the MDP into two categories:

\[
J^{(0)} := \{ i \mid 1 \leq i \leq d, \exists (\pi, \zeta) \in \mathcal{S} \text{ s.t. } J_i(\pi, \zeta) < \kappa_i \}, \quad J^{(1)} := \{ 1, 2, \ldots, d \} \setminus J^{(0)}. \tag{4.12}
\]

**Proposition 4.4.** Suppose Assumption \( \text{[4.7]} \) hold. Let \( \{ (\rho_n, h_n, \beta_n) \} \) be a maximizing sequence of the dual program \( (P^* ) \). Let \( (\mu^*, p^*) \) be any stationary optimal pair for the constrained MDP. Then the following hold:

(i) The sequence \( \{ \beta_{n,j} \}_{n \geq 0} \) is bounded for every \( j \in J^{(0)} \).

(ii) \( \lim_{n \to \infty} \beta_{n,j} = 0 \) if \( J_j(\mu^*, p^*) < \kappa_j \).

(iii) Suppose there exists \( (\pi, \zeta) \in \Pi \times \mathcal{P}(\mathcal{X}) \) such that

\[
J_j(\pi, \zeta) < \kappa_j \quad \forall j \in J^{(1)}, \quad J_j(\pi, \zeta) < \infty \quad \forall j \in J^{(0)} \cup \{ 0 \}.
\]

Then the sequence \( \{ \beta_{n,j} \}_{n \geq 0} \) is bounded.

**Remark 4.3.** An optimal solution \( (\gamma^*, \alpha^*) \) of \( (P^* ) \) corresponds to a stationary optimal pair \( (\mu^*, p^*) \) with \( \alpha_j^* = \kappa_j - J_j(\mu^*, p^*) \) for \( 1 \leq j \leq d \) (this follows from the correspondence relationship explained at the beginning of Section \( \text{[3.2]} \)). So Prop. \( \text{[4.4]} \)(ii) entails the complementarity relation \( (\alpha^*, \beta^*) = 0 \) for an optimal solution \( (\gamma^*, \alpha^*) \) of \( (P^* ) \), if we define \( \beta^* = (\beta_1^*, \ldots, \beta_d^* ) \) as follows: \( \beta_j^* = \lim_{n \to \infty} \beta_{n,j} \) if this limit exists, and assign \( \beta_j^* \) an arbitrary number otherwise.

Regarding the set \( J^{(1)} \), it can be seen that when \( \mathcal{S} \neq \emptyset \), \( J^{(1)} \) consists of all those \( i \) such that \( w.r.t. \ c_i \), every feasible pair in \( \mathcal{S} \) has the same maximally allowed average cost \( \kappa_i \).

Proposition \( \text{[4.4]} \)(iii) gives a sufficient condition under which the \( J^{(1)} \)-components of \( \{ \beta_{n,j} \} \) are also bounded—note that this condition involves non-feasible policy and initial distribution pairs and is different from the Slater condition \( J^{(1)} = \emptyset \). One exceptional case where Prop. \( \text{[4.4]} \)(iii) is inapplicable is when \( \kappa_i = 0 \) for all \( 1 \leq i \leq d \).

When \( \{ \beta_{n,j} \}_{n \geq 0} \) is bounded, as when the condition of Prop. \( \text{[4.4]} \)(iii) holds, we can choose a subsequence of the maximizing sequence \( \{ (\rho_n, h_n, \beta_n) \} \) so that \( \beta_n \) converges. The subsequence is obviously also a maximizing sequence for \( (P^* ) \). Then, with additional assumptions on the sequence of functions \( h_n \), we can derive an optimality equation for the constrained MDP that is analogous to the ACOE \( \text{[3.14]} \) in Prop. \( \text{[3.2]} \) for the unconstrained MDP. We state this result in the next proposition. It is comparable with the result of \( \text{[19, Theorem 5.2(b)]} \) for constrained lower semicontinuous MDPs.

**Proposition 4.5** (ACOE for \( p^* \)-almost all states in the constrained MDP). Suppose Assumption \( \text{[4.7]} \) hold. Let \( (\mu^*, p^*) \) be a stationary optimal pair for the constrained MDP. Let \( \{ (\rho_n, h_n, \beta_n) \} \) be a maximizing sequence of the dual program \( (P^* ) \). Assume that \( \{ \beta_n \} \) converges to some finite \( \beta^* \), and furthermore, assume that \( \{ h_n \} \) satisfies either of the following two conditions with \( h^* = \lim \sup h_n \):

(i) \( h_n \geq 0 \), \( \int h^* dp^* < +\infty \), and \( \int_{\mathcal{X}} \sup h_n(y) q(dy \mid x, a) < +\infty \) for all \( (x, a) \in \Gamma \);

(ii) \( h_n \leq 0 \) and \( \int h^* dp^* > -\infty \).
Then $h^*$ is finite $p^*$-almost everywhere, and with
\[ c^*(x,a) := c_0(x,a) - \sum_{i=1}^{d} \beta^*_i c_i(x,a), \quad \tilde{\rho}^* := \rho^*_c - \sum_{i=1}^{d} \beta^*_i \kappa_i, \]
we have that
\[ \tilde{\rho}^* + h^*(x) \leq c^*(x,a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x,a), \quad \forall (x,a) \in \Gamma, \]  \hfill (4.13)
and that for $p^*$-almost all $x \in \mathcal{X}$,
\[ \tilde{\rho}^* + h^*(x) = \inf_{a \in A(x)} \left\{ c^*(x,a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x,a) \right\} \] \hfill (4.14)
\[ = \int_{a \in A(x)} \left\{ c^*(x,a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x,a) \right\} \mu^*(da \mid x). \] \hfill (4.15)

4.3 Proofs

4.3.1 Proof of Theorem 4.2 (Outline)

The proof of Theorem 4.2 is similar to that of Theorem 2.2 on the property of stationary pairs and the existence of a stationary minimum pair for an unconstrained MDP. The latter proof is given in our prior work [40, Sec. 4.1, proofs of Prop. 3.2 and Theorem 3.3], and some of the main arguments we have also mentioned earlier in the proof of Lemma 3.6. Therefore, to avoid repetition, we will only outline the proof of Theorem 4.2 below. We will first state some results we obtained previously in the unconstrained MDP context, and we will then directly apply those results to the present case of constrained MDPs.

In [40], given a policy and initial distribution pair $(\pi, \zeta)$ with $J(\pi, \zeta) < \infty$, we consider the occupancy measures of the policy $\pi$: 
\[ \gamma_n(B) := \frac{1}{n} \sum_{k=1}^{n} P^\pi_{x_k, a_k} \{ (x_k, a_k) \in B \}, \quad B \in \mathcal{B}(\Gamma). \] \hfill (4.16)
Under the strict unboundedness condition in Assumption 2.1 (SU), $\{\gamma_n\}$ is tight, so from any subsequence of $\{\gamma_n\}$, we can extract a further subsequence $\{\gamma_{n_k}\}$ with $\gamma_{n_k} \Rightarrow \bar{\gamma} \in \mathcal{P}(\Gamma)$. It is then proved in [40, Sec. 4.1] that under the majorization condition (M) and the strict unboundedness condition (SU) in Assumption 2.1 the limiting probability measure $\bar{\gamma}$ has the following properties:

(a) $\bar{\gamma}$ corresponds to a stationary pair $(\bar{\mu}, \bar{\rho})$; i.e., $\bar{\gamma}(d(x,a)) = \bar{\mu}(da \mid x) \bar{\rho}(dx)$ and $\bar{\rho}$ is the invariant distribution of the Markov chain induced by the stationary policy $\bar{\mu}$ on $\mathcal{X}$.

(b) The average cost of the pair $(\bar{\mu}, \bar{\rho})$ satisfies
\[ J(\bar{\mu}, \bar{\rho}) = \langle \bar{\gamma}, c \rangle \leq \liminf_{k \to \infty} \langle \gamma_{n_k}, c \rangle. \] \hfill (4.17)

Similarly, in [40], when proving the existence of a stationary minimum pair, we start with a sequence of stationary pairs $(\mu_n, p_n)$ with $J(\mu_n, p_n) \downarrow \rho^*$, and instead of the occupancy measures in (4.16), we define $\gamma_n$ to be
\[ \gamma_n(d(x,a)) := \mu_n(da \mid x) p_n(dx). \] \hfill (4.18)
We then proceed similarly to the above case to extract a weakly convergent subsequence of $\{\gamma_n\}$ and prove that the limiting probability measure $\bar{\gamma}$ has the same properties (a)-(b) as given above.

We now explain how we can apply these results to prove Theorem 4.2 for the constrained MDP. To prove Theorem 4.2 (i), we consider $\{\gamma_n\}$ defined by (4.16) for a pair $(\pi, \zeta) \in \mathcal{S}$. By the feasibility of $(\pi, \zeta)$, its average costs are all finite:
\[ J_i(\pi, \zeta) = \limsup_{n \to \infty} \langle \gamma_n, c_i \rangle < \infty, \quad \forall i = 0, 1, \ldots, d. \]
Since at least one of the one-stage cost functions $c_0, c_1, \ldots, c_d$ is strictly unbounded by Assumption 4.1(SU), this implies that $\langle \gamma_n \rangle$ is a tight family of probability measures. We then proceed as discussed above to obtain the limiting probability measure $\hat{\gamma}$ from a weakly convergent subsequence $\{\gamma_n\}$ of $\{\gamma_n\}$.

Next, using the majorization condition in Assumption 4.1(M), together with Assumption 4.1(SU), it follows as before that $\hat{\gamma}$ has the property (a) and gives us a stationary pair $(\hat{\mu}, \hat{\nu})$. Moreover, since Assumption 4.1(M) is the same as Assumption 2.1(M) holding for every one-stage cost function $c_i$ in the constrained MDP, 4.1(L) in the property (b) above now holds with the function $c$ replaced by every $c_i$; that is

$$J_i(\hat{\mu}, \hat{\nu}) = \langle \hat{\gamma}, c_i \rangle \leq \liminf_{k \to \infty} \langle \gamma_n, c_i \rangle, \quad \forall i = 0, 1, \ldots, d.$$ 

Then, since $J_i(\pi, \zeta) = \limsup_{n \to \infty} \langle \gamma_n, c_i \rangle \geq \liminf_{k \to \infty} \langle \gamma_n, c_i \rangle$, it follows that

$$J_i(\hat{\mu}, \hat{\nu}) \leq J_i(\pi, \zeta), \quad \forall i = 0, 1, \ldots, d.$$ 

This proves Theorem 4.2(i).

To prove Theorem 4.2(ii), which asserts the existence of a stationary optimal pair, we consider a sequence of stationary pairs $(\mu_n, p_n) \in \mathcal{S}$ with $J_i(\mu_n, p_n) \downarrow \rho^*_c$ (there exists such a sequence by the part (i) just proved). Let $\gamma_n$ be defined as in (4.18). Then, since $J_i(\mu_n, p_n) = \langle \gamma_n, c_i \rangle$ for all $0 \leq i \leq d$, we have

$$\sup_{n \geq 0} \langle \gamma_n, c_i \rangle < \infty, \quad \forall i = 0, 1, \ldots, d.$$ 

Since one of the one-stage cost functions $c_i$ is strictly unbounded under our assumption, as in the proof of the part (i), we can extract a convergent subsequence $\{\gamma_{n_k}\}$ of $\{\gamma_n\}$. Then from its limiting probability measure $\gamma^*$, we can obtain a stationary pair $(\mu^*, p^*)$ such that for all $i = 0, 1, \ldots, d$,

$$J_i(\mu^*, p^*) = \langle \gamma^*, c_i \rangle \leq \liminf_{k \to \infty} \langle \gamma_{n_k}, c_i \rangle. \quad (4.19)$$ 

Since $\langle \gamma_{n_k}, c_i \rangle = J_i(\mu_{n_k}, p_{n_k})$, this implies that

$$J_0(\mu^*, p^*) \leq \rho^*_c, \quad J_i(\mu^*, p^*) \leq \kappa_i, \quad \forall i = 1, 2, \ldots, d,$$

and hence $(\mu^*, p^*)$ is a stationary optimal pair for the constrained MDP.

We now prove Theorem 4.2(iii), which asserts the existence of a stationary lexicographically optimal pair. First, let us define recursively sets $\mathcal{S}_i^*$ and scalars $\kappa_i^*$ as follows: Let

$$\kappa_0^* = \rho^*_c, \quad \mathcal{S}_0^* = \{ (\pi, \zeta) \in \mathcal{S} \mid J_0(\pi, \zeta) = \rho^*_c \},$$

and for $1 \leq i \leq d$, let

$$\kappa_i^* = \inf \{ J_i(\pi, \zeta) \mid (\pi, \zeta) \in \mathcal{S}_{i-1}^* \}, \quad \mathcal{S}_i^* = \{ (\pi, \zeta) \in \mathcal{S}_{i-1}^* \mid J_i(\pi, \zeta) = \kappa_i^* \}.$$

The set $\mathcal{S}_i^*$ consists of all the lexicographically optimal pairs, so to prove Theorem 4.2(iii), we need to show $\Delta_s \cap \mathcal{S}_i^* \neq \emptyset$. By Theorem 4.2(ii) just proved, $\Delta_s \cap \mathcal{S}_0^* \neq \emptyset$. Let us prove by induction that $\Delta_s \cap \mathcal{S}_i^* \neq \emptyset$ for all $i \leq d$.

Assume that for some $j \leq d$, $\mathcal{S}_{j-1}^* \neq \emptyset$. Then $\kappa_j^*$ is well-defined, and there exists a sequence of policy and initial distribution pairs $(\pi_n, \zeta_n) \in \mathcal{S}_{j-1}^*$ with

$$J_j(\pi_n, \zeta_n) \downarrow \kappa_j^*.$$

By Theorem 4.2(i) proved earlier, for each $(\pi_n, \zeta_n)$, there is a stationary pair $(\mu_n, p_n)$ with

$$J_i(\mu_n, p_n) \leq J_i(\pi_n, \zeta_n), \quad \forall i = 0, 1, \ldots, d.$$
This together with the fact \((\pi_n, \zeta_n) \in S_{j-1}^*\) implies that \((\mu_n, p_n) \in S_{j-1}^*\). Consider now the sequence \(\{(\mu_n, p_n)\}\) of stationary pairs thus constructed. Exactly the same proof arguments for establishing the part (ii) can be applied here, and they yield that there exists a stationary pair \((\mu^*, p^*)\) that satisfies (4.19). Therefore,
\[
J_i(\mu^*, p^*) = \kappa_i^*, \quad i = 0, 1, \ldots, j,
\]
and consequently, \((\mu^*, p^*) \in S_j^*\). This proves that \(\Delta_s \cap S_j^* \neq \emptyset\); then, by induction, \(\Delta_s \cap S_d^* \neq \emptyset\). Hence there is a stationary lexicographically optimal pair for the constrained MDP.

This completes the proof of Theorem 4.2.

4.3.2 Proof of Theorem 4.3 (Outline)

The consistency and solvability of (P) follow from Theorem 4.2(i)-(ii), respectively. The consistency of \((P^*)\) is trivial (e.g., let \(\rho = 0, h(\cdot) \equiv 0, \beta = 0\). Thus, \(0 \leq \sup(P^*) \leq \inf(P) = \rho^*_c\).

We now prove the absence of duality gap. This proof is similar to that of Theorem 3.1(ii) for the unconstrained MDP case. Since the value of \((P^*)\) is finite, by [2, Theorem 3.3] (cf. Theorem 2.4), the value of \((P^*)\) equals the subvalue \(\underline{\rho}\) of (P). Therefore, to prove there is no duality gap is to prove \(\underline{\rho} = \rho^*_c\). For this, it suffices to show
\[
((1, 0, \kappa), \underline{\rho}) \in H,
\]
where the set \(H\) is as defined in (2.24) and, for the case here, is given by
\[
H := \{(L(\gamma, \alpha), \langle \gamma, c_0 \rangle + r) \mid \gamma \in \mathcal{M}_w^+(\Gamma), \alpha \in \mathbb{R}_+^d, r \geq 0\}. \tag{4.20}
\]
Recall that by definition the subvalue \(\underline{\rho} = \inf \{r \mid (1, 0, \kappa), r) \in \mathcal{R}\}\) (cf. Section 2.2).

To prove \(((1, 0, \kappa), \underline{\rho}) \in H\), we will construct a stationary pair \((\bar{\mu}, \bar{\rho}) \in \mathcal{S}\) with \(J_0(\bar{\mu}, \bar{\rho}) \leq \underline{\rho}\), and the proof proceeds in four steps as in the proof of Theorem 3.1(ii). Let us outline these steps, explaining briefly some minor changes in the details of the arguments.

Step (i): From the definition of \(\underline{\rho}\), it follows that \(((1, 0, \kappa), \underline{\rho}) \in \mathcal{R}\) and there exist a direct set \(\mathcal{I}\) and a net \(\{(\gamma_i, \alpha_i)\}_{i \in \mathcal{I}}\) in \(\mathbb{M}_w^+(\Gamma) \times \mathbb{R}_+^d\) such that
\[
\gamma_i(\Gamma) \to 1, \tag{4.21}
\]
\[
\int_X h(x) \gamma_i(dx) - \int_X \int_X h(y) q(dy \mid x, a) \gamma_i(d(x, a)) \to 0, \quad \forall h \in \mathcal{F}_b(\mathcal{X}), \tag{4.22}
\]
\[
\langle \gamma_i, c_1 \rangle, \cdots, \langle \gamma_i, c_d \rangle + \alpha_i \to \kappa, \tag{4.23}
\]
\[
\langle \gamma_i, c_0 \rangle \to \underline{\rho}. \tag{4.24}
\]
As before, in view of (4.21) and the fact \(\gamma_i \in \mathbb{M}_w^+(\Gamma)\), by redefining the net \(\{(\gamma_i, \alpha_i)\}_{i \in \mathcal{I}}\) if necessary, we may assume that every \(\gamma_i\) in the above is a probability measure on \(\mathcal{B}(\Gamma)\).

Step (ii): Similarly to Lemma 3.4 we extract a sequence \(\{(\gamma_n, \alpha_n)\}_{n \geq 0} \subset \{(\gamma_i, \alpha_i)\}_{i \in \mathcal{I}}\) such that
\[
\int_X h(x) \gamma_n(dx) - \int_X \int_X h(y) q(dy \mid x, a) \gamma_n(d(x, a)) \to 0, \quad \forall h \in \mathcal{C}_b(\mathcal{X}) \cup \hat{\mathcal{F}}_b(\mathcal{X}), \tag{4.25}
\]
\[
\langle \gamma_n, c_1 \rangle, \cdots, \langle \gamma_n, c_d \rangle + \alpha_n \to \kappa, \tag{4.26}
\]
\[
\langle \gamma_n, c_0 \rangle \to \underline{\rho}, \tag{4.27}
\]
where \(\mathcal{C}_b(\mathcal{X})\) and \(\hat{\mathcal{F}}_b(\mathcal{X})\) in (4.25) are two chosen countable subsets of \(\mathcal{F}_b(\mathcal{X})\), the properties of which are needed in the subsequent two steps of our proof. In particular, the set \(\mathcal{C}_b(\mathcal{X})\) is the countable set of bounded continuous functions with the property (3.21), the same set as defined in the proof of Theorem 3.1(ii). The countable set \(\hat{\mathcal{F}}_b(\mathcal{X})\) is also defined by the equation (4.24) in that proof:
\[
\hat{\mathcal{F}}_b(\mathcal{X}) := \{1_E(\cdot) \mid E = (O \setminus D) \cap B^c\} \text{ for some } (O, D, \nu, B) \in \mathcal{W}_1 \cup \mathcal{W}_2.
\]
However, we define the set $W_1$ slightly differently here, to take into account the multiple one-stage cost functions in the constrained MDP. Specifically, in the definition of $W_1$ (cf. Lemma 3.3 and the definitions preceding this lemma), we make the following changes. We now use the sets and finite measures $(O, D, ν)$ involved in Assumption (4.1) (M) instead of Assumption (2.1) (M). We choose the sets $B_{j,m,ℓ}$ for each $j, m, ℓ ∈ ℤ_+$, such that besides the property in Lemma (3.3) (i), we have that restricted to $B_{j,m,ℓ} × F_j$, all the $(d + 1)$ truncated one-stage cost functions, $c^m_i$, $i = 0, 1, \ldots, d$, are continuous (where $c^m_i(·) = \min\{c_i(·), m\}$). This is possible by Lusin’s theorem (since we have only a finite number of these cost functions, we can apply Lusin’s theorem to each one of them and then combine the results).

**Step (iii):** This step is the same as before. The relations (4.26) - (4.27) together with Assumption (4.1) (SU) imply that $\{γ_n\}$ is a tight family of probability measures and therefore has a weakly convergent subsequence $\{γ_{nk}\}$. Consider the corresponding subsequence $(\bar{γ}_n)$ by redefining $(\{γ_n, α_n\})$ to be this subsequence. Now, denote the limit of $\{γ_n\}$ by $\bar{γ}$, and decompose $\bar{γ}$ as $\bar{γ}(d(x, a)) = \bar{μ}(dx | x) \bar{ρ}(dx)$, where $\bar{ρ}$ is the marginal of $\bar{γ}$ on $X$ and $\bar{μ}$ is a stationary policy. Then, using Assumption (4.1) (M) instead of Assumption (2.1) (M), we have that Lemma 3.3 holds as before, which gives us the desired majorization properties for $\bar{γ}_n$ and $\bar{γ}$ that we will need in the next, last step.

**Step (iv):** This step is almost the same as before, except that we apply those same arguments in the proof of (3.25) to every cost function $c_i$, $0 \leq i \leq d$, in the present constrained problem. Then, similar to Lemma 3.6 we obtain that the pair $(\bar{μ}, \bar{ρ})$ is a stationary pair and satisfies that

$$J_i(\bar{μ}, \bar{ρ}) = \langle γ_i, c_i \rangle ≤ \lim_{n→∞} \inf \langle γ_n, c_i \rangle, \quad ∀ i = 0, 1, \ldots, d.$$ 

Combining this with (4.26) and (4.27) (recall also $α_n ≥ 0$), we obtain

$$J_0(\bar{μ}, \bar{ρ}) ≤ ρ, \quad J_i(\bar{μ}, \bar{ρ}) ≤ κ_i, \quad ∀ i = 0, 1, \ldots, d.$$ 

Therefore, if we let

$$\bar{r} := ρ - J_0(\bar{μ}, \bar{ρ}) ≥ 0, \quad \bar{α} := κ - (J_1(\bar{μ}, \bar{ρ}), \ldots, J_d(\bar{μ}, \bar{ρ})) ≥ 0,$$

then

$$((1, 0, κ), ρ) = (L(\bar{γ}, \bar{α}), \langle γ, α_0 \rangle + \bar{r}) ∈ H.$$ 

This implies $ρ = ρ^*_c$ (since it implies $ρ^*_c ≤ ρ$, whereas $ρ ≤ ρ^*_c$). Hence there is no duality gap between (P) and $(P^*)$.

**4.3.3 Proofs of Props. 4.4 and 4.5**

**Proof of Prop. 4.4** (i) Consider any $j ∈ J^{(0)}$ and some pair $(π, ζ) ∈ S$ with $J_j(π, ζ) < κ_j$. By Theorem 4.2 (i), there exists a stationary pair $(\bar{μ}, \bar{ρ}) ∈ S$ with $J_j(\bar{μ}, \bar{ρ}) ≤ J_j(π, ζ)$ for all $0 ≤ i ≤ d$. Then $J_j(\bar{μ}, \bar{ρ}) < κ_j$.

Now for each $n ≥ 0$, since $(ρ_n, h_n, β_n)$ is feasible for $(P^*)$, we have from (4.10) - (4.11) that $β_n ≤ 0$ and for all $(x, a) ∈ Γ$,

$$ρ_n + h_n(x) ≤ c_0(x, a) - \sum_{i=1}^d β_n,i c_i(x, a) + \int_X h_n(y) q(dy | x, a),$$

and therefore, by adding $\sum_{i=1}^d β_n,i κ_i$ to both sides,

$$ρ_n + \sum_{i=1}^d β_n,i κ_i + h_n(x) ≤ c_0(x, a) + \sum_{i=1}^d β_n,i (κ_i - c_i(x, a)) + \int_X h_n(y) q(dy | x, a). \quad (4.28)$$
Integrate both sides of \[4.29\] w.r.t. the probability measure \(\tilde{\gamma}(d(x, a)) = \tilde{\mu}(da \mid x) \tilde{p}(dx)\). Notice that
\[
\int h_n \, d\tilde{p} = \int_{\Gamma} \int_X h_n(y) q(dy \mid x, a) \, d\tilde{\gamma}
\]
since \((\tilde{\mu}, \tilde{p})\) is a stationary pair. We thus obtain
\[
\rho_n + \sum_{i=1}^d \beta_{n,i} \kappa_i \leq J_0(\tilde{\mu}, \tilde{p}) + \sum_{i=1}^d \beta_{n,i} (\kappa_i - J_i(\tilde{\mu}, \tilde{p})).
\]
(4.29)

Take \(n \to \infty\). Since \(\{(\rho_n, h_n, \beta_n)\}\) is a maximizing sequence for \((P^*)\), \(\rho_n + \sum_{i=1}^d \beta_{n,i} \kappa_i \to \rho_c^*\) by Theorem 4.3(ii). It then follows from \(4.29\) that
\[
\rho_c^* - J_0(\tilde{\mu}, \tilde{p}) \leq \liminf_{n \to \infty} \sum_{i=1}^d \beta_{n,i} (\kappa_i - J_i(\tilde{\mu}, \tilde{p}))
\]
(4.30)

where we used the fact \(\kappa_i - J_i(\tilde{\mu}, \tilde{p}) \geq 0\) and \(\beta_{n,i} \leq 0\) for all \(i\) to derive \(4.31\). Since \(\kappa_j - J_j(\tilde{\mu}, \tilde{p}) > 0\), \(4.31\) implies \(\liminf_{n \to \infty} \beta_{n,j} > -\infty\). Hence the sequence \(\{\beta_{n,j}\}_{n \geq 0}\) is bounded.

(ii) In this case, \(j \in \mathcal{J}^{(0)}\), and \(4.31\) holds with \((\tilde{\mu}, \tilde{p}) = (\mu^*, p^*)\) and with its left-hand side equal to \(\rho_c^* - J_0(\mu^*, p^*) = 0\). Therefore, if \(J_j(\mu^*, p^*) < \kappa_j\), we must have \(\lim_{n \to \infty} \beta_{n,j} = 0\).

(iii) In this case, by assumption there is a pair \((\bar{\pi}, \bar{\zeta})\) satisfying
\[
J_j(\bar{\pi}, \bar{\zeta}) < \kappa_j \quad \forall j \in \mathcal{J}^{(1)}, \quad J_j(\bar{\pi}, \bar{\zeta}) < \infty \quad \forall j \in \mathcal{J}^{(0)} \cup \{0\}.
\]
As in the part (i), let us consider a stationary pair \((\tilde{\mu}, \tilde{p})\) with \(J_i(\tilde{\mu}, \tilde{p}) \leq J_i(\bar{\pi}, \bar{\zeta})\) for all \(0 \leq i \leq d\). Such a pair exists by Theorem 4.2(i), since we can apply this theorem with a different feasible set \(\mathcal{S}'\) instead of \(\mathcal{S}\) and in \(\mathcal{S}'\) we can use \(J_i(\bar{\pi}, \bar{\zeta})\) as the upper limits on the average costs w.r.t. \(c_i\) for \(1 \leq i \leq d\), for instance.

The average costs of this stationary pair \((\tilde{\mu}, \tilde{p})\) thus satisfy
\[
J_j(\tilde{\mu}, \tilde{p}) < \kappa_j \quad \forall j \in \mathcal{J}^{(1)}, \quad J_j(\tilde{\mu}, \tilde{p}) < \infty \quad \forall j \in \mathcal{J}^{(0)} \cup \{0\}.
\]
(4.32)

We also have, as in the part (i), that \(4.30\) holds for this pair \((\tilde{\mu}, \tilde{p})\). Now, as we proved, \(\{\beta_{n,i}\}_{n \geq 0}\) is bounded for every \(i \in \mathcal{J}^{(0)}\). This together with the second relation in \(4.32\) implies that the term
\[
\limsup_{n \to \infty} \sum_{i \in \mathcal{J}^{(0)}} \beta_{n,i} (\kappa_i - J_i(\tilde{\mu}, \tilde{p}))
\]
is finite. From \(4.30\), we have the inequality
\[
\rho_c^* - J_0(\tilde{\mu}, \tilde{p}) \leq \liminf_{n \to \infty} \sum_{i=1}^d \beta_{n,i} (\kappa_i - J_i(\tilde{\mu}, \tilde{p}))
\]
\[
\leq \limsup_{n \to \infty} \sum_{i \in \mathcal{J}^{(0)}} \beta_{n,i} (\kappa_i - J_i(\tilde{\mu}, \tilde{p})) + \liminf_{n \to \infty} \sum_{i \in \mathcal{J}^{(1)}} \beta_{n,i} (\kappa_i - J_i(\tilde{\mu}, \tilde{p})).
\]
(4.33)

In \(4.33\), since the term on the left-hand side and the first term on the right-hand side are both finite, the second term on the right-hand side must satisfy
\[
\liminf_{n \to \infty} \sum_{i \in \mathcal{J}^{(1)}} \beta_{n,i} (\kappa_i - J_i(\tilde{\mu}, \tilde{p})) > -\infty.
\]
Then, since \(\beta_n \leq 0\), in view of the first relation in \(4.32\), the preceding inequality implies that \(\{\beta_{n,i}\}_{n \geq 0}\) must be bounded for every \(i \in \mathcal{J}^{(1)}\). Combining this with the result of the part (i), we obtain that for every \(i = 1, 2, \ldots, d\), the sequence \(\{\beta_{n,i}\}_{n \geq 0}\) is bounded. Hence \(\{\beta_n\}\) is bounded. \(\square\)
Proof of Prop. 4.5. The proof arguments are similar to those of [19, Theorem 5.2(b)] for constrained MDPs and those of Prop. 3.2 for unconstrained MDPs. By the inequality (4.28), for each $n \geq 0$,

$$\rho_n + \sum_{i=1}^{d} \beta_n,\kappa_i + h_n(x) \leq c_0(x, a) + \sum_{i=1}^{d} \beta_n, (\kappa_i - c_i(x, a)) + \int_{\mathcal{X}} h_n(y) q(dy \mid x, a), \quad \forall (x, a) \in \Gamma.$$ 

Let $n \to \infty$. Since $\rho_n + \sum_{i=1}^{d} \beta_n,\kappa_i \to \rho^*_c$ by Theorem 4.3(ii) and $\beta_n \to \beta^* \leq 0$ by assumption, we obtain

$$\rho^*_c + \limsup_{n \to \infty} h_n(x) \leq c_0(x, a) + \sum_{i=1}^{d} \beta^*_i (\kappa_i - c_i(x, a)) + \limsup_{n \to \infty} \int_{\mathcal{X}} h_n(y) q(dy \mid x, a), \quad \forall (x, a) \in \Gamma.$$ 

This implies the desired inequality (4.13) if it is true that

$$\limsup_{n \to \infty} \int_{\mathcal{X}} h_n(y) q(dy \mid x, a) \leq \int_{\mathcal{X}} \limsup_{n \to \infty} h_n(y) q(dy \mid x, a), \quad \forall (x, a) \in \Gamma. \quad (4.34)$$

Now, under either of the two assumptions on $\{b_n\}$ in the proposition, the inequality (4.34) follows from Fatou’s Lemma (cf. the proof of Prop. 3.2). Therefore, the inequality (4.13) is proved; i.e.,

$$\rho^*_c + \rho^*(x) \leq c_0(x, a) + \sum_{i=1}^{d} \beta^*_i (\kappa_i - c_i(x, a)) + \int_{\mathcal{X}} \rho^*(y) q(dy \mid x, a), \quad \forall (x, a) \in \Gamma. \quad (4.35)$$

Next, integrate both sides of (4.35) w.r.t. the probability measure \(\gamma^*(d(x, a)) = \mu^*(da \mid x) p^*(dx)\), where the integrability is ensured by our assumption \(\int |h^*| dp^* < \infty\) together with the invariance property of \(p^*\), as in the proof of Prop. 3.2. Then, similarly to the derivation of (4.29) in the proof of Prop. 4.4(i), we obtain that

$$\rho^*_c \leq J_0(\mu^*, p^*) + \sum_{i=1}^{d} \beta^*_i (\kappa_i - J_i(\mu^*, p^*)).$$

Since \(J_0(\mu^*, p^*) = \rho^*_c\) and the second term in the right-hand side above is nonpositive, equality must hold in the above inequality, and this result can be equivalently expressed as

$$\int_{\mathcal{X}} \int_{\mathcal{A}} \left\{ \rho^*_c + \rho^*(x) - c_0(x, a) - \sum_{i=1}^{d} \beta^*_i (\kappa_i - c_i(x, a)) - \int_{\mathcal{X}} h^*(y) q(dy \mid x, a) \right\} \mu^*(da \mid x) p^*(dx) = 0.$$ 

Similarly to the proof of Prop. 3.2, the preceding equality together with the inequality (4.35) implies that for \(p^*\)-almost all $x \in \mathcal{X}$,

$$\rho^*_c - \sum_{i=1}^{d} \beta^*_i \kappa_i + \rho^*(x) = \inf_{a \in \mathcal{A}(x)} \left\{ c_0(x, a) - \sum_{i=1}^{d} \beta^*_i c_i(x, a) + \int_{\mathcal{X}} h^*(y) q(dy \mid x, a) \right\} \mu^*(da \mid x)$$

This gives the desired ACOE (4.14) and (4.15).

Acknowledgments

The author would like to thank Professor Eugene Feinberg for helpful discussion. This research was supported by a grant from Alberta Innovates—Technology Futures.
References

[1] Altman, E. (1999). *Constrained Markov Decision Processes*. Chapman and Hall/CRC, Boca Raton, FL.
[2] Anderson, E. J. and Nash, P. (1987). *Linear Programming in Infinite-Dimensional Spaces*. John Wiley & Sons, Chichester, UK.
[3] Bertsekas, D. P. and Shreve, S. E. (1978). *Stochastic Optimal Control: The Discrete Time Case*. Academic Press, New York.
[4] Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley & Sons, New York.
[5] Blackwell, D. and Ryll-Nardzewski, C. (1963). Non-existence of everywhere proper conditional distributions. *Ann. Math. Statist.*, 34:223–225.
[6] Borkar, V. S. (1983). Controlled Markov chains and stochastic networks. *SIAM J. Control Optim.*, 21:652–666.
[7] Borkar, V. S. (1984). On minimum cost per unit time control of Markov chains. *SIAM J. Control Optim.*, 22:965–978.
[8] Borkar, V. S. (1988). A convex analytic approach to MDPs. *Probab. Th. Rel. Fields*, 78:583–602.
[9] Denardo, E. V. (1970). On linear programing in a Markov decision problem. *Manag. Sci.*, 16:281–288.
[10] Denardo, E. V. and Fox, B. L. (1968). Multichain Markov renewal programs. *SIAM J. Appl. Math.*, 16:468–487.
[11] Derman, C. (1963). Stable sequential control rules and Markov chains. *J. Math. Anal. Appl.*, 6:257–265.
[12] Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge University Press, Cambridge, UK.
[13] Feinberg, E. A. and Shwartz, A. (1996). Constrained discounted dynamic programming. *Math. Oper. Res.*, 21:922–945.
[14] Feinberg, E. A. and Shwartz, A., editors (2002). *Handbook of Markov Decision Processes: Methods and Applications*. Springer, New York.
[15] Hernández-Lerma, O. (1993). Existence of average optimal policies in MCPs with strictly unbounded costs. *Kybernetika*, 29:1–17.
[16] Hernández-Lerma, O. and González-Hernández, J. (1998). Infinite linear programming and multichain Markov control processes in uncountable spaces. *SIAM J. Control Optim.*, 36:313–335.
[17] Hernández-Lerma, O. and González-Hernández, J. (2000). Constrained Markov control processes in Borel spaces: the discounted case. *Math. Meth. Oper. Res.*, 52:271–285.
[18] Hernández-Lerma, O., González-Hernández, J., and López-Martínez, R. R. (2003). Constrained average cost Markov control processes in Borel spaces. *SIAM J. Control Optim.*, 42:442–468.
[19] Hernández-Lerma, O. and Lasserre, J. B. (1994). Linear programming and average optimality of Markov control processes on Borel spaces-unbounded costs. *SIAM J. Control Optim.*, 32:480–500.
[20] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, New York.
[21] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.
[22] Hernández-Lerma, O. and Romera, R. (2004). Multiobjective Markov control processes: a linear programming approach. *Morfismos*, 8:3–33.
[23] Hordijk, A. and Kallenber, L. C. M. (1979). Linear programming and Markov decision chains. *Manage. Sci.*, 24:352–362.
[24] Hordijk, A. and Kallenber, L. C. M. (1984). Constrained undiscounted stochastic dynamic programming. *Math. Oper. Res.*, 9:112–141.
[25] Hordijk, A. and Kallenber, L. C. M. (1994). Linear programming formulation of MDPs in countable state space: The multichain case. *Z. Oper. Res.*, 40:91–108.
[26] Huang, Y. and Kurano, M. (1997). The LP approach in average rewards MDPs with multiple cost constraints: the countable state case. *J. Infor. Optim. Sci.*, 18:33–47.
[27] Kallenber, L. C. M. (1983). *Linear Programming and Finite Markov Control Problems*. Mathematical Centre, Amsterdam.
[28] Kurano, M. (1989). The existence of a minimum pair of state and policy for Markov decision processes
under the hypothesis of Doeblin. *SIAM J. Control Optim.*, 27:296–307.

[31] Kurano, M., Nakagami, J.-I., and Huang, Y. (2000). Constrained Markov decision processes with compact state and action spaces: The average case. *Optimization*, 48:255–269.

[32] Lasserre, J. B. (1994). Average optimal stationary policies and linear programming in countable space Markov decision processes. *J. Math. Anal. Appl.*, 183:233–249.

[33] Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, 2nd edition.

[34] Parthasarathy, K. R. (1967). *Probability Measures on Metric Spaces*. Academic Press, New York.

[35] Puterman, M. L. (1994). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, New York.

[36] Robertson, A. P. and Robertson, W. (1973). *Topological Vector Spaces*. Cambridge University Press, Cambridge, UK, 2nd edition.

[37] Vega-Amaya, O. (1999). Sample path average cost optimality of Markov control processes with strictly unbounded cost. *Applications Mathematicae*, 26:363–381.

[38] Yamada, K. (1975). Duality theorem in Markovian decision problems. *J. Math. Anal. Appl.*, 50:579–595.

[39] Yu, H. (2019a). On Markov decision processes with Borel spaces and an average cost criterion. [http://arxiv.org/abs/1901.03374](http://arxiv.org/abs/1901.03374) (38 pages).

[40] Yu, H. (2019b). On the minimum pair approach for average-cost Markov decision processes with countable discrete action space and strictly unbounded costs. [http://arxiv.org/abs/1902.10685](http://arxiv.org/abs/1902.10685) (24 pages).