

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

Huizhen Yu\*

## Abstract

We consider the linear programming approach for constrained and unconstrained Markov decision processes (MDPs) under the long-run average cost criterion, where the class of MDPs in our study have Borel state spaces and discrete countable action spaces. 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 a duality gap and other optimality results. Our results do not require a lower-semicontinuous MDP model. Thus, they can be applied to countable action space MDPs where the dynamics and one-stage costs are discontinuous in the state variable. Our proofs make use of the continuity property of Borel measurable functions asserted by Lusin's theorem.

## Keywords:

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

---

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

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Preliminaries</b>	<b>4</b>
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
<b>3</b>	<b>Linear Programming for Average-Cost MDPs</b>	<b>9</b>
3.1	Primal and Dual Linear Programs . . . . .	9
3.2	Optimality Results and Discussion . . . . .	12
<b>4</b>	<b>Extension to Constrained Average-Cost MDPs</b>	<b>14</b>
4.1	Model Assumptions and Existence of Stationary Optimal Pairs . . . . .	15
4.2	Linear Programming Formulation and Optimality Results . . . . .	16
<b>5</b>	<b>Proofs</b>	<b>19</b>
5.1	Proofs for Section 3 . . . . .	19
5.1.1	Proof of Theorem 3.1 . . . . .	20
5.1.2	Proof of Prop. 3.1 . . . . .	25
5.2	Proofs for Section 4 . . . . .	26
5.2.1	Proof of Theorem 4.1 (Outline) . . . . .	26
5.2.2	Proof of Theorem 4.2 (Outline) . . . . .	27
5.2.3	Proofs of Props. 4.1 and 4.2 . . . . .	29
	<b>References</b>	<b>31</b>

# 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 Borel state spaces, discrete countable action spaces, and unbounded one-stage costs.

LP methods for average-cost MDPs have a long history and an extensive literature. For MDPs with finite state and action spaces, see e.g., [9, 24, 25, 28]; for countable state spaces and countable or compact action spaces, see [6, 7, 26, 27, 31]; and for Borel state and action spaces, see [15, 17, 18, 21, 30, 38]. The interested reader may also consult the books [1, 13, 19, 20, 35] and their references. The third group of results deal with uncountably infinite state spaces and are most closely related to our work. In particular, using the theory of infinite-dimensional LP (Anderson and Nash [2]), Hernández-Lerma and Lasserre [18] (see also [20, Chap. 12], [21]) formulated a general LP framework for Borel space average-cost MDPs. They studied the relations between the values of the primal/dual linear programs and the minimum average cost of an MDP, proved the absence of a duality gap under certain continuity conditions on the MDP model, and related the solutions of the programs to stationary optimal policies and average cost optimality equations (ACOE) of the MDP. Much earlier than [18], Yamada [38] considered linear programs for a special class of geometrically ergodic MDPs with compact Euclidean state/action spaces and proved duality results for these problems. Building on the work [18], Hernández-Lerma and González-Hernández [15] provided additional results and generalizations. Extensions of the LP method to constrained average-cost problems were studied by Kurano et al. [30] for compact spaces and by Hernández-Lerma et al. [17] for non-compact spaces.

Another line of research that is closely related to our work, as well as to the prior work on LP mentioned above, is the minimum pair approach for average-cost Borel space MDPs ([14, 29], [19, Chap. 5.7]; see also the related convex analytic approach [6]). With this approach, one considers minimizing the average cost over all policies and initial distributions, and the interest is in the existence of an optimal pair of policy and initial distribution with the following structure. The policy is stationary, and the associated initial distribution is an invariant probability measure of the Markov chain induced by the policy. In this paper we shall call a pair with such a structure a “stationary pair” and if it attains the minimum average cost, a “stationary minimum pair.” The feasibility and solvability of the primal linear programs studied in the prior work mentioned earlier in fact depend on the existence of such pairs. Conversely, a stationary minimum pair, when it exists, can be found by solving a linear program in the space of invariant probability measures induced by stationary policies, thus providing a way to find a stationary optimal policy for a subset of states.<sup>1</sup> In some cases, with further ergodicity and regularity conditions, one can also extend the policy to an optimal one over the entire state space and establish stronger optimality, including sample-path optimality, of the policy [14, 29, 32, 37].

Our work builds upon earlier research on the LP and minimum pair methods for average-cost MDPs mentioned above. In those prior results the action space is more general than the countable action space we deal with in this paper. However, except for [38], all of those results assume a lower-semicontinuous MDP model. Namely, they require the one-stage cost functions to be lower semicontinuous and the state transition stochastic kernels to be (weakly) continuous ([38] involves different continuity conditions; see Remark 3.2 for details). Our work does not require this assumption.

We recently introduced in [40] a majorization condition on the state transition stochastic kernel to deal with Borel space MDPs that do not satisfy such continuity conditions. For the case of countable

---

<sup>1</sup>For finite state and action MDPs, Denardo [8] seems to be the first to recognize the relation between the solution of a certain linear program and a stationary minimum pair, and he proposed to find a stationary average-cost optimal policy in a multichain MDP by repeatedly solving those linear programs on subproblems with smaller state spaces. This procedure is not applicable in general when the state space is uncountably infinite, since the “chain structure” of an MDP in this case can be complicated and hard to analyze. For some results on LP for “multichain” Borel space MDPs, see [15].

action spaces (with the discrete topology), we obtained the existence of a stationary minimum pair and other average-cost optimality results analogous to those for lower-semicontinuous MDPs given by [14, 29, 32, 37]. The purpose of the majorization condition is to make use of Lusin's theorem on the continuity of Borel measurable functions [11, Thm. 7.5.2]. Roughly speaking, 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 to extract arbitrarily large (according to a given finite measure) sets on which certain Borel measurable functions involved in our analysis have desired continuity properties. With this technique, 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, although the application range is currently limited to the case of countable action spaces.

The purpose of this work is to further analyze the implications of the majorization condition and Lusin's theorem in the LP context, for both unconstrained and constrained MDPs. The main contributions of this paper are as follows.

- (i) For unconstrained average-cost MDPs, under the strictly unbounded cost condition and the majorization condition (cf. Assumption 2.1), we prove there is no duality gap between the primal and dual linear programs in an LP formulation (see Theorem 3.1).
- (ii) For constrained average-cost MDPs, under conditions similar to those in (i), we first prove the existence of a stationary optimal pair and a stationary lexicographically optimal pair (which are analogous to stationary minimum pairs for unconstrained MDPs), and we then prove the absence of a duality gap for an LP formulation (see Theorems 4.1 and 4.2, respectively).

In addition, we also discuss the maximizing sequences of dual linear programs and their relation with certain versions of ACOE (see Prop. 3.1 for unconstrained MDPs and Props. 4.1, 4.2 for constrained MDPs). Our results for unconstrained (resp., constrained) MDPs given in this paper can be compared with some of the prior results in [20, Chap. 12] and [21] (resp., [17] and [30]) for lower-semicontinuous models.

While this paper focuses on the average cost criterion, the analysis we give, with minor changes, can also be applied to constrained (or multi-objective) discounted-cost MDPs similar to those studied in [12, 16, 23], for finding constrained optimal or Pareto optimal policies (for a given initial distribution) using the LP approach, in the case of countable action spaces. In a separate recent work [39] based on similar ideas, we introduced another majorization condition for MDPs where both the state and action spaces are Borel, and used the majorization condition instead of the commonly required continuity/compactness conditions to prove the average cost optimality inequalities via the vanishing discount factor approach.

The rest of this paper is organized as follows. In Section 2 we give 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. In Section 3 we present our LP formulation and duality results for unconstrained MDPs. We then extend these results to constrained MDPs in Section 4. Proofs for the theorems in Sections 3 and 4 are given in Section 5.

## 2 Preliminaries

We start with some notations and basic definitions. 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  $\mathcal{B}(X)$ . We will refer to nonnegative or signed measures on  $\mathcal{B}(X)$  as Borel measures. A *Borel space* (a.k.a. *standard Borel space*) is a separable metrizable space that is homeomorphic to a Borel subset of some Polish space (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  $X$*  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. We denote the

stochastic kernel by  $q(dy|x)$ . When it is continuous on  $X$ , we call it a *continuous stochastic kernel* (it is also called *weakly continuous* or *weak Feller* in the literature). 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, we shall also consider other topologies in the next section when these spaces appear in infinite-dimensional linear programs.

We now introduce average-cost MDPs and the minimum pair approach, after which we will briefly review infinite-dimensional linear programs in topological vector spaces.

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

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

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

is assumed to be a *Borel subset* of  $\mathbb{X} \times \mathbb{A}$ . 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 state transition is governed by a *Borel measurable* stochastic kernel  $q(dy|x, a)$  on  $\mathbb{X}$  given  $\mathbb{X} \times \mathbb{A}$ , and that the one-stage cost function  $c : \mathbb{X} \times \mathbb{A} \rightarrow [0, +\infty]$  is *nonnegative and Borel measurable*, real-valued on  $\Gamma$ , and taking the value  $+\infty$  outside  $\Gamma$ .

A policy is a sequence of stochastic kernels on  $\mathbb{A}$  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  $\pi := (\mu_0, \mu_1, \dots)$  where for each  $n \geq 0$ ,  $\mu_n(da_n \mid x_0, a_0, \dots, a_{n-1}, x_n)$  is a Borel measurable stochastic kernel on  $\mathbb{A}$  given  $(\mathbb{X} \times \mathbb{A})^n \times \mathbb{X}$  and obeys the control constraint of the MDP:

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

Such a policy is called *nonrandomized* if in the above every measure on  $\mathbb{A}$  is a Dirac measure, and it is called *stationary* if the function  $(x_0, a_0, \dots, a_{n-1}, x_n) \mapsto \mu_n(da_n \mid x_0, a_0, \dots, a_{n-1}, x_n)$  depends only on the state  $x_n$ , in the same way for every  $n \geq 0$ . In the stationary case, we can write the policy as  $\pi = (\mu, \mu, \dots)$  for a Borel measurable stochastic kernel  $\mu(da \mid x)$  on  $\mathbb{A}$  given  $\mathbb{X}$  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 let  $\Pi_s$  be the subset of all stationary policies in  $\Pi$ . Given that the action space  $\mathbb{A}$  is countable,  $\Pi$  and  $\Pi_s$  are nonempty (see e.g., [40, Sect. 2]), and the Borel measurable policies will be adequate for our purpose—henceforth, we shall simply call them policies. We also note that although  $\mathbb{A}$  is countable, in the above and throughout the paper, we write probability measures on  $\mathbb{A}$  using the general notation for probability measures on a possibly uncountably infinite space, for notational simplicity.

### 2.1.1 Average Cost Criterion and Minimum Pair

In an MDP, a policy  $\pi \in \Pi$  and an initial (state) distribution  $\zeta \in \mathcal{P}(\mathbb{X})$  induce a stochastic process  $\{(x_n, a_n)\}_{n \geq 0}$  on the infinite product of state and action spaces,  $(\mathbb{X} \times \mathbb{A})^\infty$ . 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  $q(dy|x, a)$  [3, Prop. 7.28]. We denote this probability measure by  $\mathbf{P}_\zeta^\pi$  and the corresponding expectation operator by  $\mathbb{E}_\zeta^\pi$ . The long-run expected average cost of the policy  $\pi$  for the initial distribution  $\zeta$  is defined by

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

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, we are especially interested in the types of pairs defined below.

Let  $\rho^*$  be the *minimum average cost* over all policies and initial distributions:

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

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

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

- (a) For a stationary policy  $\mu \in \Pi_s$  and an initial distribution  $p \in \mathcal{P}(\mathbb{X})$ , if  $p$  is an invariant probability measure of the Markov chain induced by  $\mu$  on  $\mathbb{X}$ , 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.* Various terminologies are used in the literature for what we call a stationary pair  $(\mu, p)$ . In the references [17, 18, 20], the policy  $\mu$  is called a “stable policy” if  $J(\mu, p) < \infty$ . In the reference [7], the probability measure  $\gamma(d(x, a)) = \mu(da|x)p(dx)$  is called an “ergodic occupation measure”—we will discuss such measures in Section 3.1.  $\square$

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

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

**Assumption 2.1.**

- (G) For some  $\pi \in \Pi$  and  $\zeta \in \mathcal{P}(\mathbb{X})$ , the average cost  $J(\pi, \zeta) < \infty$ .
- (SU) There exists a nondecreasing sequence of compact sets  $\Gamma_j \uparrow \Gamma$  such that

$$\lim_{j \rightarrow \infty} \inf_{(x, a) \in \Gamma_j^c} c(x, a) = +\infty.$$

- (M) For each compact set  $K \in \{\text{proj}_{\mathbb{X}}(\Gamma_j)\}$ , there exist an open set  $O \supset K$ , a closed set  $D \subset \mathbb{X}$ , and a finite measure  $\nu$  on  $\mathcal{B}(\mathbb{X})$  (all of which can depend on  $K$ ) such that

$$q((O \setminus D) \cap B \mid x, a) \leq \nu(B), \quad \forall B \in \mathcal{B}(\mathbb{X}), (x, a) \in \Gamma, \quad (2.1)$$

where the closed set  $D$  (possibly empty) is such that restricted to  $D \times \mathbb{A}$ , the state transition stochastic kernel  $q(dy \mid x, a)$  is continuous and the one-stage cost function  $c(x, a)$  is lower semicontinuous.<sup>2</sup>

The first two conditions in this assumption are standard: (G) excludes vacuous problems, and (SU) defines the case of *strictly unbounded* one-stage costs. They were used in, e.g., [14, 20, 32, 37] to derive average-cost optimality and LP duality results for lower-semicontinuous MDP models with strictly unbounded costs.

When the function  $c(\cdot)$  is lower semicontinuous, (SU) is equivalent to  $c(\cdot)$  being *inf-compact* on  $\Gamma$ , i.e.,  $E_r := \{(x, a) \in \Gamma \mid c(x, a) \leq r\}$  is compact for all  $r \geq 0$ . In our case, these sets  $E_r$  need not be closed and instead, (SU) is equivalent to  $E_r$  having compact closures. Note also that the set  $\Gamma$  is  $\sigma$ -compact under (SU) and, since  $\text{proj}_{\mathbb{X}}(\Gamma_j) \uparrow \mathbb{X}$ , the space  $\mathbb{X}$  thus must also be  $\sigma$ -compact.

<sup>2</sup>Since  $\mathbb{A}$  is discrete, the continuity condition here means that for each action  $a \in \mathbb{A}$ ,  $q(dy \mid \cdot, a)$  and  $c(\cdot, a)$  are continuous and lower semicontinuous, respectively, on the set  $D$ .

Condition (M) was introduced in our recent work [40]. We use the majorization property required in (M) instead of the lower-semicontinuity model conditions commonly required in the literature. The set  $D$  in (M) is introduced to separate a “continuous part” of the model from the rest, in order to sharpen (M), although this condition can also be used with  $D = \emptyset$ . Condition (M) seems natural for problems where the probability measures  $\{q(\cdot | x, a) \mid (x, a) \in \Gamma\}$  have densities on  $\mathbb{X} \setminus D$  with respect to (w.r.t.) a common  $\sigma$ -finite reference measure and those density functions are bounded uniformly from above. For instance, if  $\mathbb{X} = \mathbb{R}^n$  and the reference measure is the Lebesgue measure, we can take  $\nu$  in (M) to be a multiple of the Lebesgue measure restricted to a bounded open set that contains  $K$ . See [40, Example 3.2 and Remark 3.3] for more specific examples that illustrate situations where (M) is naturally satisfied or cannot be satisfied.

Under the preceding assumption, the following results are proved in [40] by making use of Lusin’s theorem (see [40, Thm. 3.5] for sample-path and other optimality properties of a stationary minimum pair). They are analogous to the prior results for lower-semicontinuous MDPs [14, 20, 29], and they will serve as the starting point for the analyses we present in this paper.

**Theorem 2.1** (optimality of stationary pairs [40, Prop. 3.2, Thm. 3.3]).

*Under Assumption 2.1, the following hold:*

- (i) *For any pair  $(\pi, \zeta) \in \Pi \times \mathcal{P}(\mathbb{X})$  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

We now give a brief overview of topological vector spaces over the real field and infinite-dimensional linear programs in such spaces. The reader is referred to the books [2, 36] for in-depth studies of these subjects, and to the book [20, Chap. 12.2] for a more detailed introduction than ours. Here we shall focus on a few basic concepts and results that will be needed in this paper.

Let  $X$  and  $Y$  be two (real) vector spaces, and let  $\theta$  denote the element zero for both spaces. The pair  $(X, Y)$  is called a *dual pair* if there is a bilinear form  $\langle \cdot, \cdot \rangle : X \times Y \rightarrow \mathbb{R}$  such that

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

We recall that 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.3]). When endowed with the weak topologies given above, each space in a dual pair  $(X, Y)$  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.3]. Convergence in  $X$  under the weak topology  $\sigma(X, Y)$  can be characterized as follows: a net  $\{x_i\}_{i \in \mathcal{I}}$  in  $X$  converges to  $\bar{x} \in X$  if and only if (iff)

$$\langle x_i, y \rangle \rightarrow \langle \bar{x}, y \rangle, \quad \forall y \in Y.$$

We consider equality-constrained linear programs and their dual linear programs in topological vector spaces. The definitions of these programs involve several objects, which we introduce first:

- two dual pairs of vector spaces  $(X, Y)$  and  $(Z, W)$ , with each space endowed with its respective weak topology;
- a linear mapping  $L : X \rightarrow Z$  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  $L^*w$  on  $X$  and is defined by the identity relation (where  $\langle x, L^*w \rangle$  stands for  $(L^*w)(x)$ ):

$$\langle x, L^*w \rangle := \langle Lx, w \rangle, \quad \forall x \in X, w \in W.$$

An important property of  $L$  and  $L^*$  is given by the following proposition:

**Proposition 2.1** ([36, Chap. II, Prop. 12 and its corollary]). *A linear mapping  $L : X \rightarrow Z$  is weakly continuous if and only if  $L^*(W) \subset Y$ . If  $L$  is weakly continuous, so is  $L^*$ .*

This proposition 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  $L^* : W \rightarrow Y$ , one can define the dual of the primal linear program.

Let  $c \in Y$  and  $b \in Z$ . Consider the following equality-constrained primal linear program (P) in the space  $X$  and its dual linear program (P\*) in the space  $W$  (cf. [2, Chap. 3.3]):

$$\begin{aligned} \text{(P)} \quad & \text{minimize} \quad \langle x, c \rangle \\ & \text{subject to} \quad Lx = b, \quad x \in \Lambda. \end{aligned} \tag{2.2}$$

$$\begin{aligned} \text{(P}^*) \quad & \text{maximize} \quad \langle b, w \rangle \\ & \text{subject to} \quad -L^*w + c \in \Lambda^*, \quad w \in W. \end{aligned} \tag{2.3}$$

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  $\inf(\text{P})$  and  $\sup(\text{P}^*)$  denote the values of (P) and (P\*), respectively. The elementary duality theory (cf. [2, Chap. 3.3]) asserts that if (P) and (P\*) are both consistent, then

$$\sup(\text{P}^*) \leq \inf(\text{P}).$$

If the equality  $\sup(\text{P}^*) = \inf(\text{P})$  holds, we say there is *no duality gap*.

There are several sufficient conditions for the absence of a duality gap. For our purpose, one duality theorem—Theorem 2.2 below from [2]—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

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

Let  $\bar{H}$  denote the closure of  $H$  in the weak topology  $\sigma(Z \times \mathbb{R}, W \times \mathbb{R})$  (corresponding to the dual pair  $(Z \times \mathbb{R}, W \times \mathbb{R})$  with the bilinear form  $\langle (z, r), (w, r') \rangle = \langle z, w \rangle + rr'$ ). We call (P) *subconsistent* if there exists some  $r \in \mathbb{R}$  with  $(b, r) \in \bar{H}$ . When (P) is subconsistent, the *subvalue* of (P) is defined by

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

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



**Theorem 2.2** (subconsistency and duality [2, Thm. 3.3]). *(P) is subconsistent with a finite subvalue  $\underline{\rho}$  if and only if (P\*) is consistent with a finite value  $\underline{\rho}$ .*

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 2.1). 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 2.2). We present the LP formulation and our main duality results in Sections 3.1 and 3.2, respectively. (The proofs of the theorems are given in Section 5.)

Our formulation of the primal linear program is the same as that given by the prior work [20, Chap. 12.3]. But our dual program formulation is different; it avoids a condition on the state transition stochastic kernel used in [20, Chap. 12.3], without affecting the desired duality result (cf. Remark 3.3). This LP formulation we present is one instance of a general class of formulations discussed in the prior work [18, Sect. 4]; however, for the sake of completeness, we will give a detailed account of it using the terminologies introduced in Section 2.2.

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

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

For a Borel measure  $\gamma$  on  $\Gamma$ , let  $\hat{\gamma}$  denote the marginal of  $\gamma$  on  $\mathbb{X}$ . To define minimization problems on stationary pairs in an MDP, 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

$$\hat{\gamma}(B) = \int_{\Gamma} q(B | x, a) \gamma(d(x, a)), \quad \forall B \in \mathcal{B}(\mathbb{X}). \quad (3.1)$$

The correspondence is essentially given by

$$\gamma(d(x, a)) = \mu(da | x) p(dx), \quad (3.2)$$

and has the property that

$$J(\mu, p) = \int c d\gamma. \quad (3.3)$$

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

$$p(B) = \int_{\mathbb{X}} \int_{\mathbb{A}} q(B | x, a) \mu(da | x) p(dx), \quad \forall B \in \mathcal{B}(\mathbb{X}). \quad (3.4)$$

This is the same as (3.1) for the probability measure  $\gamma$  given by (3.2), since the marginal of  $\gamma$  is  $\hat{\gamma} = p$  and  $\mu$  obeys the control constraint of the MDP. The equality (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 (3.1), by [3, Cor. 7.27.2], we can decompose  $\gamma$  as in (3.2) with  $p = \hat{\gamma}$  and  $\mu(da | x)$  being a Borel measurable stochastic kernel on  $\mathbb{A}$  given  $\mathbb{X}$  that obeys the control constraint of the MDP. Then, since  $\gamma$  satisfies (3.1), the pair  $(\mu, p)$  with  $p = \hat{\gamma}$  satisfies (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 (3.3).

Due to this correspondence between  $(\mu, p)$  and  $\gamma$ , finding a stationary minimum pair can be expressed as an optimization problem in which one minimizes  $\int c d\gamma$  over the set of probability measures  $\gamma$  that satisfy (3.1) (a.k.a. the set of “ergodic occupation measures” [7]).

Before expressing this optimization problem as a linear 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 function and the constraints. The following definitions are introduced for this purpose. Consider a positive weight function  $w : \Gamma \rightarrow \mathbb{R}_+$ ,

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

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

$$\mathbb{M}_w(\Gamma) := \left\{ \gamma \in \mathbb{M}(\Gamma) \mid \int w d|\gamma| < \infty \right\}$$

where  $|\gamma|$  denotes the total variation of  $\gamma$ . Let  $\mathbb{F}_w(\Gamma)$  be the set of Borel measurable functions  $\phi$  on  $\Gamma$  such that

$$|\phi| \leq \ell w \quad \text{for some } \ell > 0.$$

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

We are now ready to define the primal and dual linear programs for the average-cost MDP. Let us specialize the programs (P) and (P\*) defined in Section 2.2, by identifying the objects involved in those programs as follows:

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

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

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

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

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

$$\gamma \in \mathbb{M}_w^+(\Gamma), \quad \gamma(\Gamma) = 1, \quad \hat{\gamma}(B) = \int_{\Gamma} q(B | x, a) \gamma(d(x, a)), \quad \forall B \in \mathcal{B}(\mathbb{X}), \quad (3.5)$$

where  $\hat{\gamma}$  is the marginal of  $\gamma$  on  $\mathbb{X}$ , as we recall. In other words, in accordance with the earlier discussion, 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

$$L\gamma = b := (1, \theta).$$

Here  $\theta$  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  $L : \mathbb{M}_w(\Gamma) \rightarrow \mathbb{R} \times \mathbb{M}(\mathcal{X})$  with  $L = (L_0, L_1)$  where, for  $\gamma \in \mathbb{M}_w(\Gamma)$ ,

$$L_0\gamma := \gamma(\Gamma), \tag{3.6}$$

$$(L_1\gamma)(B) := \hat{\gamma}(B) - \int_{\Gamma} q(B | x, a) \gamma(d(x, a)), \quad \forall B \in \mathcal{B}(\mathcal{X}). \tag{3.7}$$

- From the identity  $\langle \gamma, L^*(\rho, h) \rangle = \langle L\gamma, (\rho, h) \rangle$ , the adjoint  $L^*$  of  $L$  is given by the linear mapping that maps each  $(\rho, h) \in \mathbb{R} \times \mathbb{F}_b(\mathcal{X})$  to the function

$$L^*(\rho, h)(x, a) := \rho + h(x) - \int_{\mathcal{X}} h(y) q(dy | x, a), \quad (x, a) \in \Gamma. \tag{3.8}$$

Since  $L^*(\mathbb{R} \times \mathbb{F}_b(\mathcal{X})) \subset \mathbb{F}_w(\Gamma)$ , both  $L$  and  $L^*$  are weakly continuous ([36, Chap. II, Prop. 12 and its corollary]; see also Prop. 2.1). 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

$$\rho + h(x) - \int_{\mathcal{X}} h(y) q(dy | x, a) \leq c(x, a), \quad \forall (x, a) \in \Gamma. \tag{3.9}$$

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

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

$$\begin{aligned} \text{(P)} \quad & \text{minimize} \quad \langle \gamma, c \rangle \\ & \text{subject to} \quad L\gamma = (1, \theta), \quad \gamma \in \mathbb{M}_w^+(\Gamma), \end{aligned} \tag{3.10}$$

and

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

As mentioned earlier, our formulation of (P\*) is different from the one given in the book [20, Chap. 12.3]. We will explain the difference and the reason for it in detail in the next subsection (see Remark 3.3).

A few properties of (P) and (P\*) are easy to see. From the relation between stationary pairs and feasible solutions of the primal program (P), it is clear that under Assumption 2.1, the existence of a stationary minimum pair (cf. Theorem 2.1(ii)) ensures that (P) is both consistent and solvable. Moreover, the proof of Theorem 2.1(ii) (cf. [40]) shows that due to the strict unboundedness of the one-stage costs, if  $\{\gamma_n\}$  is a sequence of feasible solutions of (P) with  $\langle \gamma_n, c \rangle \downarrow \inf(\text{P}) = \rho^*$  (such a sequence is called a *minimizing sequence* of (P)), then any subsequence of  $\{\gamma_n\}$  contains a further subsequence that converges to an optimal solution of (P) in the topology of weak convergence (of probability measures). The consistency of the dual program (P\*) is trivial: since  $c \geq 0$ , a feasible solution is given by  $\rho = 0$  and  $h(\cdot) \equiv 0$ . We then have  $0 \leq \sup(\text{P}^*) \leq \inf(\text{P}) = \rho^*$  under Assumption 2.1.

Next, we will address the duality between (P) and (P\*). We will also examine a connection between (P\*) and the ACOE 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  $\rho_n \uparrow \sup(\text{P}^*)$ .

### 3.2 Optimality Results and Discussion

Our main result of this section is the absence of a duality gap stated in part (ii) of the following theorem. It can be compared with the prior result of [20, Chap. 12.3, Thm. 12.3.4] for average-cost lower-semicontinuous MDPs. In our case, without lower-semicontinuity model assumptions, we will use Lusin's theorem together with the majorization property in Assumption 2.1(M) to prove it.

**Theorem 3.1** (consistency and absence of a 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:  $\inf(P) = \sup(P^*) = \rho^*$ .

*Remark 3.1* (about the proof of Theorem 3.1). Besides the differences in assumptions as mentioned above, another difference between our proof of the absence of a duality gap and the proof given in the prior work [20, Chap. 12.3C] is the following. The approach of the latter proof is to show that the set  $H$  defined by (2.4) is weakly closed (i.e.,  $H = \overline{H}$ ). This is a sufficient condition for the absence of a duality gap, but it requires one to show that every point of  $\overline{H}$  is in  $H$ . Our proof uses the duality between the subvalue  $\underline{\rho}$  of (P) and the value of (P\*) asserted in [2, Thm. 3.3] (cf. Theorem 2.2). With this it suffices to show that a single point of  $\overline{H}$ , namely, the point  $(b, \underline{\rho}) = ((1, \theta), \underline{\rho})$ , is in  $H$ . Thus our 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 [40], 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, Thm. 3.3] instead in our proof.  $\square$

*Remark 3.2* (comparison with a duality result in [38]). For compact Euclidean state and action spaces, Yamada proved an LP duality result [38, Thm. 3] under certain continuity and ergodicity conditions on the MDP. His continuity conditions are different from the lower-semicontinuous model assumption we mentioned, but they 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, Condition (A2)]. 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, Conditions (A1), (A4)], in which he required the density function  $p(y | x, a)$  to be bounded away from zero uniformly for all  $(x, a) \in \Gamma$ . Using this condition together with the continuity and other assumptions, he proved the absence of a duality gap [38, Thm. 3]. Both his conditions and his proof arguments are very different from ours.  $\square$

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

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

need not admit a bounded solution  $h$ .

As mentioned earlier, our LP formulation is only an instance of the class of formulations discussed in [18, Sect. 4]. A different dual program (P\*) is studied in [20, Chap. 12.3]. It involves, instead of  $(\mathbb{R} \times \mathbb{M}(\mathbb{X}), \mathbb{R} \times \mathbb{F}_b(\mathbb{X}))$ , the dual pair  $(\mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}), \mathbb{R} \times \mathbb{F}_{w_0}(\mathbb{X}))$ , where the two spaces  $\mathbb{M}_{w_0}(\mathbb{X})$  and

$\mathbb{F}_{w_0}(\mathbb{X})$  are defined similarly to  $\mathbb{M}_w(\Gamma)$  and  $\mathbb{F}_w(\Gamma)$ , respectively: with  $w_0(x) := 1 + \inf_{a \in A(x)} c(x, a)$ ,  $x \in \mathbb{X}$ ,

$$\mathbb{M}_{w_0}(\mathbb{X}) := \{p \in \mathbb{M}(\mathbb{X}) \mid \int w_0 d|p| < \infty\}, \quad \mathbb{F}_{w_0}(\mathbb{X}) := \{h \in \mathbb{F}(\mathbb{X}) \mid |h| < \ell w_0 \text{ for some } \ell > 0\}.$$

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. [20, Chap. 12.3A, Assumption 12.3.1]): for some constant  $k > 0$ ,

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

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 a duality gap between  $(P)$  and  $(P^*)$ .

Also, the use of the dual pair  $(\mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}), \mathbb{R} \times \mathbb{F}_{w_0}(\mathbb{X}))$  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. [20, Chap. 12.4B, Thm. 12.4.2]). This makes it less appealing to us to have the dual pair  $(\mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}), \mathbb{R} \times \mathbb{F}_{w_0}(\mathbb{X}))$  with its extra condition (3.12) 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.  $\square$

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  $\{h_n\}$ , using essentially the same arguments as those for [20, Chap. 12.4B, Thm. 12.4.2(c)]. We include the result in the proposition below, for the sake of completeness. The first part of its condition is satisfied under Assumption 2.1 (Theorem 3.1); the second part of its condition specifies the additional conditions on  $\{h_n\}$  we need. The ACOE (3.14) in the conclusion holds for “almost all” (a.a.) states and in general, it *need not hold for all*  $x \in \mathbb{X}$  (see e.g., [40, Example 3.1]).

**Proposition 3.1** (ACOE for  $p^*$ -a.a. states). *Let  $\{(\rho_n, h_n)\}$  be a maximizing sequence of the dual program  $(P^*)$ , and let  $h^* = \limsup_{n \rightarrow \infty} h_n$ . Suppose that:*

- (i) *a stationary minimum pair  $(\mu^*, p^*)$  exists and  $\inf(P) = \sup(P^*) = \rho^* < +\infty$ ;*
- (ii) *the functions  $h_n$  satisfy that*

$$\int_{\mathbb{X}} |h^*| dp^* < +\infty, \quad \int_{\mathbb{X}} \sup_{n \geq 0} |h_n(y)| q(dy \mid x, a) < +\infty \quad \forall (x, a) \in \Gamma.$$

*Then  $h^*$  is finite everywhere,*

$$\rho^* + h^*(x) \leq c(x, a) + \int_{\mathbb{X}} h^*(y) q(dy \mid x, a) \quad \forall (x, a) \in \Gamma, \quad (3.13)$$

*and for  $p^*$ -a.a.  $x \in \mathbb{X}$ ,*

$$\rho^* + h^*(x) = \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} h^*(y) q(dy \mid x, a) \right\} \quad (3.14)$$

$$= \int_{A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} h^*(y) q(dy \mid x, a) \right\} \mu^*(da \mid x). \quad (3.15)$$

*Remark 3.4.* We discuss briefly a relation between the above ACOE and nonrandomized stationary optimal policies for the average-cost MDP. Firstly, one can find a subset  $\tilde{\mathbb{X}} \subset \mathbb{X}$  with  $p^*(\tilde{\mathbb{X}}) = 1$  and

a Borel measurable function  $f : \mathbb{X} \rightarrow \mathbb{A}$  with  $f(x) \in A(x)$  for all  $x \in \mathbb{X}$ , such that  $\hat{X}$  is absorbing w.r.t.  $f$  and  $f$  attains the minimum in the ACOE (3.14) on  $\hat{X}$ :

$$q(\hat{\mathbb{X}} | x, f(x)) = 1, \quad \rho^* + h^*(x) = c(x, f(x)) + \int_{\hat{\mathbb{X}}} h^*(y) q(dy | x, f(x)), \quad \forall x \in \hat{\mathbb{X}}. \quad (3.16)$$

More specifically, to find such  $\hat{\mathbb{X}}$  and  $f$ , consider the set  $\mathbb{X}'$  with  $p^*(\mathbb{X}') = 1$  on which (3.14)-(3.15) hold, and the Markov chain  $\{x_n\}$  induced by the policy  $\mu^*$  and the initial distribution  $p^*$ . Since  $p^*$  is an invariant probability measure of this Markov chain, one can construct a set  $\hat{\mathbb{X}} \subset \mathbb{X}'$  with  $p^*(\hat{\mathbb{X}}) = 1$  that is absorbing under  $\mu^*$  (see the proof of [22, Lem. 2.2.3(c)] or [33, Prop. 4.2.3(ii)]). Next, based on the relations (3.14)-(3.15) on  $\hat{\mathbb{X}}$ , the desired function  $f$  can be found: this can be done either directly in the special case of a countable action space we have here, or, more generally, by using the Blackwell and Ryll-Nardzewski selection theorem [5, Thm. 2] as discussed in [18, Remark 4.6].

Secondly, for  $\hat{\mathbb{X}}$  and  $f$  satisfying (3.16), one can apply standard arguments to show that under certain conditions, the nonrandomized stationary policy  $f$  is average-cost optimal for all initial states  $x \in \hat{\mathbb{X}}$ . In particular, if  $h^* \geq 0$ , it is straightforward to show that  $f$  is optimal on  $\hat{\mathbb{X}}$ . In more general cases of  $h^*$ , the optimality of  $f$  on  $\hat{\mathbb{X}}$  can be established by imposing further conditions to ensure that for all  $x \in \hat{\mathbb{X}}$ ,  $\mathbb{E}_x^f[|h^*(x_n)|] < \infty$  for  $n \geq 0$  and  $\liminf_{n \rightarrow \infty} n^{-1} \mathbb{E}_x^f[h^*(x_n)] \geq 0$ . (For derivation details, see e.g., the related discussions in [18, Sect. 3] and [19, Chap. 5.2] on canonical triplets.)  $\square$

## 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  $\mathbb{X} \times \mathbb{A}$ :  $c_0, c_1, \dots, c_d$ . 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.  $c_1, \dots, c_d$  within given limits.

More specifically, let  $\kappa := (\kappa_1, \dots, \kappa_d) \geq 0$  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.  $c_i$ ,  $i = 0, 1, \dots, d$ . Define the feasible set of policy and initial distribution pairs by

$$\mathcal{S} := \{(\pi, \zeta) \in \Pi \times \mathcal{P}(\mathbb{X}) \mid J_0(\pi, \zeta) < \infty, J_i(\pi, \zeta) \leq \kappa_i, i = 1, \dots, d\}. \quad (4.1)$$

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. 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. (What we call optimal pairs are called “constrained optimal pairs” in the prior work [30].)

**Definition 4.1** (optimal pairs).

(a) We call  $(\pi^*, \zeta^*) \in \Pi \times \mathcal{P}(\mathbb{X})$  an *optimal pair* for the constrained MDP if

$$(\pi^*, \zeta^*) \in \mathcal{S} \quad \text{and} \quad J_0(\pi^*, \zeta^*) = \rho_c^*.$$

(b) We call an optimal pair  $(\pi^*, \zeta^*)$  *lexicographically optimal* if 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  $\bar{d} \leq d$ ,

$$J_i(\pi^*, \zeta^*) = J_i(\pi, \zeta) \quad \forall i \leq \bar{d} - 1, \quad J_{\bar{d}}(\pi^*, \zeta^*) < J_{\bar{d}}(\pi, \zeta).$$

**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 what follows, we first adapt the strict unboundedness condition (SU) and the majorization condition (M) 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 theorems of this section are collected in Section 5.2.

## 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  $\mathcal{S} \neq \emptyset$ .*

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

$$\lim_{j \rightarrow \infty} \inf_{(x,a) \in \Gamma_j^c} c_i(x,a) = +\infty.$$

(M) *For each compact set  $K \in \{\text{proj}_{\mathbb{X}}(\Gamma_j)\}$ , there exist an open set  $O \supset K$ , a closed set  $D \subset \mathbb{X}$ , and a finite measure  $\nu$  on  $\mathcal{B}(\mathbb{X})$  (all of which can depend on  $K$ ) such that*

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

*where the closed set  $D$  (possibly empty) is such that restricted to  $D \times \mathbb{A}$ , 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 for the unconstrained problem. Condition (G) is to exclude vacuous problems. Condition (SU) is the same as that considered in [17] for the constrained MDP, and it differs from Assumption 2.1(SU) in that here we require *some* one-stage cost function in the constrained problem to be strictly unbounded. Condition (M) is almost identical to Assumption 2.1(M) except that here the closed set  $D$  must be such that on it, *every* one-stage cost function in the constrained problem is lower semicontinuous in the state variable. As before, having a nonempty set  $D$  in the majorization condition (M) sharpens this condition by allowing us to treat a “continuous” part of the model separately from the rest.

Theorem 4.1 below extends our earlier results for MDPs [40, Prop. 3.2, Thm. 3.3] (cf. Theorem 2.1) to constrained MDPs. In particular, its part (i) can be compared with Theorem 2.1(i), and its parts (ii)-(iii) with Theorem 2.1(ii). The proof will only be outlined in Section 5.2, as it is mostly based on the arguments given in [40]—roughly speaking, the present majorization condition allows us to apply the reasoning 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 [17, Thm. 3.2] and [30, the solvability part of Lem. 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 [23].

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

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

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

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

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

*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 [25, Sect. 4, p. 284] for an interesting counterexample (involving a multichain MDP) that is due to Derman [10]. The difference between this known fact and the existence of a stationary optimal pair in Theorem 4.1 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.  $\square$

*Remark 4.2* (pathwise average costs of  $\mu^*$ ). Suppose that in part (ii) or (iii) of Theorem 4.1, the policy  $\mu^*$  induces on  $\mathcal{X}$  a *positive Harris recurrent* Markov chain (see e.g., [33, Chap. 10.1] for definition). Then, by the ergodic properties of such Markov chains and by the same proof of [40, Thm. 3.5(b)], we have that for all initial distributions  $\zeta$ ,  $\mathbf{P}_\zeta^{\mu^*}$ -almost surely,

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

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

## 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  $J_0(\pi, \zeta)$  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 4.1 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

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

where the weight function  $w : \Gamma \rightarrow \mathbb{R}_+$  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,  $(\mathbb{M}_w(\Gamma), \mathbb{F}_w(\Gamma))$  and  $(\mathbb{R}^d, \mathbb{R}^d)$ ; i.e.,

$$\langle (\gamma, \alpha), (\phi, \alpha') \rangle := \langle \gamma, \phi \rangle + \langle \alpha, \alpha' \rangle = \int_{\Gamma} \phi d\gamma + \sum_{i=1}^d \alpha_i \alpha'_i \quad (4.2)$$

for  $\gamma \in \mathbb{M}_w(\Gamma)$ ,  $\phi \in \mathbb{F}_w(\Gamma)$ , and  $\alpha, \alpha' \in \mathbb{R}^d$  (with  $\alpha_i, \alpha'_i$  denoting their  $i$ th components).

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:

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

and

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



Note that if  $\gamma$  is a probability measure associated with some stationary pair  $(\mu, p) \in \mathcal{S}$  via (3.2), then  $\gamma$  is feasible for (P); in particular,  $\langle \gamma, w \rangle \leq 1 + \sum_{i=1}^d \langle \gamma, c_i \rangle < \infty$ , so  $\gamma \in \mathbb{M}_w^+(\Gamma)$ . 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:

$$\begin{aligned} \text{(P)} \quad & \text{minimize} \quad \langle \gamma, c_0 \rangle \\ & \text{subject to} \quad L(\gamma, \alpha) = (1, \theta, \kappa), \quad \gamma \in \mathbb{M}_w^+(\Gamma), \quad \alpha \geq 0 \end{aligned} \quad (4.3)$$

where the linear mapping  $L : \mathbb{M}_w(\Gamma) \times \mathbb{R}^d \rightarrow \mathbb{R} \times \mathbb{M}(\mathbb{X}) \times \mathbb{R}^d$  is given by  $L = (L_0, L_1, L_2)$  with

$$L_0(\gamma, \alpha) := \gamma(\Gamma), \quad (4.4)$$

$$L_1(\gamma, \alpha)(B) := \hat{\gamma}(B) - \int_{\Gamma} q(B \mid x, a) \gamma(d(x, a)), \quad \forall B \in \mathcal{B}(\mathbb{X}), \quad (4.5)$$

$$L_2(\gamma, \alpha) := (\langle \gamma, c_1 \rangle, \dots, \langle \gamma, c_d \rangle) + \alpha, \quad (4.6)$$

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

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

$$(\mathbb{R} \times \mathbb{M}(\mathbb{X}) \times \mathbb{R}^d, \mathbb{R} \times \mathbb{F}_b(\mathbb{X}) \times \mathbb{R}^d),$$

with the bilinear form defined as the sum of the bilinear forms for the three dual pairs,  $(\mathbb{R}, \mathbb{R})$ ,  $(\mathbb{M}(\mathbb{X}), \mathbb{F}_b(\mathbb{X}))$ , 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  $L^* = (L_1^*, L_2^*)$  on  $\mathbb{R} \times \mathbb{F}_b(\mathbb{X}) \times \mathbb{R}^d$  given by

$$L_1^*(\rho, h, \beta)(x, a) := \rho + h(x) - \int_{\mathbb{X}} h(y) q(dy \mid x, a) + \sum_{i=1}^d \beta_i c_i(x, a), \quad (x, a) \in \Gamma, \quad (4.7)$$

$$L_2^*(\rho, h, \beta) := \beta, \quad (4.8)$$

for  $(\rho, h, \beta) \in \mathbb{R} \times \mathbb{F}_b(\mathbb{X}) \times \mathbb{R}^d$ . Clearly,  $L^*(\mathbb{R} \times \mathbb{F}_b(\mathbb{X}) \times \mathbb{R}^d) \subset \mathbb{F}_w(\Gamma) \times \mathbb{R}^d$ , so both linear mappings  $L$  and  $L^*$  are weakly continuous ([36, Chap. II, Prop. 12 and its corollary]; cf. Prop. 2.1). The objective function of (P\*) is

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

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

$$\begin{aligned} \text{(P*)} \quad & \text{maximize} \quad \rho + \sum_{i=1}^d \beta_i \kappa_i \\ & \text{subject to} \quad L^*(\rho, h, \beta) \leq (c_0, 0), \quad \rho \in \mathbb{R}, \quad h \in \mathbb{F}_b(\mathbb{X}), \quad \beta \in \mathbb{R}^d. \end{aligned} \quad (4.9)$$

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

$$\rho + h(x) - \int_{\mathbb{X}} h(y) q(dy \mid x, a) + \sum_{i=1}^d \beta_i c_i(x, a) \leq c_0(x, a), \quad \forall (x, a) \in \Gamma, \quad (4.10)$$

$$\beta \leq 0. \quad (4.11)$$

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 4.1(ii). Moreover, the proof of Theorem 4.1(ii) also shows that any minimizing sequence  $\{(\gamma_n, \alpha_n)\}$  of (P) has a subsequence  $(\gamma_{n_k}, \alpha_{n_k}) \rightarrow (\gamma^*, \alpha^*)$ , where  $(\gamma^*, \alpha^*)$  is an optimal solution of (P) and  $\gamma_{n_k} \rightarrow \gamma^*$  in the topology of weak convergence of probability measures.

The absence of a duality gap is the main result of this section. Its proof, outlined in Section 5.2, uses essentially the same proof arguments for Theorem 3.1(ii), which handle the discontinuous MDP models by making use of Lusin's theorem together with the majorization property in Assumption 4.1(M).

**Theorem 4.2** (consistency and absence of a 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:  $\inf(P) = \sup(P^*) = \rho_c^*$ .

This theorem is comparable with the prior results [17, Thm. 4.4] and [30, Lem. 2.3] on the LP approach for constrained lower-semicontinuous MDPs ([30] considers compact spaces, and [17] non-compact spaces). Besides the differences in model assumptions, our formulation of the dual program (P\*) also differs from that in [17]. The main difference lies in the choice of the spaces  $\mathbb{M}(\mathbb{X})$  and  $\mathbb{F}_b(\mathbb{X})$  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 [17], which is the same condition (3.12) from [20, Chap. 12.3] that we discussed earlier in Remark 3.3. For the same reason as explained in Remark 3.3, the dual program (P\*) as we formulated above need not admit an optimal solution.

For completeness, in the rest of this section, we discuss some solution properties of the dual program (P\*) and derive a version of ACOE for the constrained MDP. Consider a maximizing sequence  $\{(\rho_n, h_n, \beta_n)\}$  of (P\*), i.e., feasible solutions of (P\*) with  $\rho_n + \langle \kappa, \beta_n \rangle \uparrow \sup(P^*)$ . 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:

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

When  $\mathcal{S} \neq \emptyset$ ,  $\mathcal{J}^{(1)}$  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 4.1.** *Suppose Assumption 4.1 hold. Let  $\{(\rho_n, h_n, \beta_n)\}$  be a maximizing sequence of the dual program (P\*). Then the following hold:*

- (i) The sequence  $\{\beta_{n,j}\}_{n \geq 0}$  is bounded for every  $j \in \mathcal{J}^{(0)}$ .
- (ii) For  $1 \leq j \leq d$ ,  $\lim_{n \rightarrow \infty} \beta_{n,j} = 0$  if  $J_j(\mu^*, p^*) < \kappa_j$  for some stationary optimal pair  $(\mu^*, p^*)$  of the constrained MDP.
- (iii) Suppose there exists  $(\pi, \zeta) \in \Pi \times \mathcal{P}(\mathbb{X})$  such that

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

*Then the sequence  $\{\beta_n\}_{n \geq 0}$  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 3.2). So Prop. 4.1(ii) entails the complementarity relation  $\langle \alpha^*, \beta^* \rangle = 0$  for an optimal solution  $(\gamma^*, \alpha^*)$  of (P), if we define  $\beta^* = (\beta_1^*, \dots, \beta_d^*)$  as follows:  $\beta_j^* = \lim_{n \rightarrow \infty} \beta_{n,j}$  if this limit exists, and assign  $\beta_j^*$  an arbitrary number otherwise. Proposition 4.1(iii) gives a sufficient condition under which the  $\mathcal{J}^{(1)}$ -components of  $\{\beta_n\}$  are also bounded—note that this condition involves *non-feasible* policy and initial distribution pairs and is different from the Slater condition  $J_i(\pi, \zeta) < \kappa_i$ ,  $1 \leq i \leq d$ . One exceptional case where Prop. 4.1 is inapplicable is when  $\kappa = 0$ .  $\square$

When  $\{\beta_n\}_{n \geq 0}$  is bounded, as when the condition of Prop. 4.1(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 functions  $h_n$ , we can derive an optimality equation for the constrained MDP that is analogous to the ACOE (3.14) in Prop. 3.1 for the unconstrained MDP. We state this result in the next proposition. It is comparable with the result of [17, Thm. 5.2(b)] for constrained lower-semicontinuous MDPs; in the latter reference, (4.14) is called the “constrained optimality equation.”

**Proposition 4.2** (ACOE for  $p^*$ -a.a. states in the constrained MDP).

Let  $\{(\rho_n, h_n, \beta_n)\}$  be a maximizing sequence of the dual program  $(P^*)$ , and let  $h^* = \limsup_{n \rightarrow \infty} h_n$ . Suppose that:

- (i) a stationary optimal pair  $(\mu^*, p^*)$  exists and  $\inf(P) = \sup(P^*) = \rho_c^* < +\infty$ ;
- (ii) the functions  $h_n$  satisfy that

$$\int_{\mathcal{X}} |h^*| dp^* < +\infty, \quad \int_{\mathcal{X}} \sup_{n \geq 0} |h_n(y)| q(dy | x, a) < +\infty \quad \forall (x, a) \in \Gamma;$$

- (iii) the sequence  $\{\beta_n\}$  converges to some finite  $\beta^*$ .

Then  $h^*$  is finite 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

$$\tilde{\rho}^* + h^*(x) \leq c^*(x, a) + \int_{\mathcal{X}} h^*(y) q(dy | x, a), \quad \forall (x, a) \in \Gamma, \quad (4.13)$$

and for  $p^*$ -a.a.  $x \in \mathcal{X}$ ,

$$\tilde{\rho}^* + h^*(x) = \inf_{a \in A(x)} \left\{ c^*(x, a) + \int_{\mathcal{X}} h^*(y) q(dy | x, a) \right\} \quad (4.14)$$

$$= \int_{a \in A(x)} \left\{ c^*(x, a) + \int_{\mathcal{X}} h^*(y) q(dy | x, a) \right\} \mu^*(da | x). \quad (4.15)$$

## 5 Proofs

This section collects the proofs of the theorems given in Sections 3 and 4.

### 5.1 Proofs for Section 3

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  $p \in \mathcal{P}(X)$ , denoted  $p_n \xrightarrow{w} p$ , if  $\int f dp_n \rightarrow \int f dp$  for all  $f \in \mathcal{C}_b(X)$ . If  $\mathcal{E}$  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  $p \in \mathcal{E}$ , we say that  $\mathcal{E}$  is *tight*.

By Prohorov’s theorem [4, Thm. 6.1], any sequence in a tight family  $\mathcal{E}$  has a further subsequence that converges weakly to a probability measure in  $\mathcal{P}(X)$ . We will use this fact many times in our proofs, for some family  $\mathcal{E} \subset \mathcal{P}(\Gamma)$  that satisfies  $\sup_{\gamma \in \mathcal{E}} \langle \gamma, c \rangle < \infty$ . By the strict unboundedness condition on  $c$  given in Assumption 2.1(SU), such a family  $\mathcal{E}$  must be tight (as can be seen easily from condition (SU) and the definition of tightness).

### 5.1.1 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,  $0 \leq \sup(\text{P}^*) \leq \inf(\text{P}) = \rho^*$ .

We now prove that there is no duality gap between (P) and (P\*). Our approach is to use [2, Thm. 3.3] (cf. Theorem 2.2 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 of (P) is defined as

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

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

$$H := \{(L\gamma, \langle \gamma, c \rangle + r) \mid \gamma \in \mathbb{M}_w^+(\Gamma), r \geq 0\}, \quad (5.1)$$

and  $\overline{H}$  is the closure of  $H$  in the weak topology  $\sigma(\mathbb{R} \times \mathbb{M}(\mathbb{X}) \times \mathbb{R}, \mathbb{R} \times \mathbb{F}_b(\mathbb{X}) \times \mathbb{R})$ . Since (P) and (P\*) are consistent,  $\sup(\text{P}^*)$  is finite and equals the subvalue  $\underline{\rho}$  by [2, Thm. 3.3] (cf. Theorem 2.2). So, to show  $\inf(\text{P}) = \sup(\text{P}^*)$ , we need to prove  $\rho^* = \underline{\rho}$ . In what follows, we will prove that

$$((1, 0), \underline{\rho}) \in H,$$

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

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

$$(L\gamma_i, \langle \gamma_i, c \rangle) \rightarrow ((1, 0), \underline{\rho})$$

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

$$\gamma_i(\Gamma) \rightarrow 1, \quad (5.2)$$

$$\int_{\mathbb{X}} h(x) \hat{\gamma}_i(dx) - \int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy \mid x, a) \gamma_i(d(x, a)) \rightarrow 0, \quad \forall h \in \mathbb{F}_b(\mathbb{X}), \quad (5.3)$$

$$\langle \gamma_i, c \rangle \rightarrow \underline{\rho}. \quad (5.4)$$

In view of (5.2), 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  $\gamma_i(\Gamma) \rightarrow 1$ , 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  $\mathcal{B}(\Gamma)$ :

$$\gamma_i(\Gamma) = 1, \quad \forall i \in \mathcal{I}.$$

**Step (ii):** Next, from the net  $\{\gamma_i\}_{i \in \mathcal{I}}$ , we will extract a *sequence* of probability measures with the property that the convergence in (5.3) holds for a *countable subset* of the functions in  $\mathbb{F}_b(\mathbb{X})$ . We start by defining this subset. It consists of two countable families of functions,  $\hat{\mathcal{C}}_b(\mathbb{X})$  and  $\hat{\mathbb{F}}_b(\mathbb{X})$ . The set  $\hat{\mathcal{C}}_b(\mathbb{X})$  involves continuous bounded functions that will be used to determine if two probability measures on  $\mathbb{X}$  are equal. The set  $\hat{\mathbb{F}}_b(\mathbb{X})$  involves indicator functions of certain sets in  $\mathbb{X}$  that will be important in the subsequent proof to handle the discontinuities in the MDP model by using Lusin's theorem and the majorization property in Assumption 2.1(M). The construction of  $\hat{\mathbb{F}}_b(\mathbb{X})$  will use the arguments we used in the proof of [40, Thm. 3.5(a)]. The precise definitions of these two sets are as follows.

Recall that  $\mathcal{C}_b(\mathbb{X})$  is the set of (real-valued) bounded continuous functions on  $\mathbb{X}$ . Since  $\mathbb{X}$  is metrizable, by [34, Chap. II, Thm. 6.6], there exists a countable set

$$\hat{\mathcal{C}}_b(\mathbb{X}) := \{h_1, h_2, \dots\} \subset \mathcal{C}_b(\mathbb{X})$$

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

$$\int h dp_n \rightarrow \int h dp, \quad \forall h \in \hat{\mathcal{C}}_b(\mathbb{X}).$$

Then by [11, Prop. 11.3.2], for any  $p, p' \in \mathcal{P}(\mathbb{X})$ ,

$$p = p' \iff \int h dp = \int h dp', \quad \forall h \in \hat{\mathcal{C}}_b(\mathbb{X}). \quad (5.5)$$

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

We now define the other countable family  $\hat{\mathbb{F}}_b(\mathbb{X})$  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  $\mathbb{X} \times \mathbb{A}$  (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(SU), 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  $K = \text{proj}_{\mathbb{X}}(\Gamma_j)$ . Let  $F_j := \text{proj}_{\mathbb{A}}(\Gamma_j)$ , the projection of  $\Gamma_j$  on  $\mathbb{A}$ . Then the set  $F_j$  is compact, and since  $\mathbb{A}$  is countable and discrete, this means that the set  $F_j$  is finite.

**Lemma 5.1.** *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  $\mathbb{X}$  such that the following hold:*

- (i)  $\nu_j(\mathbb{X} \setminus B_{j,m,\ell}^1) \leq \ell^{-1}$  and  $\nu_j(\mathbb{X} \setminus B_{j,\ell}^2) \leq \ell^{-1}$ ;
- (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  $q(dy | \cdot, \cdot)$  is continuous.*

*Proof.* This lemma is a consequence of Lusin's theorem (see [11, Thm. 7.5.2]), which asserts that if  $f$  is a Borel measurable function from a topological space  $X$  into a separable metric space  $S$  and  $\nu$  is a closed regular finite Borel measure on  $X$ , then for any  $\delta > 0$ , there is a closed set  $B$  such that  $\nu(X \setminus B) < \delta$  and the restriction of  $f$  to  $B$  is continuous.

We apply this theorem with  $X = \mathbb{X}$  and  $\nu = \nu_j$  for each  $j$  in the lemma. Since  $\mathbb{X}$  is a metrizable topological space, every finite Borel measure is closed regular by [11, Thm. 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  $X = \mathbb{X}$ ,  $S = \mathbb{R}$ ,  $\nu = \nu_j$  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  $\nu_j(\mathbb{X} \setminus E_a) < \delta$  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  $\nu_j(\mathbb{X} \setminus B_{j,m,\ell}^1) \leq \ell^{-1}$  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  $q(dy | x, a)$ , which is a  $\mathcal{P}(\mathbb{X})$ -valued Borel measurable function on  $\mathbb{X} \times \mathbb{A}$ . Specifically, we let  $X = \mathbb{X}$ ,  $S = \mathcal{P}(\mathbb{X})$ ,  $\nu = \nu_j$ , and  $\delta = \ell^{-1}/|F_j|$ . (Since  $\mathbb{X}$  is separable and metrizable, by [3, Prop. 7.20],  $\mathcal{P}(\mathbb{X})$  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  $\nu_j(\mathbb{X} \setminus E_a) < \delta$  and restricted to  $E_a$ ,  $q(dy | \cdot, a)$  is continuous. We then let the desired set  $B_{j,\ell}^2 = \bigcap_{a \in F_j} E_a$ .  $\square$

We group  $(O_j, D_j, \nu_j, B_{j,m,\ell}^1)$ ,  $(O_j, D_j, \nu_j, B_{j,\ell}^2)$  in the preceding proof into two countable collections  $\mathcal{W}_1$  and  $\mathcal{W}_2$ :

$$\mathcal{W}_1 := \{(O_j, D_j, \nu_j, B_{j,m,\ell}^1) \mid j, m, \ell \in \mathbb{Z}_+\}, \quad \mathcal{W}_2 := \{(O_j, D_j, \nu_j, B_{j,\ell}^2) \mid j, \ell \in \mathbb{Z}_+\}.$$

Let  $\mathbb{1}_E$  denote the indicator function for a set  $E$ . Finally, define a countable set  $\hat{\mathbb{F}}_b(\mathbb{X})$  of indicator functions on  $\mathbb{X}$  by

$$\hat{\mathbb{F}}_b(\mathbb{X}) := \{\mathbb{1}_E(\cdot) \mid E = (O \setminus D) \cap B^c \text{ for some } (O, D, \nu, B) \in \mathcal{W}_1 \cup \mathcal{W}_2\}. \quad (5.6)$$

Note that the sets  $E$  in (5.6) 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 5.2.** *There exists a sequence  $\{\gamma_n\}_{n \geq 0} \subset \{\gamma_i\}_{i \in \mathcal{I}}$  such that*

$$\int_{\mathbb{X}} h(x) \hat{\gamma}_n(dx) - \int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy \mid x, a) \gamma_n(d(x, a)) \rightarrow 0, \quad \forall h \in \hat{\mathcal{C}}_b(\mathbb{X}) \cup \hat{\mathbb{F}}_b(\mathbb{X}), \quad (5.7)$$

$$\langle \gamma_n, c \rangle \rightarrow \underline{\rho}. \quad (5.8)$$

*Proof.* Let us order the functions in the countable set  $\hat{\mathcal{C}}_b(\mathbb{X}) \cup \hat{\mathbb{F}}_b(\mathbb{X})$  as  $h_1, h_2, \dots$ . Choose any  $\bar{i}_0 \in \mathcal{I}$  and let  $\gamma_n = \gamma_{\bar{i}_0}$  for  $n = 0$ . For each  $n \geq 1$ , by (5.3)-(5.4), there exists  $\bar{i}_n \in \mathcal{I}$ ,  $\bar{i}_n \geq \bar{i}_{n-1}$  such that for all  $i \geq \bar{i}_n$ ,

$$\left| \int_{\mathbb{X}} h(x) \hat{\gamma}_i(dx) - \int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy \mid x, a) \gamma_i(d(x, a)) \right| \leq n^{-1}, \quad \forall h \in \{h_1, h_2, \dots, h_n\},$$

$$\underline{\rho} - n^{-1} \leq \langle \gamma_i, c \rangle \leq \underline{\rho} + n^{-1}.$$

Let  $\gamma_n = \gamma_{\bar{i}_n}$ . The resulting sequence  $\{\gamma_n\}_{n \geq 0}$  satisfies (5.7)-(5.8).  $\square$

**Step (iii):** Henceforth, we work with the sequence  $\{\gamma_n\}$  of probability measures given by Lemma 5.2. The relation (5.8) 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, Thm. 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  $\gamma_n \xrightarrow{w} \bar{\gamma}$ .

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

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

This gives us a stationary policy  $\bar{\mu}$ . Before we investigate the property of the pair  $(\bar{\mu}, \bar{p})$  in the next step, we need the following majorization property, which will be used to deal with the discontinuities in the MDP model:

**Lemma 5.3.** *For every  $(O, D, \nu, B) \in \mathcal{W}_1 \cup \mathcal{W}_2$ ,*

$$\limsup_{n \rightarrow \infty} \hat{\gamma}_n((O \setminus D) \cap B^c) \leq \nu(B^c), \quad \bar{p}((O \setminus D) \cap B^c) \leq \nu(B^c).$$

*Proof.* For  $(O, D, \nu, B) \in \mathcal{W}_1 \cup \mathcal{W}_2$ , let  $E = (O \setminus D) \cap B^c$  and since the indicator function  $\mathbb{1}_E \in \hat{\mathbb{F}}_b(\mathbb{X})$ , we have, by (5.7) in Lemma 5.2, that

$$\epsilon_n := \left| \hat{\gamma}_n(E) - \int_{\Gamma} q(E \mid x, a) \gamma_n(d(x, a)) \right| \rightarrow 0.$$

We also have, by Assumption 2.1(M),

$$\int_{\Gamma} q(E \mid x, a) \gamma_n(d(x, a)) \leq \int_{\Gamma} \nu(B^c) \gamma_n(d(x, a)) = \nu(B^c).$$

Hence  $\hat{\gamma}_n(E) \leq \nu(B^c) + \epsilon_n$  for all  $n \geq 0$ ; consequently,  $\limsup_{n \rightarrow \infty} \hat{\gamma}_n(E) \leq \nu(B^c)$ .

Now  $\hat{\gamma}_n \xrightarrow{w} \bar{p}$  (since  $\gamma_n \xrightarrow{w} \bar{\gamma}$ ) and  $E$  is an open set (since  $O$  is open and  $D, B$  are closed). Therefore, by [11, Thm. 11.1.1] and the first part of the proof,  $\bar{p}(E) \leq \liminf_{n \rightarrow \infty} \hat{\gamma}_n(E) \leq \nu(B^c)$ .  $\square$

**Step (iv):** We are now ready to prove that  $((1, 0), \underline{\rho}) \in H$ .

**Lemma 5.4.** *The pair  $(\bar{\mu}, \bar{p})$  is a stationary pair with  $J(\bar{\mu}, \bar{p}) = \langle \bar{\gamma}, c \rangle \leq \underline{\rho}$ .*

*Proof outline.* We will only outline the proof, because the arguments for this lemma are essentially the same as those we used in an earlier work to prove the existence and pathwise optimality properties of stationary pairs [40, Sect. 4.1 and Sect. 4.3.1]. By Lemma 5.2, it suffices to prove the inequality

$$\langle \bar{\gamma}, c \rangle \leq \lim_{n \rightarrow \infty} \langle \gamma_n, c \rangle = \underline{\rho} \quad (5.9)$$

and to prove that for all  $h \in \hat{C}_b(\mathbb{X})$ ,

$$\lim_{n \rightarrow \infty} \int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy | x, a) \gamma_n(d(x, a)) = \int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy | x, a) \bar{\gamma}(d(x, a)). \quad (5.10)$$

To see the sufficiency of (5.9) and (5.10), note that (5.10), together with (5.7) in Lemma 5.2 and the fact  $\lim_{n \rightarrow \infty} \int h d\gamma_n = \int h d\bar{p}$  for all  $h \in \hat{C}_b(\mathbb{X})$  (since  $\gamma_n \xrightarrow{w} \bar{p}$ ), will imply that

$$\int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy | x, a) \bar{\gamma}(d(x, a)) = \int_{\mathbb{X}} h \bar{p}(dx), \quad \forall h \in \hat{C}_b(\mathbb{X}).$$

In turn, this will imply that  $\bar{p}$  is identical to the probability measure  $\int_{\Gamma} q(\cdot | x, a) \bar{\gamma}(d(x, a))$  (cf. (5.5)), 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 (5.9) will give us the desired inequality  $J(\bar{\mu}, \bar{p}) = \langle \bar{\gamma}, c \rangle \leq \underline{\rho}$ .

**Proving (5.9):** The proof of (5.9) is essentially the same as that given in [40, Sect. 4, proofs of Lems. 4.3, 4.9]. Below, we sketch the main proof arguments (see the proofs in [40] for the details of each step):

1. To show (5.9), it suffices to show that for each  $m \in \mathbb{Z}_+$ ,

$$\int c^m d\bar{\gamma} \leq \liminf_{n \rightarrow \infty} \int c^m d\gamma_n. \quad (5.11)$$

(In the above, the probability measures  $\bar{\gamma}$  and  $\gamma_n$  are extended from  $\Gamma$  to  $\mathbb{X} \times \mathbb{A}$ , and  $c^m$  is the truncated one-stage cost function  $\min\{c(\cdot), m\}$ , as we recall.)

2. Fix  $m$ . To prove (5.11), consider arbitrarily small  $\epsilon = \delta = \ell^{-1}$ , for some arbitrarily large  $\ell \in \mathbb{Z}_+$ . Assumption 2.1(SU) together with (5.8) in Lemma 5.2 allows us to choose  $j \in \mathbb{Z}_+$  large enough so that for the compact set  $\Gamma_j$  in Assumption 2.1(SU), we have  $\gamma_n(\Gamma_j^c) \leq \epsilon$  for all  $n$  and  $\bar{\gamma}(\Gamma_j^c) \leq \epsilon$ . This in turn allows us to bound  $\int_{\Gamma_j^c} c^m d\gamma_n$  and  $\int_{\Gamma_j^c} c^m d\bar{\gamma}$  by  $m\epsilon$ , an negligible term when we take  $\epsilon \rightarrow 0$ . Consequently, to prove (5.11), we can focus on the integrals of  $c^m$  on the compact set  $\Gamma_j$  and on bounding the difference

$$\int_{\Gamma_j} c^m d\gamma_n - \int_{\Gamma_j} c^m d\bar{\gamma}. \quad (5.12)$$

3. We now handle the term (5.12)—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  $(O, D, \nu, B) := (O_j, D_j, \nu_j, B_{j,m,\ell}^1) \in \mathcal{W}_1$  (cf. the definition of the set  $\mathcal{W}_1$  given in Step (ii)). By the definition of the set  $B_{j,m,\ell}^1$  (cf. Lemma 5.1 in Step (ii)), the function  $c^m$  is continuous on the closed set  $B \times F$ , where  $F = \text{proj}_{\mathbb{A}}(\Gamma_j)$ , and  $\nu(B^c) \leq \delta = \ell^{-1}$ . We handle the continuous part of  $c^m$  separately from the rest of  $c^m$ . Specifically, we first consider the restriction of  $c^m$  to the closed set  $(D \cup B) \times F$ , which is a lower semicontinuous function on  $(D \cup B) \times F$  in

view of the property of  $D$  given in Assumption 2.1(M). We apply the Tietze–Urysohn extension theorem [11, Thm. 2.6.4] to extend this function to a function  $\tilde{c}^m$  on the entire space  $\mathbb{X} \times \mathbb{A}$  that is nonnegative, lower semicontinuous, and also bounded above by  $m$ . Since  $\gamma_n \xrightarrow{w} \bar{\gamma}$ , by [19, Prop. E.2],

$$\liminf_{n \rightarrow \infty} \int \tilde{c}^m d\gamma_n \geq \int \tilde{c}^m d\bar{\gamma}.$$

We then handle the difference between  $c^m$  and  $\tilde{c}^m$ . These two functions differ only outside the set  $(D \cup B) \times F$ . By using the fact  $\nu(B^c) \leq \delta$  and  $O \supset \text{proj}_{\mathbb{X}}(\Gamma_j)$  (cf. Assumption 2.1(M)), the majorization property given in Lemma 5.3, and the bounds  $\int_{\Gamma_j^c} c^m d\gamma_n \leq m\epsilon$ ,  $\int_{\Gamma_j^c} \tilde{c}^m d\bar{\gamma} \leq m\epsilon$  from Step 2, we can calculate that

$$\limsup_{n \rightarrow \infty} \left| \int_{\mathbb{X} \times \mathbb{A}} (c^m - \tilde{c}^m) d\gamma_n \right| \leq m(\delta + \epsilon), \quad \left| \int_{\mathbb{X} \times \mathbb{A}} (c^m - \tilde{c}^m) d\bar{\gamma} \right| \leq m(\delta + \epsilon).$$

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

$$\liminf_{n \rightarrow \infty} \int c^m d\gamma_n \geq \int c^m d\bar{\gamma} - 2m(\delta + \epsilon).$$

By letting  $\ell \rightarrow \infty$  so that  $\delta, \epsilon \rightarrow 0$ , the desired relation (5.11) follows for all  $m \in \mathbb{Z}_+$  and this implies (5.9).

**Proving (5.10):** The proof of (5.10) is similar to the above and essentially the same as that given in [40, Sect. 4, proofs of Lems. 4.4, 4.10]. We outline the main arguments below (see [40] for detailed derivations):

1. Consider an arbitrary  $h \in \hat{C}_b(\mathbb{X})$ . Let  $\epsilon = \delta = \ell^{-1}$ , for some arbitrarily large  $\ell \in \mathbb{Z}_+$ . Proceed as in Step 2 of the proof of (5.9) to choose  $j \in \mathbb{Z}_+$  large enough so that for the compact set  $\Gamma_j$  in Assumption 2.1(SU), we have  $\gamma_n(\Gamma_j^c) \leq \epsilon$  for all  $n$  and  $\bar{\gamma}(\Gamma_j^c) \leq \epsilon$ .
2. Define a function  $\phi(x, a) := \int_{\mathbb{X}} h(y)q(dy | x, a)$  on  $\mathbb{X} \times \mathbb{A}$ . Corresponding to the chosen  $j$  and  $\ell$ , choose the element  $(O, D, \nu, B) := (O_j, D_j, \nu_j, B_{j,\ell}^2) \in \mathcal{W}_2$  and let  $F := \text{proj}_{\mathbb{A}}(\Gamma_j)$ . By the definition of the set  $B_{j,\ell}^2$  (cf. Lemma 5.1 in Step (ii)),  $\nu(B^c) \leq \delta = \ell^{-1}$  and on the closed set  $B \times F$ ,  $q(dy | \cdot, \cdot)$  is continuous. Then, since  $q(dy | \cdot, \cdot)$  is also continuous on the closed set  $D \times \mathbb{A}$  (cf. Assumption 2.1(M)) and  $h$  is a bounded continuous function, we have, by [3, Prop. 7.30], that the function  $\phi$  is continuous on the closed set  $(D \cup B) \times F$ . We now treat the continuous part of  $\phi$  separately: by the Tietze–Urysohn extension theorem [11, Thm. 2.6.4], the restriction of  $\phi$  to  $(D \cup B) \times F$  can be extended to a bounded continuous function  $\tilde{\phi}$  on the entire space  $\mathbb{X} \times \mathbb{A}$ , with  $\|\tilde{\phi}\|_{\infty} \leq \|\phi\|_{\infty} \leq \|h\|_{\infty}$ . Since  $\gamma_n \xrightarrow{w} \bar{\gamma}$ , we have

$$\lim_{n \rightarrow \infty} \int \tilde{\phi} d\gamma_n = \int \tilde{\phi} d\bar{\gamma}.$$

We then handle the difference between  $\phi$  and  $\tilde{\phi}$ . These two functions differ only outside the set  $(D \cup B) \times F$ . By using the fact  $\nu(B^c) \leq \delta$  and  $O \supset \text{proj}_{\mathbb{X}}(\Gamma_j)$  (cf. Assumption 2.1(M)), the majorization property given in Lemma 5.3, and the bounds  $\gamma_n(\Gamma_j^c) \leq \epsilon$ ,  $\bar{\gamma}(\Gamma_j^c) \leq \epsilon$  from Step 1, we can calculate that

$$\limsup_{n \rightarrow \infty} \left| \int_{\mathbb{X} \times \mathbb{A}} (\phi - \tilde{\phi}) d\gamma_n \right| \leq 2\|h\|_{\infty} \cdot (\delta + \epsilon), \quad \left| \int_{\mathbb{X} \times \mathbb{A}} (\phi - \tilde{\phi}) d\bar{\gamma} \right| \leq 2\|h\|_{\infty} \cdot (\delta + \epsilon).$$

3. Finally, putting all the pieces together gives us the bound

$$\limsup_{n \rightarrow \infty} \left| \int \phi d\gamma_n - \int \phi d\bar{\gamma} \right| \leq 4\|h\|_{\infty} \cdot (\delta + \epsilon).$$

By letting  $\ell \rightarrow \infty$  so that  $\delta, \epsilon \rightarrow 0$ , the desired relation (5.10) follows.



The lemma now follows from (5.9)-(5.10), as discussed earlier.  $\square$

By Lemma 5.4,  $((1, \theta), \underline{\rho}) = (L\bar{\gamma}, \langle \bar{\gamma}, c \rangle + \bar{r})$  for  $\bar{r} = \underline{\rho} - \langle \bar{\gamma}, c \rangle \geq 0$ . Thus  $((1, \theta), \underline{\rho}) \in H$  and consequently,  $\underline{\rho} = \rho^*$ . This completes the proof of Theorem 3.1.

### 5.1.2 Proof of Prop. 3.1

The proof is similar to that of [20, Chap. 12.4B, Thm. 12.4.2(c)]. Since  $\{(\rho_n, h_n)\}$  is a maximizing sequence of  $(P^*)$ , for all  $n \geq 0$ ,  $(\rho_n, h_n)$  is feasible for  $(P^*)$ :

$$\rho_n + h_n(x) \leq c(x, a) + \int_{\mathbb{X}} h_n(y) q(dy | x, a), \quad \forall (x, a) \in \Gamma. \quad (5.13)$$

By assumption  $\rho_n \uparrow \rho^*$  and for each  $(x, a) \in \Gamma$ ,  $\int_{\mathbb{X}} \sup_n |h_n(y)| q(dy | x, a) < +\infty$ . The latter implies

$$\limsup_{n \rightarrow \infty} \int_{\mathbb{X}} h_n(x) q(dy | x, a) \leq \int_{\mathbb{X}} \limsup_{n \rightarrow \infty} h_n(x) q(dy | x, a) < +\infty$$

by Fatou's lemma. So, letting  $n \rightarrow \infty$  and taking limit superior on both sides of (5.13), we obtain

$$\rho^* + h^*(x) \leq c(x, a) + \int_{\mathbb{X}} h^*(y) q(dy | x, a) < +\infty, \quad \forall (x, a) \in \Gamma, \quad (5.14)$$

which is the desired inequality and also shows that  $h^*$  is finite everywhere.

Next, we prove the ACOE for  $p^*$ -a.a. states. Since  $(\mu^*, p^*)$  is a stationary minimum pair and  $\int |h^*| dp^* < \infty$  by assumption, we have

$$\begin{aligned} \rho^* &= \int_{\mathbb{X}} \int_{\mathbb{A}} c(x, a) \mu^*(da | x) p^*(dx), \\ -\infty < \int_{\mathbb{X}} h^*(x) dp^* &= \int_{\mathbb{X}} \int_{\mathbb{A}} \int_{\mathbb{X}} h^*(y) q(dy | x, a) \mu^*(da | x) p^*(dx) < +\infty, \end{aligned}$$

and hence

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

This together with (5.14) implies that for  $p^*$ -a.a.  $x \in \mathbb{X}$ ,

$$\rho^* + h^*(x) - \int_{\mathbb{A}} \left\{ c(x, a) + \int_{\mathbb{X}} h^*(y) q(dy | x, a) \right\} \mu^*(da | x) = 0,$$

which in turn implies that for  $p^*$ -a.a.  $x \in \mathbb{X}$ ,

$$\begin{aligned} \rho^* + h^*(x) &= \int_{\mathbb{A}} \left\{ c(x, a) + \int_{\mathbb{X}} h^*(y) q(dy | x, a) \right\} \mu^*(da | x) \\ &\geq \inf_{a \in A(x)} \left\{ c(x, a) + \int_{\mathbb{X}} h^*(y) q(dy | x, a) \right\}. \end{aligned} \quad (5.15)$$

Then, by (5.14), equality must hold in (5.15), and this gives the desired ACOE (3.14) and (3.15). The proof of Prop. 3.1 is now complete.

## 5.2 Proofs for Section 4

### 5.2.1 Proof of Theorem 4.1 (Outline)

The proof of Theorem 4.1 is similar to that of Theorem 2.1 on stationary minimum pairs for an unconstrained MDP. The latter proof is given in our prior work [40, Sect. 4.1, proofs of Prop. 3.2 and Thm. 3.3], and its main arguments have already been explained earlier in the proof of Lemma 5.4. So we will only outline the proof of Theorem 4.1, in order to avoid repetition. We will first state some of our prior results for unconstrained MDPs. We will then directly apply them to the present case of constrained MDPs.

In [40, Sect. 4.1] we considered two kinds of sequences  $\{\gamma_n\} \subset \mathcal{P}(\Gamma)$ . In the first case,  $\{\gamma_n\}$  are the occupancy measures of a policy  $\pi$ , for an initial distribution  $\zeta$  that satisfies  $J(\pi, \zeta) < \infty$ :

$$\gamma_n(B) := \frac{1}{n} \sum_{k=1}^n \mathbf{P}_\zeta^\pi \{(x_k, a_k) \in B\}, \quad B \in \mathcal{B}(\Gamma). \quad (5.16)$$

In the second case,  $\{\gamma_n\}$  corresponds to a sequence of stationary pairs  $(\mu_n, p_n)$  that satisfy  $\sup_n J(\mu_n, p_n) < \infty$ :

$$\gamma_n(d(x, a)) := \mu_n(da | x) p_n(dx). \quad (5.17)$$

In both cases,  $\sup_n \langle \gamma_n, c \rangle < \infty$ , which, together with the strict unboundedness condition in Assumption 2.1(SU), implies that (i)  $\{\gamma_n\}$  is tight and for the compact sets  $\Gamma_j$  in Assumption 2.1(SU), as  $j \rightarrow \infty$ ,  $\gamma_n(\Gamma_j) \rightarrow 0$  uniformly in  $n$ ; and (ii) a weakly convergent subsequence  $\{\gamma_{n_k}\}$  can be extracted from any subsequence of  $\{\gamma_n\}$ :  $\gamma_{n_k} \xrightarrow{w} \bar{\gamma} \in \mathcal{P}(\Gamma)$ . For both cases, the limiting probability measure  $\bar{\gamma}$  is proved to have the following properties, by using (i)-(ii) and the majorization condition in Assumption 2.1(M):

(a)  $\bar{\gamma}$  corresponds to a stationary pair  $(\bar{\mu}, \bar{p}) \in \Delta_s$  (i.e.,  $\bar{\gamma}(d(x, a)) = \bar{\mu}(da | x) \bar{p}(dx)$ ).

(b) The average cost of the pair  $(\bar{\mu}, \bar{p})$  satisfies

$$J(\bar{\mu}, \bar{p}) = \langle \bar{\gamma}, c \rangle \leq \liminf_{k \rightarrow \infty} \langle \gamma_{n_k}, c \rangle. \quad (5.18)$$

We now explain how we can apply these results to prove Theorem 4.1 for the constrained MDP. To prove Theorem 4.1(i), we consider  $\{\gamma_n\}$  defined by (5.16) for a pair  $(\pi, \zeta) \in \mathcal{S}$ . By the feasibility of  $(\pi, \zeta)$ , its average costs are all finite:

$$J_i(\pi, \zeta) = \limsup_{n \rightarrow \infty} \langle \gamma_n, c_i \rangle < \infty, \quad \forall i = 0, 1, \dots, d.$$

Since at least one of the one-stage cost functions  $c_0, c_1, \dots, c_d$  is strictly unbounded by Assumption 4.1(SU), this implies that  $\{\gamma_n\}$  is a tight family of probability measures on  $\Gamma$  and for the compact sets  $\Gamma_j$  in Assumption 4.1(SU), the convergence  $\gamma_n(\Gamma_j) \rightarrow 0$  as  $j \rightarrow \infty$  is uniform in  $n$ . We then proceed as in the unconstrained case to obtain, from a weakly convergent subsequence  $\{\gamma_{n_k}\}$  of  $\{\gamma_n\}$ , the limiting probability measure  $\bar{\gamma}$ . Next, using the majorization condition in Assumption 4.1(M), it follows as before that  $\bar{\gamma}$  has the property (a) given above and gives us a stationary pair  $(\bar{\mu}, \bar{p})$ . Moreover, because 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, (5.18) in the property (b) above now holds with the function  $c$  replaced by every  $c_i$ ; that is

$$J_i(\bar{\mu}, \bar{p}) = \langle \bar{\gamma}, c_i \rangle \leq \liminf_{k \rightarrow \infty} \langle \gamma_{n_k}, c_i \rangle, \quad \forall i = 0, 1, \dots, d.$$

Since  $J_i(\pi, \zeta) = \limsup_{n \rightarrow \infty} \langle \gamma_n, c_i \rangle \geq \liminf_{k \rightarrow \infty} \langle \gamma_{n_k}, c_i \rangle$ , it follows that

$$J_i(\bar{\mu}, \bar{p}) \leq J_i(\pi, \zeta), \quad \forall i = 0, 1, \dots, d.$$

This proves Theorem 4.1(i).

To prove Theorem 4.1(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  $J_0(\mu_n, p_n) \downarrow \rho_c^*$  (there exists such a sequence by the part (i) just proved). Let  $\gamma_n$  be defined as in (5.17). Then, since  $J_i(\mu_n, p_n) = \langle \gamma_n, c_i \rangle$  for all  $0 \leq i \leq d$ , we have

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

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

$$J_i(\mu^*, p^*) = \langle \gamma^*, c_i \rangle \leq \liminf_{k \rightarrow \infty} \langle \gamma_{n_k}, c_i \rangle. \quad (5.19)$$

Since  $\langle \gamma_{n_k}, c_i \rangle = J_i(\mu_{n_k}, p_{n_k})$  and  $(\mu_{n_k}, p_{n_k})$  is feasible for the constrained problem, (5.19) implies

$$J_0(\mu^*, p^*) \leq \rho_c^*, \quad J_i(\mu^*, p^*) \leq \kappa_i, \quad \forall i = 1, 2, \dots, d.$$

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

We now prove Theorem 4.1(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

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

Then  $\mathcal{S} \supset \mathcal{S}_0^* \supset \mathcal{S}_1^* \cdots \supset \mathcal{S}_d^*$  and  $\mathcal{S}_d^*$  consists of all the lexicographically optimal pairs. So, to prove Theorem 4.1(iii), we need to show  $\Delta_s \cap \mathcal{S}_d^* \neq \emptyset$ . By Theorem 4.1(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 4.1(i) proved earlier, for each  $(\pi_n, \zeta_n)$ , there is a stationary pair  $(\mu_n, p_n)$  with

$$J_i(\mu_n, p_n) \leq J_i(\pi_n, \zeta_n), \quad \forall i = 0, 1, \dots, d.$$

This together with the fact  $(\pi_n, \zeta_n) \in \mathcal{S}_{j-1}^*$  implies that  $(\mu_n, p_n) \in \mathcal{S}_{j-1}^*$ . 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 (5.19). Therefore,

$$J_i(\mu^*, p^*) = \kappa_i^*, \quad i = 0, 1, \dots, j,$$

and consequently,  $(\mu^*, p^*) \in \mathcal{S}_j^*$ . This proves  $\Delta_s \cap \mathcal{S}_j^* \neq \emptyset$ ; then, by induction,  $\Delta_s \cap \mathcal{S}_d^* \neq \emptyset$ . Hence there is a stationary lexicographically optimal pair for the constrained MDP.

This completes the proof of Theorem 4.1.

### 5.2.2 Proof of Theorem 4.2 (Outline)

The consistency and solvability of (P) follow from Theorem 4.1(i)-(ii). The consistency of (P\*) is trivial (e.g., let  $\rho = 0, h(\cdot) \equiv 0, \beta = 0$ ). Thus,  $0 \leq \sup(\text{P}^*) \leq \inf(\text{P}) = \rho_c^*$ .

We now prove the absence of a 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, Thm. 3.3] (cf. Theorem 2.2),

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

$$((1, \theta, \kappa), \underline{\rho}) \in H,$$

where the set  $H$  is as defined in (2.4) and, for the case here, is given by

$$H := \{(L(\gamma, \alpha), \langle \gamma, c_0 \rangle + r) \mid \gamma \in \mathbb{M}_w^+(\Gamma), \alpha \in \mathbb{R}_+^d, r \geq 0\}. \quad (5.20)$$

Recall that by definition the subvalue  $\underline{\rho} = \inf \{r \mid ((1, \theta, \kappa), r) \in \bar{H}\}$  (cf. Section 2.2).

To prove  $((1, \theta, \kappa), \underline{\rho}) \in H$ , we will construct a stationary pair  $(\bar{\mu}, \bar{p}) \in \mathcal{S}$  with  $J_0(\bar{\mu}, \bar{p}) \leq \underline{\rho}$ , 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  $((1, \theta, \kappa), \underline{\rho}) \in \bar{H}$  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

$$\gamma_i(\Gamma) \rightarrow 1, \quad (5.21)$$

$$\int_{\mathbb{X}} h(x) \hat{\gamma}_i(dx) - \int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy \mid x, a) \gamma_i(d(x, a)) \rightarrow 0, \quad \forall h \in \mathbb{F}_b(\mathbb{X}), \quad (5.22)$$

$$(\langle \gamma_i, c_1 \rangle, \dots, \langle \gamma_i, c_d \rangle) + \alpha_i \rightarrow \kappa, \quad (5.23)$$

$$\langle \gamma_i, c_0 \rangle \rightarrow \underline{\rho}. \quad (5.24)$$

As before, in view of (5.21) 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 5.2, we extract a *sequence*  $\{(\gamma_n, \alpha_n)\}_{n \geq 0} \subset \{(\gamma_i, \alpha_i)\}_{i \in \mathcal{I}}$  such that

$$\int_{\mathbb{X}} h(x) \hat{\gamma}_n(dx) - \int_{\Gamma} \int_{\mathbb{X}} h(y) q(dy \mid x, a) \gamma_n(d(x, a)) \rightarrow 0, \quad \forall h \in \hat{\mathcal{C}}_b(\mathbb{X}) \cup \hat{\mathbb{F}}_b(\mathbb{X}), \quad (5.25)$$

$$(\langle \gamma_n, c_1 \rangle, \dots, \langle \gamma_n, c_d \rangle) + \alpha_n \rightarrow \kappa, \quad (5.26)$$

$$\langle \gamma_n, c_0 \rangle \rightarrow \underline{\rho}, \quad (5.27)$$

where  $\hat{\mathcal{C}}_b(\mathbb{X})$  and  $\hat{\mathbb{F}}_b(\mathbb{X})$  in (5.25) are two chosen *countable* subsets of  $\mathbb{F}_b(\mathbb{X})$ , the properties of which are needed in the subsequent two steps of our proof. In particular, the set  $\hat{\mathcal{C}}_b(\mathbb{X})$  is the countable set of bounded continuous functions with the property (5.5), the same set as defined in the proof of Theorem 3.1(ii). The countable set  $\hat{\mathbb{F}}_b(\mathbb{X})$  is also defined by the equation (5.6) in that proof:

$$\hat{\mathbb{F}}_b(\mathbb{X}) := \{\mathbb{1}_E(\cdot) \mid E = (O \setminus D) \cap B^c \text{ for some } (O, D, \nu, B) \in \mathcal{W}_1 \cup \mathcal{W}_2\}.$$

However, while the set  $\mathcal{W}_2$  is defined in the same way as before, we define the set  $\mathcal{W}_1$  slightly differently here, to take into account the multiple one-stage cost functions in the constrained MDP. Specifically, in the definition of  $\mathcal{W}_1$  (cf. Lemma 5.1 and the definitions preceding this lemma), we make the following changes. We now use the sets and finite measures  $(O, D, \nu)$  involved in Assumption 4.1(M) instead of Assumption 2.1(M). We choose the sets  $B_{j,m,\ell}^1$  for each  $j, m, \ell \in \mathbb{Z}_+$  such that besides the property in Lemma 5.1(i), we have that restricted to  $B_{j,m,\ell}^1 \times F_j$ , all the  $(d+1)$  truncated one-stage cost functions,  $c_i^m$ ,  $i = 0, 1, \dots, d$ , are continuous (where  $c_i^m(\cdot) = \min\{c_i(\cdot), m\}$ ). 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 (5.26)-(5.27) together with Assumption 4.1(SU) imply that  $\{\gamma_n\}$  is a tight family of probability measures and therefore has a weakly convergent subsequence  $\{\gamma_{n_k}\}$ . Consider the corresponding subsequence  $\{(\gamma_{n_k}, \alpha_{n_k})\}$ ; for notational simplicity, we will drop the subscript  $k$  by redefining  $\{(\gamma_n, \alpha_n)\}$  to be this subsequence. Now, denote

the limit of  $\{\gamma_n\}$  by  $\bar{\gamma}$ , and decompose  $\bar{\gamma}$  as  $\bar{\gamma}(d(x, a)) = \bar{\mu}(da | x) \bar{p}(dx)$ , where  $\bar{p}$  is the marginal of  $\bar{\gamma}$  on  $\mathcal{X}$  and  $\bar{\mu}$  is a stationary policy. Then, using Assumption 4.1(M) instead of Assumption 2.1(M), we have that Lemma 5.3 holds as before, which gives us the desired majorization properties for  $\hat{\gamma}_n$  and  $\bar{p}$  that we will need in the next, last step.

**Step (iv):** This step is almost the same as before, except that we apply those arguments in the proof of (5.9) to *every* cost function  $c_i$ ,  $0 \leq i \leq d$ , in the present constrained problem. Then, similar to Lemma 5.4, we obtain that the pair  $(\bar{\mu}, \bar{p})$  is a stationary pair and satisfies that

$$J_i(\bar{\mu}, \bar{p}) = \langle \bar{\gamma}, c_i \rangle \leq \liminf_{n \rightarrow \infty} \langle \gamma_n, c_i \rangle, \quad \forall i = 0, 1, \dots, d.$$

Combining this with (5.26) and (5.27) (recall also  $\alpha_n \geq 0$ ), we obtain

$$J_0(\bar{\mu}, \bar{p}) \leq \underline{\rho}, \quad J_i(\bar{\mu}, \bar{p}) \leq \kappa_i, \quad \forall i = 0, 1, \dots, d.$$

Therefore, if we let

$$\bar{r} := \underline{\rho} - J_0(\bar{\mu}, \bar{p}) \geq 0, \quad \bar{\alpha} := \kappa - (J_1(\bar{\mu}, \bar{p}), \dots, J_d(\bar{\mu}, \bar{p})) \geq 0,$$

then

$$((1, \theta, \kappa), \underline{\rho}) = (L(\bar{\gamma}, \bar{\alpha}), \langle \bar{\gamma}, c_0 \rangle + \bar{r}) \in H.$$

This implies  $\underline{\rho} = \rho_c^*$  (since it implies  $\rho_c^* \leq \underline{\rho}$ , whereas  $\underline{\rho} \leq \rho_c^*$ ). Hence there is no duality gap between (P) and (P\*).

### 5.2.3 Proofs of Props. 4.1 and 4.2

*Proof of Prop. 4.1.* (i) Consider any  $j \in \mathcal{J}^{(0)}$  and some pair  $(\pi, \zeta) \in \mathcal{S}$  with  $J_j(\pi, \zeta) < \kappa_j$ . By Theorem 4.1(i), there exists a stationary pair  $(\bar{\mu}, \bar{p}) \in \mathcal{S}$  with  $J_i(\bar{\mu}, \bar{p}) \leq J_i(\pi, \zeta)$  for all  $0 \leq i \leq d$ . Then  $J_j(\bar{\mu}, \bar{p}) < \kappa_j$ .

Now for each  $n \geq 0$ , since  $(\rho_n, h_n, \beta_n)$  is feasible for (P\*), we have from (4.10)-(4.11) that  $\beta_n \leq 0$  and for all  $(x, a) \in \Gamma$ ,

$$\rho_n + h_n(x) \leq c_0(x, a) - \sum_{i=1}^d \beta_{n,i} c_i(x, a) + \int_{\mathcal{X}} h_n(y) q(dy | x, a),$$

and therefore, by adding  $\sum_{i=1}^d \beta_{n,i} \kappa_i$  to both sides,

$$\rho_n + \sum_{i=1}^d \beta_{n,i} \kappa_i + h_n(x) \leq c_0(x, a) + \sum_{i=1}^d \beta_{n,i} (\kappa_i - c_i(x, a)) + \int_{\mathcal{X}} h_n(y) q(dy | x, a). \quad (5.28)$$

Integrate both sides of (5.28) w.r.t. the probability measure  $\bar{\gamma}(d(x, a)) = \bar{\mu}(da | x) \bar{p}(dx)$ . Notice that  $\int h_n d\bar{p} = \int_{\Gamma} \int_{\mathcal{X}} h_n(y) q(dy | x, a) \bar{\gamma}(d(x, a))$  since  $(\bar{\mu}, \bar{p})$  is a stationary pair. We thus obtain

$$\rho_n + \sum_{i=1}^d \beta_{n,i} \kappa_i \leq J_0(\bar{\mu}, \bar{p}) + \sum_{i=1}^d \beta_{n,i} (\kappa_i - J_i(\bar{\mu}, \bar{p})). \quad (5.29)$$

Take  $n \rightarrow \infty$ . Since  $\{(\rho_n, h_n, \beta_n)\}$  is a maximizing sequence for (P\*),  $\rho_n + \sum_{i=1}^d \beta_{n,i} \kappa_i \rightarrow \rho_c^*$  by Theorem 4.2(ii). It then follows from (5.29) that

$$\rho_c^* - J_0(\bar{\mu}, \bar{p}) \leq \liminf_{n \rightarrow \infty} \sum_{i=1}^d \beta_{n,i} (\kappa_i - J_i(\bar{\mu}, \bar{p})) \quad (5.30)$$

$$\leq \liminf_{n \rightarrow \infty} \beta_{n,j} (\kappa_j - J_j(\bar{\mu}, \bar{p})), \quad (5.31)$$

where we used the fact  $\kappa_i - J_i(\bar{\mu}, \bar{p}) \geq 0$  and  $\beta_{n,i} \leq 0$  for all  $i$  to derive (5.31). Since  $\kappa_j - J_j(\bar{\mu}, \bar{p}) > 0$ , (5.31) implies  $\liminf_{n \rightarrow \infty} \beta_{n,j} > -\infty$ . Hence the sequence  $\{\beta_{n,j}\}_{n \geq 0}$  is bounded.

(ii) In this case, suppose  $j$  is such that  $J_j(\mu^*, p^*) < \kappa_j$ . Then  $j \in \mathcal{J}^{(0)}$  and (5.31) holds with  $(\bar{\mu}, \bar{p}) = (\mu^*, p^*)$  and with its left-hand side equal to  $\rho_c^* - J_0(\mu^*, p^*) = 0$ . This yields  $\lim_{n \rightarrow \infty} \beta_{n,j} = 0$ .

(iii) In this case, by assumption there is some pair  $(\bar{\pi}, \bar{\zeta}) \in \Pi \times \mathcal{P}(\mathcal{X})$  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 part (i), let us consider a stationary pair  $(\bar{\mu}, \bar{p})$  with  $J_i(\bar{\mu}, \bar{p}) \leq J_i(\bar{\pi}, \bar{\zeta})$  for all  $0 \leq i \leq d$ . Such a pair exists by Theorem 4.1(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  $(\bar{\mu}, \bar{p})$  thus satisfy

$$J_j(\bar{\mu}, \bar{p}) < \kappa_j \quad \forall j \in \mathcal{J}^{(1)}, \quad J_j(\bar{\mu}, \bar{p}) < \infty \quad \forall j \in \mathcal{J}^{(0)} \cup \{0\}. \quad (5.32)$$

We also have, as in part (i), that (5.30) holds for this pair  $(\bar{\mu}, \bar{p})$ . Now, as we proved in part (i),  $\{\beta_{n,i}\}_{n \geq 0}$  is bounded for every  $i \in \mathcal{J}^{(0)}$ . This together with the second relation in (5.32) implies that the term

$$\limsup_{n \rightarrow \infty} \sum_{i \in \mathcal{J}^{(0)}} \beta_{n,i} (\kappa_i - J_i(\bar{\mu}, \bar{p}))$$

is finite. From (5.30), we have the inequality

$$\begin{aligned} \rho_c^* - J_0(\bar{\mu}, \bar{p}) &\leq \liminf_{n \rightarrow \infty} \sum_{i=1}^d \beta_{n,i} (\kappa_i - J_i(\bar{\mu}, \bar{p})) \\ &\leq \limsup_{n \rightarrow \infty} \sum_{i \in \mathcal{J}^{(0)}} \beta_{n,i} (\kappa_i - J_i(\bar{\mu}, \bar{p})) + \liminf_{n \rightarrow \infty} \sum_{i \in \mathcal{J}^{(1)}} \beta_{n,i} (\kappa_i - J_i(\bar{\mu}, \bar{p})). \end{aligned} \quad (5.33)$$

In (5.33), 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

$$\liminf_{n \rightarrow \infty} \sum_{i \in \mathcal{J}^{(1)}} \beta_{n,i} (\kappa_i - J_i(\bar{\mu}, \bar{p})) > -\infty.$$

Then, since  $\beta_n \leq 0$ , in view of the first relation in (5.32), 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 part (i), we obtain that for every  $i = 1, 2, \dots, d$ , the sequence  $\{\beta_{n,i}\}_{n \geq 0}$  is bounded. Hence  $\{\beta_n\}$  is bounded.  $\square$

*Proof of Prop. 4.2.* The proof arguments are similar to those of [17, Thm. 5.2(b)] for constrained MDPs and those of Prop. 3.1 for unconstrained MDPs. By the feasibility of  $\{(\rho_n, h_n, \beta_n)\}$  for (P\*), we have the inequality (5.28); that is, for each  $n \geq 0$ ,

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

Let  $n \rightarrow \infty$ . Since  $\rho_n + \sum_{i=1}^d \beta_{n,i} \kappa_i \uparrow \rho_c^* < \infty$  and  $\beta_n \rightarrow \beta^* \leq 0$  by assumption, we obtain

$$\rho_c^* + \limsup_{n \rightarrow \infty} h_n(x) \leq c_0(x, a) + \sum_{i=1}^d \beta_i^* (\kappa_i - c_i(x, a)) + \limsup_{n \rightarrow \infty} \int_{\mathcal{X}} h_n(y) q(dy | x, a), \quad \forall (x, a) \in \Gamma.$$

For each  $(x, a) \in \Gamma$ , it follows from the assumption  $\int_{\mathbb{X}} \sup_{n \geq 0} |h_n(y)| q(dy | x, a) < +\infty$  and Fatou's lemma that

$$\limsup_{n \rightarrow \infty} \int_{\mathbb{X}} h_n(y) q(dy | x, a) \leq \int_{\mathbb{X}} \limsup_{n \rightarrow \infty} h_n(y) q(dy | x, a) < +\infty.$$

Combining the preceding two relations gives us the desired inequality (4.13):

$$\rho_c^* + h^*(x) \leq c_0(x, a) + \sum_{i=1}^d \beta_i^* (\kappa_i - c_i(x, a)) + \int_{\mathbb{X}} h^*(y) q(dy | x, a), \quad \forall (x, a) \in \Gamma, \quad (5.34)$$

which also shows that  $h^*$  is finite everywhere.

Next, corresponding to the stationary optimal pair  $(\mu^*, p^*)$ , let  $\gamma^*(d(x, a)) = \mu^*(da | x) p^*(dx)$  and integrate both sides of (5.34) w.r.t. the probability measure  $\gamma^*$ . As in the proof of Prop. 3.1, here the integrability is ensured by our assumption  $\int |h^*| dp^* < \infty$  and the invariance property of  $p^*$ , which also imply that  $-\infty < \int_{\mathbb{X}} h^*(x) dp^* = \int_{\Gamma} \int_{\mathbb{X}} h^*(y) q(dy | x, a) \gamma^*(d(x, a)) < +\infty$ . We thus obtain

$$\rho_c^* \leq J_0(\mu^*, p^*) + \sum_{i=1}^d \beta_i^* (\kappa_i - J_i(\mu^*, p^*)).$$

But  $J_0(\mu^*, p^*) = \rho_c^*$  and the second term in the right-hand side above is nonpositive, so equality must hold in the above inequality. This result can be equivalently expressed as

$$\int_{\mathbb{X}} \int_{\mathbb{A}} \left\{ \rho_c^* + h^*(x) - c_0(x, a) - \sum_{i=1}^d \beta_i^* (\kappa_i - c_i(x, a)) - \int_{\mathbb{X}} h^*(y) q(dy | x, a) \right\} \mu^*(da | x) p^*(dx) = 0.$$

Similarly to the proof of Prop. 3.1, the preceding equality together with the inequality (5.34) implies that for  $p^*$ -a.a.  $x \in \mathbb{X}$ ,

$$\begin{aligned} \rho_c^* - \sum_{i=1}^d \beta_i^* \kappa_i + h^*(x) &= \int_{a \in A(x)} \left\{ c_0(x, a) - \sum_{i=1}^d \beta_i^* c_i(x, a) + \int_{\mathbb{X}} h^*(y) q(dy | x, a) \right\} \mu^*(da | x) \\ &= \inf_{a \in A(x)} \left\{ c_0(x, a) - \sum_{i=1}^d \beta_i^* c_i(x, a) + \int_{\mathbb{X}} h^*(y) q(dy | x, a) \right\}. \end{aligned}$$

This gives the desired ACOE (4.14) and (4.15).  $\square$

## Acknowledgments

The author would like to thank Professor Eugene Feinberg and the anonymous reviewer for their comments that helped her improve the paper, and Dr. Martha Steenstrup for reading parts of the paper and giving her advice on improving the presentation. This research was supported by grants from DeepMind, Alberta Machine Intelligence Institute (AMII), and Alberta Innovates—Technology Futures (AITF).

## 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. (1988). A convex analytic approach to MDPs. *Probab. Th. Rel. Fields*, 78:583–602.
- [7] Borkar, V. S. (1994). Ergodic control of Markov chains with constraints—the general case. *SIAM J. Control Optim.*, 32:176–186.
- [8] Denardo, E. V. (1970). On linear programming in a Markov decision problem. *Manag. Sci.*, 16:281–288.
- [9] Denardo, E. V. and Fox, B. L. (1968). Multichain Markov renewal programs. *SIAM J. Appl. Math.*, 16:468–487.
- [10] Derman, C. (1963). Stable sequential control rules and Markov chains. *J. Math. Anal. Appl.*, 6:257–265.
- [11] Dudley, R. M. (2002). *Real Analysis and Probability*. Cambridge University Press, Cambridge, UK.
- [12] Feinberg, E. A. and Shwartz, A. (1996). Constrained discounted dynamic programming. *Math. Oper. Res.*, 21:922–945.
- [13] Feinberg, E. A. and Shwartz, A., editors (2002). *Handbook of Markov Decision Processes: Methods and Applications*. Springer Science+Business Media, New York.
- [14] Hernández-Lerma, O. (1993). Existence of average optimal policies in MCPs with strictly unbounded costs. *Kybernetika*, 29:1–17.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] Hernández-Lerma, O. and Lasserre, J. B. (1996). *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. Springer, New York.
- [20] Hernández-Lerma, O. and Lasserre, J. B. (1999). *Further Topics on Discrete-Time Markov Control Processes*. Springer, New York.
- [21] Hernández-Lerma, O. and Lasserre, J. B. (2002). The linear programming approach. In Feinberg, E. A. and Shwartz, A., editors, *Handbook of Markov Decision Processes: Methods and Applications*, chapter 12. Springer Science+Business Media, New York.
- [22] Hernández-Lerma, O. and Lasserre, J. B. (2003). *Markov Chains and Invariant Probabilities*. Birkhäuser Verlag, Basel.
- [23] Hernández-Lerma, O. and Romera, R. (2004). Multiobjective Markov control processes: A linear programming approach. *Morfismos*, 8:1–33.
- [24] Hordijk, A. and Kallenberg, L. C. M. (1979). Linear programming and Markov decision chains. *Manag. Sci.*, 24:352–362.
- [25] Hordijk, A. and Kallenberg, L. C. M. (1984). Constrained undiscounted stochastic dynamic programming. *Math. Oper. Res.*, 9:112–141.
- [26] Hordijk, A. and Lasserre, J. B. (1994). Linear programming formulation of MDPs in countable state space: The multichain case. *Z. Oper. Res.*, 40:91–108.
- [27] 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.
- [28] Kallenberg, L. C. M. (1983). *Linear Programming and Finite Markov Control Problems*. Mathematical Centre, Amsterdam.
- [29] 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.
- [30] 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.
- [31] Lasserre, J. B. (1994). Average optimal stationary policies and linear programming in countable space Markov decision processes. *J. Math. Anal. Appl.*, 183:233–249.
- [32] Lasserre, J. B. (1999). Sample-path average optimality for Markov control processes. *IEEE Trans. Automat. Contr.*, 44(10):1966–1971.
- [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. (2020a). Average cost optimality inequality for Markov decision processes with Borel spaces and universally measurable policies. *SIAM J. Control Optim.*, 58(4):2469–2502.
  - [40] Yu, H. (2020b). On the minimum pair approach for average-cost Markov decision processes with countable discrete action spaces and strictly unbounded costs. *SIAM J. Control Optim.*, 58(2):660–685.