Title of the book!

Onésimo Hernández-Lerma
Depto. de Matemáticas
CINVESTAV-IPN
Apdo. Postal 14-740
Mexico, D.F. 07000
Mexico
email: ohernand@math.cinvestav.mx

Jean B. Lasserre
LAAS-CNRS
7 Avenue du Colonel Roche
31077 Toulouse Cédex 4
France
email: lasserre@laas.fr

July 29, 1999

Chapter 1

The linear programming approach

Abstract

Abstract of the paper

1.1 Introduction

In this chapter we study the *linear programming* (LP) approach to Markov decision problems and our ultimate goal is to show how a Markov decision problem (MDP) can be approximated by *finite* linear programs.

The LP approach to Markov decision problems dates back to the early sixties with the pioneering work of De Ghellinck [10], d'Epenoux [11] and Manne [26] for MDPs with finite sate and action spaces. Among later contributions for finite or countable state and action MDPs let us mention Altman [1], Borkar [8], [9], Denardo [12], Kallenberg [24], Hordijk and Kallenberg [22], Hordijk and Lasserre [23], Lasserre [25], and for MDPs in infinite dimensional spaces and in discrete or continuous time, Bhatt and Borkar [7], Haneveld [13], Heilmann [14], [15], Hernández-Lerma [16], Hernandez-Lerma and Lasserre [17], [19], Mendiondo and Stockbridge[27], Stockbridge [32], Yamada [35]...

Among the nice features of the LP approach, the most evident is that it is valid in a very general context. For instance, for the long-run expected average cost (AC) problem, one does not need to assume that the Average Cost Optimality Equation (ACOE) holds, a restrictive assumption. Also, to our knowledge, it is the only approach that permits to handle constrained MDPs in a very natural way. Finally, it is possible to devise simple convergent numerical approximation schemes that require to solve finite LPs for which efficient codes are now available.

Let us briefly outline one simple way to see how the LP approach can be naturally introduced, although it was not the idea underlying the first papers on the LP approach to MDPs. The starting point is to observe that given a policy $\pi \in \Pi$, an initial distribution ν and a one-step cost function $c : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$, the finite-horizon functional

$$J(\nu, \pi, N, 1, c) \;:=\; N^{-1} E_\nu^\pi \sum_{t=0}^{N-1} c(x_t, a_t),$$

can be written as a *linear* functional $\int c d\mu_N^{\pi,\nu}$ with $\mu_N^{\pi,\nu}$ the expected (state-action) occupation measure

$$\mu_N^{\pi,\nu}(B) \;:=\; N^{-1} E_\nu^\pi \sum_{t=0}^{N-1} 1\{(x_t, a_t) \in B\}, \qquad B \in \mathcal{B}(\mathbb{X} \times \mathbb{A}).$$

Under some conditions, and with some limiting arguments as $N \to \infty$, one may show that, for instance, minimizing the long-run expected average cost criterion (the AC problem) reduces to solving a linear program. More precisely, the AC problem reduces to minimize the linear criterion $\int c d\mu$ over a set of probability measures μ on $\mathbb{X}\times\mathbb{A}$ that satisfy some linear "invariance" constraints involving the transition kernel P. This approach for MDPs is of course related to the Birkhoff Individual Ergodic Theorem (for noncontrolled Markov chains) which states that given an homogeneous Markov chain X_t, $t = 0, 1, \ldots$ on $\mathbb{X}$, a cost function $c : \mathbb{X} \to \mathbb{R}$, and under some conditions,

$$\lim_{N\to\infty} N^{-1} E_\nu \sum_{t=0}^{N-1} c(X_t) \;=\; \int c d\mu^\nu,$$

for some invariant probability measure μ^ν.

However, we should note that the first papers on the LP approach to MDPs used a different (in fact, dual) approach. Namely, the LP formulation was a rephrasing of the average (or discounted)-cost optimality equations. We briefly discuss this approach in Remark 13 that yields a dual linear program.

Although the LP approach is valid for several criteria, including the N-step expected total cost, the infinite-horizon expected discounted cost, the long-run expected average cost, constrained discounted and average cost problems, we have chosen to illustrate the LP approach with the AC problem. With ad hoc modifications and appropriate assumptions, the reader would easily deduce the corresponding linear programs associated with the other mentioned problems.

We shall first proceed to find a suitable linear program associated to the Markov decision problem. Here, by a "suitable" linear program we mean a linear program (P) that together with its dual (P*) satisfies that

$$\sup(\mathrm{P}^*) \le (\mathrm{MDP})^* \le \inf(\mathrm{P}), \tag{1.1}$$

where (using terminology specified in the following section)

$$\begin{aligned} \inf(\mathrm{P}) \;&:=\; \text{value of the primal program (P)},\\ \sup(\mathrm{P}^*) \;&:=\; \text{value of the dual program (P}^*\text{)},\\ (\mathrm{MDP})^* \;&:=\; \text{value function of the Markov decision problem.} \end{aligned}$$

In particular, if there is *no duality gap* for (P), so that

$$\sup(\mathrm{P}^*) = \inf(\mathrm{P}), \tag{1.2}$$

then of course the values of (P) and of (P*) yield the desired value function (MDP)*.

However, to find an *optimal policy* for the Markov decision problem, (1.1) and (1.2) are not sufficient because they do not guarantee that (P) or (P*) are *solvable*. If it can be ensured that, say, the primal (P) is solvable—in which case we write its value as min (P)—and that

$$\min(\mathrm{P}) = (\mathrm{MDP})^*, \tag{1.3}$$

then an optimal solution for (P) can be used to determine an optimal policy for the Markov decision problem. Likewise, if the dual (P*) is solvable and its value—which in this case is written as max (P*)—satisfies

$$\max(\mathrm{P}^*) = (\mathrm{MDP})^*, \tag{1.4}$$

then we can use an optimal solution for (P*) to find an optimal policy for the Markov decision problem. In fact, one of the main results in this chapter (Theorem 16) gives conditions under which (1.3) and (1.4) are both satisfied, so that in particular *strong duality* for (P) holds, that is,

$$\max(\mathrm{P}^*) = \min(\mathrm{P}). \tag{1.5}$$

Section 1.2 presents background material. It contains, in particular, a brief introduction to infinite LP. In Section 1.3 we introduce the program (P) associated to the AC problem, and we show that (P) is solvable and that there is no duality gap, so that (1.2) becomes

$$\sup(\mathrm{P}^*) = \min(\mathrm{P}).$$

Section 1.4 deals with approximating sequences for (P) and its dual (P*). In particular, it is shown that if a suitable maximizing sequence for (P*) exists, then the strong duality condition (1.5) is satisfied. Section 1.5 presents an approximation scheme for (P) using finite-dimensional programs. The scheme consists of three main steps. In step 1 we introduce an "increasing" sequence of *aggregations* of (P), each one with finitely many constraints. In step 2 each aggregation is *relaxed* (from an equality to an inequality), and, finally, in step 3, each aggregation-relaxation is combined with an *inner approximation* that has a finite number of decision variables. Thus the resulting aggregation-relaxation-inner approximation turns out to be a finite linear program, that is, a program with finitely many constraints and decision variables. The corresponding convergence theorems are stated without proof, and the reader is referred to [19] and [20] for proofs and further technical details. These approximation schemes can be extended to a very general class of infinite-dimensional linear programs (as in [18]), not necessarily related to MDPs.

1.2 Linear Programming in infinite-dimensional spaces

The material is divided into four subsections. The first two subsections review some basic definitions and facts related to dual pairs of vector spaces and linear operators whereas the last two subsections summarize the main results on infinite LP needed in later sections.

1.2.1 Dual pairs of vector spaces

Let $\mathcal{X}$ and $\mathcal{Y}$ be two arbitrary (real) vector spaces, and let $\langle\cdot,\cdot\rangle$ be a **bilinear form** on $\mathcal{X}\times\mathcal{Y}$, that is, a real-valued function on $\mathcal{X}\times\mathcal{Y}$ such that

- the map $x\mapsto\langle x,y\rangle$ is linear on $\mathcal{X}$ for every $y\in\mathcal{Y}$, and
- the map $y\mapsto\langle x,y\rangle$ is linear on $\mathcal{Y}$ for every $x\in\mathcal{X}$.

Then the pair $(\mathcal{X},\mathcal{Y})$ is called a **dual pair** if the bilinear form "separates points" in x and y, that is,

- for each $x\neq 0$ in $\mathcal{X}$ there is some $y\in\mathcal{Y}$ with $\langle x,y\rangle\neq 0$, and
- for each $y\neq 0$ in $\mathcal{Y}$ there is some $x\in\mathcal{X}$ with $\langle x,y\rangle\neq 0$.

If $(\mathcal{X},\mathcal{Y})$ is a dual pair, then so is $(\mathcal{Y},\mathcal{X})$.

If $(\mathcal{X}_1,\mathcal{Y}_1)$ and $(\mathcal{X}_2,\mathcal{Y}_2)$ are two dual pairs of vector spaces with bilinear forms $\langle\cdot,\cdot\rangle_1$ and $\langle\cdot,\cdot\rangle_2$, respectively, then the product $(\mathcal{X}_1\times\mathcal{X}_2,\mathcal{Y}_1\times\mathcal{Y}_2)$ is endowed with the bilinear form

$$\langle(x_1,x_2),(y_1,y_2)\rangle := \langle x_1,y_1\rangle_1+\langle x_2,y_2\rangle_2. \tag{1.6}$$

For MDPs, a typical dual pair of vector spaces is the following. Let S be a Borel space with Borel σ-algebra $\mathcal{B}(S)$, and let $\mathcal{X}:=\mathbb{M}(S)$ be the normed linear space of finite signed measures μ on $\mathcal{B}(S)$, with finite *w-norm*

$$\|\mu\|_w := \int_S w\,d|\mu|, \tag{1.7}$$

for some weight function $w \geq 1$. Now, let $\mathcal{Y} := \mathbb{B}_w(S)$ be the normed linear space of real-valued measurable functions on S with finite w-norm

$$\|u\|_w := \sup_S |u(s)|/w(s). \tag{1.8}$$

Then, the dual pair $(\mathcal{X}, \mathcal{Y}) = (\mathbb{M}_w(S), \mathbb{F}_w(S))$ endowed with the bilinear form

$$\langle \mu, u \rangle := \int_S u d\mu \tag{1.9}$$

is easily seen to be a dual pair. Moreover, by (1.6), the bilinear form corresponding to the dual pair $(\mathbb{R}^n \times \mathbb{M}_w(S), \mathbb{R}^n \times \mathbb{F}_w(S))$ is

$$\langle (x, \mu), (y, u) \rangle = x \cdot y + \langle \mu, u \rangle \tag{1.10}$$

where $x \cdot y := \sum_{i=1}^n x_i y_i$ denotes the usual scalar product of n-vectors.

Given a dual pair $(\mathcal{X}, \mathcal{Y})$, we denote by $\sigma(\mathcal{X}, \mathcal{Y})$ the **weak topology** on $\mathcal{X}$ (also referred to as the **σ-topology** on $\mathcal{X}$), namely, the coarsest—or weakest—topology on $\mathcal{X}$ under which all the elements of $\mathcal{Y}$ are continuous when regarded as linear forms $\langle \cdot, y \rangle$ on $\mathcal{X}$. Equivalently, the base of neighborhoods of the origin of the σ-topology is the family of all sets of the form

$$N(I, \varepsilon) := \{x \in \mathcal{X} | \langle x, y \rangle| \leq \varepsilon \quad \forall y \in I\}, \tag{1.11}$$

where $\varepsilon > 0$ and I is a *finite* subset of $\mathcal{Y}$. (See, for instance, Robertson and Robertson [30], p. 32.) In this case, if $\{x_n\}$ is a sequence or a net in $\mathcal{X}$, then x_n *converges to* x *in the weak topology* $\sigma(\mathcal{X}, \mathcal{Y})$ if

$$\langle x_n, y \rangle \to \langle x, y \rangle \quad \forall y \in \mathcal{Y}. \tag{1.12}$$

For instance, for the dual pair $(\mathbb{M}_w(S), \mathbb{F}_w(S))$, a sequence or a net of measures μ_n converges to μ in the weak topology $\sigma(\mathbb{M}_w(S), \mathbb{F}_w(S))$ if

$$\langle \mu_n, u \rangle \to \langle \mu, u \rangle \quad \forall u \in \mathbb{F}_w(S), \tag{1.13}$$

where $\langle \cdot, \cdot \rangle$ stands for the bilinear form in (1.9).

Remark 1 *(a) Let $(\mathcal{X}, \mathcal{Y})$ be a dual pair such that $\mathcal{Y}$ is a Banach space and $\mathcal{X} = \mathcal{Y}^*$ is the topological dual of $\mathcal{Y}$. In this case, the weak topology $\sigma(\mathcal{X}, \mathcal{Y})$ is called the* **weak*** *(weak-star)* **topology** *on $\mathcal{X}$, and so (1.12) is referred to as the* **weak* convergence** *of x_n to x.*

(b) For instance, with S a locally compact separable metric (LCSM) space, let $\mathcal{X} := \mathbb{M}(S)$ be the Banach space of finite signed measures on S, endowed with the total variation norm $\|\mu\|_{TV} = |\mu|(S)$, and let $\mathcal{Y} := C_0(S)$ be the (separable) Banach space of continuous functions that vanish at infinity, equipped with the sup-norm. By the Riesz Representation Theorem (see, for example, Rudin [31]), $\mathbb{M}(S)$ is the topological dual of $C_0(S)$, and so the weak topology $\sigma(\mathbb{M}(S), C_0(S))$ on $\mathbb{M}(S)$ is in fact the weak topology.*

Definition 2 *Let $(\mathcal{X}, \mathcal{Y})$ and $(\mathcal{Z}, \mathcal{W})$ be two dual pairs of vector spaces, and $G : \mathcal{X} \to \mathcal{Z}$ a linear map.*

(a) G is said to be **weakly continuous** *if it is continuous with respect to the weak topologies $\sigma(\mathcal{X}, \mathcal{Y})$ and $\sigma(\mathcal{Z}, \mathcal{W})$; that is, if $\{x_n\}$ is a net in $\mathcal{X}$ such that $x_n \to x$ in the weak topology $\sigma(\mathcal{X}, \mathcal{Y})$ [see (1.12)], then $Gx_n \to Gx$ in the weak topology $\sigma(\mathcal{Z}, \mathcal{W})$, i.e.,*

$$\langle Gx_n, v\rangle \to \langle Gx, v\rangle \quad \forall v \in \mathcal{W}. \tag{1.14}$$

(b) The **adjoint** *G^* of G is defined by the relation*

$$\langle Gx, v\rangle = \langle x, G^* v\rangle \quad \forall x \in \mathcal{X},\ v \in \mathcal{W}. \tag{1.15}$$

The following proposition gives a well-known (easy-to-use) criterion for the map G in Definition 2 to be weakly continuous—for a proof see, for instance, Robertson and Robertson [30], p. 38.

Proposition 3 *The linear map G is weakly continuous if and only if its adjoint G^* maps $\mathcal{W}$ into $\mathcal{Y}$, that is, $G^*(\mathcal{W}) \subset \mathcal{Y}$.*

Positive and dual cones. (a) Let $(\mathcal{X}, \mathcal{Y})$ be a dual pair of vector spaces, and K a *convex cone* in $\mathcal{X}$, that is, $x + x'$ and λx belong to K whenever x and x' are in K and $\lambda > 0$. Unless explicitly stated otherwise, we shall assume that K is not the whole space, that is, $K \neq \mathcal{X}$, and that the origin (that is, the zero vector, 0) is in K. In this case, K defines a partial order $\geq$ on X such that

$$x \geq x' \;\Leftrightarrow\; x - x' \in K,$$

and K will be referred to as a *positive cone*. The *dual cone* of K is the convex cone K^* in $\mathcal{Y}$ defined by

$$K^* := \{y \in \mathcal{Y} \mid \langle x, y\rangle \geq 0 \quad \forall x \in K\}. \tag{1.16}$$

(b) If $\mathcal{X} = \mathbb{M}_w(S)$, we will denote by $\mathbb{M}_w(S)_+$ the "natural" *positive cone in* $\mathbb{M}_w(S)$, which consists of all the *nonnegative* measures in $\mathbb{M}_w(S)$, that is,

$$\mathbb{M}_w(S)_+ := \{\mu \in \mathbb{M}_w(S) | \mu \geq 0\}.$$

The corresponding dual cone $\mathbb{M}_w(S)_+^*$ in $\mathbb{F}_w(S)$ coincides with the "natural" positive cone

$$\mathbb{F}_w(S)_+ := \{u \in \mathbb{F}_w(S) | u \geq 0\}.$$

1.2.2 Infinite linear programming

An infinite linear program requires the following components:

- two dual pairs $(\mathcal{X}, \mathcal{Y})$ and $(\mathcal{Z}, \mathcal{W})$ of real vector spaces;
- a weakly continuous linear map $L : \mathcal{X} \to \mathcal{Z}$, with adjoint $L^* : \mathcal{W} \to \mathcal{Y}$;
- a positive cone K in $\mathcal{X}$, with dual cone K^* in $\mathcal{Y}$ [see (1.16)]; and
- vectors $b \in \mathcal{Z}$ and $c \in \mathcal{Y}$.

Then the **primal** linear program is

$$\begin{aligned} \mathbb{P}: \ &\text{minimize } \langle x, c\rangle \\ &\text{subject to: } Lx = b, \ x \in K. \end{aligned} \tag{1.17}$$

The corresponding **dual** problem is

$$\begin{aligned} \mathbb{P}^*: \ &\text{maximize } \langle b, w\rangle \\ &\text{subject to: } c - L^*w \in K^*, \ w \in \mathcal{W}. \end{aligned} \tag{1.18}$$

An element x of $\mathcal{X}$ is called **feasible** for $\mathbb{P}$ if it satisfies (1.17), and $\mathbb{P}$ is said to be **consistent** if it has a feasible solution. If $\mathbb{P}$ is consistent, then its **value** is defined as

$$\inf \mathbb{P} := \inf\{\langle x, c\rangle | x \text{ is feasible for } \mathbb{P}\}; \tag{1.19}$$

otherwise, $\inf \mathbb{P} := +\infty$. The program $\mathbb{P}$ is **solvable** if there is a feasible solution x^* that achieves the infimum in (1.19). In this case, x^* is called an **optimal solution** for $\mathbb{P}$ and, instead of $\inf \mathbb{P}$, the value of $\mathbb{P}$ is written as

$$\min \mathbb{P} = \langle x^*, c\rangle.$$

Similarly, $v \in \mathcal{W}$ is **feasible** for the dual program $\mathbb{P}^*$ if it satisfies (1.18), and $\mathbb{P}^*$ is said to be **consistent** if it has a feasible solution. If $\mathbb{P}^*$ is consistent, then its **value** is defined as

$$\sup \mathbb{P}^* := \sup\{\langle b, v\rangle \mid v \text{ is feasible for } \mathbb{P}^*\}; \tag{1.20}$$

otherwise, $\sup \mathbb{P}^* := -\infty$. The dual $\mathbb{P}^*$ is **solvable** if there is a feasible solution v^* that attains the supremum in (1.20), in which case we write the value of $\mathbb{P}^*$ as

$$\max \mathbb{P}^* = \langle b, w^*\rangle.$$

The next theorem can be proved as in elementary (finite-dimensional) LP.

Theorem 4 *(a) (***Weak duality.***) If $\mathbb{P}$ and $\mathbb{P}^*$ are both consistent, then their values are finite and satisfy*

$$\sup \mathbb{P}^* \leq \inf \mathbb{P}. \tag{1.21}$$

*(b) (***Complementary slackness.***) If x is feasible for $\mathbb{P}$, v is feasible for $\mathbb{P}^*$, and*

$$\langle x, c - L^* v\rangle = 0, \tag{1.22}$$

then x is optimal for $\mathbb{P}$ and w is optimal for $\mathbb{P}^*$.

The converse of Theorem 4(b) does not hold in general. It does hold, however, if there is **no duality gap** for $\mathbb{P}$, which means that equality holds in (1.21), i.e.,

$$\sup \mathbb{P}^* = \inf \mathbb{P}. \tag{1.23}$$

On the other hand, it is said that the **strong duality** condition for $\mathbb{P}$ holds if $\mathbb{P}$ and its dual $\mathbb{P}^*$ are both solvable and

$$\max \mathbb{P}^* = \min \mathbb{P}. \tag{1.24}$$

The following theorem gives conditions under which $\mathbb{P}$ is solvable and there is no duality gap—for a proof see Anderson and Nash [1, Theorem 3.9].

Theorem 5 *Let H be the set in $\mathcal{Z} \times \mathbb{R}$ defined as*

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

If $\mathbb{P}$ is consistent and H is weakly closed [that is, closed in the weak topology $\sigma(\mathcal{Z} \times \mathbb{R}, \mathcal{W} \times \mathbb{R})$], then $\mathbb{P}$ is solvable and there is no duality gap, so that (1.23) becomes

$$\sup \mathbb{P}^* = \min \mathbb{P}.$$

1.2.3 Approximation of linear programs

An important practical question is how to obtain—or at least estimate—the value of a linear program. In later sections we shall consider two approaches related to the following definitions.

Definition 6 (Minimizing and maximizing sequences.)

(a) A sequence $\{x_n\}$ in $\mathcal{X}$ is called a **minimizing sequence** *for $\mathbb{P}$ if each x_n is feasible for $\mathbb{P}$ and $\langle x_n, c\rangle \downarrow \inf \mathbb{P}$.*

(b) A sequence $\{v_n\}$ in $\mathcal{W}$ is called a **maximizing sequence** *for the dual problem $\mathbb{P}^*$ if each v_n is feasible for $\mathbb{P}^*$ and $\langle b, v_n\rangle \uparrow \sup \mathbb{P}^*$.*

Note that if $\mathbb{P}$ is consistent with a finite value $\inf \mathbb{P}$, then [by definition (1.19) of $\inf \mathbb{P}$] there exists a minimizing sequence. A similar remark holds for $\mathbb{P}^*$.

The equality $Lx = b$ in (1.17) is of course equivalent to write $Lx - b = 0$, or $\langle Lx - b, v\rangle = 0$ for all $v \in \mathcal{W}$. If the latter equality is required to hold only in a subset W of $\mathcal{W}$, we then have an *aggregation* of constraints of $\mathbb{P}$. On the other hand, if (1.17) holds only in a subset K' of K, we obtain an *inner approximation* of $\mathbb{P}$. The corresponding linear programs become as follows.

Definition 7 (Aggregations and inner approximations.)

(a) Let W be a subset of $\mathcal{W}$. Then the linear program

$$\begin{aligned} \mathbb{P}(W) :\ & \textit{minimize } \langle x, c\rangle \\ & \textit{subject to: } \langle Lx - b, w\rangle = 0 \quad \forall w \in W,\ x \in K, \end{aligned} \tag{1.25}$$

is called an **aggregation** *(of constraints) of $\mathbb{P}$.*

(b) If $K' \subset K$ is a subset of the positive cone $K \subset \mathcal{X}$, then the program

$$\begin{aligned} \mathbb{P}(K') \ :\ & \textit{minimize } \langle x, c\rangle \\ & \textit{subject to: } Lx = b,\ x \in K', \end{aligned} \tag{1.26}$$

is called an **inner approximation** *of $\mathbb{P}$.*

As K' is contained in K, we have $\inf \mathbb{P} \leq \inf \mathbb{P}(K')$. On the other hand, if x satisfies (1.17), then it satisfies (1.25), and so $\inf \mathbb{P}(W) \leq \inf \mathbb{P}$. Hence

$$\inf \mathbb{P}(W) \leq \inf \mathbb{P} \leq \inf \mathbb{P}(K').$$

Thus, we can use an aggregation (of constraints) to approximate $\inf \mathbb{P}$ *from below*, whereas an inner approximation can be used to approximate $\inf \mathbb{P}$ *from above*. One can also easily get the following (for a proof see Hernández-Lerma and Lasserre [18]):

Proposition 8 *Suppose that $\mathbb{P}$ is solvable.*

(a) If W is weakly dense in $\mathcal{W}$, then $\mathbb{P}(W)$ is equivalent to $\mathbb{P}$ in the sense that $\mathbb{P}(W)$ is also solvable and

$$\min \mathbb{P}(W) = \min \mathbb{P}.$$

(b) If K' is weakly dense in K, then there is a sequence $\{x_n\}$ in K' such that

$$\langle x_n, c\rangle \to \min \mathbb{P}.$$

1.3 Linear programing formulation of the AC-problem

Let $(\mathbb{X}, \mathbb{A}, \{\mathbb{A}(x), x \in \mathbb{X}\}, P, c)$ be an MDP with Borel state and action spaces $\mathbb{X}$, $\mathbb{A}$, one-step cost function $c : \mathbb{X} \times \mathbb{A} \to \mathbb{R}$, transition kernel P, and with the long-run expected average cost criterion, that is, given a policy $\pi \in \Pi$ and an initial distribution $\nu \in \mathcal{P}(\mathbb{X})$ (the space of probability measures on $\mathcal{B}(\mathbb{X})$), its long-run expected average cost $J(\pi, \nu)$ is given by:

$$J(\pi, \nu) := \limsup_{N\to\infty} N^{-1} E_\nu^\pi \sum_{t=0}^{N-1} c(x_t, a_t).$$

We recall that to the state space $\mathbb{X}$ and the action space $\mathbb{A}$ is associated the space $\mathbb{K}$ of *feasible state-action pairs*, i.e.,

$$\mathbb{K} := \{(x, a) \in \mathbb{X} \times \mathbb{A} \mid x \in \mathbb{X},\, a \in \mathbb{A}(x)\}. \tag{1.27}$$

It is assumed that $\mathbb{K}$ is a Borel subset of $\mathbb{X} \times \mathbb{A}$ and contains the graph of a measurable function from $\mathbb{X}$ to $\mathbb{A}$. This implies in particular that the set of stationary deterministic policies $\mathbb{F}$ is not empty. Let Φ be the set of stochastic kernels φ on $\mathbb{A}$ given $\mathbb{X}$ for which $\varphi(\mathbb{A}(x)|x) = 1$ for all $x \in \mathbb{X}$. In other words, Φ stands for the family of randomized stationary policies.

Remark 9 *Every p.m. μ on $\mathbb{X}\times\mathbb{A}$ concentrated on $\mathbb{K}$ can be "disintegrated" as*

$$\mu(B \times C) = \int_B \varphi(C \mid x)\, \widehat{\mu}(dx) \quad \forall B \in \mathcal{B}(\mathbb{X}),\, C \in \mathcal{B}(\mathbb{A}), \tag{1.28}$$

for some $\varphi \in \Phi$, where $\widehat{\mu}$ is the marginal of μ on $\mathbb{X}$, that is, $\widehat{\mu}(B) := \mu(B \times \mathbb{A})$ for all $B \in \mathcal{B}(\mathbb{X})$. Sometimes we shall write the disintegration (1.28) of μ as $\mu = \widehat{\mu} \cdot \varphi$.

Throughout the rest of this chapter we suppose that the following assumption is satisfied.

Assumption A1. *(a) $J(\widehat{\pi}, \widehat{x}) < \infty$ for some policy $\widehat{\pi}$ and some initial state $\widehat{x}$.*

(b) The one-stage cost function $c(x, a)$ is nonnegative.

(c) $c(x, a)$ is inf-compact, that is, the set $C_r := \{(x, a) \in \mathbb{K} \,|\, c(x, a) \leq r\} \subset \mathbb{K}$ is compact for each $r \in \mathbb{R}$.

(d) The transition law P is weakly continuous, that is,

$$(x, a) \mapsto \int u(y) P(dy|x, a)$$

is a continuous bounded function on $\mathbb{K}$ for every continuous bounded function u on $\mathbb{X}$.

Remark 10 *(a) Assumption A1(c) clearly implies that the one-stage cost c is lower semicontinuous, that is, the set C_r is closed for each $r \in \mathbb{R}$. It also implies that the set $A_r(x) := \{a \in \mathbb{A}(x) \,|\, c(x, a) \leq r\}$ is compact for each $x \in \mathbb{X}$ and $r \in \mathbb{R}$. These facts, together with Assumption A1(b), yield, in particular, that the function $x \mapsto \min_{a \in \mathbb{A}(x)} c(x, a)$ is measurable (see Rieder [29]).*

(b) Another important consequence of Assumption A1(c) is as follows. Let $M = \{\mu_i, i \in I\}$ be an arbitrary family of p.m.'s on $\mathbb{X} \times \mathbb{A}$, concentrated on $\mathbb{K}$. If

$$k := \sup_{i \in I} \langle \mu_i, c \rangle < \infty,$$

the M is tight, that is, for each $\epsilon > 0$ there is a compact subset $K = K_\epsilon$ of $\mathbb{K}$ such that

$$\sup_{i \in I} \mu_i(K^c) < \epsilon,$$

where K^c stands for the complement of K. Indeed, let C_r be as in Assumption A1(c), with $r > 0$, and note that, for all $i \in I$,

$$\begin{aligned} k \geq \langle \mu_i, c \rangle &\geq \int_{C_r^c} c(x, a)\, \mu_i(d(x, a)) \\ &\geq \mu_i(C_r^c) \cdot \inf\{c(x, a) \,|\, (x, a) \notin C_r\} \\ &\geq \mu_i(C_r^c) \cdot r. \end{aligned}$$

Hence $\sup_{i\in I}\mu_i(C_r^c)\le k/r$ *for all* $r>0$*, which implies that* M *is tight.Note also that the latter condition is obviously true if* M *is a bounded set of measures (that is,* $\sup_{i\in I}\mu_i(\mathbb{K})\le m$ *for some* $m>0$*) rather than p.m.'s*

Let $\rho_{\min}$ be defined as

$$\rho_{\min} := \inf_{\mathcal{P}(\mathbb{X})} J^*(\nu) = \inf_{\mathcal{P}(\mathbb{X})} \inf_{\Pi} J(\pi,\nu). \tag{1.29}$$

A pair $(\pi^*,\nu^*)\in\Pi\times\mathcal{P}(\mathbb{X})$ that satisfies

$$J(\pi^*,\nu^*) \;=\; \rho_{\min} \tag{1.30}$$

is called a "minimum pair".

In this section we introduce a linear program (P) such that

$$\sup(\mathrm{P}^*) \le \rho_{\min} \le \inf(\mathrm{P}). \tag{1.31}$$

Then we will show that (P) is *solvable* and that there is *no duality gap*, so that instead of (1.31) we will have the stronger relation

$$\sup(\mathrm{P}^*) = \rho_{\min} = \min(\mathrm{P}). \tag{1.32}$$

Moreover, disintegrating any optimal solution μ^* of (P) as $\mu^*=\widehat{\mu^*}\cdot\varphi_*$ (see Remark 9) for some $\varphi_*\in\Phi$, yields that $(\varphi_*,\widehat{\mu^*})$ satisfies (1.30), that is, $(\varphi_*,\widehat{\mu^*})$ is a minimum pair, and, in addition,

$$J(\varphi_*,x) \;=\; \rho_{\min} \qquad \widehat{\mu^*}\text{-a.e.} \tag{1.33}$$

1.3.1 The linear programs

We first introduce the components of the linear program (P) as in §1.2.2.

The dual pairs. Let $\mathbb{K}\subset\mathbb{X}\times\mathbb{A}$ be the set defined in (1.27), and let $w(x,a)$ and $w_0(x)$ be the weight functions on $\mathbb{K}$ and $\mathbb{X}$, respectively, defined as

$$w(x,a) := 1 + c(x,a), \quad w_0(x) := \min_{\mathbb{A}(x)} w(x,a). \tag{1.34}$$

By Remark 10(a), $w_0(x)$ is measurable. The dual pairs we are concerned with are

$$(\mathcal{X},\mathcal{Y}) := (\mathbb{M}_w(\mathbb{K}),\mathbb{B}_w(\mathbb{K})) \tag{1.35}$$

and

$$(\mathcal{Z}, \mathcal{W}) := (\mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}), \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X})). \tag{1.36}$$

In particular, the bilinear form on $(\mathbb{M}_w(\mathbb{K}), \mathbb{B}_w(\mathbb{K}))$ is as in (1.9), namely,

$$\langle \mu, u \rangle := \int_{\mathbb{K}} u \, d\mu, \tag{1.37}$$

and on $(\mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}), \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X}))$ is

$$\langle (r, \nu), (\rho, v) \rangle := r \cdot \rho + \int_{\mathbb{X}} v \, d\nu. \tag{1.38}$$

Note that, since $c(x, a)$ is nonnegative [Assumption A1(b)], (1.34) yields

$$0 \leq c(x, a) \leq w(x, a) \quad \forall (x, a) \in \mathbb{K},$$

which implies that *the cost-per-stage function c is in* $\mathbb{B}_w(\mathbb{K})$, and, on the other hand,

$$1 \leq w_0(x) \leq w(x, a) \quad \forall (x, a) \in \mathbb{K}. \tag{1.39}$$

Moreover, the policy $\widehat{\pi}$ and the initial state $\widehat{x}$ in Assumption A1(a) satisfy

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{t=0}^{n-1} E_{\widehat{x}}^{\widehat{\pi}}[w(x_t, a_t)] = 1 + J(\widehat{\pi}, \widehat{x}) < \infty. \tag{1.40}$$

We also need the following additional assumption.

Assumption A2. *There is a constant k such that*

$$\int_{\mathbb{X}} w_0(y) P(dy|x, a) \leq k w(x, a) \quad \forall (x, a) \in \mathbb{K}.$$

In other words, Assumption A2 states that the function

$$(x, a) \mapsto \int_{\mathbb{X}} w_0(y) P(dy|x, a) \quad \text{is in} \quad \mathbb{B}_w(\mathbb{K}).$$

The linear maps. Let L_0 and L_1 be the linear maps

$$L_0 : \mathbb{M}_w(\mathbb{K}) \to \mathbb{R} \quad \text{and} \quad L_1 : \mathbb{M}_w(\mathbb{K}) \to \mathbb{M}_{w_0}(X),$$

with

$$L_0\mu := \langle \mu, 1\rangle = \mu(\mathbb{K}) \tag{1.41}$$

and

$$(L_1\mu)(B) := \widehat{\mu}(B) - \int_{\mathbb{K}} P(B|x,a)\mu(d(x,a)) \quad \text{for} \quad B \in \mathcal{B}(\mathbb{X}), \tag{1.42}$$

where $\widehat{\mu}$ denotes the marginal of μ on $\mathbb{X}$. Finally, let

$$L : \mathbb{M}_w(\mathbb{K}) \to \mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X})$$

be the linear map

$$L\mu := (L_0\mu, L_1\mu) \quad \text{for} \quad \mu \in \mathbb{M}_w(\mathbb{K}), \tag{1.43}$$

with adjoint

$$L^* : \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X}) \to \mathbb{B}_w(\mathbb{K})$$

given by

$$L^*(\rho, u)(x,a) := \rho + u(x) - \int_{\mathbb{X}} u(y)P(dy|x,a) \tag{1.44}$$

for every pair (ρ, u) in $\mathbb{R} \times \mathbb{B}_{w_0}(X)$ and (x,a) in $\mathbb{K}$. Hence, (1.39), Assumption A2 and Proposition 3 yield that

$$\textit{the linear map } L \textit{ in (1.43) is weakly continuous,} \tag{1.45}$$

that is, continuous with respect to the weak topologies

$$\sigma(\mathbb{M}_w(\mathbb{K}), \mathbb{B}_w(\mathbb{K})) \;\text{ and }\; \sigma(\mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}), \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X})).$$

The linear programs. Consider the vectors

$$b := (1,0) \;\text{ in }\; \mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}), \;\text{ and }\; c \;\text{ in }\; \mathbb{B}_w(\mathbb{K}),$$

where c is the cost-per-stage function, as well as the positive cone

$$K := \mathbb{M}_w(\mathbb{K})_+, \tag{1.46}$$

whose dual cone is

$$K^* := \mathbb{B}_w(\mathbb{K})_+. \tag{1.47}$$

Then the **primal** linear program is

(P) minimize $\langle \mu, c\rangle$

$$\text{subject to: } L\mu = (1,0), \quad \mu \in \mathbb{M}_w(\mathbb{K})_+. \tag{1.48}$$

More explicitly, by (1.41)–(1.43), the constraint (1.48) is satisfied if

$$\mu(\mathbb{K}) = 1 \quad \text{with} \quad \mu \in \mathbb{M}_w(\mathbb{K})_+, \tag{1.49}$$

and $L_1\mu = 0$, i.e.,

$$\widehat{\mu}(B) - \int_{\mathbb{K}} P(B|x,a)\mu(d(x,a)) = 0 \quad \forall B \in \mathcal{B}(\mathbb{X}), \tag{1.50}$$

with $\mu \in \mathbb{M}_w(\mathbb{K})_+$. Observe that (1.49) requires μ to be a *probability measure* (p.m.). Moreover, disintegrating μ into $\varphi \cdot \hat{\mu}$ (see Remark 9) for some $\varphi \in \Phi$, and using the notation

$$P_\varphi(\bullet|x) := \int_{\mathbb{A}} P(\bullet|x,a)\varphi(da|x),$$

(1.50) can be written as

$$\widehat{\mu}(B) = \int_{\mathbb{X}} P(B|x,\varphi)\widehat{\mu}(dx) \quad \forall B \in \mathcal{B}(\mathbb{X}),$$

which means that μ *is feasible for* (P) *if* μ *is a p.m. on* $\mathbb{K}$ *such that its marginal* $\widehat{\mu}$ *on* $\mathbb{X}$ *is an i.p.m. for the transition kernel* $P_\varphi(\bullet|\bullet) = P(\varphi)$.

On the other hand, observe that

$$\langle b, v\rangle = \langle (1,0), (\rho, u)\rangle = \rho \quad \forall v = (\rho, u) \in \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X}).$$

Hence, by (1.47) and (1.44), the **dual** of (P) is

(P*) maximize ρ

$$\begin{aligned} \text{subject to: } \; & \rho + u(x) - \int_{\mathbb{X}} u(y)P(dy|x,a) \le c(x,a) \\ & \forall (x,a) \in \mathbb{K}, \text{ with } (\rho, u) \in \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X}). \end{aligned} \tag{1.51}$$

This completes the specification of the linear programs associated to the AC problem.

1.3.2 Solvability of (P)

Before proceeding to verify (1.31) and (1.32), let us note the following.

Remark 11 *We will use the following conventions:*

(a) A measure μ on $\mathbb{K} \subset \mathbb{X} \times \mathbb{A}$ may (and will) be viewed as a measure on all of $\mathbb{X} \times \mathbb{A}$ by defining $\mu(\mathbb{K}^c) := 0$, where $\mathbb{K}^c := \mathbb{X} \times \mathbb{A} \setminus \mathbb{K}$.

(b) We will regard $c : \mathbb{K} \to \mathbb{R}_+$ as a function on all of $\mathbb{X} \times \mathbb{A}$ with $c(x,a) := +\infty$ if (x,a) is in $\mathbb{K}^c$. Observe that this convention is consistent with Assumption A1(c), and, moreover, the weight function $w = +\infty$ on $\mathbb{K}^c$. Any other function u in $\mathbb{B}_w(\mathbb{K})$ can be arbitrarily extended to $\mathbb{X} \times \mathbb{A}$, for example, as $u := 0$ on $\mathbb{K}^c$.

(c) $0 \cdot (+\infty) := 0$

(d) A function u in $\mathbb{B}_{w_0}(\mathbb{X})$ will also be seen as the function in $\mathbb{B}_w(\mathbb{K})$ given by $u(x,a) := u(x)$ for all (x,a) in $\mathbb{K}$.

Then, in particular, we may write the bilinear form in (1.37) as

$$\langle \mu, u \rangle = \int_{\mathbb{X} \times \mathbb{A}} u \, d\mu$$

for any measure μ in $\mathbb{M}_w(\mathbb{K})$ and any function u in $\mathbb{B}_w(\mathbb{K})$ or in $\mathbb{B}_{w_0}(\mathbb{X})$.

We will next show that (P) is **solvable** and

$$\sup(\mathrm{P}^*) \leq \rho_{\min} = \min(\mathrm{P}). \tag{1.52}$$

To do this let us first recall that a randomized stationary policy $\varphi \in \Phi$ is said to be *stable* if there exists an invariant probability measure (i.p.m.) p_φ for the transition kernel $P_\varphi(B|x) := \int_{\mathbb{A}} P(B|x,a)\varphi(da|x)$, i.e.,

$$p_\varphi(B) := \int P_\varphi(B|x) \, p_\varphi(dx),$$

and, in addition, the average cost $J(\varphi, p_\varphi)$ satisfies that

$$J(\varphi, p_\varphi) = \int c_\varphi(x) \, p_\varphi(dx),$$

where $c_\varphi(x) = \int_{\mathbb{A}} c(x,a)\varphi(da|x)$.

Theorem 12 *Suppose that Assumptions A1 and A2 are satisfied. Then:*

(a) **[Solvability of (P)].** *There exists an optimal solution μ^* for (P), and*

$$\min(P) = \rho_{\min} = \langle \mu^*, c \rangle. \tag{1.53}$$

The disintegration of μ^* as $\widehat{\mu^*} \cdot \varphi_*$ for some $\varphi_* \in \Phi$ yields that $(\varphi_*, \widehat{\mu^*})$ is a minimum pair and, in addition,

$$J(\varphi_*, x) \; = \; \rho_{\min} \quad \widehat{\mu^*}\text{-}a.e.. \tag{1.54}$$

(b) **[Consistency of ($\mathbf{P}^*$)].** *The dual problem (P^*) is consistent and it satisfies the inequality in (1.52).*

Proof. (a) By Theorem 5.7.9(a) in [17], there exists a stable randomized stationary policy φ_* such that $(\varphi_*, p_{\varphi_*})$ is a minimum pair. That is, p_{φ_*} is an i.p.m. for the transition kernel

$$P_{\varphi_*}(B|x) := \int_{\mathbb{A}} P(B|x,a)\varphi_*(da|x),$$

and

$$J(\varphi^*, p_{\varphi_*}) = \int_{\mathbb{X}} c_{\varphi_*}(x) p_{\varphi_*}(dx) = \rho_{\min} < \infty, \tag{1.55}$$

where

$$c_{\varphi_*}(x) := \int_{\mathbb{A}} c(x,a)\varphi_*(da|x).$$

Furthermore, as p_{φ_*} is an i.p.m. for P_{φ_*}, for every B in $\mathcal{B}(\mathbb{X})$ we have

$$p_{\varphi_*}(B) = \int_{\mathbb{X}} P_{\varphi_*}(B|x) p_{\varphi_*}(dx),$$

i.e.,

$$p_{\varphi_*}(B) = \int_{\mathbb{X}} \int_{\mathbb{A}} P(B|x,a)\varphi_*(da|x) p_{\varphi_*}(dx). \tag{1.56}$$

Now let μ^* be the measure on $\mathbb{X} \times \mathbb{A}$ defined as

$$\mu^*(B \times C) := \int_B \varphi_*(C|x) p_{\varphi_*}(dx) \quad \forall B \in \mathcal{B}(\mathbb{X}),\ C \in \mathcal{B}(\mathbb{A}).$$

Then, μ^* is a p.m. on $\mathbb{X} \times \mathbb{A}$, concentrated on $\mathbb{K}$, and its marginal on $\mathbb{X}$ coincides with p_{φ_*}:

$$\widehat{\mu}^*(B) := \mu^*(B \times \mathbb{A}) = p_{\varphi_*}(B) \quad \forall B \in \mathcal{B}(\mathbb{X}).$$

It follows that we may rewrite (1.56) and (1.55) as

$$\widehat{\mu}^*(B) - \int_{\mathbb{K}} P(B|x,a)\mu^*(d(x,a)) = 0 \quad \forall B \in \mathcal{B}(\mathbb{X}),$$

and

$$J(\varphi^*, p_{\varphi_*}) = \langle \mu^*, c \rangle = \rho_{\min} < \infty, \tag{1.57}$$

which means that we already have the second equality in (1.53), as well as the equalities $\mu^*(\mathbb{K}) = 1$ and $L_1\mu^* = 0$ in (1.49) and (1.50).

Therefore, to prove (1.53) in part (a) it suffices to show that

(i) μ^* is in $\mathbb{M}_w(\mathbb{K})$ so that μ^* is indeed feasible for (P); and

(ii) $\langle \mu, c \rangle \geq \rho_{\min}$ for any feasible solution μ for (P), which would yield $\inf(\mathrm{P}) \geq \rho_{\min}$.

In other words, (i), (ii) and (1.57) will give that μ^* is feasible for (P) and

$$\rho_{\min} = \langle \mu^*, c \rangle \geq \inf(\mathrm{P}) \geq \rho_{\min}, \quad \text{i.e.,} \quad \langle \mu^*, c \rangle = \rho_{\min}.$$

Proof of (i). This is easy because, by (1.34) and (1.57),

$$\langle \mu^*, w \rangle = 1 + \langle \mu^*, c \rangle < \infty.$$

Proof of (ii). If μ satisfies (1.49) and (1.50), then, in particular, μ is a probability measure on $\mathbb{X} \times \mathbb{A}$ concentrated on $\mathbb{K}$. Thus, μ can be "disintegrated" as $\widehat{\mu} \cdot \varphi$ for some $\varphi \in \Phi$ (see Remark 9). Furthermore, taking $(\varphi, p_\varphi) := (\varphi, \widehat{\mu})$, (1.50) gives that φ is a stable randomized stationary policy, and, therefore, by the definition of $\rho_{\min}$,

$$\langle \mu, c \rangle = J(\varphi, \widehat{\mu}) \geq \rho_{\min}.$$

This proves (ii).

To complete the proof of part (a), observe that from $\langle c, \mu^* \rangle = \rho_{\min}$, it follows that $c_{\varphi_*} \in L_1(\widehat{\mu^*})$. Therefore, by Birkhoff's Individual Ergodic Theorem

$$\int_{\mathbb{X}} J(\varphi_*, x)\, d\widehat{\mu^*} = \int_{\mathbb{X}} c_{\varphi_*}\, d\widehat{\mu^*} = \rho_{\min},$$

and (1.54) follows from combining the above equality with $J(\varphi_*, x) \geq \rho_{\min}$.

(b) By (a) and the weak duality property (1.21), to prove (b) it suffices to show that (P^*) is consistent. This, however, is obvious: for example, the pair (ρ, u) with $\rho = u(\cdot) \equiv 0$ satisfies (1.51). ∎

Remark 13 *As mentioned in the Introduction, the LP formulation of MDPs began in the early 1960s as a way to solve the associated optimality (or dynamic programming) equation. In particular, for the average cost problem the question was to find a solution (ρ, u) to the Average Cost Optimality Equation (ACOE) studied in previous chapters, that is, a number ρ and a function u on $\mathbb{X}$ such that*

$$\rho + u(x) \;=\; \min_{a \in \mathbb{A}(x)} [c(x,a) + \int_{\mathbb{X}} u(y) P(dy|x,a)] \quad \forall (x,a) \in \mathbb{K}. \tag{1.58}$$

The idea was the following. If (ρ, u) satisfies (1.58), then we obviously have

$$\rho + u(x) \;\leq\; c(x,a) + \int_{\mathbb{X}} u(y) P(dy|x,a) \quad \forall (x,a) \in \mathbb{K},$$

or, equivalently,

$$\rho + u(x) - \int_{\mathbb{X}} u(y) P(dy|x,a) \;\leq\; c(x,a) \quad \forall (x,a) \in \mathbb{K}, \tag{1.59}$$

which is exactly the same as (1.51). On the other hand, from (1.39) and Assumption A2, it is easy to check that

$$\lim_{n \to \infty} \frac{1}{n} E_\nu^\pi u(x_n) \;=\; 0$$

for any function $u \in \mathbb{B}_{w_0}(\mathbb{X})$, any policy π and any initial distribution ν for which $J(\pi, \nu) < \infty$, which in turn, by (1.59), yields that

$$\rho \;\leq\; J(\pi, \nu).$$

As this holds for any pair $(\rho, u) \in \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X})$ that satisfies (1.59), that is, for any feasible solution for (P^*), we obtain the inequality $\sup(P^*) \leq \rho_{\min}$, already obtained in Theorem 12(b) using standard LP arguments. Thus, the essence of the original LP approach to MDPs was to give conditions for the dual (P^*) to be solvable, and for the absence of a duality gap. We will now address these questions in Theorems 14 and 16.

1.3.3 Absence of duality gap

We now prove (1.32).

Theorem 14 (Absence of duality gap.) *If Assumptions A1 and A2 are satisfied, then (1.32) holds.*

Proof. We wish to use Theorem 5 with $\mathcal{Z}$ and L as in (1.36) and (1.43), respectively. Hence, we wish to show that the set

$$H := \{(L\mu, \langle \mu, c\rangle + r) | \mu \in \mathbb{M}_w(\mathbb{K})_+,\ r \geq 0\}$$

is closed in the weak topology

$$\sigma(\mathbb{R} \times \mathbb{M}_{w_0}(\mathbb{X}) \times \mathbb{R},\ \mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X}) \times \mathbb{R}).$$

Let $(D, \leq)$ be a directed set, and consider a net $\{(\mu_\alpha, r_\alpha), \alpha \in D\}$ in $\mathbb{M}_w(\mathbb{K})_+ \times \mathbb{R}_+$ such that

$$L_0\mu_\alpha := \mu_\alpha(\mathbb{K}) \to r_* \tag{1.60}$$

$$\langle L_1\mu_\alpha, u\rangle \to \langle \nu_*, u\rangle \qquad \forall u \in \mathbb{B}_{w_0}(\mathbb{X}), \text{ and} \tag{1.61}$$

$$\langle \mu_\alpha, c\rangle + r_\alpha \to \rho_*. \tag{1.62}$$

We will show that $((r_*, \nu_*), \rho_*)$ is in H; that is, there exists a measure μ in $\mathbb{M}_w(\mathbb{K})_+$ and a number $r \geq 0$ such that

$$r_* = L_0\mu := \mu(\mathbb{K}), \tag{1.63}$$

$$\nu_* = L_1\mu, \text{ and} \tag{1.64}$$

$$\rho_* = \langle \mu, c\rangle + r. \tag{1.65}$$

We shall consider two cases, $r_* = 0$ and $r_* > 0$.

Case 1: $r_* = 0$. By definition (1.41) of L_0,

$$L_0\mu_\alpha = \mu_\alpha(\mathbb{K}). \tag{1.66}$$

Therefore, if $r_* = 0$ in (1.60), it follows easily that (1.63)–(1.65) hold with $\mu(\cdot) = 0$ and $r = \rho_*$.

Case 2: $r_* > 0$. By (1.60) [together with (1.66)] and (1.62), there exists α_0 in D such that

$$0 < \mu_\alpha(\mathbb{K}) \leq 2r_* \text{ and } \langle \mu_\alpha, c\rangle \leq 2\rho_* \quad \forall \alpha \geq \alpha_0. \tag{1.67}$$

Hence, as $\langle \mu_\alpha, c+1\rangle = \langle \mu_\alpha, c\rangle + \mu_\alpha(\mathbb{K})$, we get that $\Gamma := \{\mu_\alpha, \alpha \geq \alpha_0\}$ is a *bounded* set of measures, which combined with Assumption A1(c) yields that Γ is *tight* (see Remark 10(b)). Moreover, if $\mu_\alpha(\mathbb{K}) > 0$, we may "normalize" μ_α rewriting it as $\mu_\alpha(\cdot)/\mu_\alpha(\mathbb{K})$, and so we may assume that Γ is a (tight) family of probability measures. Then, by Prohorov's Theorem (see e.g. [4]), for each sequence $\{\mu_n\}$ in Γ there is a subsequence $\{\mu_m\}$ and a p.m. μ on $\mathbb{K}$ such that

$$\langle \mu_m, v\rangle \to \langle \mu, v\rangle \quad \forall v \in C_b(\mathbb{K}), \tag{1.68}$$

where $C_b(\mathbb{K})$ denotes the space of continuous bounded functions on $\mathbb{K}$.

In particular, taking $v(\cdot) \equiv 1$, (1.60) yields that μ *satisfies (1.63)*. We will next show that

(i) μ is in $\mathbb{M}_w(\mathbb{K})_+$, that is, $\|\mu\|_w := \langle \mu, w\rangle < \infty$ [see (1.7)], and

(ii) μ satisfies (1.64).

Proof of (i). As $w := 1 + c$, to prove (i) we need to show that $\langle \mu, c\rangle$ is finite. We will prove the latter by showing that

$$[(1.68),\ c \geq 0 \text{ and l.s.c.}] \Rightarrow \liminf \langle \mu_m, c\rangle \geq \langle \mu, c\rangle. \tag{1.69}$$

Indeed, if $c \geq 0$ and l.s.c. [as in Assumption A1(b)], then there exists an increasing sequence of functions v_k in $C_b(\mathbb{K})$ such that $v_k \uparrow c$. It follows from (1.68) that for each k

$$\liminf_{m\to\infty} \langle \mu_m, c\rangle \geq \liminf_{m\to\infty} \langle \mu_m, v_k\rangle = \langle \mu, v_k\rangle.$$

Thus, letting $k \to \infty$, the Monotone Convergence Theorem gives (1.69).

Proof of (ii). The weak continuity condition on P [Assumption A1(d)] implies that the adjoint of L_1, namely,

$$(L_1^* u)(x,a) := u(x) - \int_{\mathbb{X}} u(y) P(dy|x,a),$$

maps $C_b(\mathbb{X})$ into $C_b(\mathbb{K})$. Therefore, (1.68) and (1.61) yield that for any function u in $C_b(\mathbb{X})$

$$\begin{aligned}
\langle L_1\mu, u\rangle = \langle \mu, L_1^* u\rangle &= \lim_{m\to\infty} \langle \mu_m, L_1^* u\rangle \ \text{[by (1.68)]} \\
&= \lim_{m\to\infty} \langle L_1 \mu_m, u\rangle \\
&= \langle \nu_*, u\rangle \ \text{[by (1.61)]}.
\end{aligned}$$

That is, $\langle L_1\mu, u\rangle = \langle \nu_*, u\rangle$ for any function u in $C_b(\mathbb{X})$, which implies (1.64). This proves (ii).

Summarizing, we have shown that μ is a measure in $\mathbb{M}_w(\mathbb{K})_+$ that satisfies (1.63) and (1.64). Finally, from (1.69) and (1.62) we see that

$$\rho_* \geq \langle \mu, c\rangle + \liminf_{m\to\infty} r_m \geq \langle \mu, c\rangle \quad \text{as} \quad r_m \geq 0 \ \forall m.$$

Thus, defining $r := \rho_* - \langle \mu, c\rangle$ (≥ 0), we conclude that μ and r satisfy (1.63), (1.64) and (1.65). This shows that H is indeed weakly closed, and so (1.32) follows. ∎

Having (1.32), in the following sections we consider conditions for the solvability of the dual problem (P^*) and for the convergence of approximations to the optimal values max(P^*) and min(P).

1.4 Approximating sequences and strong duality

In the rest of this chapter we are mainly interested in the approximation of the AC-related linear program (P) and its dual (P^*). In this section we first study minimizing sequences for (P), and then maximizing sequences for (P^*).

1.4.1 Minimizing sequences for (P)

By Definition 6(a), a sequence of measures μ_n in $\mathbb{M}_w(\mathbb{K})_+$ is a **minimizing sequence** for (P) if each μ_n is feasible for (P), that is, it satisfies (1.48), and in addition

$$\langle \mu_n, c\rangle \downarrow \min(\mathrm{P}), \tag{1.70}$$

where we have used that (P) is *solvable* [Theorem 12(a)] to write its value as min (P) rather than inf (P).

Theorem 15 *Suppose that Assumptions A1 and A2 are satisfied. If $\{\mu_n\}$ is a minimizing sequence for* (P)*, then there exists a subsequence $\{j\}$ of $\{n\}$ such that $\{\mu_j\}$ converges in the weak topology $\sigma(\mathbb{M}(\mathbb{K}), C_b(\mathbb{K}))$ to an optimal solution for* (P)*.*

Proof. Let $\{\mu_n\}$ be a minimizing sequence for (P); that is [by (1.48)],

$$\langle \mu_n, 1 \rangle = 1 \quad \text{and} \quad L_1 \mu_n = 0 \quad \forall n, \tag{1.71}$$

and (1.70) holds. In particular, (1.70) implies that for any given $\varepsilon > 0$ there exists $n(\varepsilon)$ such that

$$\min(\mathrm{P}) \leq \langle \mu_n, c \rangle \leq \min(\mathrm{P}) + \varepsilon \quad \forall n \geq n(\varepsilon). \tag{1.72}$$

By the second inequality [together with Assumption A1(c)], the sequence $\{\mu_n\}$ is tight (see Remark 10(b)), so that there exists a p.m. μ^* on $\mathbb{K}$ and a subsequence $\{j\}$ of $\{n\}$ such that

$$\langle \mu_j, v \rangle \to \langle \mu^*, v \rangle \quad \forall v \in C_b(\mathbb{K}). \tag{1.73}$$

Moreover, by (1.69),

$$\langle \mu^*, c \rangle \leq \liminf_{j \to \infty} \langle \mu_j, c \rangle \leq \min(\mathrm{P}) + \varepsilon. \tag{1.74}$$

Thus, as ε was arbitrary, the latter inequality and (1.72) yield

$$\min(\mathrm{P}) = \langle \mu^*, c \rangle. \tag{1.75}$$

This will prove that μ^* is optimal for (P) provided that μ^* is *feasible* for (P); in other words, provided that μ^* is a measure in $\mathbb{M}_w(\mathbb{K})_+$ and that

$$L\mu^* = (L_0 \mu^*, L_1 \mu^*) = (1, 0). \tag{1.76}$$

This, however, is obvious because (1.74) yields $\langle \mu^*, w \rangle = 1 + \langle \mu^*, c \rangle < \infty$, whereas (1.76) follows from (1.71) and (1.73). ∎

1.4.2 Maximizing sequences for $(\mathbf{P}^*)$

By Definition 6(b) and the definition of the dual program (P^*), a sequence (ρ_n, u_n) in $\mathbb{R} \times \mathbb{B}_{w_0}(\mathbb{X})$ is a maximizing sequence for (P^*) if

$$\rho_n + u_n(x) \leq c(x, a) + \int_{\mathbb{X}} u_n(y) Q(dy|x, a) \tag{1.77}$$

for all n and $(x, a) \in \mathbb{K}$, and, in addition,

$$\rho_n = \langle (1, 0), (\rho_n, u_n) \rangle \uparrow \sup(\mathrm{P}^*). \tag{1.78}$$

The following theorem shows that the existence of a suitable maximizing sequence for (P^*) implies, in particular, that the *strong duality* condition for (P) holds [see (1.24)].

Theorem 16 [Solvability of (P*), strong duality and the ACOE.] *Suppose that Assumptions A1 and A2 are satisfied, and, furthermore, there exists a maximizing sequence* (ρ_n, u_n) *for* (P*) *with* $\{u_n\}$ *bounded in the* w_0*-norm, that is,*

$$\|u_n\|_{w_0} \leq k \quad \forall n, \tag{1.79}$$

for some constant k. *Then:*

(a) The dual problem (P*) *is solvable.*

(b) The strong duality condition holds, that is, $\max(\mathrm{P}^*) = \min(\mathrm{P})$.

(c) If μ^* *is an optimal solution for the primal program* (P), *then the ACOE holds* $\widehat{\mu}^*$*-a.e., where* $\widehat{\mu}^*$ *is the marginal of* μ^* *on* $\mathbb{X}$*; in fact, there is a function* h^* *in* $\mathbb{B}_{w_0}(\mathbb{X})$ *and a deterministic stationary policy* f_* *such that*

$$\begin{aligned} \rho^* + h^*(x) &= \min_{\mathbb{A}(x)} \left[c(x,a) + \int_{\mathbb{X}} h^*(y) P(dy|x,a) \right] \\ &= c(x, f_*) + \textstyle\int_{\mathbb{X}} h^*(y) P(dy|x, f_*) \end{aligned} \tag{1.80}$$

for $\widehat{\mu}^*$*-almost all* $x \in \mathbb{X}$.

Proof. (a) By Theorem 14 we have

$$\sup(\mathrm{P}^*) = \rho^* = \min(\mathrm{P}) \tag{1.81}$$

and, moreover, we can write (1.78) as

$$\rho_n \uparrow \rho^*. \tag{1.82}$$

Now define the function

$$h^*(x) := \limsup_{n\to\infty} u_n(x),$$

which belongs to $\mathbb{B}_{w_0}(\mathbb{X})$, by (1.79). Therefore [by (1.82) and Fatou's Lemma], taking $\limsup_n$ in (1.77) we obtain

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

This yields that (ρ^*, h^*) is feasible for (P*) [see (1.51)], which together with the first equality in (1.81) shows that (ρ^*, h^*) is in fact *optimal for* (P*).

(b) This part follows from (a) and (1.81).

(c) Let us first note that if μ is feasible for (P) and (ρ, u) is feasible for (P*), then

$$\langle L\mu, (\rho, u)\rangle = \langle (1, 0), (\rho, u)\rangle = \rho,$$

or, equivalently,

$$\langle \mu, L^*(\rho, u)\rangle = \rho, \tag{1.83}$$

where L^* is the adjoint of L, in (1.44). Now let μ^* be an optimal solution for (P), and (ρ^*, h^*) an optimal solution for (P*). By part (b) we have

$$\langle \mu^*, c\rangle = \rho^*,$$

whereas (1.83) gives

$$\langle \mu^*, L^*(\rho^*, h^*)\rangle = \rho^*.$$

Thus, subtracting the last two equalities we get

$$\langle \mu^*, c - L^*(\rho^*, h^*)\rangle = 0,$$

i.e.,

$$\int_{\mathbb{X}\times\mathbb{A}} [c(x, a) - L^*(\rho^*, h^*)(x, a)]\mu^*(d(x, a)) = 0. \tag{1.84}$$

We may disintegrate μ^* as $\mu^*(d(x, a)) = \varphi(da|x)\widehat{\mu}^*(dx)$ for some stochastic kernel $\varphi \in \Phi$, and then [using (1.44)] we can rewrite (1.84) as

$$\int_{\mathbb{X}} \left[c(x, \varphi) - \rho^* - h^*(x) + \int_{\mathbb{X}} h^*(y)P(dy|x, \varphi)\right] \widehat{\mu}^*(dx) = 0.$$

Therefore, as the integrand is *nonnegative* [by (1.51)], we get that for $\widehat{\mu}^*$-a.a. (almost all) x in $\mathbb{X}$

$$\begin{aligned}\rho^* + h^*(x) &= c(x, \varphi) + \int_{\mathbb{X}} h^*(y)P(dy|x, \varphi)\\ &= \int_{\mathbb{A}} \left[c(x, a) + \int_{\mathbb{X}} h^*(y)P(dy|x, a)\right] \varphi(da|x),\end{aligned}$$

and so

$$\rho^* + h^*(x) \geq c(x, f_*) + \int_{\mathbb{X}} h^*(y)P(dy|x, f_*) \quad \widehat{\mu}^* - a.a.\ x \in \mathbb{X} \tag{1.85}$$

for some decision function $f_* \in \mathbb{F}$ whose existence is guaranteed by a measurable selection result (see Lemma 15.1 in Hinderer [21]). Finally, as (1.51) implies

$$\rho^* + h^*(x) \leq \min_{\mathbb{A}(x)} \left[c(x,a) + \int_{\mathbb{X}} h^*(y) P(dy|x,a) \right] \text{ for all } (x,a) \in \mathbb{K},$$

we get that, by (1.85), for $\widehat{\mu}^*$-a.a. $x \in \mathbb{X}$

$$\begin{aligned} \rho^* + h^*(x) &\geq c(x, f_*) + \int_{\mathbb{X}} h^*(y) P(dy|x, f_*) \\ &\geq \min_{\mathbb{A}(x)} \left[c(x,a) + \int_{\mathbb{X}} h^*(y) P(dy|x,a) \right] \\ &\geq \rho^* + h^*(x), \end{aligned}$$

and (1.80) follows. ∎

For MDPs with *finite* state and action spaces it is well known that the *policy iteration* (or Howard's) algorithm is equivalent to solving the primal program (P) by the simplex method. For non-finite MDPs there is nothing similar. In fact, for general infinite-dimensional linear programs it is not even known what the "simplex method" is! However, every algorithm that would produce a minimizing sequence for (P) can be interpreted as a "policy iteration" method since every feasible point μ of (P), when disintegrated as $\widehat{\mu} \cdot \varphi$ for some $\varphi \in \Phi$, can be associated with the stationary randomized policy φ. Similarly, every "policy iteration" algorithm moving in the space of stationary policies with an i.p.m. would produce a minimizing sequence for (P).

On the other hand, by "duality", one would expect that *value iteration* should be somehow related to solving the dual program (P^*). This relation can be seen as follows.

Remark 17 [($\mathbf{P}^*$) vs. value iteration.] *Consider the n-step cost*

$$E_x^\pi \left[\sum_{t=0}^{n-1} c(x_t, a_t) \right]$$

and the corresponding value function $v_n(x)$, which can be computed recursively by

$$v_n(x) = \min_{a \in \mathbb{A}(x)} \left[c(x,a) + \int_{\mathbb{X}} v_{n-1}(y) P(dy|x,a) \right] \quad \text{for } n = 1, 2, \ldots, \tag{1.86}$$

with $v_0(.) \equiv 0$. *Define* $m_0 := 0$, *and for* $n = 1, 2, \ldots,$

$$\begin{aligned} m_n &:= \inf_{\mathbb{X}} [v_n(x) - v_{n-1}(x)] + m_{n-1}, \\ \rho_n &:= m_n - m_{n-1} = \inf_{\mathbb{X}} [v_n(x) - v_{n-1}(x)], \\ u_n(\cdot) &:= v_n(\cdot) - m_n. \end{aligned}$$

Then (1.86) can be rewritten as

$$\rho_n + u_n(x) = \min_{a \in \mathbb{A}(x)} [c(x,a) + \int_{\mathbb{X}} u_{n-1}(y) P(dy|x,a)], \tag{1.87}$$

which yields

$$\rho_n + u_n(x) \leq c(x,a) + \int_{\mathbb{X}} u_{n-1}(y) P(dy|x,a).$$

Moreover, as

$$v_n(\cdot) - v_{n-1}(\cdot) \geq \rho_n,$$

the sequence $\{u_n(\cdot)\}$ *is nondecreasing, and so from (1.87) we obtain*

$$\rho_n + u_n(x) \leq c(x,a) + \int_{\mathbb{X}} u_n(y) P(dy|x,a),$$

which means that the pairs (ρ_n, u_n) *are feasible for* (P^*); *see (1.77). In addition, the sequence* $\{\rho_n\}$ *is nondecreasing, and, therefore, there exists a number* $\widehat{\rho} \leq \sup(P^*)$ *such that*

$$\rho_n = \langle (1,0), (\rho_n, u_n) \rangle \uparrow \widehat{\rho}. \tag{1.88}$$

Thus, comparing (1.88) and (1.78) we conclude that the pairs (ρ_n, u_n) *form a maximizing sequence for* (P^*) *provided that* $\widehat{\rho} = \rho_{\min}$ *(see (1.32)) or, equivalently, provided that the value iteration algorithm converges.*

1.5 Finite LP approximations

We will now show a procedure to approximate the AC-related primal linear program (P) by *finite-dimensional* linear programs. We will work in essentially the same setting of the previous sections except that now we shall require the spaces $\mathbb{X}$ and $\mathbb{K}$ to be *locally compact separable metric* (LCSM) spaces. Hence throughout the following we suppose:

Assumption A3. Assumptions A1 and A2 are satisfied, and in addition $\mathbb{X}$ and $\mathbb{K}$ are LCSM spaces.

A sufficient condition for $\mathbb{K}$ to be LCSM is that $\mathbb{X}$ and $\mathbb{A}$ are LCSM spaces and that $\mathbb{K}$ is either open or closed in $\mathbb{X} \times \mathbb{A}$. On the other hand, the hypothesis that $\mathbb{X}$ and $\mathbb{K}$ are LCSM spaces ensures that $C_0(\mathbb{X})$ and $C_0(\mathbb{K})$ are both *separable* Banach spaces [See Remark 1(b).] In particular, $C_0(\mathbb{X})$ contains a *countable* subset $\mathcal{C}(\mathbb{X})$ which is dense in $C_0(\mathbb{X})$. This is a key fact to proceed with the first step of our approximation procedure.

1.5.1 Aggregation

Let $\mathcal{P}_w(\mathbb{K})$ be the family of *probability measures* (p.m.'s) in $\mathbb{M}_w(\mathbb{K})_+$, that is, the family of measures μ that satisfy (1.49). Thus, by (1.48), we may rewrite (P) as:

$$\text{(P)} \qquad \text{minimize } \langle \mu, c \rangle$$

$$\text{subject to: } L_1\mu = 0, \quad \mu \in \mathcal{P}_w(\mathbb{K}), \tag{1.89}$$

where $L_1\mu$ is the signed measure in $\mathbb{M}_{w_0}(\mathbb{X}) \subset \mathbb{M}(\mathbb{X})$ defined by (1.42). We also have:

Lemma 18 *Let $\mathcal{C}(\mathbb{X}) \subset C_0(\mathbb{X})$ be a countable dense subset of $C_0(\mathbb{X})$. Then the following are equivalent conditions for μ in $\mathcal{P}_w(\mathbb{K})$:*

(a) $L_1\mu = 0$.

(b) $\langle L_1\mu, u \rangle = 0 \quad \forall u \in C_0(\mathbb{X})$.

(c) $\langle L_1\mu, u \rangle = 0 \quad \forall u \in \mathcal{C}(\mathbb{X})$.

Proof. The equivalence of (a) and (b) is due to the fact that $(\mathbb{M}(\mathbb{X}), C_0(\mathbb{X}))$ is a dual pair—in fact, $\mathbb{M}(\mathbb{X})$ is the topological dual of $C_0(\mathbb{X})$ [Remark 1(b)]. Finally, the implication (b) $\Rightarrow$ (c) is obvious, whereas the converse follows from the denseness of $\mathcal{C}(\mathbb{X})$ in $C_0(\mathbb{X})$. ∎

By (1.89) and Lemma 18, we may further rewrite (P) in the equivalent form:

$$\text{(P)} \qquad \text{minimize } \langle \mu, c \rangle$$

$$\text{subject to: } \langle L_1\mu, u \rangle = 0 \quad \forall u \in \mathcal{C}(X); \; \mu \in \mathcal{P}_w(\mathbb{K}). \tag{1.90}$$

Observe that (1.90) defines an *aggregation* (of constraints) of (P); see Definition 7(a). In other words, the constraint $L_1\mu = 0$ in (1.89) is "aggregated" into *countably* many constraints $\langle L_1\mu, u\rangle = 0$ with u in $\mathcal{C}(\mathbb{X})$. We next reaggregate (1.90) into *finitely* many constraints as follows.

Let $\{\mathcal{C}_k\}$ be an increasing sequence of *finite* sets $\mathcal{C}_k \uparrow \mathcal{C}(\mathbb{X})$. For each k, consider the aggregation

$$\mathbb{P}(\mathcal{C}_k) \qquad \text{minimize } \langle \mu, c\rangle$$

$$\text{subject to: } \langle L_1\mu, u\rangle = 0 \quad \forall u \in \mathcal{C}_k;\ \mu \in \mathcal{P}_w(\mathbb{K}). \tag{1.91}$$

This linear program has indeed a *finite number of constraints*, namely, the cardinality $|\mathcal{C}_k|$ of $\mathcal{C}_k$. We also have our first approximation result:

Theorem 19 *Suppose that Assumption A3 is satisfied. Then*

(a) $\mathbb{P}(\mathcal{C}_k)$ *is solvable for each* $k = 1, 2, \ldots$; *in fact, the aggregation* $\mathbb{P}(W)$ *is solvable for any subset* W *of* $C_0(\mathbb{X})$.

(b) For each $k = 1, 2, \ldots$, *let* μ_k *be an optimal solution for* $\mathbb{P}(\mathcal{C}_k)$, *i.e.,*

$$\langle \mu_k, c\rangle = \min \mathbb{P}(\mathcal{C}_k).$$

Then

$$\langle \mu_k, c\rangle \uparrow \min(P) = \rho_{\min}, \tag{1.92}$$

where the equality is due to Theorem 12(a).

Furthermore, there is a subsequence $\{\mu_m\}$ *of* $\{\mu_k\}$ *that converges in the weak topology* $\sigma(\mathbb{M}(\mathbb{K}), C_b(\mathbb{K}))$ *to an optimal solution* μ^* *for* (P), *i.e.,*

$$\langle \mu_m, v\rangle \to \langle \mu^*, v\rangle \quad \forall v \in C_b(\mathbb{K}); \tag{1.93}$$

in fact, any weak-$\sigma(\mathbb{M}(\mathbb{K}), C_b(\mathbb{K}))$ *accumulation point of* $\{\mu_k\}$ *is an optimal solution for* (P).

1.5.2 Aggregation-relaxation

The *equality* constraint $\langle L_1\mu, u\rangle = 0$ in (1.91) will now be "relaxed" to inequalities of the form $|\langle L_1\mu, u\rangle| \le \varepsilon$ with $\varepsilon > 0$.

Let $\mathcal{C}_k \uparrow \mathcal{C}(\mathbb{X})$ be as in (1.91), and let $\{\varepsilon_k\}$ be a sequence of numbers $\varepsilon_k \downarrow 0$. For each $k = 1, 2, \ldots$, consider the linear program

$$\mathbb{P}(\mathcal{C}_k, \varepsilon_k) \qquad \text{minimize } \langle \mu, c\rangle$$

$$\text{subject to: } |\langle L_1\mu, u\rangle| \le \varepsilon_k \quad \forall u \in \mathcal{C}_k;\ \mu \in \mathcal{P}_w(\mathbb{K}). \tag{1.94}$$

Remark 20 *If $\varepsilon > 0$ and $I \subset C_0(\mathbb{X})$ is a finite subset of $C_0(\mathbb{X})$, then [by (1.11)] the set*

$$N(I, \varepsilon) := \{\nu \in \mathbb{M}(\mathbb{X}) \,|\, |\langle \nu, u \rangle| \leq \varepsilon \;\; \forall u \in I\}$$

defines a (closed) weak—actually weak—neighborhood of the "origin" (that is, the null measure) in $\mathbb{M}(\mathbb{X})$. In particular, if we take ε and I as ε_k and $\mathcal{C}_k$, respectively, then the constraint (1.94) states that $L_1\mu$ is in the weak* neighborhood $N(\mathcal{C}_k, \varepsilon_k)$, i.e.,*

$$L_1\mu \in N(\mathcal{C}_k, \varepsilon_k). \tag{1.95}$$

This provides a natural interpretation of $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$ as an approximation of the original program (P) in the weak topology $\sigma(\mathbb{M}(\mathbb{X}), C_0(\mathbb{X}))$.*

The following result states that Theorem 19 remains basically unchanged when $\mathbb{P}(\mathcal{C}_k)$ is replaced by $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$.

Theorem 21 *Suppose that Assumption A3 is satisfied. Then*

(a) $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$ *is solvable for each* $k = 1, 2, \ldots$.

(b) If μ_k is an optimal solution for $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$, i.e.,

$$\langle \mu_k, c \rangle = \min \mathbb{P}(\mathcal{C}_k, \varepsilon_k) \;\; \textit{for} \;\; k = 1, 2, \ldots,$$

then $\{\mu_k\}$ satisfies the same conclusion of Theorem 19(b); in particular,

$$\langle \mu_k, c \rangle \uparrow \min(P) = \rho_{\min}. \tag{1.96}$$

1.5.3 Aggregation-relaxation-inner approximations

The programs $\mathbb{P}(\mathcal{C}_k)$ and $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$ have a *finite number of constraints* and give "nice" approximation results—Theorems 19 and 21. However, they are still not good enough for our present purpose because the "decision variable" μ lies in the *infinite-dimensional* space $\mathbb{M}_w(\mathbb{K}) \subset \mathbb{M}(\mathbb{K})$. (For the latter spaces to be finite-dimensional we would need the state and action sets, $\mathbb{X}$ and $\mathbb{A}$, to be both finite sets.) Now, to obtain *finite-dimensional* approximations of (P) we will combine $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$ with a suitable sequence of *inner approximations* [see Definition 7(b)]. These are based on the following well-known result (for a proof see, for instance, Theorem 4, p. 237, in Billingsley [4], or Theorem 6.3, p. 44, in Parthasarathy [28]).

Proposition 22 [Existence of a weakly dense set in $\mathcal{P}(S)$.] *Let S be a separable metric space and $D \subset S$ a countable dense subset of S. Then the family of p.m.'s whose supports are finite subsets of D is dense in $\mathcal{P}(S)$ in the weak topology $\sigma(\mathbb{M}(S), C_b(S))$.*

We will now apply Proposition 22 to the space $S := \mathbb{K}$. Let $D \subset \mathbb{K}$ be a *countable* dense subset of $\mathbb{K}$, and let $\{D_n\}$ be an increasing sequence of *finite* sets $D_n \uparrow D$. For each $n = 1, 2, \ldots$, let $\Delta_n := \mathcal{P}(D_n)$ be the family of p.m.'s on D_n; that is, *an element of Δ_n is a convex combination of the Dirac measures concentrated at points of D_n*. Then, as $D_n \uparrow D$, the sets Δ_n for an *increasing sequence* (of sets of p.m.'s) whose limit

$$\Delta := \bigcup_{n=1}^{\infty} \Delta_n \tag{1.97}$$

is dense in $\mathcal{P}(\mathbb{K})$ in the weak topology $\sigma(\mathbb{M}(\mathbb{K}), C_b(\mathbb{K}))$; that is, for each p.m. μ in $\mathcal{P}(\mathbb{K})$, there is a sequence $\{\nu_k\}$ in Δ such that

$$\langle \nu_k, v\rangle \to \langle \mu, v\rangle \quad \forall v \in C_b(\mathbb{K}). \tag{1.98}$$

Let us now consider a linear program as $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$ except that the p.m.'s μ in (1.94) are replaced by p.m.'s in $\Delta_n \cap \mathcal{P}_w(\mathbb{K})$. That is, instead of $\mathbb{P}(\mathcal{C}_k, \varepsilon_k)$ consider the *finite* program

$$\mathbb{P}(\mathcal{C}_k, \varepsilon_k, \Delta_n): \qquad \text{minimize } \langle \mu, c\rangle$$

$$\text{subject to: } |\langle L_1\mu, u\rangle| \le \varepsilon_k \quad \forall u \in \mathcal{C}_k,\ \mu \in \Delta_n \cap \mathcal{P}_w(\mathbb{K}). \tag{1.99}$$

This is indeed a *finite* linear program because it has a finite number $|\mathcal{C}_k|$ of constraints, and a finite number $|D_n|$ of "decision variables", namely, the coefficients of a measure in $\Delta_n \cap \mathcal{P}_w(\mathbb{K})$.

The corresponding approximation result is as follows.

Theorem 23 [Finite approximations for (P).] *If Assumption A3 is satisfied then:*

(a) For each $k = 1, 2, \ldots$, there exists $n(k)$ such that, for all $n \ge n(k)$, the finite linear program $\mathbb{P}(\mathcal{C}_k, \varepsilon_k, \Delta_n)$ is solvable and

$$\min \mathbb{P}(\mathcal{C}_k, \varepsilon_k) \le \min \mathbb{P}(\mathcal{C}_k, \varepsilon_k, \Delta_n). \tag{1.100}$$

(b) *Suppose that, in addition, the cost-per-stage function $c(x, a)$ is continuous. Then for each $k = 1, 2, \ldots$ there exists $n^*(k)$ such that*

$$\min \mathbb{P}(\mathcal{C}_k, \varepsilon_k, \Delta_n) \leq \min(P) + \varepsilon_k \quad \forall n \geq n^*(k); \tag{1.101}$$

hence [by (1.100) and (1.96)]

$$\min \mathbb{P}(\mathcal{C}_k, \varepsilon_k, \Delta_n) \to \min(P) = \rho_{\min} \ \ as \ \ k \to \infty, \tag{1.102}$$

where of course the limit is taken over values of $n \geq n^*(k)$. Moreover, if μ_{kn} [for $k \geq 1$ and $n \geq n^*(k)$] is an optimal solution for $\mathbb{P}(\mathcal{C}_k, \varepsilon_k, \Delta_n)$, then every weak accumulation point of $\{\mu_{kn}\}$ is an optimal solution for (P).

The approximation results in this section are based on Hernández-Lerma and Lasserre [19]. A similar approach, combining aggregations, relaxations and inner approximations, can be used to approximate general (not necessarily MDP-related) infinite linear programs, as in Hernández-Lerma and Lasserre [18]. These two papers provide many related references.

The approximation schemes in Section 1.5 are somewhat similar in spirit to schemes proposed by Vershik [33] and Vershik and Temel't [34], but with a basic difference. Namely, we use *weak* and *weak** topologies (see Remark 20 and Lemma 22), whereas Vershik and Temel't use stronger—for instance, normed—topologies. This is a key fact because we only need "reasonable" things, whereas their context would require convergence in the *total variation norm*, which is obviously too restrictive. For instance, for an uncountable metric space, the density result in Proposition 22—with finitely supported measures—is, in general, virtually impossible to get in the total variation norm.

Finally, it is worth noting that the approach in this section can be used to approximately compute an i.p.m. for a noncontrolled Markov chain on a LCSM space whose transition kernel satisfies the (weak) Feller condition in Assumption A1(d). The idea would be to introduce an "artificial" MDP with a *singleton* control set $\mathbb{A}$ and with a continuous "cost" function that satisfies the hypothesis of Theorem 23.

1.6 Conclusion

In this chapter we have developed an LP approach for MDPs with the average cost criterion. As mentioned in the introduction, this LP approach can be adapted to other optimality criteria, and to constrained MDPS with ad hoc modifications left to the reader.

From an approximation point of view, it was shown how finite LP approximations schemes can be designed to approximate the optimal value. Validating such numerical schemes on a significant sample of problems remains to be done. In addition, an important computational issue is to construct control *policies* that are ϵ-optimal, i.e., whose cost is within ϵ of the optimal value. The numerical schemes for the optimal value might provide a valuable tool. Indeed, from the converging sequence of finite LPs, one may construct (incomplete) stationary "policies" defined only at some points of the state space $\mathbb{X}$. Extending such policies to the whole space $\mathbb{X}$ and proving their ϵ-optimality is a topic for further research. Another interesting isssue is to compare the LP approach with others, notably those that approximate from the beginning the original problem with a finite or countable state and action model.

Finally, as already noted, such numerical schemes might prove to be a valuable tool to compute invariant probability distributions for (noncontrolled) Markov chains.

Bibliography

[1] E. Altman, *Constrained Markov Decision Processes*, Chapman & Hall/CRC, Boca Raton, 1999.

[2] E.J. Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces*, Wiley, Chichester, U.K., 1987.

[3] R.B. Ash, *Real Analysis and Probability*, Academic Press, New York, 1972.

[4] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.

[5] N. Bourbaki, *Integration*, Chap. IX. Hermann, Paris, 1969.

[6] H. Brézis, *Analyse Fonctionnelle: Théorie and Applications*, 4th Ed., Masson, Paris, 1993.

[7] A. Bhatt and V. Borkar, "Occupation measures for controlled Markov processes: Characterization and optimality," *Ann. Probab.* **24**, 1531-1562, 1996.

[8] V. Borkar, "A convex analytic approach to Markov decision processes," *Probab. Theory Relat. Fields* **78**, 583-602, 1988.

[9] V. Borkar, "Ergodic control of Markov chains with constraints – the general case," *SIAM J. Control Optimization* **32**, 176-186, 1994.

[10] G.T. De Ghellinck, "Les problèmes de décisions séquentielles," *Cahiers du Centre d'Etudes de Recherche Opérationnelle* **2**, 161-179, 1960.

[11] F. D'Epenoux, "Sur un problème de production et de stockage dans l'aléatoire," *Revue Francaise de Recherche Opérationnelle* **14**, 3-16, 1960.

[12] E.V. Denardo, "On linear programming in a Markov decision problem," *Mnagement Science* **16**, 281-288, 1970.

[13] W.K. Haneveld, *Duality in Stochastic Linear and Dynamic Programming*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, 1986.

[14] W.R. Heilmann, "Solving stochastic dynamic programming problems by linear programming - an annotated bibliography," *Z. Oper. Res. Ser. A* **22**, 43-53, 1978.

[15] W.R. Heilmann, "Solving a general discounted dynamic program by linear programming," *Z. Wahrsch. Gebiete* **48**, 339-346, 1979.

[16] O. Hernández-Lerma and J. González–Hernández, "Infinite linear programming and multichain Markov control processes in uncountable spaces," *SIAM J. Contr. Optim.* **36**, 313-335, 1998.

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

[18] O. Hernández-Lerma and J.B. Lasserre, "Approximation schemes for infinite linear programs," *SIAM J. Optim.* **8**, 973-988, 1998.

[19] O. Hernández-Lerma and J.B. Lasserre, 'Linear programming approximations for Markov control processes in metric spaces," *Acta Appl. Math.* **51**, 123-139, 1998.

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

[21] K. Hinderer, *Foundations of Non-stationary Dynamic Programming with Discrete-Time Parameter*, Lecture Notes in Oper. Res. and Math. Syst. **33**, Springer-Verlag, Berlin, 1970.

[22] A. Hordijk and L.C.M. Kallenberg, "Linear programming and Markov decision chains," *Management Science* **25**, 352-362, 1979.

[23] A. Hordijk and J.B. Lasserre, "Linear programming formulation of MDPs in countable state space: the multichain case," *Zeitschrift fur Oper. Res.* **40**, 91-108, 1994.

[24] L.C.M. Kallenberg, *Linear Programming and Finite Markovian Control Problems*, Mathematical Centre Tracts 148, Amsterdam, 1983.

[25] J.B. Lasserre, "Average optimal stationary policies and linear programming in countable space Markov decision processes," *J. Math. Anal. Appl.* **183**, 233-249, 1994.

[26] A.S. Manne, "Linear programming and sequential decisions," *Management Science* **6**, 259-267, 1960.

[27] M.S. Mendiondo and R. Stockbridge, "Approximation of infinite-dimensional linear programming problems which arise in stochastic control," *SIAM J. Contr. Optim.* **36**, 1448-1472, 1998.

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

[29] U. Rieder, "Measurable selection theorems for optimization problems," *Manuscripta Math.* **24**, 115-131, 1978.

[30] A.P. Robertson and W. Robertson, *Toplogogical Vector Spaces*, Cambridge University Press, Cambridge, U.K., 1964.

[31] W. Rudin, *Real and Complex Analysis*, 3rd ed. McGraw-Hill, New York, 1986.

[32] R. Stockbridge, "Time-average control of martingale problems: A linear programming formulation," *Ann. Prob.* **18**, 190-205, 1990.

[33] A.M. Vershik, "Some remarks on the infinite-dimensional problems of linear programming," *Russian Math. Surveys* **29**, 117-124, 1970.

[34] A.M. Vershik and V. Temel't, "Some questions concerning the approximation of the optimal value of infinite-dimensional problems in linear programming," *Siberian Math. J.* **9**, 591-601, 1968.

[35] K. Yamada, "Duality theorem in Markovian decision problems," *J. Math. Anal. Appl.* **50**, 579-595, 1975.