12 THE LINEAR PROGRAMMING APPROACH

Onésimo Hernández-Lerma
Jean B. Lasserre

Abstract: This chapter is concerned with the Linear Programming (LP) approach to MDPs in general Borel spaces, valid for several criteria, including the finite horizon and long run expected average cost, as well as the infinite horizon expected discounted cost.

12.1 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 [30] for MDPs with finite state 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 [28], Hordijk and Kallenberg [25], Hordijk and Lasserre [26], Lasserre [29], and for MDPs in general Borel spaces and in discrete or continuous time, Bhatt and Borkar [7], Haneveld [13], Heilmann [14], [15], Hernández-Lerma and González-Hernández [16], Hernandez-Lerma and Lasserre [19], [21], Mendiondo and Stockbridge[31], Stockbridge [39], Yamada [42].

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. Under weak hypotheses, one obtains the existence of a stationary optimal policy (possibly on a subset S of the state space). The LP approach permits to identify this set S which is an ergodic class of minimum expected average-cost. Getting an expected AC optimal policy for all initial states (for the unichain case as well as the multichain case) requires much stronger assumptions. However, the LP approach is still possible via the introduction of additional variables. Also, it permits to handle some constrained MDPs in a very natural form. Finally, it is possible to devise simple convergent numerical approximation schemes that require to solve finite LPs for which efficient codes are now available. However, if convergence to the optimal value is obtained, it remains to devise a convergent approximation scheme for policies, as done in alternative methods like for instance in Hernández-Lerma [17] or Sennott [37], [38] for control of queues.

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 a 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 12.5 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 control up to an exit time, the long-run expected average cost, the constrained discounted and average cost problems, we have chosen to illustrate the LP approach with the AC problem. With ad hoc suitable modifications and appropriate assumptions, the reader would easily deduce the corresponding linear programs associated with the other mentioned problems. For instance, with respect to constrained MDPs, the reader is referred to Huang and Kurano [27], Altman [1], Piunovskiy [33] and Hernández-Lerma and Gonzalez-Hernández [18]. Similarly, for multiobjective MDPs, see for instance, Hernández-Lerma and Romera [23].

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

$$\sup(\mathrm{P}^*) \leq (\mathrm{MDP})^* \leq \inf(\mathrm{P}), \tag{12.1}$$

where (using terminology specified in the following section)

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

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

$$\sup(\mathrm{P}^*) = \inf(\mathrm{P}), \tag{12.2}$$

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, (12.1) and (12.2) 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

$$\min(\mathrm{P}) = (\mathrm{MDP})^*, \tag{12.3}$$

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

$$\max(\mathrm{P}^*) = (\mathrm{MDP})^*, \tag{12.4}$$

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 12.6) gives conditions under which (12.3) and (12.4) are both satisfied, so that in particular *strong duality* for (P) holds, that is,

$$\max(\mathrm{P}^*) = \min(\mathrm{P}). \tag{12.5}$$

Section 12.2 presents background material. It contains, in particular, a brief introduction to infinite LP. In Section 12.3 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 (12.2) becomes

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

Section 12.4 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 (12.5) is satisfied. Section 12.5 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 [21] and [22] for proofs and further technical details. These approximation schemes can be extended to a very general class of infinite-dimensional linear programs (as in [20]), not necessarily related to MDPs.

12.2 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.

12.2.1 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

$$\langle (x_1, x_2), (y_1, y_2) \rangle := \langle x_1, y_1 \rangle_1 + \langle x_2, y_2 \rangle_2. \tag{12.6}$$