11 CONVEX ANALYTIC METHODS IN MARKOV DECISION PROCESSES

Vivek S. Borkar

Abstract: This article describes the convex analytic approach to classical Markov decision processes wherein they are cast as a static convex programming problem in the space of measures. Applications to multiobjective problems are described.

11.1 INTRODUCTION

11.1.1 Background

Markov decision processes optimize phenomena evolving with time and thus are intrinsically dynamic optimization problems. Nevertheless, they can be cast as abstract 'static' optimization problems over a closed convex set of measures. They then become convex programming (in fact, infinite dimensional linear programming) problems for which the enormous machinery of the latter fields can be invoked and used to advantage. Logically, these are extensions of the linear programming approach to finite state finite action space problems due to Manne [43]. (Further references are given in the 'bibliographical note' at the end.) The attraction of this approach lies in the following:

(i) It leads to elegant alternative derivations of known results, sometimes under weaker hypotheses, from a novel perspective.

(ii) It brings to the fore the possibility of using convex/linear programming techniques for computing near-optimal strategies.

(iii) It allows one to handle certain unconventional problems (such as control under additional constraints on secondary 'costs') where traditional dynamic programming paradigm turns out to be infeasible or awkward.

While the convex analytic formulation is available for a variety of cost criteria and fairly general state spaces, our primary focus will be on the 'pathwise ergodic control' problem on a countable state space, for which the theory is the most elegant. This is done in the next section. Section 3 sketches extensions to other cost criteria and general state spaces, and the dual problem. Section 4 considers multiobjective problems, such as the problem of control under constraints. An 'Appendix' discusses stability of controlled Markov chains.

11.1.2 Notation

Initially we shall work with a denumerable state space, identified with the set $\{0, 1, 2, \cdots\}$ without any loss of generality. For a Polish space X, $\mathcal{P}(X)$ will denote the Polish space of probability measures on X with the Prohorov topology. (See, e.g. [15], Chapter 2.) Recall the various cost criteria commonly considered:

1. *Finite horizon cost*: $\mathbb{E}[\sum_{n=0}^{T} r(x_n, a_n)]$ for some finite $T > 0$.
2. *Discounted cost*: $\mathbb{E}[\sum_{n=0}^{\infty} \delta^n r(x_n, a_n)]$ for some discount factor $\delta \in (0, 1)$.
3. *Total cost*: Same as above for $\delta = 1$.
4. *Cost till exit time*: $\mathbb{E}[\sum_{n=0}^{\tau-1} r(x_n, a_n)]$ with $\tau = \min\{n \geq 0 : x_n \notin D\}$ for a prescribed $D \subset \mathbb{X}$.
5. *Expected ergodic cost*: $\limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} \mathbb{E}[r(x_m, a_m)]$
6. *Pathwise ergodic cost*: Minimize 'almost surely'
$$\limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} r(x_n, a_n). \tag{11.1}$$
7. *Risk-sensitive cost*: $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}[e^{\alpha \sum_{m=0}^{n-1} r(x_n, a_n)}]$

All except the last one (which has per stage costs entering in a multiplicative, as opposed to additive, fashion) are amenable to a convex analytic formulation. Nevertheless, as already mentioned, we shall confine ourselves to the pathwise ergodic cost for the most part as a representative case. (See [49] for a traditional dynamic programming based perspective.) Section 3.1 will briefly mention the counterparts of these results for the other costs. Note that if $\pi \in \Pi^{RS}$ is used and x_0 is in the support of an ergodic distribution $\eta \in \mathcal{P}(\mathbb{X})$ for the corresponding time-homogeneous Markov chain $\{x_n\}$, then (11.1) will a.s. equal
$$\sum_i \eta(i)\bar{r}(i, \pi(i))$$
where $\bar{r} : \mathbb{X} \times \mathcal{P}(\mathbb{A}) \to \mathbb{R}$ is defined by
$$\bar{r}(i, u) = \int r(i, a)u(da), \ (i, u) \in \mathbb{X} \times \mathcal{P}(\mathbb{A}).$$
This will be the starting point of our convex analytic formulation of the pathwise ergodic control problem, which we take up next.