CENTRO DE INVESTIGACION Y DE ESTUDIOS
AVANZADOS DEL INSTITUTO POLITECNICO
NACIONAL

DEPARTAMENTO DE MATEMATICAS

Procesos de control markovianos con ganancia promedio por trayectorias

TESIS

que para obtener el grado de Doctor en Ciencias en la especialidad de Matemáticas, presenta

Armando Felipe Mendoza Pérez

Asesor
Dr. Onésimo Hernández Lerma

CENTRO DE INVESTIGACION Y DE ESTUDIOS
AVANZADOS DEL INSTITUTO POLITECNICO
NACIONAL

DEPARTMENT OF MATHEMATICS

Pathwise average reward Markov control processes

THESIS

to obtain the degree of Doctor in Science in the speciality of Mathematics, presented by

Armando Felipe Mendoza Pérez

Thesis advisor
Dr. Onésimo Hernández Lerma

Contents

Abstract **iii**

Abbreviations **iv**

Notation **v**

1 Introduction **1**
1.1 Introduction 1
1.2 Summary 2
1.3 Preliminaries 3
1.4 The canonical construction 5
1.5 Weighted-norm spaces 7

2 The optimality equation **9**
2.1 The Poisson equation 10
2.2 Proof of Theorem 2.1.4 12
2.3 Uniqueness of solutions to the P.E. 17
2.4 The optimality equation 19
2.5 Preliminary results 21
2.6 Proof of Theorem 2.4.3 23

3 Pathwise Average Reward Optimality **27**
3.1 Definitions and a preliminary result 27
3.2 Technical preliminaries 28
3.3 Pathwise average optimal policies 34

4 Variance minimization **36**
4.1 Definitions 36
4.2 Preliminary results 37

4.3 Main result 42
4.4 Asymptotic normality 43

5 Constrained MCPs **52**
5.1 Expected constraints 53
5.2 Technical preliminaries 56
5.3 Optimal policies 61
5.4 A parametric family of AROEs 67
5.5 Existence of pathwise constrained optimal policies 69

6 Examples **72**
6.1 Introduction 72
6.2 A LQ system 73
6.3 An inventory system 79

7 Conclusions and open problems **84**

References **85**

Abstract

In this dissertation we study Markov control processes on Borel spaces, with possibly unbounded rewards, and a long-run pathwise (or sample-path) average reward criterion. The first problem we are concerned with is to show the existence, under suitable assumptions, of stationary policies that maximize the pathwise average reward. In a second problem we take the latter set of average optimal policies and show that it contains policies that *minimize* the asymptotic variance, and under which the pathwise rewards are asymptotically normal. In the remainder of our work we consider constrained problems. Firstly, we study the case with *expected* average constraints. We show the existence of optimal policies, and also that the problem with expected constraints can be solved by means of a parametric family of so-called optimality equations. Finally, the latter results on expected constraints problems are extended to the case with *pathwise* constraints. To conclude, we illustrate our results with a detailed analysis of a linear-quadratic problem, and an inventory system.

Abbreviations

a.a.	almost all
a.e.	almost everywhere
a.s.	almost surely
i.i.d.	independent and identically distributed
i.p.m.	invariant probability measure
l.s.c.	lower semicontinuous
u.s.c.	upper semicontinuous
CP	constrained problem
EAR	expected average reward
LQ	linear quadratic (problem)
MCM	Markov control model
MCPs	Markov control processes
OE	optimality equation
PAR	pathwise average reward
P.E.	Poisson equation

Notation

$\blacksquare$	end of a proof
$:=$	equality by definition
1_B	indicator function of a set B
$\mathbb{N}$	the set of positive integers $\{1, 2, \cdots\}$
$\mathbb{N}_0$	the set of nonnegative integers $\{0, 1, 2, \cdots\}$
$\mathbb{R}$	the set of real numbers
$\mathbb{K}$	set of feasible state-acttions pairs
φ	randomized stationary policy
Φ	set of randomized stationary policies
$\mathbb{F}$	set of decision functions
X	Borel (state) space
$\mathcal{B}(X)$	Borel σ-algebra of subsets of X
$C_b(X)$	Banach space of continuous bounded functions on X
$B_b(X)$	Banach space of measurable bounded functions on X
$B_W(X)$	Banach space of W-bounded measurable functions on X
$M(X)$	Banach space of finite signed measures on $\mathcal{B}(X)$
$\mathcal{P}(X)$	family of Borel probability measures on X

Chapter 1

Introduction

1.1 Introduction

This thesis deals with discrete-time Markov control processes in Borel spaces, with unbounded rewards. The criterion to be optimized is a long-run pathwise average reward subject to constraints on a finite numbers of long-run pathwise average costs. These problems form an important class of stochastic control problems with applications in many areas, including mathematical economics, queueing systems, epidemic processes, etc.; see, for instance [3, 7, 12, 13, 24] as well as the books [1] and [23].

However, most of the related literature is concentrated on expected average constraints. In contrast, for problems with pathwise constraints there are a lot fewer works. For instance, for finite state MCPs, we should mention the article by Haviv [12], and the works by Ross and Varadarajan [26, 27]. For MCPs on Borel spaces we only know the recent work by Vega-Amaya [31]. The article by Haviv shows, by means of examples, that pathwise constraints are in general, more "natural" than expected constraints, because MCPs with constraints on the *expected* state-action frequencies can lead to optimal policies that do not satisfy certain principles of optimality (as Bellman's principle). In contrast, the model with pathwise constraints leads to feasible optimal policies which satisfy these principles of optimality.

As can be seen in the related literature, there are several standard techniques to analyze the expected constraints problem. For example, the so-called *direct method*, where the idea is to transform the problem into an equivalent optimization problem in a suitable space of measures. Moreover,

under appropiate hypotheses, the latter problem can be transformed into either a *convex-analytical* problem or an infinite-dimensional *linear program* depending on the underlying assumptions. In this work, to obtain our main results, we use the direct metod in combination with other techniques such as *convex analysis*, *Lagrange multipliers* and *dynamic programming*.

We extend our results on the expected case to the pathwise problem using a strong law of large number for Markov chains and the so-called stability theorem for martingales. In particular, we prove that optimal policies in the former case are also optimal for the pathwise problem.

For the uncounstrained case, we prove the equivalence between pathwise average reward optimal policies and expected average reward optimal policies. Moreover, we study the existence of canonical policies that minimizes the limiting average variance. We also prove that under appropiate growth conditions on the reward, this canonical policies have an asymptotic normality behavior.

1.2 Summary

The material in this thesis is organized as follows.

In the remainder of this chapter we introduce some background material on Markov control models (MCM) (Section 1.3 and Section 1.4).

In Chapter 2 we give our preliminary results used throughout this work. Under fixed point arguments, we consider the uncounstrained expected average reward MCPs. The motivation of this chapter is to give explicit expressions for the invariant measures, also for the functions h^*_φ that solve the P.E., and the functions h^* that solve an average reward optimality equation. This fact will be particularly useful to prove boundedness conditions, necessary for asymptotic behaviors (law of large numbers, asymptotic normality) and to prove compactness conditions.

In Chapter 3 we establish the existence of unconstrained pathwise average optimal policies assuming additional assumptions that guarantee the application of the martingale stability theorem to obtain certain pathwise ergodic limits. Moreover, under our hypotheses, the equivalence between sample-path average optimal policies and expected average optimal policies is assured.

In Chapter 4 we study the existence of a stationary canonical policy that minimizes the limiting average variance in the class $\mathbb{F}_{cp}$. As a consequence,

we have that under certain growth condition, we prove that these canonical policies have an asymptotic normality behavior.

In Chapter 5 we study constrained MCPs. Here, we use a conjunction of techniques, including the *direct method*, *convex analysis*, *Lagrange multipliers* and *dynamic programming*, to establish the existence of solutions of certain average reward optimality equation, which in particular provides optimal policies to our problem with expected constraints. We also show that the expected constrained problem (CP) can be solved by means of a parametric family of AROEs, which do not depend on unknown parameters. Furthermore, we extend these results to MCPs with pathwise constraints.

In Chapter 6 we illustrate with some examples the results obtained in the previous chapters.

Finally, in Chapter 7 we state some general conclusions of our work, as well as some open problems.

1.3 Preliminaries

The material in this section is quite standard and we refer the reader to the books [14, 15] for a detailed description.

Consider a discrete-time Markov control model (MCM)

$$(\mathbf{X}, A, \{A(x) : x \in \mathbf{X}\}, Q, r),$$

with *state space* $\mathbf{X}$ and *control* (or *action*) set A, both assumed to be Borel spaces with Borel σ-algebras $\mathcal{B}(\mathbf{X})$ and $\mathcal{B}(A)$, respectively. The family $\{A(x) : x \in \mathbf{X}\}$ consists of nonempty sets $A(x) \in \mathcal{B}(A)$, with $A(x)$ being the set of *feasible controls* (or *actions*) in the state $x \in \mathbf{X}$. The set

$$\mathbb{K} := \{(x, a) : x \in \mathbf{X}, a \in A(x)\} \tag{1.3.1}$$

of *feasible state-actions pairs* is supposed to be a Borel subset of $\mathbf{X} \times A$. Moreover, the *transition law*

$$Q = \{Q(B|x, a) : B \in \mathcal{B}(\mathbf{X}),\ (x, a) \in \mathbb{K}\} \tag{1.3.2}$$

is a stochastic kernel on $\mathbf{X}$ given $\mathbb{K}$, whereas $r : \mathbb{K} \to \mathbb{R}$ is a measurable function called the *reward-per-stage*. Throughout the remainder of this work, a fixed MCM is assumed to be given.

Definition 1.3.1 *Let* $\mathbb{F}$ *be the set of all* decision functions *or* selectors*, i.e., measurable functions* $f : \mathbf{X} \to A$ *such that* $f(x)$ *is in* $A(x)$ *for all* $x \in \mathbf{X}$*, and let* Φ *be for the set of stochastic kernels* φ *on* A *given* $\mathbf{X}$ *for which* $\varphi(A(x)|x) = 1$*.*

Remark 1.3.2 *A selector* $f \in \mathbb{F}$ *may be identified with the stochastic kernel* $\varphi \in \Phi$ *for which* $\varphi(\cdot|x)$ *is the Dirac measure at* $f(x)$ *for all* $x \in \mathbf{X}$*. Hence, we have* $\mathbb{F} \subset \Phi$*.*

We shall assume that $\mathbb{F}$ is nonempty, or equivalently, that the set $\mathbb{K}$ in (1.3.1) contains the graph of a measurable function from $\mathbf{X}$ to A. This assumptions ensures that the set of control policies, defined below, is nonempty (see, for instance, [14, Chapter 2]).

Let $\mathbb{N}_0 := \{0, 1, \cdots\}$ and $\mathbb{N} := \{1, 2, \cdots\}$.

Definition 1.3.3 (**Control Policies**)*. For every* $n \in \mathbb{N}$*, let* H_n *be the family of admissible histories up to time* n*; that is,* $H_0 := \mathbf{X}$*, and* $H_n := \mathbb{K} \times H_{n-1}$ *if* $n \geq 1$*. A* (randomized) control policy *is a sequence* $\pi = \{\pi_n\}$ *of stochastic kernels* π_n *on* A *given* H_n *such that*

$$\pi_n(A(x_n)|h_n) = 1 \tag{1.3.3}$$

for every n*-history* $h_n = (x_0, a_0, \cdots, x_{n-1}, a_{n-1}, x_n)$ *in* H_n*. Let* Π *denote the set of all control policies. Moreover, a control policy* $\pi = \{\pi_n\}$ *is said to be a*

(a) randomized Markov policy *if there exists a sequence* $\{\varphi_n\}$ *of stochastic kernels* $\varphi_n \in \Phi$ *such that*

$$\pi_n(\cdot|h_n) = \varphi_n(\cdot|x_n) \quad \forall h_n \in H_n, n \in \mathbb{N}_0; \tag{1.3.4}$$

(b) (randomized) stationary policy *if there exists a stochastic kernel* $\varphi \in \Phi$ *such that*

$$\pi_n(\cdot|h_n) = \varphi(\cdot|x_n) \quad \forall h_n \in H_n, n \in \mathbb{N}_0; \tag{1.3.5}$$

(c) deterministic stationary policy *if there is a selector* $f \in \mathbb{F}$ *such that* $\pi_n(\cdot|h_n)$ *is the Dirac measure at* $f(x_n) \in A(x_n)$ *for all* $h_n \in H_n$ *and* $n \in \mathbb{N}_0$*.*

The set of all randomized Markov policies is denoted by Π_{RM}. As usual, we identify Φ, the set of stochastic kernels on A given $\mathbf{X}$, with the set of all randomized stationary policies, and $\mathbb{F}$ with the set of all deterministic stationary policies. Note that

$$\mathbb{F} \subset \Phi \subset \Pi_{RM} \subset \Pi.$$

If $\pi = \{\varphi\}$ is a stationary policy, abusing the notation we write $\pi = \varphi$.

1.4 The canonical construction

For future reference, in this section we present the canonical construction of the underlying probability space.

Let $(\Omega, \mathcal{F})$ be the (canonical) measurable space consisting of the sample space $\Omega := (\mathbf{X} \times A)^\infty$ and its product σ-algebra $\mathcal{F}$. The elements of Ω are sequences of the form $\omega = (x_0, a_0, x_1, a_1, \cdots)$ with x_n in $\mathbf{X}$ and a_n in A for all $n = 0, 1, \cdots$; the projections x_n and a_n from Ω to the sets $\mathbf{X}$ and A are called *state* and *control* (or *action*) variables, respectively. Observe that Ω contains the space $H_\infty := \mathbb{K}^\infty$ of admisible histories $(x_0, a_0, x_1, a_1, \cdots)$ with $(x_n, a_n) \in \mathbb{K}$ for each $n \in \mathbb{N}_0$.

Let $\pi = \{\pi_n\}$ be an arbitrary control policy and ν an arbitrary probability measure on $\mathbf{X}$, referred to as the "initial distribution". Then, by a theorem of C. Ionescu-Tulcea (see, for instance, [14, Proposition C.10 and Remark C.11]) there exists a unique probability measure P_ν^π defined on the sample space $(\Omega, \mathcal{F})$, which by (1.3.3) is supported on H_∞, namely, $P_\nu^\pi(H_\infty) = 1$, and, moreover, for all $B \in \mathcal{B}(\mathbf{X})$, $C \in \mathcal{B}(A)$, and $h_n \in H_n$, $n = 0, 1, \cdots$:

$$P_\nu^\pi(x_0 \in B) = \nu(B), \tag{1.4.1}$$

$$P_\nu^\pi(a_n \in C | h_n) = \pi_n(C|h_n), \tag{1.4.2}$$

$$P_\nu^\pi(x_{n+1} \in B | h_n, a_n) = Q(B|x_n, a_n). \tag{1.4.3}$$

From the theorem of C. Ionescu-Tulcea mentioned above also ensures that the measure P_ν^π can be written in the form

$$P_\nu^\pi(dx_0, da_0, dx_1, da_1, dx_2, \cdots) = \nu(dx_0)\pi_0(da_0|x_0)Q(dx_1|x_0, a_0)\cdot$$

$$\cdot\pi_1(da_1|x_0, a_0, x_1)Q(dx_2|x_1, a_1)\cdots. \tag{1.4.4}$$

Definition 1.4.1 *The stochastic process* $(\Omega, \mathcal{F}, P_\nu^\pi, \{x_n\})$ *is called a* discrete-time Markov control Process *(MCP), which is also known as a* Markov decision process.

Remark 1.4.2 *Notation*

(a) *The expectation operator with respect to* P_ν^π *is denoted by* E_ν^π. *If* ν *is concentrated at the "initial state"* $x \in \mathbf{X}$, *then we write* P_ν^π *and* E_ν^π *as* P_x^π *and* E_x^π, *respectively. Moreover, if* $\pi = \varphi$ *is a stationary policy, then we denote* P_ν^π *and* E_ν^π *as* P_ν^φ *and* E_ν^φ, *respectively.*

(b) *Let* $\varphi \in \Phi$ *be a stochastic kernel on* A *given* $\mathbf{X}$, c *a measurable function on* $\mathbb{K}$, *and* Q *the transition law in (1.3.2). Then we define, for every* $x \in \mathbf{X}$,

$$c_\varphi(x) := \int_A c(x,a)\varphi(da|x) \tag{1.4.5}$$

and

$$Q_\varphi(\cdot|x) := \int_A Q(\cdot|x,a)\varphi(da|x). \tag{1.4.6}$$

In particular, for a function $f \in \mathbb{F}$, *(1.4.5)-(1.4.6) become*

$$c_f(x) = c(x, f(x)) \quad \text{and} \quad Q_f(B|x) = Q(B|x, f(x)).$$

Proposition 1.4.3 *Let* ν *be an arbitrary initial distribution. If* $\pi = \{\varphi_n\}$ *is a randomized Markov policy, then* $\{x_n\}$ *is a nonhomogeneus Markov process with transition kernels* $\{Q_{\varphi_n}(\cdot|\cdot)\}$, *that is, for every* $B \in \mathcal{B}(\mathbf{X})$ *and* $n = 0, 1, \cdots$,

$$\begin{aligned} P_\nu^\pi(x_{n+1} \in B|x_0, \cdots, x_n) &= P_\nu^\pi(x_{n+1} \in B|x_n) \\ &= Q_{\varphi_n}(B|x_n) \end{aligned} \tag{1.4.7}$$

For a proof of Proposition 1.4.3 see [14, p. 19-20].

In Proposition 1.4.3, let $\pi = \varphi$ be a stationary policy. The n-step transition probabilities are denoted by Q_φ^n, that is

$$Q_\varphi^n(B|x) := P_x^\varphi(x_n \in B), \quad n \in \mathbb{N}_0, B \in \mathcal{B}(\mathbf{X}), x \in \mathbf{X}, \tag{1.4.8}$$

with $Q^1_\varphi(\cdot|x) := Q_\varphi(\cdot|x)$ and $Q^0_\varphi(\cdot|x) = \delta_x$, the Dirac measure concentrated at the initial state x. We can write Q^n_φ recursively as

$$\begin{aligned} Q^n_\varphi(B|x) &= \int_{\mathbf{X}} Q_\varphi(B|y) Q^{n-1}_\varphi(dy|x) \\ &= \int_{\mathbf{X}} Q^{n-1}_\varphi(B|y) Q_\varphi(dy|x), \quad n \geq 1. \end{aligned} \tag{1.4.9}$$

1.5 Weighted-norm spaces

Let X be a metric space, and let $B_b(X)$ be the Banach space of real-valued measurable bounded functions u on X, with the *supremum norm*

$$\|u\| := \sup_{x \in X} |u(x)|.$$

We denote by $C_b(X)$ the closed subspace of $B_b(X)$ of all continuous bounded functions on X.

We assume throughout the following that $W : X \to [\theta, \infty)$ denotes a given measurable function that will be referred to as a *weight function*, where $\theta > 0$. If u is a real-valued function on X, we define its *W-norm* as

$$\|u\|_W := \sup_{x \in X} |u(x)|/W(x). \tag{1.5.1}$$

Of course, if W is the constant function $W(\cdot) \equiv 1$, the W-norm and the supremum norm coincide.

A real-valued function u on X is said to be *bounded* if $\|u\| < \infty$ and *W-bounded* if $\|u\|_W < \infty$. In general, the weight function W will be unbounded, although it is obviously W-bounded since $\|W\|_W = 1$. On the other hand, if u is a bounded function then it is W-bounded, since $W \geq \theta$ yields

$$\|u\|_W \leq \frac{1}{\theta}\|u\| < \infty \quad \forall u \in B_b(X). \tag{1.5.2}$$

Let $B_W(X)$ be the normed linear space of W-bounded real-valued measurable functions u on X. This space is also a Banach space because if $\{u_n\}$ is a Cauchy sequence in the W-norm, then $\{u_n/W\}$ is Cauchy in the supremum norm; hence, as $B_b(X)$ is a Banach space, one can deduce the existence of a function u in $B_W(X)$ that is the W-limit of $\{u_n\}$. Combining this fact and (1.5.2) we obtain the following:

Proposition 1.5.1 *$B_W(X)$ is a Banach space that contains $B_b(X)$.*

We denote by $C_W(X)$ the linear subspace of $B_W(X)$ that consists of the continuous function on X.

Chapter 2

The optimality equation

For every policy $\pi \in \Pi$ and initial state $x \in \mathbf{X}$, let

$$J(\pi, x) := \liminf_{n\to\infty} \frac{1}{n} E_x^\pi \Big[\sum_{k=0}^{n-1} r(x_k, a_k) \Big]$$

be the corresponding long-run expected average reward (EAR). In this chapter we consider the so-called *EAR control problem* in which we wish to maximize $\pi \mapsto J(\pi, x)$ over all $\pi \in \Pi$. More precisely, we wish to find $\pi^* \in \Pi$ such that

$$J(\pi^*, x) = \sup_{\pi\in\Pi} J(\pi, x) \quad \forall x \in \mathbf{X}.$$

Our goal is to characterize such EAR-optimal policies π^* and to give conditions ensuring the existence of an EAR-optimal stationary policy. This is a standard result that can be obtained in a variety of ways. Here, we follow the approach by Vega-Amaya [30], based on "fixed point arguments" to obtain solutions to the Poisson equation (P.E.) (see Theorem 2.1.4) and to the Average Reward Optimality Equation (AROE) (see Theorem 2.4.3). These results are used troughout the rest of this work. We extend the results obtained by Vega-Amaya [30] to the set of all randomized stationary policies Φ. For completeness we prove these results, although the proofs are just slight modifications of those in [30].

2.1 The Poisson equation

In this section we study the Poisson equation (P.E.) in (2.1.4) below. In particular, we prove, under suitable conditions, the existence of solutions to the P.E. To this end, we shall introduce two sets of hypotheses. The first one, Assumption 2.1.1 below, uses a weight function W to impose a growth condition on the reward function. The second one, Assumption 2.1.2, imposes a Lyapunov condition that will yield that certain operator is a contraction on the space $B_W(\mathbf{X})$, defined in Section 1.5. The Assumptions 2.1.1, 2.1.2 will ensure the existence of a fixed point for this operator (see Lemma 2.2.1), which in turn yields the solution to the P.E.

Assumption 2.1.2 was previously used for Markov control processes on Borel spaces (see, for instance, [9], [10], [20] and [19]) but it was combined with additional conditions that imply W-geometric ergodicity. The approach in this section is quite different, because the basic idea is to use Banach's fixed point theorem and we do not need to introduce W-geometric ergodicity.

Let $(\mathbf{X}, A, \{A(x) : x \in \mathbf{X}\}, Q, r)$ be a Markov control model as defined in Section 1.3. The function W in the following assumption will play the role of a weight function, as in Section 1.5.

Assumption 2.1.1 *There exist a constant $K > 0$ and a measurable function $W(\cdot)$ on $\mathbf{X}$ such that:*

(a) *W is bounded below by a constant $\theta > 0$.*

(b) *$|r(x, a)| \leq KW(x)$ for all $(x, a) \in \mathbb{K}$.*

Let $\gamma(\cdot)$ be a measure on $\mathbf{X}$. We write

$$\gamma(u) := \int_{\mathbf{X}} u(x)\gamma(dx)$$

whenever the integral is well-defined. We will now state our second main assumption:

Assumption 2.1.2 *There exists a non-trivial finite measure $\nu(\cdot)$ on $\mathbf{X}$, a nonnegative measurable function $l(\cdot, \cdot)$ on $\mathbb{K}$ and a positive constant $\lambda < 1$ such that:*

(a) *$\nu(W) < \infty$.*

(b) $Q(\cdot|x,a) \geq l(x,a)\nu(\cdot)$ *for each* $(x,a) \in \mathbb{K}$.

(c) $\int_{\mathbf{X}} W(y)Q(dy|x,a) \leq \lambda W(x) + l(x,a)\nu(W)$ *for each* $(x,a) \in \mathbb{K}$.

(d) $\nu(l_\varphi) > 0$ *for each* $\varphi \in \Phi$, *with* $l_\varphi(\cdot)$ *as in (1.4.5) with* l *in lieu of* c.

Remark 2.1.3 *Assumption 2.1.1*(a) *and iterations of the inequality in Assumption 2.1.2*(c) *yield, for every* $x \in \mathbf{X}, \pi \in \Pi$, *and* $n = 0, 1, \cdots$,

$$\theta \leq E_x^\pi W(x_n) \leq \lambda^n W(x) + \frac{\nu(W)}{(1-\lambda)\nu(\mathbf{X})}. \tag{2.1.1}$$

This fact and (1.5.1) yield that for every $u \in B_W(\mathbf{X})$

$$\lim_{n\to\infty} \frac{1}{n} E_x^\pi |u(x_n)| = 0. \tag{2.1.2}$$

We can now state our first main result, where we use the notation r_φ and Q_φ introduced in (1.4.5) and (1.4.6), respectively.

Theorem 2.1.4 *Under Assumptions 2.1.1 and 2.1.2 the following facts hold for each* $\varphi \in \Phi$*:*

(i) *The Markov chain defined by* $Q_\varphi(\cdot|\cdot)$ *is* ν*-irreducible and positive Harris recurrent; hence it admits a unique invariant probability measure (i.p.m.), say* μ_φ.

(ii) $\mu_\varphi(W) < \infty$*; thus we have*

$$\rho_\varphi := \mu_\varphi(r_\varphi) < \infty, \quad \rho^* := \sup_{\varphi\in\Phi} \rho_\varphi < \infty. \tag{2.1.3}$$

(iii) *There exist a function* h_φ^* *in* $B_W(\mathbf{X})$ *such that the pair* $(\rho_\varphi, h_\varphi^*)$ *satisfies the P.E.*

$$h_\varphi^*(x) = r_\varphi(x) - \rho_\varphi + \int_{\mathbf{X}} h_\varphi^*(y) Q_\varphi(dy|x) \quad \forall x \in \mathbf{X}, \tag{2.1.4}$$

and, moreover, $\nu(h_\varphi^*) = 0$.

Since the proof of Theorem 2.1.4 is a bit long, we postpone it to the next section.

2.2 Proof of Theorem 2.1.4

Before proving Theorem 2.1.4 we shall introduce some concepts and preliminary results.

Define

$$\widehat{Q}(B|x,a) := Q(B|x,a) - \nu(B)l(x,a), \tag{2.2.1}$$

for each $B \in \mathcal{B}(\mathbf{X})$ and $(x,a) \in \mathbb{K}$. Under Assumption 2.1.2(b), $\widehat{Q}$ is a non-negative kernel on $\mathbf{X}$ given $\mathbb{K}$, and from Assumption 2.1.2(c) $\widehat{Q}$ is contractive in the sense that

$$\int_{\mathbf{X}} W(y)\widehat{Q}(dy|x,a) \leq \lambda W(x) \quad \forall (x,a) \in \mathbb{K}. \tag{2.2.2}$$

Let us fix $\varphi \in \Phi$ and $v \in B_W(\mathbf{X})$ and define $L^v_\varphi : B_W(\mathbf{X}) \to B_W(\mathbf{X})$ by

$$\begin{aligned} L^v_\varphi u(x) &:= v(x) + \int_{\mathbf{X}} u(y)\widehat{Q}_\varphi(dy|x) \\ &= v(x) + \int_{\mathbf{X}} u(y)Q_\varphi(dy|x) - \nu(u)l_\varphi(x) \end{aligned} \tag{2.2.3}$$

for every $x \in \mathbf{X}, u \in B_W(\mathbf{X})$.

Lemma 2.2.1 *Suppose that Assumptions 2.1.1 and 2.1.2 hold. Then for each $\varphi \in \Phi$ and $v \in B_W(\mathbf{X})$, the operator L^v_φ is a contraction with modulus λ on the Banach space $B_W(\mathbf{X})$. Hence, by Banach's fixed point theorem, there exists a unique function $h^v_\varphi \in B_W(\mathbf{X})$ such that $L^v_\varphi h^v_\varphi = h^v_\varphi$, i.e.,*

$$h^v_\varphi(x) = v(x) + \int_{\mathbf{X}} h^v_\varphi(y)Q_\varphi(dy|x) - \nu(h^v_\varphi)l_\varphi(x) \quad \forall x \in \mathbf{X}. \tag{2.2.4}$$

Proof. From (2.2.2), it can be verified that L^v_φ is a contraction on the Banach space $B_W(\mathbf{X})$. Therefore, by Banach's fixed point theorem there is a unique function $h^v_\varphi \in B_W(\mathbf{X})$ satisfying (2.2.4). ■

The following remark is used throughout the rest of this work.

Remark 2.2.2 *Banach's fixed point theorem gives an explicit formula for h^v_φ in Lemma 2.2.1:*

$$h^v_\varphi(x) = \sum_{n=0}^{\infty} \int_{\mathbf{X}} v(y)\widehat{Q}^n_\varphi(dy|x) \quad \forall x \in \mathbf{X}. \tag{2.2.5}$$

The series in (2.2.5) converges absolutely in the Banach space $B_W(\mathbf{X})$. Moreover, we can define an endomorphism $H_\varphi : B_W(\mathbf{X}) \to B_W(\mathbf{X})$, with $H_\varphi v := h^v_\varphi$. By (2.2.5), we have

$$H_\varphi v(x) = h^v_\varphi(x) = \sum_{n=0}^{\infty} \int_{\mathbf{X}} v(y) \widehat{Q}^n_\varphi(dy|x) \quad \forall x \in \mathbf{X}, v \in B_W(\mathbf{X}). \tag{2.2.6}$$

On the other hand, one can show that, for each $x \in \mathbf{X}$, the series

$$P_\varphi(dy|x) := \sum_{n=0}^{\infty} \widehat{Q}^n_\varphi(dy|x) \tag{2.2.7}$$

converges in the Banach space of finite signed measures on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$. Thus we may rewrite (2.2.6) as

$$H_\varphi v(x) = h^v_\varphi(x) = \int_{\mathbf{X}} v(y) P_\varphi(dy|x) \quad \forall x \in \mathbf{X}, v \in B_W(\mathbf{X}). \tag{2.2.8}$$

where $P_\varphi(dy|x)$ is the kernel on $\mathbf{X}$ defined in (2.2.7).

Note that

$$\|H_\varphi\|_W \leq 1/(1-\lambda). \tag{2.2.9}$$

Finally, H_φ preserves order, that is, if $u \leq v$ then $h^u_\varphi \leq h^v_\varphi$ for all u and v in $B_W(\mathbf{X})$.

Lemma 2.2.3 *If Assumptions 2.1.1 and 2.1.2 are satisfied, then for each $\varphi \in \Phi$*

$$\lim_{n\to\infty} \frac{1}{n} E^\varphi_x \sum_{k=0}^{n-1} l_\varphi(x_k) = \frac{1}{\nu(h^1_\varphi)} > 0, \tag{2.2.10}$$

where h^1_φ is the function h^v_φ in (2.2.4) with $v \equiv 1$, i.e. by (2.2.6)–(2.2.7),

$$h^1_\varphi(\cdot) := H_\varphi(1_{\mathbf{X}})(\cdot) = P_\varphi(\mathbf{X}|\cdot). \tag{2.2.11}$$

Proof. Let $\varphi \in \Phi$ be arbitrary and take $v \equiv 1$. By Lemma 2.2.1, there exists a unique function $h^1_\varphi \in B_W(\mathbf{X})$ such that

$$h^1_\varphi(x) = 1 + \int_{\mathbf{X}} h^1_\varphi(y) Q_\varphi(dy|x) - \nu(h^1_\varphi) l_\varphi(x) \quad \forall x \in \mathbf{X}.$$

By an iteration procedure we obtain

$$h^1_\varphi(x) = n - \nu(h^1_\varphi)E^\varphi_x \sum_{k=0}^{n-1} l_\varphi(x_k) + E^\varphi_x h^1_\varphi(x_n) \quad \forall x \in \mathbf{X}, n = 1, \cdots.$$

Multiplying by $1/n$ and letting $n \to \infty$, it follows from (2.1.2) that

$$\nu(h^1_\varphi) \lim_{n\to\infty} \frac{1}{n} E^\varphi_x \sum_{k=0}^{n-1} l_\varphi(x_k) = 1 \quad \forall x \in \mathbf{X},$$

which implies (2.2.10). ■

We are ready for the proof of Theorem 2.1.4.

Proof of Theorem 2.1.4 Consider an arbitrary stationary policy $\varphi \in \Phi$.

(i) By Lemma 2.2.1, for each $v \in B_W(\mathbf{X})$, there exists a unique function $h^v_\varphi \in B_W(\mathbf{X})$ satisfying (2.2.4), i.e.,

$$h^v_\varphi(x) = v(x) + \int_{\mathbf{X}} h^v_\varphi(y) Q_\varphi(dy|x) - \nu(h^v_\varphi) l_\varphi(x) \quad \forall x \in \mathbf{X}.$$

Thus, by iteration, we have

$$h^v_\varphi(x) = E^\varphi_x \sum_{k=0}^{n-1} v(x_k) - \nu(h^v_\varphi) E^\varphi_x \sum_{k=0}^{n-1} l_\varphi(x_k) + E^\varphi_x h^v_\varphi(x_n)$$

for every $x \in \mathbf{X}$, and $n = 1, \cdots$. Hence, by (2.1.2) and Lemma 2.2.3, we obtain

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} E^\varphi_x \sum_{k=0}^{n-1} v(x_k) &= \nu(h^v_\varphi) \lim_{n\to\infty} \frac{1}{n} E^\varphi_x \sum_{k=0}^{n-1} l_\varphi(x_k) \\ &= \frac{1}{\nu(h^1_\varphi)} \nu(H_\varphi v) < \infty \end{aligned} \tag{2.2.12}$$

for all $v \in B_W(\mathbf{X})$. Now, in (2.2.12) take $v \equiv 1_B$, with $B \in \mathcal{B}(\mathbf{X})$. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Q^k_\varphi(B|x) = \lim_{n\to\infty} \frac{1}{n} E^\varphi_x \sum_{k=0}^{n-1} 1_B(x_k) = \frac{\nu(h^{1_B}_\varphi)}{\nu(h^1_\varphi)} < \infty \quad \forall x \in \mathbf{X}.$$

This result and [16, Theorem 4.3.1] give the following facts:

(I) The transition probability $Q_\varphi(\cdot|\cdot)$ is positive Harris recurrent; hence, it is irreducible and it has a unique i.p.m. μ_φ.

(II) For any bounded measurable function v on $\mathbf{X}$,

$$\lim_{n\to\infty}\frac{1}{n}E_x^\varphi\sum_{k=0}^{n-1}v(x_k)=\mu_\varphi(v)\quad\forall x\in\mathbf{X}, \tag{2.2.13}$$

which implies, with $v\equiv l_\varphi$,

$$\lim_{n\to\infty}\frac{1}{n}E_x^\varphi\sum_{k=0}^{n-1}l_\varphi(x_k)=\mu_\varphi(l_\varphi).$$

By (2.2.10), we obtain

$$\mu_\varphi(l_\varphi)=\frac{1}{\nu(h_\varphi^1)}>0. \tag{2.2.14}$$

Therefore, to complete the proof of part (i) it suffices to show that $\nu(\cdot)$ is an irreducibility measure. In fact, by a characterization of ν-irreducibility (see, for instance,[22, Proposition 4.2.1]), we only need to prove that for all $x\in\mathbf{X}$, whenever $\nu(B)>0$, there exists some $m>0$, possibly depending on φ, B and x, such that $Q_\varphi^m(B|x)>0$. Let $B\in\mathcal{B}(\mathbf{X})$ be such that $\nu(B)>0$. By Assumption 2.1.2(b) and the invariance of μ_φ yield

$$\mu_\varphi(B)\geq\nu(B)\mu_\varphi(l_\varphi).$$

This inequality together with $\mu_\varphi(l_\varphi)>0$ (see (2.2.14)) gives that $\mu_\varphi(B)>0$. Hence, by (2.2.13) with $v\equiv 1_B$, we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}Q_\varphi^k(B|x)=\mu_\varphi(B)>0.$$

This implies the existence of $m>0$ such that $Q_\varphi^m(B|x)>0$, and so the desired result follows.

(ii) By (2.2.12) and (2.2.13), we see that every bounded measurable function v on $\mathbf{X}$ satisfies the formula

$$\mu_\varphi(v)=\frac{\nu(h_\varphi^v)}{\nu(h_\varphi^1)}=\frac{1}{\nu(h_\varphi^1)}\nu(H_\varphi v). \tag{2.2.15}$$

Hence, for every nonnegative bounded measurable function w satisfying $0 \leq w \leq W$, with W as in Assumption 2.1.1

$$\mu_\varphi(w) = \frac{1}{\nu(h^1_\varphi)}\nu(H_\varphi w) \leq \frac{1}{\nu(h^1_\varphi)}\nu(H_\varphi W)$$

because H_φ preserves order. Since any nonnegative measurable function is the limit of a nondecreasing sequence of nonnegative bounded measurable functions, together with the monotone convergence theorem (see, for instance, [25]), we obtain

$$\mu_\varphi(W) \leq \frac{1}{\nu(h^1_\varphi)}\nu(H_\varphi W) < \infty.$$

Thus, to complete the proof of part (ii), it only remains to verify (2.1.3). Assumption 2.1.2(b) yields $l(x,a) \leq 1/\nu(\mathbf{X})$ for each $(x,a) \in \mathbb{K}$, and so

$$\mu_\varphi(l_\varphi) \leq \frac{1}{\nu(\mathbf{X})} \quad \forall \varphi \in \Phi.$$

On the other hand, Assumption 2.1.2(c) implies $\int_{\mathbf{X}} W(y)Q_\varphi(dy|x) \leq \lambda W(x) + l_\varphi(x)\nu(W)$. Integrating both sides of this inequality with respect to μ_φ, and using that $\mu_\varphi(W) < \infty$, we obtain

$$\mu_\varphi(W) \leq \frac{\mu_\varphi(l_\varphi)\nu(W)}{1-\lambda} \leq \frac{\nu(W)}{(1-\lambda)\nu(\mathbf{X})}. \tag{2.2.16}$$

Assumption 2.1.1(b) yields

$$\rho_\varphi = \mu_\varphi(r_\varphi) \leq K\mu_\varphi(W) \quad \forall \varphi \in \Phi. \tag{2.2.17}$$

Thus, by (2.2.16),

$$\rho^* := \sup_{\varphi\in\Phi} \rho_\varphi \leq K \sup_{\varphi\in\Phi} \mu_\varphi(W) \leq \frac{K\nu(W)}{(1-\lambda)\nu(\mathbf{X})} < \infty.$$

This completes the proof of part (ii).

(iii) To prove this part let us take $v = r_\varphi - \rho_\varphi$ in Lemma 2.2.1 to obtain a function h^v_φ satisfying (2.2.4). Define $h^*_\varphi := h^v_\varphi$ for this particular v. Then we obtain

$$h^*_\varphi(x) = r_\varphi(x) - \rho_\varphi + \int_{\mathbf{X}} h^*_\varphi(y)Q_\varphi(dy|x) - \nu(h^*_\varphi)l_\varphi(x) \tag{2.2.18}$$

for every $x \in \mathbf{X}$. Integrating both sides of (2.2.18) with respect to the i.p.m. μ_φ, we can see that

$$\nu(h_\varphi^*)\mu_\varphi(l_\varphi) = 0.$$

However, since $\mu_\varphi(l_\varphi) > 0$ (see (2.2.14)), the latter relation yields that $\nu(h_\varphi^*) = 0$ and, hence, (2.2.18) reduces to the P.E. (2.1.4). ∎

Remark 2.2.4 *The unique i.p.m. μ_φ in Theorem 2.1.4*(i) *can be expressed as*

$$\mu_\varphi(\cdot) = \frac{1}{\nu(h_\varphi^1)} \int_{\mathbf{X}} P_\varphi(\cdot|x)\nu(dx). \tag{2.2.19}$$

Indeed, since $\mu_\varphi(W) < \infty$, from Lebesgue's dominated convergence theorem, we can see that (2.2.15) is satisfied for every function v in $B_W(\mathbf{X})$:

$$\begin{aligned} \mu_\varphi(v) &= \frac{1}{\nu(h_\varphi^1)} \int_{\mathbf{X}} \left[\int_{\mathbf{X}} v(y) P_\varphi(dy|x) \right] \nu(dx) \\ &= \frac{1}{\nu(h_\varphi^1)} \nu(H_\varphi v). \end{aligned} \tag{2.2.20}$$

Moreover, comparing (2.2.20) and (2.2.12), we obtain the limit

$$\lim_{n\to\infty} \frac{1}{n} E_x^\varphi \sum_{k=0}^{n-1} v(x_k) = \mu_\varphi(v) \quad \forall v \in B_W(\mathbf{X}), \tag{2.2.21}$$

which gives (2.2.19).

2.3 Uniqueness of solutions to the P.E.

To prove uniqueness of the functions h_φ^* satisfying the P.E. (2.1.4), we need the following lemma.

Lemma 2.3.1 *Suppose that Assumptions 2.1.1 and 2.1.2 hold. Let $v, h, \tilde{h}$ be functions belonging to $B_W(\mathbf{X})$, and $\varphi \in \Phi$. Suppose that*

$$h(x) = v(x) + \int_{\mathbf{X}} h(y) Q_\varphi(dy|x) \quad \forall x \in \mathbf{X},$$

and

$$\tilde{h}(x) = v(x) + \int_{\mathbf{X}} \tilde{h}(y) Q_\varphi(dy|x) \quad \forall x \in \mathbf{X}.$$

Then $h(\cdot)$ and $\tilde{h}(\cdot)$ differ by a constant, i.e.,

$$h(x) - \tilde{h}(x) = c(\varphi) \quad \forall x \in \mathbf{X},$$

where $c(\varphi)$ is the constant $c(\varphi) := \mu_\varphi(h - \tilde{h})$.

Proof. Let $u := h - \tilde{h}$. The hypotheses yield $u(x) = \int_{\mathbf{X}} u(y) Q_\varphi(dy|x)$ for all $x \in \mathbf{X}$. Hence, by induction,

$$u(x) = \int_{\mathbf{X}} u(y) Q_\varphi^n(dy|x) \quad \forall x \in \mathbf{X}, n = 0, 1, \cdots. \tag{2.3.1}$$

By (2.2.21), as $n \to \infty$ we obtain

$$u(x) = \frac{1}{n} E_x^\varphi \sum_{k=0}^{n-1} u(x_k) \to \mu_\varphi(u) \quad \forall x \in \mathbf{X}.$$

Thus, $u(x) = h(x) - \tilde{h}(x) = \mu_\varphi(u)$ for all $x \in \mathbf{X}$. ■

Proposition 2.3.2 *With the notation of Theorem 2.1.4, if $h \in B_W(\mathbf{X})$ is a function satisfying the P.E. (2.1.4), then $h(x) - h_\varphi^*(x) = c(\varphi)$ for each $x \in \mathbf{X}$ and some constant $c(\varphi)$. Moreover, $h = h_\varphi^*$ iff $\nu(h) = 0$. In particular, h_φ^* is the unique function in $B_W(\mathbf{X})$ satifying the P.E. and such that $\nu(h_\varphi^*) = 0$.*

Proof. This result follows from Lemma 2.3.1

Remark 2.3.3 *The functions h_φ^* in Theorem 2.1.4*(iii)*, satisfying the P.E. (2.1.4), are defined in similar form as the ones in [18, Lemma 4.1]. Actually, by (2.2.5), we can write h_φ^* as*

$$h_\varphi^*(x) = \sum_{n=0}^{\infty} \int_{\mathbf{X}} [r_\varphi(y) - \rho_\varphi] \widehat{Q}_\varphi^n(dy|x) = \int_{\mathbf{X}} [r_\varphi(y) - \rho_\varphi] P_\varphi(dy|x) \tag{2.3.2}$$

for all $x \in \mathbf{X}$, with $P_\varphi(dy|x)$ as in (2.2.7), while in [18], due to W-geometric ergodicity, the corresponding function is given by

$$h_\varphi^*(x) = \sum_{n=0}^{\infty} \int_{\mathbf{X}} [r_\varphi(y) - \rho_\varphi] Q_\varphi^n(dy|x) \quad \forall x \in \mathbf{X}.$$

2.4 The optimality equation

In this section, we characterize optimal policies by means of the average reward optimality equation (AROE). To this end, we introduce the *long-run expected average reward* per unit-time criterion, hereafter abbreviated *average reward criterion*, which is defined as follows.

Definition 2.4.1 *Let* $(\mathbf{X}, A, \{A(x) : x \in \mathbf{X}\}, Q, r)$ *be a given MCM (see Section 1.3 above), and let*

$$J_n(\pi, x) := E_x^\pi\left[\sum_{k=0}^{n-1} r(x_k, a_k)\right] \tag{2.4.1}$$

be the total expected n-stage reward *when using the policy* π*, given the initial state* $x_0 = x$*. Then the* long-run expected average reward (EAR) *when using* $\pi \in \Pi$*, given* $x_0 = x$*, is*

$$J(\pi, x) := \liminf_{n\to\infty} J_n(\pi, x)/n. \tag{2.4.2}$$

The EAR problem *is to find a policy* π^* *such that*

$$J(\pi^*, x) := \sup_{\pi\in\Pi} J(\pi, x) =: J^*(x) \quad \forall x \in \mathbf{X}. \tag{2.4.3}$$

A policy π^* *that satisfies (2.4.3) is said to be* EAR-optimal *and* $J^*(\cdot)$ *is called the* EAR-value function*.*

In contrast to (2.4.2), we define

$$\bar{J}(\pi, x) := \limsup_{n\to\infty} J_n(\pi, x)/n. \tag{2.4.4}$$

Note that

$$\bar{J}(\pi, x) \geq J(\pi, x)$$

for every control policy $\pi \in \Pi$ and state $x \in \mathbf{X}$.

In addition to Assumptions 2.1.1 and 2.1.2, we next impose other conditions on the control model. Several versions of these conditions have appeared in the literature (see, for instance, [18, 20, 19, 29, 30]), but the main ideas go back to [9, 10].

Assumption 2.4.2 *For each* $x \in \mathbf{X}$*:*

(a) *$A(x)$ is a (nonempty) compact subset of $\mathbf{A}$.*

(b) *$r(x,\cdot)$ is upper semicontinuous (u.s.c.) on $A(x)$.*

(c) *$Q(\cdot|x,\cdot)$ is strongly continuous on $A(x)$, that is, the mapping*

$$a \to \int_{\mathbf{X}} u(y)Q(dy|x,a)$$

is continuous on $A(x)$ for each bounded measurable function u on $\mathbf{X}$.

(d) *The mapping $a \to \int_{\mathbf{X}} W(y)Q(dy|x,a)$ is continuous on $A(x)$, with W as in Assumption 2.1.1.*

(e) *$l(x,\cdot)$ is continuous on $A(x)$, with $l(\cdot,\cdot)$ as in Assumption 2.1.2.*

The next theorem establishes the existence of solutions to the so–called *average reward optimality equation* (AROE) in (2.4.5) below. Moreover, it characterizes optimal policies by means of the AROE.

Theorem 2.4.3 *Suppose that Assumptions 2.1.1, 2.1.2 and 2.4.2 hold. Then:*

(i) *There exists a triplet (h^*, f^*, ρ^*), with $h^* \in B_W(\mathbf{X})$, $f^* \in \mathbb{F}$, and ρ^* as in (2.1.3), that satisfies the AROE*

$$\begin{aligned} h^*(x) &= \sup_{a\in A(x)} \left[r(x,a) - \rho^* + \int_{\mathbf{X}} h^*(y)Q(dy|x,a) \right] \qquad (2.4.5) \\ &= r_{f^*}(x) - \rho^* + \int_{\mathbf{X}} h^*(y)Q_{f^*}(dy|x) \quad \forall x \in \mathbf{X}. \end{aligned}$$

(ii) *Moreover,*

$$\rho^* = J(f^*, x) \geq \bar{J}(\pi, x) \geq J(\pi, x)$$

for all $x \in \mathbf{X}$ and $\pi \in \Pi$. Hence, the constant $\rho^* = J^*(x)$ is the EAR-value and f^* is an EAR-optimal policy.

We give a proof (in Section 2.6) taken from [30, Theorems 3.10 and 3.12].

2.5 Preliminary results

We define, for each $u \in B_W(\mathbf{X})$ and $x \in \mathbf{X}$

$$\widehat{T}_* u(x) := \sup_{a \in A(x)} \left[r(x,a) - \rho^* + \int_{\mathbf{X}} u(y) \widehat{Q}(dy|x,a) \right] \tag{2.5.1}$$

with $\widehat{Q}$ as in (2.2.1).

To prove Theorem 2.4.3 we need the two following lemmas. These are standard dynamic programming results, but we state them here (including the proof of Lemma 2.5.2) for completeness and ease of reference.

Lemma 2.5.1 *Suppose that Assumptions 2.1.1, 2.1.2 and 2.4.2 hold. Then for each $u \in B_W(\mathbf{X})$ there exists $f \in \mathbb{F}$ such that*

$$\widehat{T}_* u(x) = r_f(x) - \rho^* + \int_{\mathbf{X}} u(y) \widehat{Q}_f(dy|x) \quad \forall x \in \mathbf{X}. \tag{2.5.2}$$

Hence, $\widehat{T}_* u$ is measurable and it belongs to $B_W(\mathbf{X})$.

Proof. See [17, Proposition 2.6]. ∎

Lemma 2.5.2 *Suppose that Assumptions 2.1.1, 2.1.2 and 2.4.2 hold. In addition, suppose that there exists a function $h^* \in B_W(\mathbf{X})$ and a constant ρ^0 satisfying*

$$h^*(x) = \sup_{a \in A(x)} \left[r(x,a) - \rho^0 + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) \right] \quad \forall x \in \mathbf{X}. \tag{2.5.3}$$

Then we have

$$\rho^0 \geq \bar{J}(\pi, x) \geq J(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi. \tag{2.5.4}$$

Proof. Let $\pi \in \Pi$ be an arbitrary policy. Recall from (1.4.4) that

$$P^\pi_{\nu_0}(dx_0, da_0, dx_1, da_1, \cdots) = \nu_0(dx_0)\pi_0(da_0|x_0)Q(dx_1|x_0,a_0)\pi_1(da_1|h_1)\cdots$$

where ν_0 is the initial distribution.

The formula (2.5.3) gives us

$$h^*(x) + \rho^0 \geq r(x,a) + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) \quad \forall (x,a) \in \mathbb{K}.$$

Taking $\nu_0(dx_0) = \delta_x(dx_0)$, the latter inequality yields

$$\int_{\mathbf{X}} h^*(x_0)\delta_x(dx_0)\pi_0(da_0|x_0)+\rho^0 \geq \int_{\mathbf{X}} r(x_0,a_0)\delta_x(dx_0)\pi_0(da_0|x_0) + \int_{\mathbf{X}} h^*(x_1)\delta_x(dx_0)\pi_0(da_0|x_0)Q(dx_1|x_0,a_0).$$

That is

$$E_x^\pi h^*(x_0) + \rho^0 \geq E_x^\pi r(x_0,a_0) + E_x^\pi h^*(x_1).$$

In general, a similar procedure for each $n = 1, 2, \cdots$, gives,

$$E_x^\pi h^*(x_{n-1}) + \rho^0 \geq E_x^\pi r(x_{n-1},a_{n-1}) + E_x^\pi h^*(x_n).$$

Iterations of this inequality yield

$$E_x^\pi \sum_{k=0}^{n-1} h^*(x_k) + n\rho^0 \geq E_x^\pi \sum_{k=0}^{n-1} r(x_k,a_k) + E_x^\pi \sum_{k=1}^{n} h^*(x_k),$$

or, equivalently,

$$h^*(x) - E_x^\pi h^*(x_n) + n\rho^0 \geq E_x^\pi \sum_{k=0}^{n-1} r(x_k,a_k).$$

Multiplying by $1/n$ both sides of the latter inequality , gives

$$\frac{1}{n}h^*(x) - \frac{1}{n}E_x^\pi h^*(x_n) + \rho^0 \geq \frac{1}{n}E_x^\pi \sum_{k=0}^{n-1} r(x_k,a_k).$$

Therefore, as $n \to \infty$, (2.1.2) gives

$$\rho^0 \geq \limsup_{n\to\infty} \frac{1}{n}E_x^\pi \sum_{k=0}^{n-1} r(x_k,a_k) \geq \liminf_{n\to\infty} \frac{1}{n}E_x^\pi \sum_{k=0}^{n-1} r(x_k,a_k).$$

Thus

$$\rho^0 \geq \bar{J}(\pi,x) \geq J(\pi,x)$$

for each $x \in \mathbf{X}$ and $\pi \in \Pi$. ■

2.6 Proof of Theorem 2.4.3

We are ready for the proof of Theorem 2.4.3

Proof of Theorem 2.4.3 (i) By Lemma 2.5.1, $\widehat{T}_*$ is a mapping from $B_W(\mathbf{X})$ into itself. We assert that $\widehat{T}_*$ is a contraction. To this end, let u be an arbitrary function in $B_W(\mathbf{X})$ and define

$$Lu(x,a) := r(x,a) - \rho^* + \int_{\mathbf{X}} u(y)\widehat{Q}(dy|x,a) \quad \forall (x,a) \in \mathbb{K}.$$

If $v \in B_W(\mathbf{X})$ is another function, then

$$\begin{aligned} |Lu(x,a) - Lv(x,a)| &\leq \|u-v\|_W \int_{\mathbf{X}} W(y)\widehat{Q}(dy|x,a) \\ &\leq \lambda \|u-v\|_W W(x) \end{aligned}$$

This implies

$$\widehat{T}_* u(x) = \sup_{a\in A(x)} Lu(x,a) \leq \sup_{a\in A(x)} Lv(x,a) + \lambda\|u-v\|_W W(x);$$

hence

$$\widehat{T}_* u(x) \leq \widehat{T}_* v(x) + \lambda\|u-v\|_W W(x).$$

By symmetry

$$\widehat{T}_* v(x) \leq \widehat{T}_* u(x) + \lambda\|u-v\|_W W(x).$$

These two inequalities imply

$$|\widehat{T}_* u(x) - \widehat{T}_* v(x)| \leq \lambda\|u-v\|_W W(x);$$

therefore

$$\|\widehat{T}_* u - \widehat{T}_* v\|_W \leq \lambda\|u-v\|_W \quad \forall u, v \in B_W(\mathbf{X}). \tag{2.6.1}$$

Thus, $\widehat{T}_*$ is a contraction on $B_W(\mathbf{X})$, and again by Banach's fixed point theorem there exists a unique $h^* \in B_W(\mathbf{X})$ satisfying $\widehat{T}_* h^* = h^*$, that is

$$h^*(x) = \sup_{a\in A(x)} \left[r(x,a) - \rho^* + \int_{\mathbf{X}} h^*(y)\widehat{Q}(dy|x,a) \right] \quad \forall x \in \mathbf{X}.$$

On the other hand, by Lemma 2.5.1 again, there exists a function $f^* \in \mathbb{F}$ such that

$$\begin{aligned} h^*(x) &= \sup_{a \in A(x)} \left[r(x,a) - \rho^* + \int_{\mathbf{X}} h^*(y) \widehat{Q}(dy|x,a) \right] \\ &= r_{f^*}(x) - \rho^* + \int_{\mathbf{X}} h^*(y) \widehat{Q}_{f^*}(dy|x) \quad \forall x \in \mathbf{X}; \end{aligned}$$

thus

$$h^*(x) = r_{f^*}(x) - \rho^* + \int_{\mathbf{X}} h^*(y) Q_{f^*}(dy|x) - \nu(h^*) l_{f^*}(x) \quad \forall x \in \mathbf{X}.$$

Integrating both sides of the latter relation with respect to the invariant probability measure μ_{f^*}, we obtain $\nu(h^*)\mu_{f^*}(l_{f^*}) = \rho_{f^*} - \rho^* \leq 0$. By Assumptions 2.1.1(a) and 2.1.2(c) note that

$$\inf_{\varphi \in \Phi} \mu_\varphi(l_\varphi) \geq \frac{(1-\lambda)\theta}{\nu(W)} > 0. \tag{2.6.2}$$

Then $\nu(h^*) \leq 0$ because $\mu_{f^*}(l_{f^*}) > 0$. On the other hand, note that

$$h^*(x) \geq r(x,a) - \rho^* + \int_{\mathbf{X}} h^*(y) \widehat{Q}(dy|x,a) \quad \forall (x,a) \in \mathbb{K};$$

which in turn implies that

$$h^*(x) \geq r_\varphi(x) - \rho^* + \int_{\mathbf{X}} h^*(y) Q_\varphi(dy|x) - \nu(h^*) l_\varphi(x) \quad \forall x \in \mathbf{X}, \varphi \in \Phi.$$

Integrating again but now with respect to μ_φ we get

$$\nu(h^*)\mu_\varphi(l_\varphi) \geq \rho_\varphi - \rho^* \quad \forall \varphi \in \Phi;$$

by inequality (2.6.2) and the fact that $\nu(h^*) \leq 0$ we obtain

$$\nu(h^*) \geq \frac{(\rho_\varphi - \rho^*)\nu(W)}{(1-\lambda)\theta} \quad \forall \varphi \in \Phi$$

hence

$$\nu(h^*)(1-\lambda)\theta/\nu(W) + \rho^* \geq \sup_{\varphi \in \Phi} \rho_\varphi = \rho^*.$$

The latter inequality implies that $\nu(h^*) \geq 0$. It follows that $\nu(h^*) = 0$, and so the triplet (h^*, f^*, ρ^*) satisfies (2.4.5).

(ii) Integrating both sides of the AROE (2.4.5) with respect to the i.p.m. μ_{f^*}, we can see that $\rho^* = \rho_{f^*}$.

By (2.2.13)

$$J(\varphi, x) = \liminf_{n\to\infty} \frac{1}{n} E_x^{\varphi} \sum_{k=0}^{n-1} r_{\varphi}(x_k) = \mu_{\varphi}(r_{\varphi}) = \rho_{\varphi} \quad \forall x \in \mathbf{X}.$$

This fact and Lemma 2.5.2 yield

$$J(f^*, x) = \rho_{f^*} = \rho^* \geq \bar{J}(\pi, x) \geq J(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi;$$

therefore, $\rho^* = \sup_{\pi\in\Pi} J(\pi, x) = J^*(x)$ is the optimal value and f^* is EAR-optimal. ■

Remark 2.6.1 *With the notation of Theorem 2.4.3, the function $h^* \in B_W(\mathbf{X})$ and the invariant measure $\mu_{f^*} \in M_W(\mathbf{X})$ have explicit forms (see (2.3.2) and (2.2.19)). Indeed, because $\nu(h^*) = 0$, Proposition 2.3.2 yields $h^* = h_{f^*}$. Hence, by (2.3.2),*

$$h^*(x) = \sum_{n=0}^{\infty} \int_{\mathbf{X}} [r_{f^*}(y) - \rho^*] \widehat{Q}_{f^*}^n(dy|x) = \int_{\mathbf{X}} [r_{f^*}(y) - \rho^*] P_{f^*}(dy|x)$$

for all x in $\mathbf{X}$, and

$$\mu_{f^*}(\cdot) = \frac{1}{\nu(h_{f^*}^1)} \sum_{n=0}^{\infty} \int_{\mathbf{X}} \widehat{Q}_{f^*}^n(\cdot|x)\nu(dx) = \frac{1}{\nu(h_{f^*}^1)} \int_{\mathbf{X}} P_{f^*}(\cdot|x)\nu(dx)$$

with P_{f^*} as in (2.2.7), with $\varphi = f^*$.

Definition 2.6.2 *A (randomized) stationary policy $\tilde{\varphi} \in \Phi$ is called* canonical *if there exists a constant $\tilde{\rho}$ and a measurable function $\tilde{h} \in B_W(\mathbf{X})$ such that*

$$\tilde{\rho} + \tilde{h}(x) = \sup_{a\in A(x)} \Big[r(x, a) + \int_{\mathbf{X}} \tilde{h}(y) Q(dy|x, a) \Big] \quad \forall x \in \mathbf{X}, \tag{2.6.3}$$

and

$$\tilde{\rho} + \tilde{h}(x) = r_{\tilde{\varphi}}(x) + \int_{\mathbf{X}} \tilde{h}(y) Q_{\tilde{\varphi}}(dy|x) \quad \forall x \in \mathbf{X}. \tag{2.6.4}$$

If (2.6.3) and (2.6.4) are satisfied, then the triplet $(\tilde{\rho}, \tilde{h}, \tilde{\varphi})$ *is called a* canonical triplet. *Let* Φ_{cp} *be the class of canonical policies, and* Φ_{ear} *the class of (randomized) stationary EAR-optimal policies. We also define the sets of deterministic policies* $\mathbb{F}_{cp} = \mathbb{F} \cap \Phi_{cp}$ *and* $\mathbb{F}_{ear} = \mathbb{F} \cap \Phi_{ear}$.

From Theorem 2.4.3, the classes Φ_{cp} and Φ_{ear} are nonempty. Furthermore, the triplet (h^*, f^*, ρ^*) in Theorem 2.4.3 is canonical.

Proposition 2.6.3 *Under the assumptions of Theorem 2.4.3*

$$\Phi_{cp} \subset \Phi_{ear}. \tag{2.6.5}$$

Furthermore, if $(\tilde{\rho}, \tilde{h}, \tilde{\varphi})$ *is a canonical triplet, then* $\tilde{\rho}$ *is the optimal value, that is,* $\tilde{\rho} = \rho^*$.

Proof. Let $(\tilde{\rho}, \tilde{h}, \tilde{\varphi})$ a canonical triplet, so that the relationships (2.6.3) and (2.6.4) hold. Integrating both sides of (2.6.4) with respect to $\mu_{\tilde{\varphi}}$ we obtain $\tilde{\rho} = \mu_{\tilde{\varphi}}(r_{\tilde{\varphi}}) = \rho_{\tilde{\varphi}}$. On the other hand, by (2.6.4) and Lemma 2.5.2,

$$\tilde{\rho} = \rho_{\tilde{\varphi}} \geq \bar{J}(\pi, x) \geq J(\pi, x) \quad \forall x \in \mathbf{X}, \pi \in \Pi.$$

Thus

$$\rho^* = \tilde{\rho} = \rho_{\tilde{\varphi}} = J(\tilde{\varphi}, x) = J^*(x) \quad \forall x \in \mathbf{X};$$

therefore, $\tilde{\varphi}$ is an EAR-optimal policy. ■

Concluding remarks. In this chapter we establish preliminary results used troughout the rest of this work. We extend the results obtained by Vega-Amaya [30] to the set of all randomized stationary policies Φ. Furthermore, we give explicit expressions for the invariant measures (see equation (2.2.19)), and also for the functions h^*_φ (see equation (2.3.2)) that solve the P.E., and the functions h^* that solve the AROE (2.4.5). This fact will be particularly useful to prove boundedness conditions, which are necessary for "nice" asymptotic results (law of large numbers, asymptotic normality) and to prove compactness conditions.

Chapter 3

Pathwise Average Reward Optimality

In this chapter we study pathwise average reward (PAR) optimality for the general MCP introduced in Section 1.3. Our main objective is to show the existence of PAR-optimal policies under our assumptions introduced in Chapter 2. This problem is reduced to the context of Chapter 2 because in fact we prove (in Theorem 3.3.2) that a stationary policy is PAR-optimal if and only if it is EAR-optimal as in Definition 2.4.1. To this end we use the law of large numbers for martingales, also known as the martingale stability theorem (see Lemma 3.2.3 below).

As was already mentioned in Section 1.1, pathwise average optimality has been studied under some hypotheses. For instance, Hernandez-Lerma et al. [18] impose conditions ensuring w-geometric ergodicity. Here we follow the fixed point approach initiated in Chapter 2.

This chapter is based on the works of Hernandez-Lerma et al. [15, Chapter 11] and [18]. Similar results for continuous-time Markov chains can be found in [24].

3.1 Definitions and a preliminary result

Definition 3.1.1 *Let* $(\mathbf{X}, A, \{A(x) : x \in \mathbf{X}\}, Q, r)$ *be a general MCM as in Section 1.3, and let*

$$S_n(\pi, x) := \sum_{k=0}^{n-1} r(x_k, a_k) \tag{3.1.1}$$

be the pathwise n-stage total reward *when using the control policy* $\pi \in \Pi$, *given the initial state* $x \in \mathbf{X}$. *We define the* pathwise average reward*:*

$$S(\pi, x) := \liminf_{n\to\infty} S_n(\pi, x)/n. \tag{3.1.2}$$

Here we use the convention that the sequence $\{x_0, a_0, x_1, a_1, \cdots\}$ in (3.1.1) and (3.1.2) corresponds to the state-action process when using the policy π, given the initial state x.

In the following proposition we show that if in (3.1.2) π is a stationary policy $\varphi \in \Phi$, then $S(\varphi, \cdot)$ coincides with the expected average reward $\rho_\varphi = \mu_\varphi(r_\varphi)$ defined in (2.1.3).

Proposition 3.1.2 *Suppose that Assumptions 2.1.1 and 2.1.2 hold. Then for each* $\varphi \in \Phi$ *and each initial state* $x \in \mathbf{X}$

$$S(\varphi, x) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} r_\varphi(x_k) = \rho_\varphi \quad P_x^\varphi - a.s.$$

Proof. This result is a consequence of the strong law of large numbers for Markov chains (see, for instance, [22, p. 411] or [15, Theorem 11.2.1(a)]). Indeed, by Theorem 2.1.4 above, the Markov processes $\{x_k\}$ associated to the kernel $Q_\varphi(\cdot|x)$ is positive Harris recurrent and the strong law of large numbers holds:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} r_\varphi(x_k) = \mu_\varphi(r_\varphi) \quad P_x^\varphi - a.s.$$

since the function r_φ is μ_φ-integrable. Thus, from (3.1.1) and (3.1.2), we obtain the desired result. ∎

3.2 Technical preliminaries

Let W be as in Assumption 2.1.1.

Assumption 3.2.1 *There exists a positive constant* K_2 *such that*

$$r(x, a)^2 \leq K_2 W(x) \quad \forall (x, a) \in \mathbb{K}.$$

Let us consider a randomized policy $\pi \in \Pi$, and h such that $h^2 \in B_W(\mathbf{X})$; equivalently, $h \in B_w(\mathbf{X})$ where $w(x) := \sqrt{W(x)}$ for all $x \in \mathbf{X}$. For $k, n = 1, 2, \cdots$, let

$$Y_k := h(x_k) - E_x^\pi[h(x_k)|x_{k-1}] = h(x_k) - E_x^\pi[h(x_k)|h_{k-1}] \tag{3.2.1}$$

with h_{k-1} being the admissible history up to time $k-1$, and

$$M_n := \sum_{k=1}^{n} Y_k, \tag{3.2.2}$$

In particular, if $\pi = \varphi \in \Phi$, then Y_k is given by

$$Y_k = h(x_k) - \int_{\mathbf{X}} h(y) Q_\varphi(dy|x_{k-1}).$$

The next two lemmas are taken from the work of Hernandez-Lerma et al. [18]:

Lemma 3.2.2 *Suppose that Assumptions 2.1.1, 2.1.2 and 3.2.1 hold, and let $x_0 = x$ be an (arbitrary) initial state. Then $\{M_n\}_{n\geq 1}$ is a square integrable P_x^π-martingale with respect to the σ–algebra*

$$\mathcal{F}_n = \sigma\{x_0, a_0, \cdots, x_{n-1}, a_{n-1}, x_n\} = \sigma\{h_n\}.$$

Moreover,

(i) $E_x^\pi \sum_{k=1}^{\infty} k^{-2} W(x_k) < \infty$;

(ii) $\sum_{k=1}^{\infty} k^{-2} W(x_k) < \infty \quad P_x^\pi - a.s.$;

(iii) $k^{-2} W(x_k) \to 0 \quad P_x^\pi - a.s.$;

(iv) $k^{-1} w(x_k) \to 0 \quad P_x^\pi - a.s.$, *with* $w(\cdot) = \sqrt{W(\cdot)}$.

Proof. It suffices to prove (i):

Since $|h(x)| \leq \|h\|_w \sqrt{W(x)}$ for each $x \in \mathbf{X}$, from (3.2.1) we obtain

$$|Y_k| \leq \|h\|_w \{w(x_k) + E_x^\pi[w(x_k)|x_{k-1}]\}.$$

Thus

$$Y_k^2 \leq 2\|h\|_w^2 \{W(x_k) + E_x^\pi[W(x_k)|x_{k-1}]\}, \tag{3.2.3}$$

and so Y_k is square integrable with respect to the probability measure P_x^π. Hence, noting that $M_n - M_{n-1} = Y_n$,

$$\begin{aligned} E_x^\pi[M_n - M_{n-1}|\mathcal{F}_{n-1}] &= E_x^\pi[Y_n|h_{n-1}] \\ &= E_x^\pi[h(x_n)|h_{n-1}] - E_x^\pi[h(x_n)|h_{n-1}] = 0, \end{aligned}$$

i.e.,

$$E_x^\pi[M_n|\mathcal{F}_{n-1}] = M_{n-1}.$$

Then $\{M_n\}_{n\geq 1}$ is a square integrable P_x^π-martingale.

On the other hand, by Assumption 2.1.2(c) we can see that

$$\int_{\mathbf{X}} W(y)Q(dy|x,a) \leq \lambda W(x) + b \quad \forall x \in \mathbf{X}, \tag{3.2.4}$$

where b is a constant. By an iteration procedure, we obtain

$$E_x^\pi W(x_k) \leq \lambda^k W(x) + b\frac{1-\lambda^k}{1-\lambda} \quad \text{for } k = 0, 1\cdots,$$

which in turn gives (i). ■

Lemma 3.2.3 *Under the assumptions of Lemma 3.2.2,*

$$\lim_{n\to\infty} \frac{1}{n} M_n = 0 \quad P_x^\pi - a.s.$$

Proof. Inequality (3.2.3) implies

$$E_x^\pi[Y_k^2|\mathcal{F}_{k-1}] \leq 4||h||_w^2 E_x^\pi[W(x_k)|\mathcal{F}_{k-1}]$$

because $E_x^\pi[W(x_k)|x_{k-1}] = E_x^\pi[W(x_k)|\mathcal{F}_{k-1}]$. By Assumption 2.1.1(a) and (3.2.4), we obtain

$$\begin{aligned} E_x^\pi[W(x_k)|\mathcal{F}_{k-1}] &= E_x^\pi[W(x_k)|h_{k-1}] \\ &= \int_{\mathbf{X}}\int_A W(y)Q(dy|x_{k-1},a)\pi_{k-1}(da|h_{k-1}) \\ &\leq (\lambda + \frac{b}{\theta})W(x_{k-1}); \end{aligned}$$

Therefore, by Lemma 3.2.2(ii)

$$\sum_{k=1}^{\infty} k^{-2} E_x^\pi[Y_k^2|\mathcal{F}_{k-1}] \leq 4||h||_w^2(\lambda + \frac{b}{\theta})\Big[W(x_0) + \sum_{k=1}^{\infty} k^{-2}W(x_k)\Big] < \infty \quad P_x^\pi - a.s.$$

Finally, by the Martingale Stability Theorem (see, for instance, [15, p. 173], we obtain

$$\lim_{n\to\infty}\frac{1}{n}M_n = 0 \quad P_x^\pi - a.s.$$

■

Lemma 3.2.4 *Suppose that Assumptions 2.1.1, 2.1.2 and 3.2.1 hold. Let h_φ^* be the function in Theorem 2.1.4(iii). Then h_φ^* is a w-bounded function, with $w = \sqrt{W}$.*

Proof. Let $P_\varphi(dy|x)$ be the kernel on $\mathbf{X}$ defined in (2.2.7), and let $\widehat{Q}$ be as in (2.2.1). By the Cauchy-Schwartz inequality,

$$\begin{aligned}\int_{\mathbf{X}}\sqrt{W(y)}\widehat{Q}(dy|x,a) &\leq \sqrt{\int_{\mathbf{X}} W(y)\widehat{Q}(dy|x,a)}\sqrt{\widehat{Q}(\mathbf{X}|x,a)}\\ &\leq \sqrt{\lambda}\sqrt{W(x)}.\end{aligned}$$

That is

$$\int_{\mathbf{X}}\sqrt{W(y)}\widehat{Q}(dy|x,a) \leq \eta\sqrt{W(x)} \quad \forall (x,a)\in\mathbb{K},$$

where $\eta := \sqrt{\lambda} < 1$. Thus

$$\int_{\mathbf{X}}\sqrt{W(y)}\widehat{Q}_\varphi^n(dy|x) \leq \eta^n\sqrt{W(x)} \quad \forall x\in\mathbf{X}, n\in\mathbb{N}, \varphi\in\Phi.$$

This inequality implies that

$$\int_{\mathbf{X}}\sqrt{W(y)}P_\varphi(dy|x) \leq \frac{\sqrt{W(x)}}{1-\eta} \quad \forall x\in\mathbf{X}. \tag{3.2.5}$$

On the other hand, by Assumptions 2.1.1(b) and 3.2.1 it follows that $|r_\varphi(x)| \leq \sqrt{K_2 W(x)}$. By (2.2.16) and (2.2.17) we also have

$$|\rho_\varphi| \leq K\mu_\varphi(W) \leq K_3 := \frac{K\nu(W)}{(1-\lambda)\nu(X)}.$$

Consequently, from the explicit form (2.3.2) of the function h_φ^*

$$|h_\varphi^*(x)| \leq \int_{\mathbf{X}}|r_\varphi(y)-\rho_\varphi|P_\varphi(dy|x)$$

$$\begin{aligned} &\leq \int_{\mathbf{X}} |r_\varphi(y)| P_\varphi(dy|x) + \frac{|\rho_\varphi|}{\sqrt{\theta}} \int_{\mathbf{X}} \sqrt{W(y)} P_\varphi(dy|x) \\ &\leq (\sqrt{K_2} + \frac{|\rho_\varphi|}{\sqrt{\theta}}) \int_{\mathbf{X}} \sqrt{W(y)} P_\varphi(dy|x), \end{aligned} \tag{3.2.6}$$

with θ as in Assumption 2.1.1(a).

Combining (3.2.5) and (3.2.6), we obtain

$$|h^*_\varphi(x)| \leq K_4 \sqrt{W(x)} \quad \forall x \in \mathbf{X},$$

where

$$K_4 := \frac{(\sqrt{K_2} + \frac{K_3}{\sqrt{\theta}})}{1 - \eta}$$

does not depend on φ. Thus $(h^*_\varphi)^2$ is in $B_W(\mathbf{X})$. ∎

For future reference, we note the following.

Remark 3.2.5 *We define, for each $v \in B_w(\mathbf{X})$,*

$$R_\varphi v(x) := \int_{\mathbf{X}} v(y) Q_\varphi(dy|x) \quad \forall x \in \mathbf{X}. \tag{3.2.7}$$

We claim that $R_\varphi v \in B_w(\mathbf{X})$. Indeed, (3.2.4) implies

$$\int_{\mathbf{X}} W(y) Q(dy|x,a) \leq K'^2 W(x) \quad \forall (x,a) \in \mathbb{K},$$

with $K'^2 := \lambda + b/\theta$. Hence, by the Cauchy-Schwartz inequality

$$\int_{\mathbf{X}} \sqrt{W(y)} Q(dy|x,a) \leq \sqrt{\int_{\mathbf{X}} W(y) Q(dy|x,a)} \quad \forall (x,a) \in \mathbb{K}.$$

Therefore

$$\int_{\mathbf{X}} \sqrt{W(y)} Q(dy|x,a) \leq K' \sqrt{W(x)} \quad \forall (x,a) \in \mathbb{K}. \tag{3.2.8}$$

This inequality implies that $R_\varphi v$ is in $B_w(\mathbf{X})$ for each $v \in B_w(\mathbf{X})$, with $w(\cdot) = \sqrt{W(\cdot)}$.

Lemma 3.2.6 *Let $w(\cdot) = \sqrt{W(\cdot)}$. Suppose that Assumptions 2.1.1, 2.1.2 and 3.2.1 are satisfied. Given a randomized policy $\pi \in \Pi$, an initial state $x \in \mathbf{X}$ and an arbitrary $h \in B_w(\mathbf{X})$, we have*

$$\frac{1}{n}\sum_{k=0}^{n-1} L^\pi h(x_k) \to 0 \text{ as } n \to \infty \quad P_x^\pi - a.s. \tag{3.2.9}$$

with

$$L^\pi h(x_k) := E_x^\pi[h(x_{k+1})|h_k] - h(x_k),$$

where h_k is the admissible history up to time k.

For a stationary policy $\pi = \varphi \in \Phi$,

$$(L^\varphi h)(x) := \int_{\mathbf{X}} h(y) Q_\varphi(dy|x) - h(x) \quad \forall x \in \mathbf{X}.$$

Proof. Notice that

$$M_n = [h(x_n) - h(x)] - \sum_{k=0}^{n-1} L^\pi(x_k), \tag{3.2.10}$$

with M_n as in (3.2.2). From Lemma 3.2.2(iv) and Lemma 3.2.3 we have

$$\frac{1}{n}h(x_n) \to 0 \quad \text{and} \quad \frac{1}{n}M_n \to 0 \quad P_x^\pi - a.s.$$

as $n \to \infty$. These limits and (3.2.10) imply (3.2.9). ∎

Lemma 3.2.7 *Suppose that the hypotheses of Theorem 2.4.3 and Assumption 3.2.1 are satisfied, and let $(h^*, \varphi^*, \rho^*) \in B_W(\mathbf{X}) \times \Phi \times \mathbb{R}$ be a canonical triplet, that is, a solution to the AROE (2.4.5). Then $h^*(x) - h^*_{\varphi^*}(x) = c$ for all $x \in \mathbf{X}$, where c is a constant that may depend on φ^* and $h^*_{\varphi^*}$, the function in Theorem 2.1.4(iii) corresponding to the policy φ^*. Furthermore, h^* is w-bounded, with $w = \sqrt{W}$.*

This result follows from Lemma 2.3.1 and Lemma 3.2.4.

3.3 Pathwise average optimal policies

In this section we study a class of Markov control problems for which there exists a pathwise average reward (PAR) optimal policy as defined below. The main result in this section is Theorem 3.3.2 which together with Theorem 2.4.3 gives the existence of pathwise average optimal policies in $\mathbb{F}$.

Definition 3.3.1 *A randomized stationary policy $\varphi^* \in \Phi$ is said to be* pathwise average reward optimal (PAR-optimal) *if for each randomized policy $\pi \in \Pi$ and each $x \in \mathbf{X}$,*

$$S(\pi, x) \leq \rho_{\varphi^*} \quad P_x^\pi - a.s.$$

with $\rho_{\varphi^*} = \mu_{\varphi^*}(r_{\varphi^*})$.

Theorem 3.3.2 *Suppose that the hypotheses of Theorem 2.4.3 and Assumption 3.2.1 are satisfied. Then a stationary policy is PAR-optimal if and only if is expected average optimal.*

Proof. As a consequence of Proposition 3.1.2 and Theorem 2.4.3, pathwise average optimal policies are necessarily expected average optimal.

Conversely, let $(\rho^*, h^*) \in \mathbb{R} \times B_W(\mathbf{X})$ be a solution of the AROE (2.4.5). By Lemma 3.2.7, we have that h^* is in $B_w(\mathbf{X})$. Now, let $\varphi^* \in \Phi$ be an expected average optimal policy, that is, $\rho_{\varphi^*} = \rho^*$. From the AROE (2.4.5)

$$\rho_{\varphi^*} \geq r(x,a) + \int_{\mathbf{X}} h^*(y) Q(dy|x,a) - h^*(x) \quad \forall (x,a) \in \mathbb{K}.$$

Hence, for an arbitrary policy $\pi \in \Pi$ and initial state x, we have

$$\begin{aligned} \rho_{\varphi^*} &\geq r(x_k,a_k) + \int_{\mathbf{X}} h^*(y) Q(dy|x_k,a_k) - h^*(x_k) \\ &= r(x_k,a_k) + E_x^\pi[h^*(x_{k+1})|h_k,a_k] - h^*(x_k) \quad P_x^\pi - a.s. \end{aligned}$$

for all $k = 0, 1, \cdots$. Taking conditional expectation with respect to h_k,

$$\begin{aligned} \rho_{\varphi^*} &\geq E_x^\pi[r(x_k,a_k)|h_k] + E_x^\pi[h^*(x_{k+1})|h_k] - h^*(x_k) \\ &= E_x^\pi[r(x_k,a_k)|h_k] + L^\pi h^*(x_k) \quad P_x^\pi - a.s. \end{aligned}$$

with $L^\pi h^*(x_k)$ defined as in Lemma 3.2.6. Hence

$$n\rho_{\varphi^*} \geq S_n(\pi, x) + \sum_{k=0}^{n-1} L^\pi h^*(x_k) \quad P_x^\pi - a.s.$$

Finally, multiplying both sides of the latter inequality by $1/n$ and taking lim sup as $n \to \infty$, and use Lemma 3.2.6 to obtain

$$\rho_{\varphi^*} \geq \limsup_{n\to\infty} \frac{1}{n} S_n(\pi, x) \geq S(\pi, x) \quad P_x^\pi - a.s.$$

for every policy $\pi \in \Pi$. Thus, φ^* is sample-path average optimal. ■

Concluding remarks. In this chapter we have studied pathwise average reward optimality. It is proved that under our fixed-point approach and under a growth condition on the reward (Assumption 3.2.1), pathwise average reward optimality and expected average optimality are equivalent. To this end we verified that, under our hypotheses, we can use a law of large numbers for martingales, also known as the martingale stability theorem. These techniques have been used in previous works like [18] assuming w-geometric ergodicity.

Chapter 4

Variance minimization

In this chapter we study the existence of a stationary policy that minimizes the limiting average variance in the class $\mathbb{F}_{cp}$ of deterministic canonical policies (recall Definition 2.6.2). Under our assumptions, we extend the results in the works of Hernández-Lerma et al. [18] and [15, Chapter 11], which require w-geometric ergodicity. Our procedure does not need w-geometric ergodicity, and it is a consequence of our results in Chapters 2 and 3. Moreover, we show that under an appropiate growth condition on the reward, the MCP satisfies an asymptotic normality condition, which is very useful in adaptive control problems.

4.1 Definitions

Definition 4.1.1 *Let $J_n(\varphi, x)$ and $S_n(\varphi, x)$ be as in Definitions 2.4.1 and 3.1.1. For every $\varphi \in \Phi$ and initial state x we define the* limiting average variance

$$V(\varphi, x) := \limsup_{n\to\infty} \frac{1}{n}\mathrm{var}[S_n(\varphi, x)] \tag{4.1.1}$$

where (by definition of variance of a random variable)

$$\mathrm{var}[S_n(\varphi, x)] = E_x^{\varphi}[S_n(\varphi, x) - J_n(\varphi, x)]^2.$$

We introduce some notation: in the remainder of this chapter we define $h_1 := h^*$, where h^* as in (2.4.5). For each $x \in \mathbf{X}$, let $A^*(x) \subset A(x)$ be the set of control actions that attain the maximun in (2.4.5), that is,

$$A^*(x) := \left\{ a \in A(x) : \rho^* + h_1(x) = r(x, a) + \int_{\mathbf{X}} h_1(y)Q(dy|x, a) \right\} \tag{4.1.2}$$

Remark 4.1.2 *Observe that by (2.6.4), a deterministic policy $f \in \mathbb{F}$ is canonical if and only if $f(x) \in A^*(x)$ for all $x \in \mathbf{X}$.*

Let

$$\Lambda(x,a) := \int_{\mathbf{X}} {h_1}^2(y) Q(dy|x,a) - \left[\int_{\mathbf{X}} h_1(y) Q(dy|x,a)\right]^2 \tag{4.1.3}$$

Under the assumptions of Lemma 3.2.7, the function Λ on $\mathbb{K}$ is well defined.

4.2 Preliminary results

To state our variance-minimization result, we need the following lemmas.

Lemma 4.2.1 *Under Assumptions 2.1.1, 2.1.2 and 2.4.2, let us consider h_1 satisfying (2.4.5) in Theorem 2.4.3, φ an EAR-optimal policy, and h^*_φ the functions defined in Theorem 2.1.4-(iii). Then*

(a) *$h^*_\varphi(\cdot) = h_1(\cdot) + c_\varphi \quad \mu_\varphi - a.e.$ for some constant c_φ.*

(b) *There exists a canonical policy $\hat{\varphi} \in \Phi_{cp}$ such that $(\rho^*, h_1, \hat{\varphi})$ is a canonical triplet, $\hat{\varphi}(\cdot|x) = \varphi(\cdot|x)$ $\mu_\varphi - a.e.$, and $\mu_{\hat{\varphi}} = \mu_\varphi$. Moreover, $h^*_{\hat{\varphi}}(x) = h_1(x) + c_{\hat{\varphi}}$ for all $x \in \mathbf{X}$.*

Proof. (a) From the AROE (2.4.5), we have

$$\rho^* + h_1(x) \geq r_\varphi(x) + \int_{\mathbf{X}} h_1(y) Q_\varphi(dy|x) \quad \forall x \in \mathbf{X}. \tag{4.2.1}$$

Since φ is EAR-optimal, we have $\rho_\varphi = \mu_\varphi(r_\varphi) = \rho^*$. The corresponding Poisson equation (2.1.4) is

$$\rho^* + h^*_\varphi(x) = r_\varphi(x) + \int_{\mathbf{X}} h^*_\varphi(y) Q_\varphi(dy|x) \quad \forall x \in \mathbf{X}. \tag{4.2.2}$$

By (4.2.1) and (4.2.2), it follows that the function $u(\cdot) := h^*_\varphi(\cdot) - h_1(\cdot)$ in $B_W(\mathbf{X})$ is subharmonic with respect to Q_φ, i.e.

$$\int_{\mathbf{X}} u(y) Q_\varphi(dy|x) \geq u(x) \quad \forall x \in \mathbf{X}.$$

By an iteration procedure, we get

$$\int_{\mathbf{X}} u(y)Q_\varphi^n(dy|x) \geq u(x) \quad \forall x \in \mathbf{X}, n = 0, 1, \cdots.$$

From this inequality, we see that

$$\frac{1}{n}E_x^\varphi \sum_{k=0}^{n-1} u(x_k) \geq u(x) \quad \forall n \in \mathbb{N},$$

and letting $n \to \infty$ we obtain

$$\int_{\mathbf{X}} u(y)\mu_\varphi(dy) = \mu_\varphi(u) \geq u(x) \quad \forall x \in \mathbf{X}. \tag{4.2.3}$$

So, u is bounded above. We define $c_\varphi := \sup_{x\in\mathbf{X}} u(x)$. By (4.2.3), we have $c_\varphi = \int_{\mathbf{X}} u(y)\mu_\varphi(dy)$, which implies

$$u(\cdot) = h_\varphi^*(\cdot) - h_1(\cdot) = c_\varphi \quad \mu_\varphi - a.e.$$

That is,

$$h_\varphi^*(\cdot) = h_1(\cdot) + c_\varphi \quad \mu_\varphi - a.e.,$$

with $c_\varphi = \sup_{x\in\mathbf{X}} u(x) = \int_{\mathbf{X}} u(y)\mu_\varphi(dy)$.

(b) Notice that

$$\int_{\mathbf{X}} \Big[\rho^* + h_1(x) - r_\varphi(x) - \int_{\mathbf{X}} h_1(y)Q_\varphi(dy|x)\Big]\mu_\varphi(dx) = 0,$$

because $\varphi \in \Phi_{ear}$, i.e., $\rho_\varphi = \rho^*$. By Inequality (4.2.1) we have

$$\rho^* + h_1(x) = r_\varphi(x) + \int_{\mathbf{X}} h_1(y)Q_\varphi(dy|x) \quad \mu_\varphi - a.e. \tag{4.2.4}$$

Hence there exists a Borel set $N \in \mathcal{B}(\mathbf{X})$ such that $\mu_\varphi(N) = 0$ and

$$\rho^* + h_1(x) = r_\varphi(x) + \int_{\mathbf{X}} h_1(y)Q_\varphi(dy|x) \quad \forall x \in N^c := \mathbf{X} \setminus N \tag{4.2.5}$$

On the other hand, we consider a canonical policy $\varphi^* \in \Phi_{cp}$ such that (ρ^*, h_1, φ^*) is a canonical triplet, and define the new policy

$$\hat{\varphi}(\cdot|x) = 1_N(x)\varphi^*(\cdot|x) + 1_{N^c}(x)\varphi(\cdot|x) \quad \forall x \in \mathbf{X}.$$

Notice that

$$\varphi(\cdot|x) = \hat{\varphi}(\cdot|x) \quad \text{and} \quad Q_{\varphi}(\cdot|x) = Q_{\hat{\varphi}}(\cdot|x) \quad \mu_{\varphi} - a.e. \tag{4.2.6}$$

Moreover

$$\varphi(\cdot|x) = \hat{\varphi}(\cdot|x) \quad \text{and} \quad Q_{\varphi}(\cdot|x) = Q_{\hat{\varphi}}(\cdot|x) \quad \forall x \in N^c.$$

Hence, (4.2.6) implies

$$\mu_{\varphi}(\cdot) = \mu_{\hat{\varphi}}(\cdot). \tag{4.2.7}$$

Actually, we have that $Q_{\varphi}(B|x) = Q_{\hat{\varphi}}(B|x)$ $\mu_{\varphi} - a.e.$, for all B in $\mathcal{B}(\mathbf{X})$. Integrating both sides with respect to μ_{φ}, we obtain

$$\mu_{\varphi}(B) = \int_{\mathbf{X}} Q_{\hat{\varphi}}(B|x)\mu_{\varphi}(dx).$$

So, μ_{φ} is an invariant measure for $Q_{\hat{\varphi}}(B|x)$. By uniqueness of the i.p.m., we obtain $\mu_{\varphi}(\cdot) = \mu_{\hat{\varphi}}(\cdot)$.

Next we show that $(\rho^*, h_1, \hat{\varphi})$ is a canonical triplet: Let $x \in \mathbf{X}$ be arbitrary.
(i) If $x \in N$, then $\hat{\varphi}(\cdot|x) = \varphi^*(\cdot|x)$. This implies

$$\begin{aligned} \rho^* + h_1(x) &= \sup_{a \in A(x)} \Big[r(x,a) + \int_{\mathbf{X}} h_1(y) Q(dy|x,a) \Big] \\ &= r_{\varphi^*}(x) + \int_{\mathbf{X}} h_1(y) Q_{\varphi^*}(dy|x) \\ &= r_{\hat{\varphi}}(x) + \int_{\mathbf{X}} h_1(y) Q_{\hat{\varphi}}(dy|x). \end{aligned}$$

(ii) If $x \in N^c$, then $\hat{\varphi}(\cdot|x) = \varphi(\cdot|x)$. By (4.2.5)

$$\begin{aligned} \rho^* + h_1(x) &= \sup_{a \in A(x)} \Big[r(x,a) + \int_{\mathbf{X}} h_1(y) Q(dy|x,a) \Big] \\ &= r_{\varphi}(x) + \int_{\mathbf{X}} h_1(y) Q_{\varphi}(dy|x) \\ &= r_{\hat{\varphi}}(x) + \int_{\mathbf{X}} h_1(y) Q_{\hat{\varphi}}(dy|x). \end{aligned}$$

Combining (i) and (ii) we have that $(\rho^*, h_1, \hat{\varphi})$ is a canonical triplet and $\hat{\varphi}$ is a canonical policy such that $\hat{\varphi}(\cdot|x) = \varphi(\cdot|x)$ $\mu_{\varphi} - a.e.$, and $\mu_{\hat{\varphi}} = \mu_{\varphi}$.

Finally, from Lemma 2.3.1, we obtain $h^*_{\hat{\varphi}}(x) = h_1(x) + c_{\hat{\varphi}}$ for all $x \in \mathbf{X}$.

■

Lemma 4.2.2 *Suppose that Assumptions 2.1.1, 2.1.2, 2.4.2 and also 3.2.1 hold. Let $\varphi \in \Phi$ be arbitrary and let h^*_φ be a function as in Theorem 2.1.4(iii). We define the function*

$$\Psi_\varphi(x) := \int_{\mathbf{X}} {h^*_\varphi}^2(y) Q_\varphi(dy|x) - \Big[\int_{\mathbf{X}} h^*_\varphi(y) Q_\varphi(dy|x)\Big]^2 \quad \forall x \in \mathbf{X}. \tag{4.2.8}$$

Then:

(a) *The functions ${h^*_\varphi}^2$, Ψ_φ and ${h_1}^2$ belong to $B_W(\mathbf{X})$, where h_1 as in Theorem 2.4.3 satisfying the AROE (2.4.5);*

(b) *The limiting average variance satisfies*

$$V(\varphi, x) = \lim_{n\to\infty} \frac{1}{n} E^\varphi_x \sum_{k=0}^{n-1} \Psi_\varphi(x_k) = \sigma^2_\varphi \quad \forall x \in \mathbf{X}, \tag{4.2.9}$$

where $\sigma^2_\varphi = \mu_\varphi(\Psi_\varphi)$;

(c) *For each φ EAR-optimal policy in Φ_{ear} there exists a canonical policy $\hat{\varphi} \in \Phi_{cp}$ such that $\Psi_{\hat{\varphi}} = \Psi_\varphi \quad \mu_\varphi - a.e.$. Hence*

$$V(\hat{\varphi}, x) = V(\varphi, x) = \sigma^2_\varphi \quad \forall x \in \mathbf{X}.$$

In particular, for each f stationary policy in $\mathbb{F}_{ear}$ there exists a canonical policy $\hat{f} \in \mathbb{F}_{cp}$ such that

$$V(\hat{f}, x) = V(f, x) = \sigma^2_f \quad \forall x \in \mathbf{X}.$$

Proof. (a) This part follows from Lemma 3.2.4 and Lemma 3.2.7 above.

(b) This part is a consequence of [15, Theorem 11.2.4].

(c) From Lemma 4.2.1(b), for each φ in Φ_{ear} there exists a canonical policy $\hat{\varphi}$ in Φ_{cp}, such that $(\rho^*, h_1, \hat{\varphi})$ is a canonical triplet, $\hat{\varphi}(\cdot|x) = \varphi(\cdot|x) \quad \mu_\varphi - a.e.$, $\mu_{\hat{\varphi}} = \mu_\varphi$ and $h_1(x) = h^*_{\hat{\varphi}}(x) + c_{\hat{\varphi}}$ for all $x \in \mathbf{X}$. Hence

$$\Psi_{\hat{\varphi}}(x) = \int_{\mathbf{X}} {h_1}^2(y) Q_{\hat{\varphi}}(dy|x) - \Big[\int_{\mathbf{X}} h_1(y) Q_{\hat{\varphi}}(dy|x)\Big]^2 \quad \forall x \in \mathbf{X}. \tag{4.2.10}$$

Furthermore, by Lemma 4.2.1(a), there exist a subset N in $\mathcal{B}(\mathbf{X})$ such that $\mu_\varphi(N) = 0$. Moreover

(i) $h^*_\varphi(x) = h_1(x) + c_\varphi \quad \forall x \notin N$, where c_φ is a constant.

(ii) $\hat{\varphi}(\cdot|x) = \varphi(\cdot|x) \quad \forall x \notin N$.

Notice that

$$0 = \mu_\varphi(N) = \int_{\mathbf{X}} Q_\varphi(N|x)\mu_\varphi(dx).$$

Thus $Q_\varphi(N|x) = 0 \quad \mu_\varphi - a.e.$ Hence, there exists N' in $\mathcal{B}(\mathbf{X})$ such that $\mu_\varphi(N') = 0$ and

$$Q_\varphi(N|x) = 0 \quad \forall x \in N'^c. \tag{4.2.11}$$

By (i) and (4.2.11), we have that

$$\begin{aligned}\int_{\mathbf{X}} h^{*\,2}_\varphi(y)Q_\varphi(dy|x) &= \int_N h^{*\,2}_\varphi(y)Q_\varphi(dy|x) + \int_{N^c} h^{*\,2}_\varphi(y)Q_\varphi(dy|x) \\ &= \int_{N^c} (h_1(y) + c_\varphi)^2 Q_\varphi(dy|x) \\ &= \int_{\mathbf{X}} (h_1(y) + c_\varphi)^2 Q_\varphi(dy|x) \quad \forall x \notin N \cup N'.\end{aligned}$$

Similarly

$$\int_{\mathbf{X}} h^*_\varphi(y)Q_\varphi(dy|x) = \int_{\mathbf{X}} (h_1(y) + c_\varphi)Q_\varphi(dy|x) \quad \forall x \notin N \cup N'.$$

Hence

$$\Psi_\varphi(x) = \int_{\mathbf{X}} {h_1}^2(y)Q_\varphi(dy|x) - \Big[\int_{\mathbf{X}} h_1(y)Q_\varphi(dy|x)\Big]^2 \quad \forall x \notin N \cup N'. \tag{4.2.12}$$

By (ii) and (4.2.10), we have

$$\Psi_{\hat{\varphi}}(x) = \int_{\mathbf{X}} {h_1}^2(y)Q_\varphi(dy|x) - \Big[\int_{\mathbf{X}} h_1(y)Q_\varphi(dy|x)\Big]^2 \quad \forall x \notin N. \tag{4.2.13}$$

Comparing (4.2.12) and (4.2.13), we have that $\Psi_{\hat{\varphi}} = \Psi_\varphi \quad \mu_\varphi - a.e.$ Since $\mu_{\hat{\varphi}} = \mu_\varphi$, then

$$\sigma^2_{\hat{\varphi}} = \mu_{\hat{\varphi}}(\Psi_{\hat{\varphi}}) = \mu_\varphi(\Psi_\varphi) = \sigma^2_\varphi.$$

So, from part (b) of this lemma we obtain

$$V(\hat{\varphi}, x) = V(\varphi, x) \quad \forall x \in \mathbf{X}. \quad \blacksquare$$

Remark 4.2.3 *From the proof of Lemma 4.2.2(c) (see equation (4.2.12)), we can see that if $f \in \mathbb{F}_{ear}$ then*

$$\Psi_f = \Lambda_f \quad \mu_f - a.e.$$

where $\Lambda(x,a)$ as defined in (4.1.3). Consequently, by Lemma 4.2.2(b)

$$V(f,x) = \sigma_f^2 = \mu_f(\Psi_f) = \mu_f(\Lambda_f) \quad \forall x \in \mathbf{X}.$$

4.3 Main result

In this section we prove that, under the hypotheses of Theorem 3.3.2, there exists a deterministic canonical policy f^* in $\mathbb{F}_{cp}$ such that

$$V(f^*,x) = \inf_{f \in \mathbb{F}_{ear}} V(f,x) \quad \forall x \in \mathbf{X}. \tag{4.3.1}$$

Theorem 4.3.1 *Suppose that Assumptions 2.1.1, 2.1.2, 2.4.2 and also 3.2.1 hold. Then there exists a constant $\sigma_*^2 \geq 0$, a deterministic canonical policy $f^* \in \mathbb{F}_{cp}$, and a function $h_2(\cdot)$ in $B_W(\mathbf{X})$ such that, for each $x \in \mathbf{X}$,*

$$\begin{aligned} \sigma_*^2 + h_2(x) &= \min_{a \in A^*(x)} \Big[\Lambda(x,a) + \int_{\mathbf{X}} h_2(y) Q(dy|x,a)\Big] \\ &= \Lambda_{f^*}(x) + \int_{\mathbf{X}} h_2(y) Q_{f^*}(dy|x) \end{aligned} \tag{4.3.2}$$

Furthermore, f^* satisfies (4.3.1) and $V(f^*,\cdot) = \sigma_*^2$; in fact

$$V(f^*,x) = \mu_{f^*}(\Lambda_{f^*}) = \sigma_*^2 \quad \forall x \in \mathbf{X} \tag{4.3.3}$$

and

$$\sigma_*^2 \leq V(f,x) \quad \forall f \in \mathbb{F}_{ear}, x \in \mathbf{X}. \tag{4.3.4}$$

Proof. Let $A^*(x)$ and $\Lambda(x,a)$ be as in (4.1.2) and (4.1.3), respectively, and consider the new Markov control model

$$(\mathbf{X}, A, \{A^*(x) : x \in \mathbf{X}\}, Q, \widehat{C})$$

with $\widehat{C}(x,a) := \Lambda(x,a)$. We can check that this control model satisfies the assumptions of Theorem 2.4.3, i.e., Assumptions 2.1.1, 2.1.2 and 2.4.2. In this case, J is replaced by

$$\widetilde{V}(\pi,x) = \limsup_{n \to \infty} \frac{1}{n} E_x^\pi \Big[\sum_{k=0}^{n-1} \Lambda(x_k,a_k)\Big].$$

Note that

$$\widetilde{V}(f,x) = \mu_f(\Lambda_f) \quad \text{for each } f \in \mathbb{F}. \tag{4.3.5}$$

On the other hand, by Theorem 2.4.3, and by Remark 4.1.2, there exists a triplet (h_2, f^*, σ_*^2) with $h_2 \in B_W(\mathbf{X})$, $f^* \in \mathbb{F}_{cp}$ and $\sigma_*^2 := \inf_{f\in\mathbb{F}} \mu_f(\Lambda_f)$, such that

$$\begin{aligned} h_2(x) &= \min_{a\in A^*(x)} \left[\Lambda(x,a) - \sigma_*^2 + \int_{\mathbf{X}} h_2(y) Q(dy|x,a)\right] \\ &= \Lambda_{f^*} - \sigma_*^2 + \int_{\mathbf{X}} h_2(y) Q_{f^*}(dy|x) \quad \forall x \in \mathbf{X}. \end{aligned}$$

From this equation and by Remark 4.2.3

$$\sigma_*^2 = \mu_{f^*}(\Lambda_{f^*}) = V(f^*, x) \quad \forall x \in \mathbf{X},$$

Moreover, by (4.3.5) and Remark 4.2.3 again

$$\sigma_*^2 \leq \widetilde{V}(\hat{f}, x) = \mu_{\hat{f}}(\Lambda_{\hat{f}}) = V(\hat{f}, x) \quad \forall \hat{f} \in \mathbb{F}_{cp}, x \in \mathbf{X}.$$

By Lemma 4.2.2(c) we get that for each $f \in \mathbb{F}_{ear}$ there exists $\hat{f} \in \mathbb{F}_{cp}$ such that

$$V(\hat{f}, x) = V(f, x) = \sigma_f^2 \quad \forall x \in \mathbf{X}.$$

Hence

$$\sigma_*^2 \leq \sigma_f^2 = \mu_f(\Lambda_f) = V(f, x) \quad \forall f \in \mathbb{F}_{ear}, x \in \mathbf{X}. \quad \blacksquare$$

4.4 Asymptotic normality

In this section we study asymptotic normality of MCPs in Borel spaces with unbounded rewards. In [21], Mandl study the asymptotic normality for finite state MCPs. Following Mandl's approach it is possible to prove asymptotic normality for MCPs in Borel spaces.

We show that for every canonical policy $f^* \in \mathbb{F}_{cp}$ satisfying Theorem 4.3.1, the asymptotic distribution of $(S_n(f^*, x) - n\rho*)/\sqrt{n}$ for $n \to \infty$ is normal $N(0, \sigma_*^2)$. To do this we introduce the next assumption

Assumption 4.4.1 *There exists a positive constant K_3 such that*

$$|r(x,a)| \leq K_3 \sqrt[4]{W(x)} \quad \forall (x,a) \in \mathbb{K}.$$

Remark 4.4.2 *Assumption 4.4.1 implies Assumptions 2.1.1-(b) and 3.2.1*

We shall begin with some preliminary results.

Lemma 4.4.3 *Suppose that Assumptions 2.1.1, 2.1.2 and 4.4.1 hold. Let h^*_φ be the function in Theorem 2.1.4-(iii). Then h^*_φ is a $\sqrt[4]{W}$-bounded function.*

Proof. The proof is similar to the proof of Lemma 3.2.4.

The next lemma follows from the Cauchy-Schwartz inequality.

Lemma 4.4.4 *Let $v \in B_{\sqrt{W}}(\mathbf{X})$ and define Rv as*

$$Rv(x,a) := \int_{\mathbf{X}} v(y)Q(dy|x,a) \quad \forall (x,a) \in \mathbb{K}.$$

Then Rv is in $B_{\sqrt{W}}(\mathbf{X})$. Moreover, if $v \in B_{\sqrt[4]{W}}(\mathbf{X})$ then so is Rv.

Lemma 4.4.5 *Suppose that the hypotheses of Theorem 2.4.3 and Assumption 4.4.1 hold, and let h_1 be a function satisfying the AROE (2.4.5). Then h_1 is in $B_{\sqrt[4]{W}}(\mathbf{X})$*

Proof. This lemma is a direct consequence of Lemmas 2.3.1 and 4.4.3. ∎

Lemma 4.4.6 *Suppose that the hypotheses of Theorem 4.3.1 and Assumption 4.4.1 hold. Then the functions $\Lambda(\cdot,\cdot)$ and h_2 are $\sqrt{W}$-bounded.*

Proof. By Lemma 4.4.6 the function h_1 satisfying the AROE (2.4.5) is in $B_{\sqrt[4]{W}}(\mathbf{X})$. Then h_1^2 is in $B_{\sqrt{W}}(\mathbf{X})$. By Lemma 4.4.4 the functions

$$\int_{\mathbf{X}} h_1^2(y)Q(dy|\cdot,\cdot) \quad \text{and} \quad \int_{\mathbf{X}} h_1(y)Q(dy|\cdot,\cdot)$$

are $\sqrt{W}$-bounded, hence

$$\Lambda(\cdot,\cdot) = \int_{\mathbf{X}} h_1^2(y)Q(dy|\cdot,\cdot) - \left[\int_{\mathbf{X}} h_1(y)Q(dy|\cdot,\cdot)\right]^2$$

is $\sqrt{W}$-bounded.

Finally, by Lemma 4.2.2-(a) applied to the MCM

$$(\mathbf{X}, A, \{A^*(x) : x \in \mathbf{X}\}, Q, \widehat{C})$$

we have that h_2 is in $B_{\sqrt{W}}(\mathbf{X})$. ∎

Theorem 4.4.7 *Suppose that Assumptions 2.1.1, 2.1.2, 2.4.2 and 4.4.1 hold. Let $f^* \in \mathbb{F}_{cp}$ be a canonical policy satisfying Theorem 4.3.1. Then for every initial state x*

$$\frac{S_n(f^*, x) - n\rho^*}{\sqrt{n}} \tag{4.4.1}$$

has asymptotically normal distribution $N(0, \sigma_*^2)$ as $n \to \infty$.

Proof. We define

$$\tau_1(x,a) = \int_{\mathbf{X}} h_1(y)Q(dy|x,a) - h_1(x) + r(x,a) - \rho^*$$

and

$$\tau_2(x,a) = \int_{\mathbf{X}} h_2(y)Q(dy|x,a) - h_2(x) + \Lambda(x,a) - \sigma_*^2$$

for all $(x,a) \in \mathbb{K}$. We also introduce

$$\begin{aligned}
\psi_l(x,a) &= \int_{\mathbf{X}} h_l(y)Q(dy|x,a) - h_l(x) \quad \forall x \in \mathbf{X}, l = 1,2, \\
\chi_n(u) &= \exp\{iu(S_n(f^*,x) - n\rho^*)\} \quad \text{for } n = 1,2,\cdots; u \in \mathbb{R}, \\
\chi_0(u) &= 1, \\
e_1(z) &= \exp\{iz\} - iz - 1, \\
e_2(z) &= \exp\{iz\} + \frac{z^2}{2} - iz - 1.
\end{aligned}$$

Observe that

$$\tau_1(x,a) = \psi_1(x,a) + r(x,a) - \rho^*, \tag{4.4.2}$$

and

$$\tau_2(x,a) = \psi_2(x,a) + \Lambda(x,a) - \sigma_*^2 \tag{4.4.3}$$

for all $(x,a) \in \mathbb{K}$.

To prove the theorem we have to verify

$$\lim_{n\to\infty} E_x^{f^*} \chi_n\Big(\frac{u}{\sqrt{n}}\Big) = \exp\{-\frac{1}{2}\sigma_*^2 u^2\}. \tag{4.4.4}$$

Notice that $\psi_l(x_m, a_m)$ is the conditional expectation of $h_l(x_{m+1}) - h_l(x_m)$ given x_m, a_m for $l = 1,2$, that is,

$$\psi_l(x_m, a_m) = E_x^{f^*}[h_l(x_{m+1}) - h_l(x_m)|x_m, a_m].$$

This yields, with $\chi_m := \chi_m(u)$, the following equations

$$0 = iuE_x^{f^*}\Big[\sum_{m=0}^{n-1}\chi_m\psi_1(x_m,a_m) - \sum_{m=0}^{n-1}\chi_m\Big(h_1(x_{m+1}) - h_1(x_m)\Big)\Big] \qquad (4.4.5)$$

and

$$0 = \frac{u^2}{2}E_x^{f^*}\Big[\sum_{m=0}^{n-1}\chi_m\Big(h_2(x_{m+1}) - h_2(x_m)\Big) - \sum_{m=0}^{n-1}\chi_m\psi_2(x_m,a_m)\Big]. \qquad (4.4.6)$$

Furthermore, letting $r := r(x_m,a_m)$, $e_1 := e_1\Big(u(r-\rho^*)\Big)$ and $e_2 := e_2\Big(u(r-\rho^*)\Big)$,

$$\begin{aligned} E_x^{f^*}\chi_n - 1 &= E_x^{f^*}\sum_{m=0}^{n-1}(\chi_{m+1}-\chi_m) \\ &= E_x^{f^*}\sum_{m=0}^{n-1}\Big[iu(r-\rho^*) - \frac{1}{2}u^2(r-\rho^*)^2 + e_2\Big]\chi_m, \quad (4.4.7) \end{aligned}$$

$$-iuE_x^{f^*}\sum_{m=0}^{n-1}\chi_m\Big(h_1(x_{m+1}) - h_1(x_m)\Big) =$$

$$iuE_x^{f^*}\Big[h_1(x_0) - \chi_n h_1(x_n) + \sum_{m=0}^{n-1}h_1(x_{m+1})\Big(\chi_{m+1}-\chi_m\Big)\Big] =$$

$$iuE_x^{f^*}\Big[h_1(x_0) - \chi_n h_1(x_n) + \sum_{m=0}^{n-1}h_1(x_{m+1})\Big(iu(r-\rho^*)+e_1\Big)\chi_m\Big], \qquad (4.4.8)$$

$$\frac{u^2}{2}E_x^{f^*}\sum_{m=0}^{n-1}\chi_m\Big(h_2(x_{m+1}) - h_2(x_m)\Big) =$$

$$-\frac{u^2}{2}E_x^{f^*}\Big[h_2(x_0) - \chi_n h_2(x_n) + \sum_{m=0}^{n-1}h_2(x_{m+1})\Big(\exp\{iu(r-\rho^*)\} - 1\Big)\chi_m\Big]. \qquad (4.4.9)$$

Adding (4.4.5)-(4.4.9) and using (4.4.2)

$$E_x^{f^*}\chi_n - 1 =$$

$$iuE_x^{f^*}\Big[h_1(x_0) - \chi_n h_1(x_n) + \sum_{m=0}^{n-1}\chi_m\tau_1(x_m,a_m) + \sum_{m=0}^{n-1}e_1h_1(x_{m+1})\chi_m\Big]$$

$$-\frac{u^2}{2}E_x^{f^*}\sum_{m=0}^{n-1}\chi_m\Big\{\psi_2(x_m,a_m) + 2h_1(x_{m+1})(r-\rho^*) + (r-\rho^*)^2\Big\}$$

$$-\frac{u^2}{2}E_x^{f^*}\Big[h_2(x_0) - \chi_n h_2(x_n) + \sum_{m=0}^{n-1}h_2(x_{m+1})\Big(\exp\{iu(r-\rho^*)\} - 1\Big)\chi_m\Big]$$

$$+E_x^{f^*}\sum_{m=0}^{n-1}e_2\chi_m.$$

Hence

$$E_x^{f^*}\chi_n - 1 =$$

$$\kappa''(n,u) - \frac{u^2}{2}E_x^{f^*}\sum_{m=0}^{n-1}\chi_m\Big\{\psi_2(x_m,a_m) + 2h_1(x_{m+1})(r-\rho^*) + (r-\rho^*)^2\Big\} \quad (4.4.10)$$

with

$$\kappa''(n,u) =$$

$$iuE_x^{f^*}\Big[h_1(x_0) - \chi_n h_1(x_n) + \sum_{m=0}^{n-1}\chi_m\tau_1(x_m,a_m) + \sum_{m=0}^{n-1}e_1h_1(x_{m+1})\chi_m\Big]$$

$$-\frac{u^2}{2}E_x^{f^*}\Big[h_2(x_0) - \chi_n h_2(x_n) + \sum_{m=0}^{n-1}h_2(x_{m+1})\Big(\exp\{iu(r-\rho^*)\} - 1\Big)\chi_m\Big]$$

$$+E_x^{f^*}\sum_{m=0}^{n-1}e_2\chi_m. \quad (4.4.11)$$

Observing that

$$\Lambda(x_m,a_m) = E_x^{f^*}[h_1^2(x_{m+1})|x_m,a_m] - \Big(E_x^{f^*}[h_1(x_{m+1})|x_m,a_m]\Big)^2$$

and in view of (4.4.3), we can express (4.4.10) as

$$E_x^{f^*}\chi_n - 1 =$$

$$\kappa''(n,u) - \frac{u^2}{2}E_x^{f^*}\sum_{m=0}^{n-1}\chi_m\Big\{\sigma_*^2 + \tau_2(x_m,a_m) - h_1^2(x_{m+1}) + \Big(E_x^{f^*}[h_1(x_{m+1})|x_m,a_m] + r(x_m,a_m) - \rho^*\Big)^2\Big\} =$$

$$\kappa''(n,u) - \frac{u^2}{2}E_x^{f^*}\sum_{m=0}^{n-1}\chi_m\Big\{\sigma_*^2 + \tau_2(x_m,a_m) - h_1^2(x_{m+1}) + \Big(\int_{\mathbf{X}} h_1(y)Q(dy|x_m,a_m) + r(x_m,a_m) - \rho^*\Big)^2\Big\}$$

Since f^* is a canonical policy, from Remark 4.1.2 we have

$$E_x^{f^*}\chi_n - 1 =$$

$$\kappa''(n,u) - \frac{u^2}{2}E_x^{f^*}\sum_{m=0}^{n-1}\chi_m\Big\{\sigma_*^2 + \tau_2(x_m,a_m) - h_1^2(x_{m+1}) + h_1^2(x_m)\Big\} =$$

$$= \kappa''(n,u) - \frac{u^2\sigma_*^2}{2}\sum_{m=0}^{n-1}E_x^{f^*}\chi_m - \frac{u^2}{2}E_x^{f^*}\Big[h_1^2(x_0) - \chi_n h_1^2(x_n) + \sum_{m=0}^{n-1}\chi_m\tau_2(x_m,a_m) + \sum_{m=0}^{n-1}h_1^2(x_{m+1})\Big(\exp\{iu(r-\rho^*)\} - 1\Big)\chi_m\Big].$$

Hence

$$E_x^{f^*}\chi_n = 1 - \frac{u^2\sigma_*^2}{2}\sum_{m=0}^{n-1}E_x^{f^*}\chi_m + \kappa'(n,u) \tag{4.4.12}$$

with

$$\kappa'(n,u) = \kappa''(n,u) - \frac{u^2}{2}E_x^{f^*}\Big[h_1^2(x_0) - \chi_n h_1^2(x_n) + \sum_{m=0}^{n-1}\chi_m\tau_2(x_m,a_m) + \sum_{m=0}^{n-1}h_1^2(x_{m+1})\Big(\exp\{iu(r-\rho^*)\} - 1\Big)\chi_m\Big]. \tag{4.4.13}$$

Let us rewrite (4.4.12) as

$$E_x^{f^*}\chi_n = 1 + \Big(\exp\{-\frac{u^2\sigma_*^2}{2}\} - 1\Big)\sum_{m=0}^{n-1} E_x^{f^*}\chi_m + \kappa(n,u) \tag{4.4.14}$$

with

$$\kappa(n,u) = \kappa'(n,u) + \Big[1 - \frac{u^2\sigma_*^2}{2} - \exp\{-\frac{u^2\sigma_*^2}{2}\}\Big]\sum_{m=0}^{n-1} E_x^{f^*}\chi_m. \tag{4.4.15}$$

From (4.4.14), an induction argument gives

$$E_x^{f^*}\chi_n(u) = \exp\{-\frac{n\sigma_*^2u^2}{2}\}+$$

$$\Big[\exp\{-\frac{\sigma_*^2u^2}{2}\}-1\Big]\sum_{m=0}^{n-1}\exp\Big\{-\frac{\sigma_*^2u^2}{2}(n-1-m)\Big\}\kappa(m,u)+\kappa(n,u). \tag{4.4.16}$$

Observe that the proof of the limit (4.4.4) and consequently this theorem follows from (4.4.16) if we show

$$\max_{1\le m\le n}|\kappa(m,\frac{u}{\sqrt{n}})| \to 0 \quad \text{as} \quad n\to\infty. \tag{4.4.17}$$

This relation is obtained by an inspection of the differents terms of $\kappa(m, u/\sqrt{n})$:

(i) Since f^* is a canonical policy satisfying Theorem 4.3.1, we have $\tau_1(x_m, a_m) = 0$ for $m = 0, 1, \cdots$ in (4.4.11). Similarly, $\tau_2(x_m, a_m) = 0$ in (4.4.13).

(ii) By (2.1.1) we have that

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}E_x^{f^*}h(x_n) = 0 \quad \text{and} \quad \lim_{n\to\infty}\frac{1}{n}E_x^{f^*}h(x_n) = 0$$

for every h in $B_W(\mathbf{X})$. This limit appears in (4.4.11) and (4.4.13) when we replace u by $u/\sqrt{n}$.

(iii) In this part we prove the limit

$$\lim_{n\to\infty}\frac{1}{\sqrt{n}}E_x^{f^*}\sum_{m=0}^{n-1} e_1h_1(x_{m+1})\chi_m = 0;$$

see (4.4.11).

From the fact $|e_1(z)| \leq z^2/2$ for all z in $\mathbb{R}$, we obtain

$$\Big|\frac{1}{\sqrt{n}}E_x^{f^*}\sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m\Big| \leq \frac{1}{2\sqrt{n}}E_x^{f^*}\sum_{m=0}^{n-1}\frac{u^2}{n}|h_1(x_{m+1})|(r(x_m,a_m)-\rho^*)^2$$

$$= \frac{u^2}{2n^{3/2}}E_x^{f^*}\sum_{m=0}^{n-1}|\int_{\mathbf{X}} h_1(dy|x_m,f^*(x_m))|(r(x_m,f^*(x_m))-\rho^*)^2.$$

By Lemma 4.4.5, h_1 is $\sqrt[4]{W}$-bounded, in particular h_1 is w-bounded. Hence, by Lemma 4.4.4, the function $\int_{\mathbf{X}} h_1(y)Q(dy|x,f^*(x))$ is w-bounded. On the other hand, by Assumption 4.4.1 $(r(x,f^*(x))-\rho^*)^2$ is w-bounded. Therefore

$$\Big|\frac{1}{\sqrt{n}}E_x^{f^*}\sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m\Big| \leq \frac{C'u^2}{2n^{3/2}}E_x^{f^*}\sum_{m=0}^{n-1} W(x_m)$$

where C' is a constant depending on h_1 and r. By (2.1.1) we obtain

$$\Big|\frac{1}{\sqrt{n}}E_x^{f^*}\sum_{m=0}^{n-1} e_1 h_1(x_{m+1})\chi_m\Big| \leq \frac{C'u^2}{2n^{3/2}}n\Big(\lambda W(x)+\frac{\nu(W)}{(1-\lambda)\nu(\mathbf{X})}\Big).$$

which converges to zero as $n\to\infty$.

(iv) We shall next prove

$$\lim_{n\to\infty}\frac{1}{n}E_x^{f^*}\sum_{m=0}^{n-1} e_2\chi_m = 0.$$

This limit appears in (4.4.11) when we replace u by $u/\sqrt{n}$.

Notice that $|e_2(z)| \leq |z|^3/6$ for all z in $\mathbb{R}$. So, by Assumptions 2.1.1-(a) and 4.4.1, together with (2.1.1),

$$\begin{aligned}
\Big|\frac{1}{n}E_x^{f^*}\sum_{m=0}^{n-1} e_2\chi_m\Big| &\leq \frac{|u|^3}{6n^{5/2}}E_x^{f^*}\sum_{m=0}^{n-1}|r(x_m,f^*(x_m))-\rho^*|^3\\
&\leq \frac{k'^3|u|^3}{6n^{5/2}}E_x^{f^*}\sum_{m=0}^{n-1} W(x_m)^{3/4}\\
&\leq \frac{k^3|u|^3}{6n^{5/2}}E_x^{f^*}\sum_{m=0}^{n-1} W(x_m)\\
&\leq \frac{k^3|u|^3}{6n^{3/2}}\Big(\lambda W(x)+\frac{\nu(W)}{(1-\lambda)\nu(\mathbf{X})}\Big)
\end{aligned}$$

which converges to zero as $n \to \infty$, with k and k' some constants.

(v) Let h be in $B_w(\mathbf{X})$. Then

$$\lim_{n\to\infty} \frac{1}{n} E_x^{f^*} \sum_{m=0}^{n-1} h(x_{m+1}) \Big(\exp\{i \frac{u}{\sqrt{n}}(r - \rho^*)\} - 1 \Big) \chi_m = 0.$$

This limit appears in (4.4.11) and (4.4.13) when u is replaced by $u/\sqrt{n}$.

It follows from the relation $e_1(z) = \exp\{iz\} - iz - 1$ that

$$\exp\{i \frac{u}{\sqrt{n}}(r - \rho^*)\} - 1 = i \frac{u}{\sqrt{n}}(r - \rho^*) + e_1 \Big(\frac{u}{\sqrt{n}}(r - \rho^*) \Big).$$

So

$$|\frac{1}{n} E_x^{f^*} \sum_{m=0}^{n-1} h(x_{m+1}) \Big(\exp\{i \frac{u}{\sqrt{n}}(r - \rho^*)\} - 1 \Big) \chi_m| \leq$$

$$\frac{|u|}{n^{3/2}} E_x^{f^*} \sum_{m=0}^{n-1} |h(x_{m+1})| |(r(x_m, f^*(x_m)) - \rho^*)| + \frac{1}{n} E_x^{f^*} \sum_{m=0}^{n-1} |h(x_{m+1})| |e_1|.$$

This gives the desired conclusion by similar arguments to those in (iii).

(vi) The absolute value of the expression within brackets in (4.4.15) is majorized by $\sigma_*^4 u^4/8$, then the corresponding term in $\kappa(n, u/\sqrt{n})$ is majorized by $\sigma_*^4 u^4/8n^2$.

The statements (i)-(vi) imply (4.4.17) and consequently prove the theorem. ∎

Concluding remarks. Our motivation for this chapter was to extend, under our fixed-point approach, the results concerning the variance-minimization problem studied by Hernández-Lerma, Vega-Amaya and Carrasco [18].

The minimization of variance is motivated by the fact that among the optimal policies for which the control problem is solved, those with minimal variance are preferable. This situation is verified because these policies imply asymptotic normality. Furthermore, the examples in Chapter 6 show that our assumptions to solve the variance-minimization problem are indeed verifiable.

Chapter 5

Constrained MCPs

The problem we are concerned with in this chapter is to maximize a long-run sample-path (or pathwise) average reward for the given discrete-time MCM, subject to constraints on a given finite number of long-run pathwise average costs. To this end, we give conditions for the existence of optimal policies for the problem with expected constraints (see Theorem 5.3.1). Moreover, in Theorem 5.4.1 we can show that the expected case can be solved by means of a parametric family of AROEs. Finally, we extend the results in the former steps to our problem with pathwise constraints (see Theorem 5.5.2).

For finite state MCPs, we should mention the article by Haviv [12], and the works by Ross and Varadarajan [26, 27]. For MCPs on Borel spaces we only know the recent work by Vega-Amaya [31]. The article by Haviv shows, by means of examples, that pathwise constraints are in general, more "natural" than expected constraints and because MCPs with constraints on the *expected* state-action frequencies can lead to optimal policies that do not satisfy certain principles of optimality (as Bellman's principle). In contrast, the model with pathwise constraints leads to feasible optimal policies which satisfy these principles.

The article by Vega-Amaya [31] shows, under appropiate assumptions such as positive Harris recurrence, that there exists a randomized stationary policy and an initial distribution that solve the constrained expected average cost. Such a policy also minimizes the sample path average costs for every initial distribution measure. These results are used to solve control problems with constraints on the state occupation measures.

As can be seen in the paper by Prieto-Rumeau and Hernandez-Lerma [24], previous attempts have been made to solve this problem for continuous-time

denumerable-state controlled Markov chains. We extend the results obtained in this work to the discrete-time MCM.

5.1 Expected constraints

Fix numbers $\theta_1, \cdots, \theta_q$ in $\mathbb{R}$, and measurable functions $c_1, \cdots, c_q$ in $B_W(\mathbb{K})$ interpreted as cost-per-stage functions. Now, we are concerned with the maximization, for every initial state $x \in \mathbf{X}$, of

$$J(\pi, x) := \liminf \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} r(x_k, a_k)$$

over the set of all control policies Π that satisfy the constraints

$$J_i(\pi, x) := \limsup \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} c_i(x_k, a_k) \leq \theta_i \quad \forall i = 1, \cdots, q.$$

In short, our problem is

$$\text{maximize } J(\pi, x) \tag{5.1.1}$$

$$\text{subject to: } \quad \pi \in \Pi \quad \text{and} \quad J_i(\pi, x) \leq \theta_i \quad \forall x \in \mathbf{X}, i = 1, \cdots, q. \tag{5.1.2}$$

Observe that $J(\pi, x)$ is defined as a "lim inf" whereas $J_i(\pi, x)$ is a "lim sup". This is because the function r is interpreted as a reward-per-stage function, and the functions c_i as cost-per-stage functions.

Definition 5.1.1 *A policy $\pi \in \Pi$ is said to be* feasible *for the constrained problem (CP) (5.1.1)-(5.1.2) if it satisfies the constraints in (5.1.2), that is, $J_i(\pi, x) \leq \theta_i$ for all x in $\mathbf{X}$, $i = 1, \cdots, q$. Moreover, a feasible policy π^* is called* optimal *for (5.1.1)-(5.1.2) if $J(\pi, x) \leq J(\pi^*, x)$ for every feasible π.*

Let Φ_{feas} be the class of feasible randomized stationary policies, i.e.,

$$\Phi_{feas} := \{\varphi \in \Phi : J_i(\varphi, x) \leq \theta_i \quad \forall x \in \mathbf{X}, i = 1, \cdots, q\}.$$

Henceforth $V^*(\theta_1, \cdots, \theta_q, x)$ will designate the optimal value function of (5.1.1)-(5.1.2) on the set of randomized stationary policies Φ, that is,

$$V^*(\theta_1, \cdots, \theta_q, x) := \sup_{\varphi \in \Phi_{feas}} J(\varphi, x), \tag{5.1.3}$$

for every $x \in \mathbf{X}$. Moreover, under the assumptions of Theorem 2.4.3, for $i = 1, \cdots, q$, we can define

$$\theta_{i,min} := \min_{\varphi \in \Phi} \int_{\mathbf{X}} c_{i\varphi}(y)\mu_\varphi(dy) \quad \text{and} \quad \theta_{i,max} := \max_{\varphi \in \Phi} \int_{\mathbf{X}} c_{i\varphi}(y)\mu_\varphi(dy),$$

which are finite numbers. To avoid trivial situations, we will assume that the constants θ_i in (5.1.2) verify that

$$\theta_{i,min} < \theta_i < \theta_{i,max} \quad \forall i = 1, \cdots, q. \tag{5.1.4}$$

Now let W be as in Assumption 2.1.1, and $w := \sqrt{W}$. We denote by $\mathcal{P}_w(\mathbb{K})$ the set of all Borel probability measure μ on $\mathbb{K}$ such that

$$\int_{\mathbb{K}} w(y)\mu(d(y,a)) < \infty.$$

Finally, for each $\varphi \in \Phi$, we define $\hat{\mu}_\varphi \in \mathcal{P}_w(\mathbb{K})$ as follows

$$\hat{\mu}_\varphi(B \times C) := \int_B \varphi(C|x)\mu_\varphi(dx) \quad \forall B \in \mathcal{B}(\mathbf{X}), C \in \mathcal{B}(A); \tag{5.1.5}$$

where μ_φ is the unique invariant probability measure for the transition kernel $Q_\varphi(\cdot|\cdot)$. By (2.2.16) we can see that

$$\int_{\mathbb{K}} w(y)\hat{\mu}_\varphi(d(y,a)) = \int_{\mathbf{X}} w(y)\mu_\varphi(dy) < \infty.$$

We denote by Γ the set of all these measures $\hat{\mu}_\varphi$, i.e.,

$$\Gamma := \{\hat{\mu}_\varphi : \varphi \in \Phi\} \subset \mathcal{P}_w(\mathbb{K}). \tag{5.1.6}$$

Let $B_b(\mathbf{X})$ be the set of bounded measurable functions on $\mathbf{X}$. We shall denote by R the operator defined for each v in $B_b(\mathbf{X})$ as

$$(Rv)(x,a) := \int_{\mathbf{X}} v(y)Q(dy|x,a) - v(x) \quad \forall (x,a) \in \mathbb{K}. \tag{5.1.7}$$

Following [22], we will refer to R as a drift operator.

The study of the general properties of the measures in Γ is based on the following Lemma 5.1.2:

Lemma 5.1.2 *Consider a probability measure $\hat{\mu}$ in $\mathcal{P}_w(\mathbb{K})$. A necessary and sufficient condition for $\hat{\mu}$ to be in Γ is that*

$$\int_{\mathbb{K}} Rv \, d\hat{\mu} = 0 \quad \textit{for every } v \in B_b(\mathbf{X}). \tag{5.1.8}$$

Proof. (*Necessity.*) Fix $\hat{\mu} \in \Gamma$ and $v \in B_b(\mathbf{X})$. By (5.1.6), there exist $\varphi \in \Phi$ such that $\hat{\mu} = \hat{\mu}_\varphi$. We have

$$\begin{aligned} \int_{\mathbb{K}} Rv \, d\hat{\mu} &= \int_{\mathbf{X}} \left[\int_A (Rv)(x,a)\varphi(da|x) \right] \mu_\varphi(dx) \\ &= \int_{\mathbf{X}} \left[\int_{\mathbf{X}} v(y) Q_\varphi(dy|x) - v(x) \right] \mu_\varphi(dx) \end{aligned} \tag{5.1.9}$$

Since μ_φ is an invariant probability measure for the transition kernel $Q_\varphi(\cdot|\cdot)$, we obtain

$$\int_{\mathbf{X}} \left[\int_{\mathbf{X}} v(y) Q_\varphi(dy|x) - v(x) \right] \mu_\varphi(dx) = 0$$

which proves (5.1.8).

(*Sufficiency.*) Suppose now that (5.1.8) holds for some $\hat{\mu} \in \mathcal{P}_w(\mathbb{K})$. Therefore, by a standard result on the disintegration of measures [14, Proposition D.8], there exists $\varphi \in \Phi$ such that $\hat{\mu}$ can be "disintegrated" as

$$\hat{\mu}(B \times C) = \int_B \varphi(C|x)\tilde{\mu}(dx) \quad \forall B \in \mathcal{B}(\mathbf{X}), C \in \mathcal{B}(A), \tag{5.1.10}$$

where $\tilde{\mu}(B) := \hat{\mu}(B \times A)$ for all $B \in \mathcal{B}(\mathbf{X})$ is the *marginal* (or *projection*) of $\hat{\mu}$ on $\mathbf{X}$. Letting $v(\cdot) = 1_B(\cdot)$, with B in $\mathcal{B}(\mathbf{X})$ and by a similar procedure as in (5.1.9), we obtain

$$\begin{aligned} 0 &= \int_{\mathbb{K}} Rv \, d\hat{\mu} \\ &= \int_{\mathbf{X}} \left[\int_{\mathbf{X}} v(y) Q_\varphi(dy|x) - v(x) \right] \tilde{\mu}(dx) \\ &= \int_{\mathbf{X}} \Big[Q_\varphi(B|x) - 1_B(x) \Big] \tilde{\mu}(dx), \end{aligned}$$

i.e., $\int_{\mathbf{X}} Q_\varphi(B|x)\tilde{\mu}(dx) = \tilde{\mu}(B)$ for every $B \in \mathcal{B}(\mathbf{X})$. Thus $\tilde{\mu}$ is an invariant probability measure for the kernel $Q_\varphi(\cdot|\cdot)$. By uniqueness of the i.p.m. (see Theorem 2.1.4(i)), $\tilde{\mu} = \mu_\varphi$ and hence, $\hat{\mu} = \hat{\mu}_\varphi$ as we wanted to prove. ∎

5.2 Technical preliminaries

Assumption 5.2.1 **(a)** *$Q(\cdot|\cdot,\cdot)$ is strongly continuous on $\mathbb{K}$, that is, the mapping*

$$(x,a) \mapsto \int_{\mathbf{X}} v(y)Q(dy|x,a)$$

is continuous on $\mathbb{K}$ for every measurable bounded function v on $\mathbf{X}$.

(b) *The cost functions $c_i(\cdot,\cdot) \in B_W(\mathbb{K})$ are nonnegative (or bounded below) and lower semicontinuous (l.s.c.).*

(c) *$r(\cdot,\cdot)$ is u.s.c. on $\mathbb{K}$.*

(d) *Let W be as in Assumption 2.1.1. The function $w = \sqrt{W}$, seen as a function $(x,a) \mapsto w(x)$ on $\mathbb{K}$, is continuous. Moreover, w is a moment function on $\mathbb{K}$, that is, there exists a nondecreasing sequence of compact sets $K_n \uparrow \mathbb{K}$ such that*

$$\lim_{n\to\infty} \inf\{w(x) : (x,a) \notin K_n\} = \infty.$$

(e) *The state space $\mathbf{X}$ and the control set A are separable and metrizable spaces. In particular, the set $\mathbb{K}$ of feasible state-actions pairs, is separable and metrizable.*

Notice that Assumption 5.2.1(a) implies Assumption 2.4.2(c), and Assumption 5.2.1(c) implies Assumption 2.4.2(b).

Under Assumptions 5.2.1(d)-(e), throughout the remainder of this chapter, we consider the w-weak topology on $\mathcal{P}_w(\mathbb{K})$, i.e., the smallest topology for which the mapping

$$\hat{\mu} \mapsto \int_{\mathbb{K}} v d\hat{\mu}$$

on $\mathcal{P}_w(\mathbb{K})$ is continuous for every $v \in C_w(\mathbb{K})$, where $C_w(\mathbb{K})$ is the linear subspace of $B_w(\mathbb{K})$ that consists of the continuous functions on $\mathbb{K}$.

Lemma 5.2.2 *Under Assumptions 2.1.1-(a), 2.1.2, 5.2.1-(a) and 5.2.1-(d), the set Γ is convex and compact in the w-weak topology. Let $\theta_1, \cdots, \theta_q$ be as in (5.1.4). Then the set*

$$I := \{\hat{\mu} \in \Gamma : \int_{\mathbb{K}} c_i d\hat{\mu} \le \theta_i \text{ for } i = 1, \cdots, q\} \subset \mathcal{P}_w(\mathbb{K}) \tag{5.2.1}$$

is a convex and closed (with respect to the w-weak topology) subset of Γ. *In particular, I is compact.*

Proof. The convexity property follows directly from Lemma 5.1.2, because any convex combination of measures in Γ lies in Γ.

To prove that Γ is compact we will first show that Γ is closed in the w-weak topology. Let us consider a sequence $\{\hat{\mu}_{\varphi_n}\}$ in Γ that converges to $\hat{\mu} \in \mathcal{P}_w(\mathbb{K})$ in the w-weak topology. We need to show that $\hat{\mu} \in \Gamma$. From Lemma 5.1.2 we have

$$\int_{\mathbb{K}} Rv \, d\hat{\mu}_{\varphi_n} = 0 \quad \text{for every } v \in B_b(\mathbf{X}), n \in \mathbb{N}.$$

Moreover, by the Assumption 5.2.1(a), if v is a continuous bounded function on $\mathbf{X}$, then $(x,a) \mapsto (Rv)(x,a) = \int_{\mathbf{X}} v(y)Q(dy|x,a) - v(x)$ is a continuous bounded function on $\mathbb{K}$. This implies

$$\int_{\mathbb{K}} Rv \, d\hat{\mu} = 0 \quad \forall v \in C_b(\mathbf{X}),$$

where $C_b(\mathbf{X})$ is the linear space of all bounded continuous functions on $\mathbf{X}$. On the other hand, by the argument leading to (5.1.10), there exists $\varphi \in \Phi$ such that

$$\hat{\mu}(B \times C) = \int_B \varphi(C|x)\tilde{\mu}(dx) \quad \forall B \in \mathcal{B}(\mathbf{X}), C \in \mathcal{B}(A),$$

where $\tilde{\mu}$ is the marginal of $\hat{\mu}$ on $\mathbf{X}$. Combining these facts, we have

$$\begin{aligned} 0 &= \int_{\mathbb{K}} Rv \, d\hat{\mu} \qquad (5.2.2) \\ &= \int_{\mathbf{X}} \left[\int_{\mathbf{X}} v(y) Q_\varphi(dy|x) - v(x) \right] \tilde{\mu}(dx) \quad \forall v \in C_b(\mathbf{X}). \end{aligned}$$

Now consider an arbitrary closed subset F of the metric space $\mathbf{X}$. Then there exists a sequence $\{v_n\}$ of continuous bounded functions on $\mathbf{X}$ such that $0 \le v_n \le 1$ and

$$\lim_{n\to\infty} v_n(x) = 1_F(x) \quad \forall x \in \mathbf{X}.$$

Taking $v = v_n$ in (5.2.2) and using Lebesgue's dominated convergence theorem we obtain

$$\int_{\mathbf{X}} \Big[Q_\varphi(F|x) - 1_F(x) \Big] \tilde{\mu}(dx) = 0.$$

Equivalently,

$$\tilde{\mu}(F) = \int_{\mathbf{X}} Q_\varphi(F|x)\tilde{\mu}(dx) \quad \text{for every closed subset } F \subset \mathbf{X},$$

and

$$\tilde{\mu}(G) = \int_{\mathbf{X}} Q_\varphi(G|x)\tilde{\mu}(dx) \quad \text{for every open subset } G \subset \mathbf{X}.$$

Because every finite Borel measure on the metric space $\mathbf{X}$ is regular (see, for instance, [5, Theorem 7.1.3]), for each $B \in \mathcal{B}(\mathbf{X})$ and for any closed subset $F \subset B$, and for any open subset G such that $B \subset G$:

$$\tilde{\mu}(F) = \int_{\mathbf{X}} Q_\varphi(F|x)\tilde{\mu}(dx) \le \int_{\mathbf{X}} Q_\varphi(B|x)\tilde{\mu}(dx) \le \int_{\mathbf{X}} Q_\varphi(G|x)\tilde{\mu}(dx) = \tilde{\mu}(G).$$

This yields, by the regularity of $\tilde{\mu}$,

$$\tilde{\mu}(B) = \sup_{F\subset B \text{ closed}} \tilde{\mu}(F) \le \int_{\mathbf{X}} Q_\varphi(B|x)\tilde{\mu}(dx) \le \inf_{B\subset G \text{ open}} \tilde{\mu}(G) = \tilde{\mu}(B);$$

hence

$$\tilde{\mu}(B) = \int_{\mathbf{X}} Q_\varphi(B|x)\tilde{\mu}(dx) \quad \text{for every Borel subset } B \subset \mathbf{X}.$$

Thus $\tilde{\mu}$ is the unique invariant probability measure for $Q_\varphi(\cdot|\cdot)$, i.e., $\tilde{\mu} = \mu_\varphi$ and so $\hat{\mu} = \hat{\mu}_\varphi$ in Γ. In conclusion, Γ is closed.

To prove compactness, it suffices to show that Γ is relatively compact in the w-weak topology. By (2.2.16) and the fact $w(x) \le W(x)/\sqrt{\theta}$ for all $x \in \mathbf{X}$, we have

$$\sup_{\hat{\mu}\in\Gamma} \int_{\mathbb{K}} w d\hat{\mu} = \sup_{\varphi\in\Phi} \int_{\mathbf{X}} w d\mu_\varphi \le \frac{\nu(W)}{(\nu(\mathbf{X})\cdot\sqrt{\theta})(1-\lambda)} < \infty. \tag{5.2.3}$$

On the other hand, from Assumption 5.2.1(d), there exists a nondecreasing sequence of compact sets $K_n \uparrow \mathbb{K}$ such that

$$\lim_{n\to\infty} \inf\{w(x) : (x,a) \notin K_n\} = \infty.$$

Define $w_n := \inf\{w(x) : (x,a) \notin K_n\}$, then by (2.2.16) again

$$\begin{aligned} w_n \int_{K_n^c} w d\hat{\mu}_\varphi &\le \int_{\mathbb{K}} w^2 d\hat{\mu}_\varphi = \int_{\mathbf{X}} W d\mu_\varphi \\ &\le \frac{\nu(W)}{\nu(\mathbf{X})(1-\lambda)}, \quad \text{for all } \varphi \in \Phi. \end{aligned}$$

This inequality implies that for each $\varepsilon > 0$ there exists a compact subset K of $\mathbb{K}$ such that

$$\sup_{\hat{\mu}\in\Gamma} \int_{K^c} w d\hat{\mu} \leq \varepsilon. \tag{5.2.4}$$

By (5.2.3), (5.2.4), and Prohorov's theorem (see, for instance, [8, Appendix A.5]), we conclude that Γ is relatively compact in the w-weak topology. Therefore, since in addition Γ is closed, it is compact.

It only remains to prove the compactness of I. Actually, we only have to prove that I is sequentially closed. Let $\{\hat{\mu}_{\varphi_n}\}$ be a sequence in I that converges to a measure $\hat{\mu} \in \mathcal{P}_w(\mathbb{K})$ with respect to the w-weak topology. By the arguments above,we know that $\hat{\mu}$ is in Γ. Because $C_b(\mathbb{K}) \subset C_w(\mathbb{K})$, we also have that $\{\hat{\mu}_{\varphi_n}\}$ converges weakly to $\hat{\mu}$. Therefore, since $c_i,\ i = 1, \cdots, q$, is bounded below and l.s.c. (see [14, Appendix E.2])

$$\int_{\mathbb{K}} c_i d\hat{\mu} \leq \liminf_{n\to\infty} \int_{\mathbb{K}} c_i d\hat{\mu}_{\varphi_n} \leq \theta_i \quad \text{for } i = 1, \cdots, q.$$

Thus $\hat{\mu}$ is in I, and so I is a compact set in $\mathcal{P}_w(\mathbb{K})$ provided with the w-weak topology. ■

Notation. Let $\theta := (\theta_1, \cdots, \theta_q)$, $\theta_{min} := (\theta_{1,min}, \cdots, \theta_{q,min})$, and $\theta_{max} := (\theta_{1,max}, \cdots, \theta_{q,max})$. We define

$$[\theta_{min}, \theta_{max}] := [\theta_{1,min}, \theta_{1,max}] \times \cdots \times [\theta_{q,min}, \theta_{q,max}] \subset \mathbb{R}^q.$$

Moreover, if $\theta, \theta' \in \mathbb{R}^q$, then $\theta \leq \theta'$ means that

$$\theta_i \leq \theta'_i \quad \forall i = 1, \cdots, q,$$

and $\theta << \theta'$ means

$$\theta_i < \theta'_i \quad \forall i = 1, \cdots, q.$$

Recall that a real-valued function V on $\mathbb{R}^q$ is nondecreasing if:

$$\forall \theta, \theta' \in \mathbb{R}^q \text{ such that } \theta \leq \theta', \quad \text{we have} \quad V(\theta) \leq V(\theta').$$

MCPs with expected constraints. Our next result is a direct consequence of the convexity of Γ (Lemma 5.2.2), and it is stated without proof.

Lemma 5.2.3 *The function*

$$\theta \mapsto V(\theta) := \sup\Big\{ \int_{\mathbb{K}} r d\hat{\mu} : \hat{\mu} \in \Gamma, \int_{\mathbb{K}} c_i d\hat{\mu} \leq \theta_i, \textit{for } i = 1, \cdots, q \Big\} \tag{5.2.5}$$

is concave and nondecreasing on $[\theta_{min}, \theta_{max}] \subset \mathbb{R}^q$.

Remark 5.2.4 *Note that for each* $\varphi \in \Phi$

$$\begin{aligned} J(\varphi, x) &= \lim_{n\to\infty} \frac{1}{n} E_x^{\varphi} \sum_{k=0}^{n-1} r_{\varphi}(x_k) = \int_{\mathbf{X}} r_{\varphi} d\mu_{\varphi} = \int_{\mathbb{K}} r d\hat{\mu}_{\varphi}, \\ J_i(\varphi, x) &= \lim_{n\to\infty} \frac{1}{n} E_x^{\varphi} \sum_{k=0}^{n-1} c_{i\varphi}(x_k) = \int_{\mathbf{X}} c_{i\varphi} d\mu_{\varphi} = \int_{\mathbb{K}} c_i d\hat{\mu}_{\varphi}, \end{aligned}$$

for $i = 1, \cdots, q$ *and all* $x \in \mathbf{X}$. *Comparing (5.2.5) and (5.1.3) above, we have*

$$V(\theta) = V^*(\theta, x) \quad \forall x \in \mathbf{X}, \tag{5.2.6}$$

for all $\theta_{min} << \theta << \theta_{max}$.

Lemma 5.2.5 *Under Assumptions 2.1.1, 2.1.2, 3.2.1 and 5.2.1, the mapping*

$$\hat{\mu} \mapsto \int_{\mathbb{K}} r d\hat{\mu} \in \mathbb{R}$$

on $\mathcal{P}_w(\mathbb{K})$ *is u.s.c. on* $\mathcal{P}_w(\mathbb{K})$ *with respect to the* w*-weak topology.*

Proof. First, we prove that if $g : \mathbb{K} \to \mathbb{R}$ is bounded below and l.s.c., then the mapping

$$\mathcal{G} : \hat{\mu} \mapsto \int_{\mathbb{K}} g \, d\hat{\mu}$$

is l.s.c. on $\mathcal{P}_w(\mathbb{K})$. Indeed, without loss of generality, suppose that g is a nonnegative l.s.c. function. Let $\{\hat{\mu}_n\}$ and $\hat{\mu}$ probability measures in $\mathcal{P}_w(\mathbb{K})$ such that $\{\hat{\mu}_n\}$ converges w-weakly to $\hat{\mu}$. Because $C_b(\mathbb{K}) \subset C_w(\mathbb{K})$, then $\{\hat{\mu}_n\}$ converges weakly to $\hat{\mu}$. From a well-known result (see [14, Appendix E.2]), we obtain

$$\liminf_{n\to\infty} \int_{\mathbb{K}} g \, d\hat{\mu}_n \geq \int_{\mathbb{K}} g \, d\hat{\mu}.$$

That is

$$\liminf_{n\to\infty} \mathcal{G}(\hat{\mu}_n) \geq \mathcal{G}(\hat{\mu}).$$

This means that the mapping $\mathcal{G}$ is l.s.c. with respect to the w-weak topology on $\mathcal{P}_w(\mathbb{K})$.

On the other hand, by the Assumptions 3.2.1 and 5.2.1, the function $g(x,a) := K_2' w(x) - r(x,a)$ is a nonnegative l.s.c. function on $\mathbb{K}$, with $K_2' := \sqrt{K_2}$. From the previous paragraph, the mapping

$$\hat{\mu} \mapsto \int_{\mathbb{K}} (K_2' w - r) d\hat{\mu} = K_2' \int_{\mathbb{K}} w d\hat{\mu} - \int_{\mathbb{K}} r d\hat{\mu}$$

is l.s.c. and finite-valued on $\mathcal{P}_w(\mathbb{K})$. Since $w \in C_w(\mathbb{K})$, the mapping

$$\hat{\mu} \mapsto -K_2' \int_{\mathbb{K}} w d\hat{\mu}$$

on $\mathcal{P}_w(\mathbb{K})$ is l.s.c. Because the addition of two l.s.c. functions is also l.s.c., we obtain that the mapping

$$\hat{\mu} \mapsto - \int_{\mathbb{K}} r d\hat{\mu}$$

on $\mathcal{P}_w(\mathbb{K})$ is l.s.c., and therefore

$$\hat{\mu} \mapsto \int_{\mathbb{K}} r d\hat{\mu}$$

is u.s.c. on $\mathcal{P}_w(\mathbb{K})$. ■

5.3 Optimal policies

The following theorem states the existence of an optimal policy for the constrained problem (5.1.1)-(5.1.2). Furthermore, it establishes the existence of a solution to the average reward optimality equation (AROE) (5.3.2).

Optimal policies for the CP (5.1.1)-(5.1.2) satisfying the AROE (5.3.1), are called *constrained canonical policies* (ccp).

Theorem 5.3.1 *Suppose that Assumptions 2.1.1, 2.1.2, 2.4.2, 3.2.1 and 5.2.1 are satisfied. Let $\theta_0 = (\theta_1, \cdots, \theta_q)$ be such that $\theta_{min} << \theta_0 << \theta_{max}$. Then:*

(i) *The value function $V^*(\theta_0, x)$ does not depend on $x \in \mathbf{X}$ and $V(\theta_0) = V^*(\theta_0) \equiv V^*(\theta_0, x)$, with $V(\theta_0)$ as in (5.2.5).*

(ii) *There exist* $\Lambda_0 = (\lambda_{01}, \cdots, \lambda_{0q}) \leq 0$ *and* $h \in B_W(\mathbf{X})$ *such that the AROE*

$$V^*(\theta_0) + h(x) = \sup_{a \in A(x)} \Big[r(x,a) + \sum_{i=1}^{q} \lambda_{0i} \Big(c_i(x,a) - \theta_i \Big) + \int_{\mathbf{X}} h(y) Q(dy|x,a) \Big] \tag{5.3.1}$$

holds for every $x \in \mathbf{X}$. *If in addition,* c_i *is in* $B_w(\mathbf{X})$ *for* $i = 1, \cdots, q$, *then the function* h *belongs to* $B_w(\mathbf{X})$.

(iii) *There exists an optimal randomized stationary policy* $\varphi^* \in \Phi$ *for the constrained problem (5.1.1)-(5.1.2) that attains the maximum in the right-side of (5.3.1), i.e.,*

$$V^*(\theta_0) + h(x) = r_{\varphi^*}(x) + \sum_{i=1}^{q} \lambda_{0i} \Big(c_{i\varphi^*}(x) - \theta_i \Big) + \int_{\mathbf{X}} h(y) Q_{\varphi^*}(dy|x) \tag{5.3.2}$$

for all $x \in \mathbf{X}$. *Moreover, if*

$$\theta := \Big(\int_{\mathbb{K}} c_1 d\hat{\mu}_{\varphi^*}, \cdots, \int_{\mathbb{K}} c_q d\hat{\mu}_{\varphi^*} \Big) = (J_1(\varphi^*, x), \cdots, J_q(\varphi^*, x)),$$

the following "orthogonality" property is satisfied

$$(\theta - \theta_0) \cdot \Lambda_0 = 0; \tag{5.3.3}$$

where " $\cdot$ " denotes the usual scalar product in $\mathbb{R}^q$.

Proof. *Proof of (i).* This part follows from Remark 5.2.4.

Proof of (ii). By Lemma 5.2.3, the function V, defined on the q-dimensional closed bounded interval $[\theta_{min}, \theta_{max}]$, is concave and nondecreasing. Note that for every $\hat{\mu} \in \Gamma$ and if we define

$$\theta := \Big(\int_{\mathbb{K}} c_1 d\hat{\mu}, \cdots, \int_{\mathbb{K}} c_q d\hat{\mu} \Big) \equiv \int_{\mathbb{K}} \vec{c} d\hat{\mu} \quad \text{and} \quad \zeta := \int_{\mathbb{K}} r d\hat{\mu},$$

then the point (θ, ζ) belongs to the hypograph of V.

Let $\Lambda_0 := (\lambda_{01}, \cdots, \lambda_{0q})$ with $\Lambda_0 \leq 0$, be a vector in $\in \mathbb{R}^q$ such that $-\Lambda_0$ is a superdifferential of V at θ_0 (see, for instance, [6, Chapter 1]). Then

$$\zeta + (\theta - \theta_0) \cdot \Lambda_0 \leq V(\theta) + (\theta - \theta_0) \cdot \Lambda_0 \leq V(\theta_0)$$

In particular, letting $\vec{c} := (c_1, \cdots, c_q)$,

$$\int_{\mathbb{K}} \Big(r + (\vec{c} - \theta_0) \cdot \Lambda_0\Big) d\hat{\mu} \leq V(\theta_0) \quad \forall \hat{\mu} \in \Gamma.$$

Thus

$$\sup_{\hat{\mu} \in \Gamma} \int_{\mathbb{K}} \Big(r + (\vec{c} - \theta_0) \cdot \Lambda_0\Big) d\hat{\mu} \leq V(\theta_0).$$

By the definition(5.2.1) of I and the fact that $\int_{\mathbb{K}} [(\vec{c} - \theta_0) \cdot \Lambda_0] d\hat{\mu} \geq 0$ for all $\hat{\mu} \in I$, we obtain

$$V(\theta_0) = \sup_{\hat{\mu} \in I} \int_{\mathbb{K}} r d\hat{\mu} \leq \sup_{\hat{\mu} \in \Gamma} \int_{\mathbb{K}} \Big(r + (\vec{c} - \theta_0) \cdot \Lambda_0\Big) d\hat{\mu} \leq V(\theta_0).$$

Hence

$$V(\theta_0) = \sup_{\hat{\mu} \in \Gamma} \int_{\mathbb{K}} \Big(r + (\vec{c} - \theta_0) \cdot \Lambda_0\Big) d\hat{\mu}. \tag{5.3.4}$$

The expression (5.3.4) tells us that $V(\theta_0)$ is the optimal value of an expected average reward MCP with reward-per-stage function in $B_W(\mathbb{K})$ given by

$$r + (\vec{c} - \theta_0) \cdot \Lambda_0,$$

and which satisfies the assumptions of Theorem 2.4.3. Hence, there exists a function $h \in B_W(\mathbf{X})$ such that $(V(\theta_0), h)$ is a solution of the corresponding AROE, i.e.,

$$h(x) + V(\theta_0) =$$

$$\sup_{a \in A(x)} \Big[r(x,a) + \sum_{i=1}^{q} \lambda_{0i} \Big(c_i(x,a) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q(dy|x,a)\Big] \tag{5.3.5}$$

for every $x \in \mathbf{X}$.

If in addition $c_i \in B_w(\mathbf{X})$ for $i = 1, \cdots, q$, from Lemma 4.2.2, we have that $h \in B_w(\mathbf{X})$. This completes the proof of statement (ii).

Proof of (iii). We know from part (ii) and from Theorem 2.4.3 that there exists a stationary deterministic policy $f \in \mathbb{F}$ such that

$$\begin{aligned} h(x) + V(\theta_0) &= \sup_{a \in A(x)} \Big[r(x,a) + \sum_{i=1}^{q} \lambda_{0i} \Big(c_i(x,a) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q(dy|x,a)\Big] \\ &= r_f(x) + \sum_{i=1}^{q} \lambda_{0i} \Big(c_{if}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_f(dy|x) \quad \forall x \in \mathbf{X}. \end{aligned} \tag{5.3.6}$$

On the other hand, note that

$$V(\theta_0) = \sup_{\hat{\mu} \in I} \int_{\mathbb{K}} r d\hat{\mu}.$$

Because the mapping $\hat{\mu} \mapsto \int_{\mathbb{K}} r d\hat{\mu}$ is u.s.c. on $\mathcal{P}_w(\mathbb{K})$ with respect to the w-weak topology (Lemma 5.2.5), the maximum is attained on the compact set I (Lemma 5.2.2). Thus, there exists $\varphi \in \Phi$ such that

$$\int_{\mathbb{K}} c_i d\hat{\mu}_\varphi \leq \theta_i \quad \forall i = 1, \cdots, q, \quad \text{and} \quad V(\theta_0) = \int_{\mathbb{K}} r d\hat{\mu}_\varphi. \tag{5.3.7}$$

We claim that φ is an optimal policy for the constrained problem (5.1.1)-(5.1.2) with $V(\theta_0)$ as its optimal value. To do this we will prove that for every feasible policy $\pi \in \Pi$ for the constrained problem (5.1.1)-(5.1.2) we have

$$V(\theta_0) \geq \bar{J}(\pi, x) \geq J(\pi, x). \tag{5.3.8}$$

From the AROE (5.3.5), we have

$$h(x) + V(\theta_0) - \sum_{i=1}^{q} \lambda_{0i} \Big(c_i(x,a) - \theta_i \Big) \geq r(x,a) + \int_{\mathbf{X}} h(y) Q(dy|x,a),$$

for every feasible action-pair $(x,a) \in \mathbb{K}$. Iteration of this inequality (as in the proof of Lemma 2.5.2) yields

$$E_x^\pi \sum_{k=0}^{n-1} h(x_k) + nV(\theta_0) - \sum_{i=1}^{q} \lambda_{0i} \Big(E_x^\pi \sum_{k=0}^{n-1} c_i(x_k, a_k) - n\theta_i \Big)$$

$$\geq E_x^\pi \sum_{k=0}^{n-1} r(x_k, a_k) + E_x^\pi \sum_{k=1}^{n} h(x_k),$$

or equivalently

$$\frac{1}{n} h(x) - \frac{1}{n} E_x^\pi h(x_n) + V(\theta_0) - \sum_{i=1}^{q} \lambda_{0i} \Big(\frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} c_i(x_k, a_k) - \theta_i \Big)$$

$$\geq \frac{1}{n} E_x^\pi \sum_{k=0}^{n-1} r(x_k, a_k).$$

By (2.1.2) in Remark 2.1.3, and the fact that $\lambda_{0i} \leq 0$ for $i = 1, \cdots, q$, taking the lim sup in both sides of the latter inequality, we have

$$V(\theta_0) - \sum_{i=1}^{q} \lambda_{0i}\Big(J_i(\pi, x) - \theta_i\Big) \geq \bar{J}(\pi, x),$$

that is

$$V(\theta_0) \geq \bar{J}(\pi, x) + \sum_{i=1}^{q} \lambda_{0i}\Big(J_i(\pi, x) - \theta_i\Big). \tag{5.3.9}$$

Since $\pi \in \Pi$ is a feasible policy for the constrained problem (5.1.1)-(5.1.2), then

$$\sum_{i=1}^{q} \lambda_{0i}\Big(J_i(\pi, x) - \theta_i\Big) \geq 0,$$

because $\lambda_{0i} \leq 0$. Thus (5.3.9) gives (5.3.8).

It follows that the randomized stationary policy φ satisfying (5.3.7) is an optimal policy for the constrained problem (5.1.1)-(5.1.2) with $V(\theta_0)$ as its optimal value.

On the other hand, observe that

$$\int_{\mathbb{K}} [(\vec{c} - \theta_0) \cdot \Lambda_0] d\hat{\mu}_\varphi = \int_{\mathbb{K}} \Big[\sum_{i=1}^{q} \lambda_{0i}\Big(c_i(x, a) - \theta_i\Big)\Big] \hat{\mu}_\varphi(d(x, a)) = 0. \tag{5.3.10}$$

By the definition of $h \in B_W(\mathbf{X})$ in (5.3.5), we get:

$$h(x) + V(\theta_0) \geq r_\varphi(x) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_{i\varphi}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_\varphi(dy|x), \tag{5.3.11}$$

for all x in $\mathbf{X}$. From (5.3.10),

$$\begin{aligned}
\int_{\mathbf{X}} \Big[h(x) + V(\theta_0) \;&-\; r_\varphi(x) - \sum_{i=1}^{q} \lambda_{0i}\Big(c_{i\varphi}(x) - \theta_i\Big) - \int_{\mathbf{X}} h(y) Q_\varphi(dy|x)\Big] \mu_\varphi(dx) \\
&= -\int_{\mathbb{K}} \Big[\sum_{i=1}^{q} \lambda_{0i}\Big(c_i(x, a) - \theta_i\Big)\Big] \hat{\mu}_\varphi(d(x, a)) \\
&= 0.
\end{aligned}$$

By (5.3.11) there exists a Borel subset $N \in \mathcal{B}(\mathbf{X})$ such that $\mu_\varphi(N) = 0$ and

$$h(x) + V(\theta_0) = r_\varphi(x) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_{i\varphi}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_\varphi(dy|x) \quad (5.3.12)$$

for all $x \in N^c = \mathbf{X} \setminus N$. We define the new policy

$$\varphi^*(\cdot|x) := 1_N(x)\delta_{f(x)}(\cdot) + 1_{N^c}(x)\varphi(\cdot|x) \quad \forall x \in \mathbf{X}, \quad (5.3.13)$$

where $\delta_{f(x)}(\cdot)$ is the Dirac measure concentrated at $f(x)$. It follows that $\varphi^*(\cdot|x) = \varphi(\cdot|x)$ $\mu_\varphi - a.e.$ Notice that

$$\varphi^*(\cdot|x) = \varphi(\cdot|x) \quad \text{and} \quad Q_{\varphi^*}(\cdot|x) = Q_\varphi(\cdot|x) \quad \forall x \in \mathbf{X} \setminus N. \quad (5.3.14)$$

Therefore

$$\mu_{\varphi^*} = \mu_\varphi. \quad (5.3.15)$$

In fact, if $u(\cdot)$ is a bounded measurable function on $\mathbb{K}$, then $u_{\varphi^*}(x) = u_\varphi(x)$ for all $x \in \mathbf{X} \setminus N$. Consequently, we have

$$\int_{\mathbb{K}} u \, d\hat{\mu}_{\varphi^*} = \int_{\mathbf{X}} u_{\varphi^*} \, d\mu_{\varphi^*} = \int_{\mathbf{X}} u_\varphi \, d\mu_\varphi = \int_{\mathbb{K}} u \, d\hat{\mu}_\varphi \quad \forall u \in B_1(\mathbb{K}).$$

Thus

$$\hat{\mu}_{\varphi^*} = \hat{\mu}_\varphi. \quad (5.3.16)$$

This equality of measures, gives us that $\varphi^* \in \Phi$ as defined in (5.3.13) is an optimal policy for the constrained problem (5.1.1)-(5.1.2). Moreover, by (5.3.10) and (5.3.16), φ^* satisfies

$$\int_{\mathbb{K}} [(\vec{c} - \theta_0) \cdot \Lambda_0] d\hat{\mu}_{\varphi^*} = 0.$$

Finally, we need to prove that

$$\begin{aligned} h(x) + V(\theta_0) &= \sup_{a \in A(x)} \Big[r(x,a) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_i(x,a) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q(dy|x,a) \Big] \\ &= r_{\varphi^*}(x) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_{i\varphi^*}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_{\varphi^*}(dy|x) \end{aligned} \quad (5.3.17)$$

for all $x \in \mathbf{X}$. Let us take $x \in \mathbf{X}$. If $x \in N$, we have $\varphi^*(\cdot|x) = \delta_{f(x)}(\cdot)$. By (5.3.6),

$$\begin{aligned} h(x) + V(\theta_0) &= r_f(x) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_{if}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_f(dy|x) \\ &= r_{\varphi^*}(x) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_{i\varphi^*}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_{\varphi^*}(dy|x). \end{aligned}$$

If $x \in \mathbf{X} \setminus N$, by (5.3.14) we have $\varphi^*(\cdot|x) = \varphi(\cdot|x)$ and, by (5.3.12),

$$\begin{aligned} h(x) + V(\theta_0) &= r_{\varphi}(x) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_{i\varphi}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_{\varphi}(dy|x) \\ &= r_{\varphi^*}(x) + \sum_{i=1}^{q} \lambda_{0i}\Big(c_{i\varphi^*}(x) - \theta_i\Big) + \int_{\mathbf{X}} h(y) Q_{\varphi^*}(dy|x). \end{aligned}$$

This complete the proof of (5.3.17). ■

5.4 A parametric family of AROEs

Theorem 5.3.1 shows that the constrained control problem (5.1.1)-(5.1.2) induces a non-constrained problem depending on a q-vector $\Lambda_0 \in \mathbb{R}^q$ such that $\Lambda_0 \leq 0$. However, Λ_0 is unknown and its value is obtained from the function V, which is precisely the function that we want to determine. The next theorem shows that the constrained problem (5.1.1)-(5.1.2) can be solved by means of a parametric family of AROEs.

Theorem 5.4.1 *Suppose that the same hypotheses as in Theorem 5.3.1 are satisfied, and consider the CP (5.1.1)-(5.1.2). For each q-vector $\Lambda \leq 0$, let $(\rho(\Lambda), h_\Lambda) \in \mathbb{R} \times B_W(\mathbf{X})$ be a solution to the AROE*

$$h_\Lambda(x) + \rho(\Lambda) = \sup_{a \in A(x)} \Big[r(x,a) + (\vec{c}(x,a) - \theta_0) \cdot \Lambda + \int_{\mathbf{X}} h_\Lambda(y) Q(dy|x,a) \Big] \quad (5.4.1)$$

for every $x \in \mathbf{X}$, where $\vec{c}(x,a) = (c_1(x,a), \cdots, c_q(x,a))$. Then

$$V(\theta_0) \equiv V^*(\theta_0) = \min_{\Lambda \leq 0} \rho(\Lambda). \quad (5.4.2)$$

Proof. From the proof of Theorem 5.3.1, there exists an optimal policy $\varphi^* \in \Phi$ for the constrained control problem (5.1.1)-(5.1.2) such that $V(\theta_0) = \int_{\mathbb{K}} r d\hat{\mu}_{\varphi^*}$, with $V(\theta_0)$ as in (5.2.5). Notice that for each q-vector $\Lambda \leq 0$

$$V(\theta_0) = \int_{\mathbb{K}} r d\hat{\mu}_{\varphi^*} \leq \int_{\mathbb{K}} [r + (\vec{c} - \theta_0) \cdot \Lambda] d\hat{\mu}_{\varphi^*}.$$

By (5.4.1), for all $x \in \mathbf{X}$ we have

$$h_\Lambda(x) + \rho(\Lambda) \geq r_{\varphi^*}(x) + (\vec{c}_{\varphi^*}(x) - \theta_0) \cdot \Lambda + \int_{\mathbf{X}} h_\Lambda(y) Q_{\varphi^*}(dy|x)$$

Integrating both sides of this inequality with respect to the i.p.m. μ_{φ^*}

$$\begin{aligned} \rho(\Lambda) &\geq \int_{\mathbf{X}} [r_{\varphi^*}(x) + (\vec{c}_{\varphi^*}(x) - \theta_0) \cdot \Lambda] \mu_{\varphi^*}(dx) \\ &= \int_{\mathbb{K}} [r + (\vec{c} - \theta_0) \cdot \Lambda] d\hat{\mu}_{\varphi^*} \geq V(\theta_0) \quad \forall \Lambda \leq 0. \end{aligned}$$

Therefore

$$\inf_{\Lambda \leq 0} \rho(\Lambda) \geq V(\theta_0). \tag{5.4.3}$$

By the AROE (5.4.1) again, with $\Lambda = \Lambda_0$ as in Theorem 5.3.1, and by Theorem 2.4.3(i), there exists $f \in \mathbb{F}$ such that

$$\begin{aligned} h_{\Lambda_0}(x) + \rho(\Lambda_0) &= \sup_{a \in A(x)} \Big[r(x,a) + (\vec{c}(x,a) - \theta_0) \cdot \Lambda_0 + \int_{\mathbf{X}} h_{\Lambda_0}(y) Q(dy|x,a) \Big] \\ &= r_f(x) + (\vec{c}_f(x) - \theta_0) \cdot \Lambda_0 + \int_{\mathbf{X}} h_{\Lambda_0}(y) Q_f(dy|x) \end{aligned}$$

for all $x \in \mathbf{X}$. Let us consider $h \in B_W(\mathbf{X})$ as in the optimality equation in Theorem 5.3.1(ii), i.e.,

$$h(x) + V^*(\theta_0) = \sup_{a \in A(x)} \Big[r(x,a) + (\vec{c}(x,a) - \theta_0) \cdot \Lambda_0 + \int_{\mathbf{X}} h(y) Q(dy|x,a) \Big]$$

for all $x \in \mathbf{X}$. This implies

$$h(x) + V^*(\theta_0) \geq r_f(x) + (\vec{c}_f(x) - \theta_0) \cdot \Lambda_0 + \int_{\mathbf{X}} h(y) Q_f(dy|x),$$

and so

$$(h(x) - h_{\Lambda_0}(x)) + (V^*(\theta_0) - \rho(\Lambda_0)) \geq \int_{\mathbf{X}} [h(y) - h_{\Lambda_0}(y)] Q_f(dy|x)$$

for all $x \in \mathbf{X}$. Integrating both sides by the i.p.m. μ_f, we have

$$V(\theta_0) = V^*(\theta_0) \geq \rho(\Lambda_0)$$

Comparing with inequality (5.4.3), we obtain the result desired (5.4.2) ■

5.5 Existence of pathwise constrained optimal policies

MCPs with pathwise constraints. With the notation above, and recalling Definition 3.1.1, we want to maximize (with probability one) the pathwise average reward

$$S(\pi, x) = \liminf_{n\to\infty} \frac{1}{n} S_n(\pi, x)$$

over the set of all policies $\pi \in \Pi$ satisfying that

$$S_i(\pi, x) = \limsup_{n\to\infty} \frac{1}{n} S_{i,n}(\pi, x) \leq \theta_i \quad \text{for } i = 1, \cdots, q, \quad P_x^\pi - a.s.$$

with $S_{i,n}(\pi, x) := \sum_{k=0}^{n-1} E_x^\pi[c_i(x_k, a_k)|h_k]$ for $n = 1, 2, \cdots$, $i = 1, \cdots, q$. In short, we will write

$$\text{maximize } S(\pi, x) \tag{5.5.1}$$

$$\text{subject to:} \quad \pi \in \Pi \quad \text{and} \quad S_i(\pi, x) \leq \theta_i \quad \forall x \in \mathbf{X}, i = 1, \cdots, q. \tag{5.5.2}$$

Definition 5.5.1 *We say that a policy $\varphi^* \in \Phi$ such that $g_i(\varphi^*) := \mu_{\varphi^*}(c_{i\varphi^*}) \leq \theta_i$ for $i = 1, \cdots, q$, is optimal for the pathwise CP (5.5.1)-(5.5.2) if for each $x \in \mathbf{X}$ and every $\pi \in \Pi$ such that $S_i(\pi, x) \leq \theta_i$ for $i = 1, \cdots, q$, $P_x^\pi - a.s.$, we have*

$$S(\pi, x) \leq g(\varphi^*) := \mu_{\varphi^*}(r_{\varphi^*}) \quad P_x^\pi - a.s.$$

By Theorem 5.3.1, we know that there exists an optimal (randomized) stationary policy for the "expected constrained problem" (5.1.1)-(5.1.2), which will be denoted by $\varphi^* \in \Phi$. Thus, we have

$$g(\varphi^*) = \int_{\mathbb{K}} r d\hat{\mu}_{\varphi^*} = V^*(\theta_0) \quad \text{and} \quad g_i(\varphi^*) = \int_{\mathbb{K}} c_i d\hat{\mu}_{\varphi^*} \leq \theta_i. \tag{5.5.3}$$

The following theorem proves that this optimal policy φ^* is also optimal for (5.5.1)-(5.5.2).

Theorem 5.5.2 *Suppose that Assumptions 2.1.1, 2.1.2, 2.4.2, 3.2.1 and 5.2.1 hold. Moreover, suppose that the cost functions c_i belong to $B_w(\mathbf{X})$, where $w(x) = \sqrt{W(x)}$ for all $x \in \mathbf{X}$. If $\varphi^* \in \Phi$ is an optimal policy for the (expected) CP (5.1.1)-(5.1.2), then φ^* is also optimal for the pathwise CP (5.5.1)-(5.5.2).*

Proof. Consider φ^* as in (5.5.3), that is, an optimal policy for the constrained problem (5.1.1)-(5.1.2). Let Λ_0 be as in Theorem 5.3.1 and fix an initial state $x \in \mathbf{X}$ and an arbitrary randomized policy $\pi \in \Pi$. Define

$$\widetilde{r}(x,a) := r(x,a) + (\vec{c}(x,a) - \theta_0) \cdot \Lambda_0 \quad \forall (x,a) \in \mathbb{K}.$$

From Theorem 5.3.1(ii), there exists $h \in B_w(\mathbf{X})$ such that

$$V^*(\theta_0) \geq \widetilde{r}(x,a) + \int_{\mathbf{X}} h(y) Q(dy|x,a) - h(x) \quad \forall (x,a) \in \mathbb{K}.$$

Therefore, for every $k = 0, 1, \cdots$,

$$\begin{aligned} V^*(\theta_0) &\geq \widetilde{r}(x_k,a_k) + \int_{\mathbf{X}} h(y) Q(dy|x_k,a_k) - h(x_k) \\ &= \widetilde{r}(x_k,a_k) + E_x^\pi[h(x_{k+1})|h_k,a_k] - h(x_k) \end{aligned}$$

with h_k the admissible history up to time k. Taking conditional expectations with respect to h_k, we have

$$\begin{aligned} V^*(\theta_0) &\geq E_x^\pi[\widetilde{r}(x_k,a_k)|h_k] + E_x^\pi[h(x_{k+1})|h_k] - h(x_k) \\ &= E_x^\pi[\widetilde{r}(x_k,a_k)|h_k] + L^\pi h(x_k) \quad P_x^\pi - a.s. \end{aligned}$$

with $L^\pi h(x_k)$ as in Lemma 3.2.6. Hence, for each $n = 1, 2, \cdots$,

$$V^*(\theta_0) \geq \frac{1}{n} S_n(\pi, x) + \sum_{i=1}^{q} \lambda_{0i} \Big(\frac{1}{n} S_{i,n}(\pi, x) - \theta_i \Big) + \frac{1}{n} \sum_{k=0}^{n-1} L^\pi h(x_k) \quad P_x^\pi - a.s.$$

Taking the limit as $n \to \infty$, and taking into account the limit given by Lemma 3.2.6, together with the fact that $\lambda_{0i} \leq 0$,

$$\begin{aligned} V^*(\theta_0) - \sum_{i=1}^{q} \lambda_{0i}\Big(S_i(\pi,x) - \theta_i\Big) &\geq \limsup_{n\to\infty} \Big[\frac{1}{n} S_n(\pi,x) + \frac{1}{n}\sum_{k=0}^{n-1} L^\pi h(x_k)\Big] \\ &\geq \liminf_{n\to\infty} \frac{1}{n} S_n(\pi,x) + \lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} L^\pi h(x_k) \\ &= S(\pi,x) \quad P_x^\pi - a.s. \end{aligned}$$

That is

$$V^*(\theta_0) \geq S(\pi,x) + \sum_{i=1}^{q} \lambda_{0i}\Big(S_i(\pi,x) - \theta_i\Big) \quad P_x^\pi - a.s. \tag{5.5.4}$$

for all $x \in \mathbf{X}$ and each $\pi \in \Pi$. If π verifies $S_i(\pi,x) \leq \theta_i \quad P_x^\pi - a.s.$, and recalling that $\Lambda_0 \leq 0$, then $S(\pi,x) \leq V^*(\theta_0)$ $P_x^\pi - a.s.$, which together with (5.5.3) completes the proof. ■

Concluding remarks. In this chapter we have studied pathwise average reward discrete-time MCM with pathwise constraints on Borel spaces. Under suitable assumptions we have shown the existence of optimal policies. To this end, we give conditions for the existence of optimal policies for the problem with expected constraints. In addition, we have shown that the expected problem can be solve by means of a parametric family of AROEs. Furthermore, the examples in Chapter 6 show that our assumptions are satisfied with no special degree of difficulty.

An open question is whether our results here can be extended for more general MCMs (no necessarily under our fixed point approach). A second open question is the minimization of variance for the pathwise constrained problem, that is, we would like to prove the existence, within the class of stationary optimal policies for the pathwise CP, of one with minimal limiting *average variance*. Another question is to find approximation schemes where the optimum value as well as the optimal policy can be approximated for the problem with pathwise constraints.

Chapter 6

Examples

6.1 Introduction

In this chapter we illustrate the results in Theorems 5.3.1, 5.4.1 and 5.5.2.

We consider an $\mathbf{X}$-valued controlled processes $\{x_t\}$ of the form

$$x_{t+1} = F(x_t, a_t, z_t), \quad t = 0, 1, \cdots, \tag{6.1.1}$$

and we always suppose the following:

Assumption 6.1.1

(a) The disturbance sequence $\{z_t\}$ in (6.1.1) consists of independent and identically distributed (i.i.d.) random variables with values in a Borel space Z, and $\{z_t\}$ is independent of the initial state x_0. The common distribution of the z_t is denoted by G.

(b) $F : \mathbb{K} \times Z \to \mathbf{X}$ is a given measurable function, where $\mathbb{K} \subset \mathbf{X} \times A$ is the set defined in (1.3.1).

Let π be an arbitrary control policy. By Assumption 6.1.1(a), the variables (x_t, a_t) and z_t are independent for each $t = 0, 1, \cdots$. Then the transition law Q is given by

$$\begin{aligned} Q(B|x, a) &= \text{Prob}(x_{t+1} \in B | x_t = x, a_t = a) \\ &= \int_Z 1_B[F(x, a, z)] G(dz) \end{aligned} \tag{6.1.2}$$

for every $B \in \mathcal{B}(\mathbf{X}), (x,a) \in \mathbb{K}$, and $t = 0, 1, \cdots$. Moreover, for every bounded measurable function u on $\mathbf{X}$, we have

$$\begin{aligned} u'(x,a) &:= \int_{\mathbf{X}} u(dy)Q(dy|x,a) = E[u(x_{t+1})|x_t = x, a_t = a] \\ &= \int_Z u[F(x,a,z)]G(dz). \end{aligned} \tag{6.1.3}$$

6.2 A LQ system

In this section we present a Linear-Quadratic system that satisfies all the hypotheses of Theorems 5.3.1, 5.4.1 and 5.5.2, that is, Assumptions 2.1.1, 2.1.2, 2.4.2, 3.2.1 and 5.2.1.

Consider the linear system

$$x_{t+1} = k_1 x_t + k_2 a_t + z_t, \quad t = 0, 1, \cdots, \tag{6.2.1}$$

with state space $\mathbf{X} := \mathbb{R}$ and positive coefficients k_1, k_2. The control set is $A := \mathbb{R}$, and the set of admisible controls in each state x is the interval

$$A(x) := [-k_1|x|/k_2, k_1|x|/k_2]. \tag{6.2.2}$$

The disturbance z_t in (6.2.1) consists of i.i.d. random variables with values in $Z := \mathbb{R}$, and with zero mean and finite variance, that is,

$$E(z_t) = 0 \quad \text{and} \quad \sigma^2 := E(z_t^2) < \infty. \tag{6.2.3}$$

To complete the description of our constrained control model we introduce the quadratic reward-per-stage function

$$r(x,a) := B - (r_1 x^2 + r_2 a^2) \quad \forall (x,a) \in \mathbb{K}, \tag{6.2.4}$$

with positive coefficients B, r_1, and r_2, and the cost-per-stage function

$$c(x,a) := c_1 x^2 + c_2 a^2 \quad \forall (x,a) \in \mathbb{K}, \tag{6.2.5}$$

with positive coefficients c_1, c_2. We also define

$$W(x) := \exp[\gamma|x|] \quad \text{for all} \quad x \in \mathbf{X}, \tag{6.2.6}$$

with $\gamma \geq 2$. Observe that $W \geq 1$ and Assumptions 2.1.1 and 3.2.1 hold. Moreover, we can see that $c^2 \in B_W(\mathbf{X})$. Note that $w := \sqrt{W}$ is continuous on $\mathbb{K}$ and that it is a moment function on $\mathbb{K}$ as in Assumption 5.2.1-(d). Moreover, let $\hat{s} > 0$ be such that

$$\gamma \hat{s} < \log(\gamma/2 + 1)$$

which implies

$$\lambda := \frac{2}{\gamma}(\exp[\gamma \hat{s}] - 1) < 1. \tag{6.2.7}$$

Throughout the rest of this chapter, we suppose the following Assumptions taken from [19, Section 5]:

Assumption 6.2.1 $0 < k_1 < 1/2$.

Assumption 6.2.2 *The i.i.d. disturbances z_t have a common density g, which is a continuous bounded function supported on the interval $S := [-\hat{s}, \hat{s}]$. Moreover, there exists a positive number ε such that $g(s) \geq \varepsilon$ for all $s \in S$.*

Let $S_0 := [0, \hat{s}]$, and let Υ be the Lebesgue measure on $\mathbf{X} = \mathbb{R}$. We define

$$l(x, a) := 1_{S_0}(x) \ \forall (x, a) \in \mathbb{K}, \text{ and } \nu(B) := \varepsilon \Upsilon(B \cap S_0) \ \forall B \in \mathcal{B}(\mathbf{X}). \tag{6.2.8}$$

Then, from [19, Propositions 23 and 24] we have the following.

Proposition 6.2.3 *Under the Assumptions 6.2.1 and 6.2.2, the LQ system (6.2.1)-(6.2.5) satisfies the Assumptions 2.1.2, 2.4.2 and 5.2.1.*

Remark 6.2.4 *Assumptions 2.1.2 and 2.4.2 are used in [19] to get conditions for bias optimality and strong 0-discount optimality to be equivalent.*

Proposition 6.2.5 *Suppose that Assumptions 6.2.1 and 6.2.2 hold. Then the LQ system (6.2.1)-(6.2.5) has a constrained optimal policy which is also a pathwise constrained optimal policy. Moreover, for each $\Lambda \leq 0$ let $(\rho(\Lambda), h_\Lambda) \in \mathbb{R} \times B_W(\mathbf{X})$ be a solution to the AROE*

$$h_\Lambda(x) + \rho(\Lambda) = \sup_{a \in A(x)} \Big[r_\Lambda(x, a) + \int_{\mathbf{X}} h_\Lambda(y) Q(dy|x, a) \Big], \tag{6.2.9}$$

with $r_\Lambda(x, a) := r_1(\Lambda)x^2 + r_2(\Lambda)a^2 + b$, where $r_i(\Lambda) := \Lambda \cdot c_i - r_i < 0$, $i = 1, 2$, and $b := B - \Lambda \cdot \theta_0$, then the constrained optimal value $V^*(\theta_0)$ satisfies

$$V^*(\theta_0) = \min_{\Lambda \leq 0} \rho(\Lambda) \tag{6.2.10}$$

Proof. From Proposition 6.2.3, the assumptions in Theorems 5.3.1, 5.4.1 and 5.5.2 are satisfied. Hence, the stated result follows from these theorems. ∎

Example 6.2.6 Now we analyse a particular case in which the reward-per-stage function (6.2.4) and the cost-per-stage function (6.2.5) satisfy $r_1 = r_2$ and $c_1 = c_2$, respectively, and $k_2 = 1$ in (6.2.1). For this case, we will find the optimal value and the optimal policy for the LQ model above, with expected and pathwise constraints.

Note that

$$r_1(\Lambda) = r_2(\Lambda) \quad \forall \Lambda \leq 0. \tag{6.2.11}$$

On the other hand, in [19, Section 5], it is proved, under the Assumptions 6.2.1 and 6.2.2, that $\rho(\Lambda)$ in the AROE (6.2.9) has the form

$$\rho(\Lambda) = b - v_0\sigma^2, \tag{6.2.12}$$

with σ as in (6.2.3), and v_0 is the unique positive solution to the quadratic (so-called Riccati) equation

$$k_2^2 v_0^2 + [k_2^2 r_1(\Lambda) + k_1^2 r_2(\Lambda) - r_2(\Lambda)]v_0 - r_1(\Lambda)r_2(\Lambda) = 0. \tag{6.2.13}$$

Note that v_0 dependes on Λ. Moreover, if we define

$$f_\Lambda(x) := -\widehat{f}_0 x \quad \forall x \in \mathbf{X}, \quad \text{with} \quad \widehat{f}_0 := (k_2^2 v_0 - r_2(\Lambda))^{-1} k_1 k_2 v_0. \tag{6.2.14}$$

and

$$h_\Lambda(x) := -v_0 x^2 \tag{6.2.15}$$

then, by a direct calculation we can show that $(h_\Lambda, f_\Lambda, \rho(\Lambda))$ is a canonical triplet that satisfies the AROE (6.2.9).

Since $r_2(\Lambda) < 0$, we have $|f_\Lambda(x)| \leq k_1/k_2|x|$, and so, $f_\Lambda(x) \in A(x)$ for all $x \in \mathbf{X}$, that is, f_Λ is in $\mathbb{F}$.

By (6.2.11), the positive solution of (6.2.13) is

$$v_0 = -k r_1(\Lambda) \quad \text{with} \quad k = \frac{k_1^2 + \sqrt{k_1^4 + 4}}{2}. \tag{6.2.16}$$

Inserting this values in (6.2.12) and using the definition of the constant b, we obtain the explicit form of $\rho(\Lambda)$

$$\rho(\Lambda) = B - (\sigma^2 k) \cdot r_1 + [(\sigma^2 k) \cdot c_1 - \theta_0]\Lambda \tag{6.2.17}$$

which is the equation of a straigth line with slope $(\sigma^2 k)\cdot c_1 - \theta_0$. Because we need to choose θ_0 satisfying the relation (6.2.10), then we have to impose the following assumption:

$$(\sigma^2 k)\cdot c_1 < \theta_0. \tag{6.2.18}$$

Under this condition, we have that

$$\begin{aligned} V^*(\theta_0) &= \min_{\Lambda\leq 0}\rho(\Lambda) \\ &= \min_{\Lambda\leq 0}\Big(B-(\sigma^2 k)\cdot r_1 + [(\sigma^2 k)\cdot c_1 - \theta_0]\Lambda\Big) \\ &= B-(\sigma^2 k)\cdot r_1 = \rho(0). \end{aligned} \tag{6.2.19}$$

Thus, the minimun is attained in $\Lambda = 0$ and $V^*(\theta_0) = \rho(0)$. Furthermore, inserting $\Lambda = 0$ in (6.2.14) and (6.2.15), we obtain

$$f_0(x) = -\widehat{f}_0 x \quad \text{with} \quad \widehat{f}_0 := \frac{kk_1}{1+k}, \tag{6.2.20}$$

for all $x \in \mathbf{X}$, and

$$h_0(x) = -kr_1x^2 \quad \forall x \in \mathbf{X}. \tag{6.2.21}$$

Since $(h_0, f_0, V^*(\theta_0))$ is a canonical triplet, then the following average reward optimality equation is satisfied:

$$\begin{aligned} V^*(\theta_0) + h_0(x) &= \sup_{a\in A(x)}\left[r(x,a) + \int_{\mathbf{X}} h_0(y)Q(dy|x,a)\right] \\ &= r_{f_0}(x) + \int_{\mathbf{X}} h(y)Q_{f_0}(dy|x) \quad \forall x \in \mathbf{X}, \end{aligned} \tag{6.2.22}$$

which is the equation (5.3.2) in Theorem 5.3.1. Moreover, we assert that the deterministic policy f_0 defined in (6.2.20) is an optimal policy for the constrained LQ system. To do this, we present the following result which is a slight variation of Lemma 6.5 in [13].

Lemma 6.2.7 *Let $f \in \mathbb{F}$ be a deterministic policy given by $f(x) := -\widehat{f}x$ for all $x \in \mathbf{X}$, and let $\widehat{k} := k_1 - k_2\widehat{f}$, where k_1, k_2 are the coefficients in (6.2.1). Here $\widehat{f}$ is a constant. Suppose that $|\widehat{k}| < 1$. Then, for all $x \in \mathbf{X}$*

$$J(f,x) = \liminf_{n\to\infty}\frac{1}{n}E_x^f\sum_{k=0}^{n-1} r_f(x_k) = B - (r_1 + r_2\widehat{f}^2)\sigma^2/(1-\widehat{k}^2), \tag{6.2.23}$$

and

$$J_1(f,x) = \limsup_{n\to\infty} \frac{1}{n} E_x^f \sum_{k=0}^{n-1} c_f(x_k) = (c_1 + c_2\widehat{f}^2)\sigma^2/(1-\widehat{k}^2). \qquad (6.2.24)$$

with r and c as defined in (6.2.4) and (6.2.5), respectively.

Proof. Replacing a_t in (6.2.1) with $a_t := f(x_t) = -\widehat{f}x_t$, we obtain

$$x_t = (k_1 - k_2\widehat{f})x_{t-1} + z_{t-1} = \widehat{k}x_{t-1} + z_{t-1} \quad \forall t = 1, 2, \cdots.$$

By an induction procedure, for all $t = 1, 2, \cdots$,

$$x_t = \widehat{k}^t x_0 + \sum_{j=0}^{t-1} \widehat{k}^j z_{t-1-j}.$$

From this relation, we obtain

$$E_x^f(x_t^2) = \widehat{k}^2 + (\sigma^2(1-\widehat{k}^{2t}))/(1-\widehat{k}^2).$$

This yields that

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E_x^f(x_t^2) = \liminf_{n\to\infty} \frac{1}{n} \sum_{t=0}^{n-1} E_x^f(x_t^2) = \sigma^2/(1-\widehat{k}^2). \qquad (6.2.25)$$

Since $a = f(x) = -\widehat{f}x$, we obtain

$$r_f(x) = B - (r_1 + r_2\widehat{f}^2)x^2 \quad \text{and} \quad c_f(x) = (c_1 + c_2\widehat{f}^2)x^2 \qquad (6.2.26)$$

for all $x \in \mathbf{X}$. Finally, inserting (6.2.25) in (6.2.26) we obtain (6.2.23) and (6.2.24). ∎

If f_0 is as in (6.2.20), and recalling that $r_1 = r_2$ and $c_1 = c_2$, $k_2 = 1$, we have that $|\widehat{k}| = k_1/(1+k) < 1$, with $\widehat{k} := k_1 - \widehat{f_0}$ and k as in (6.2.16). Hence, by Lemma 6.2.7, a direct calculation yields that, for all $x \in \mathbf{X}$,

$$J(f_0,x) = B - (\sigma^2 k)r_1 \quad \text{and} \quad J_1(f_0,x) = (\sigma^2 k)c_1.$$

Finally, from (6.2.18) and (6.2.19), we have

$$J(f_0,x) = V^*(\theta_0) \quad \text{and} \quad J_1(f_0,x) < \theta_0,$$

that is, f_0 is a constrained optimal policy, which is also pathwise constrained optimal for the LQ system (6.2.1)-(6.2.5).

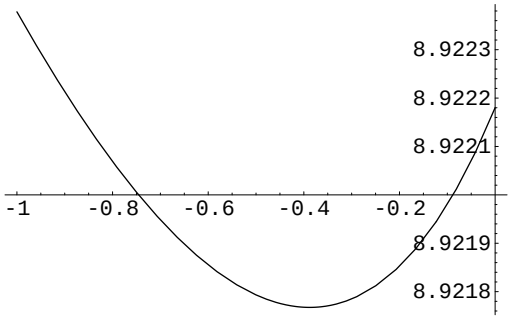

Figure 6.1: *Graph of* $\rho(\Lambda)$ *as a function on* $\Lambda \leq 0$.

We will illustrate our LQ system (6.2.1)-(6.2.5) with the following numerical special case. Suppose that the reward-per-stage function (6.2.4) and the cost-per-stage function (6.2.5) satisfy $r_1 = 1, r_2 = 2$, $B = 10$, and $c_1 = c_2 = 1$, respectively. Moreover, assume that $k_1 = 1/3$, $k_2 = 1$ in (6.2.1) and $\theta_0 := 191/180$.

In this particular case, solving the Riccati equation (6.2.13), and inserting the corresponding value in (6.2.12), we obtain

$$\rho(\Lambda) = \left(187 - 18.1\Lambda - \sqrt{325\Lambda^2 - 958\Lambda + 697}\right)/18 \quad \forall \Lambda \leq 0. \tag{6.2.27}$$

As can be seen from the graph of $\rho(\Lambda)$ for $\Lambda \leq 0$, the function $\rho(\Lambda)$ has a minimum (see Fig. 6.1). By elementary calculus, we get that $\rho(\Lambda)$ has a unique minimun in

$$\Lambda_{min} = -0.38767819\cdots,$$

with minimun value

$$\rho(\Lambda_{min}) = 8.921767464\cdots.$$

From Proposition 6.2.5, $\rho(\Lambda_{min})$ is the optimal value for the constrained problem, that is

$$V^*(\theta_0) = \rho(\Lambda_{min}) = 8.921767464\cdots, \quad \text{with} \quad \theta_0 = 191/180.$$

In addition

$$v_0 \equiv v_0(\Lambda_{min}) = 1.48960217\cdots.$$

By (6.2.14) and (6.2.15), we have that

$$f_{\Lambda_{min}}(x) = -\widehat{f}_0 x \quad \forall x \in \mathbb{R}, \quad \text{with} \quad \widehat{f}_0 = 0.12806246\cdots.$$

and

$$h(x) \equiv h_{\Lambda_{min}}(x) = -v_0 x^2.$$

By a straigthforward calculation, we can check that $(V^*(\theta_0), f_{\Lambda_{min}}, h)$ is a canonical triplet that satisfies the AROE (5.3.1) in Theorem 5.3.1. Moreover, we assert that $f_{\Lambda_{min}}$ is a constrained optimal policy, and therefore by Proposition 6.2.5, it is also a pathwise constrained optimal policy. Indeed, by Lemma 6.2.7, we have

$$J(f_{\Lambda_{min}}, x) = 8.9217674\cdots,$$

which coincides with the optimal value $V^*(\theta_0)$, with $\theta_0 = 191/180$. Finally, by a similar calculation, we obtain

$$J_1(f_{\Lambda_{min}}, x) = 1.061111\cdots = 191/180.$$

Hence

$$J(f_{\Lambda_{min}}, x) = V^*(\theta_0) \quad \text{and} \quad J_1(f_{\Lambda_{min}}, x) = \theta_0,$$

and so the constrained problem is solved.

6.3 An inventory system

The inventory-production system in this section has been studied by Vega-Amaya [28, 29] and by Hernández-Lerma and Vega-Amaya in [17].

We consider an inventory-production system in which the stock level x_t evolves in $\mathbf{X} := [0, \infty)$ according to the equation

$$x_{t+1} = \max(x_t + a_t - z_t, 0), \quad t = 0, 1, \cdots, \tag{6.3.1}$$

for some given initial stock level x_0. Here a_t is the amount of product ordered (and immediately supplied) at the beginning of each period t, whereas z_t denotes the product's demand during that period. The production variables a_t are supposed to take values in the interval $A := [0, \Omega]$, for some given constant $\Omega > 0$ irrespective of the stock level, that is,

$$A(x) = A \quad \forall x \in \mathbf{X}. \tag{6.3.2}$$

Additionally, we suppose that the demand process $\{z_t\}$ satisfies Assumption 6.1.1 with $Z := [0, \infty)$, and that the demand distribution G satisfies the following assumption:

Assumption 6.3.1 **(a)** *G has a continuous bounded density g;*

(b) *G has a finite mean value* $\bar{z}$*, i.e.,* $\bar{z} := E(z_0) = \int_0^\infty zG(dz) < \infty$*, where E denotes the expectation with respect to G.*

(c) $\Omega < \bar{z}$*.*

To complete our control model, we introduce a reward-per-stage function r that represents a net reward of the form

sales revenue–(production cost+maintenance cost)

given by

$$r(x,a) := s \cdot E\min(x+a, z_0) - [p \cdot a + m \cdot (x+a)] \quad \forall (x,a) \in \mathbb{K}, \quad (6.3.3)$$

where p, m and s are positive constants. The unit production p and the unit maintenance cost m do not exceed the unit sale price, i.e.,

$$p, m \leq s, \quad (6.3.4)$$

and the cost-per-stage function c of the form

production cost+maintenance cost

given by

$$c(x,a) := p \cdot a + m \cdot (x+a) \quad \forall (x,a) \in \mathbb{K}. \quad (6.3.5)$$

We can verify that

$$E\min(x+a, z_0) = (x+a)[1 - G(x+a)] + \int_0^{x+a} zG(dz). \quad (6.3.6)$$

Thus, the functions r and c in (6.3.3) and (6.3.5) are continuous on $\mathbb{K} = \mathbf{X} \times A$.

On the other hand, consider the moment generating function Ψ of the random variable $\Omega - z_0$, $\Psi(r) := E\exp[r(\Omega - z_0)]$, for $r \geq 0$. Note that $\Psi(0) = 1$ and by Assumption 6.3.1(c), $\Psi'(0) = E(\Omega - z_0) = \Omega - \bar{z} < 0$. Hence, there is a positive number $\widehat{r}$ such that

$$\lambda := \Psi(\widehat{r}) < 1. \quad (6.3.7)$$

We define the weigth function

$$W(x) := \exp[\widehat{r}(x + 2\bar{z})], \quad \forall x \in \mathbf{X}. \tag{6.3.8}$$

Then $W \geq 1$ and by a straigthforward calculation using Assumption 6.3.1(c), (6.3.4) and (6.3.6), we can see that there exist constants K and ω such that

$$|r(x,a)| \leq KW(x) \quad \text{and} \quad |c(x,a)| \leq \omega W(x) \quad \forall (x,a) \in \mathbb{K}.$$

Moreover, we can check that r^2, c^2 belong to $B_W(\mathbf{X})$. Hence, Assumptions 2.1.1 and 3.2.1 are satisfied.

We shall now proceed to verify the Assumptions 2.1.2, 2.4.2 and 5.2.1. To do this, note that from (6.3.1) and (6.1.3) we obtain

$$\begin{aligned} \bar{u}(x,a) &:= \int_{\mathbf{X}} u(y)Q(dy|x,a) \\ &= u(0)[1 - G(x+a)] + \int_0^{x+a} u(x+a-z)g(z)dz \\ &= u(0)[1 - G(x+a)] + \int_0^{x+a} u(z)g(x+a-z)dz. \end{aligned} \tag{6.3.9}$$

for every bounded measurable function u on $\mathbf{X}$.

We also define for every $(x,a) \in \mathbb{K}$ and $B \in \mathcal{B}(\mathbf{X})$

$$l(x,a) := 1 - G(x+a) \quad , \quad \text{and} \quad \nu(B) := \delta_0(B), \tag{6.3.10}$$

with δ_0 the Dirac measure at $x = 0$.

Proposition 6.3.2 *With the notation above and under Assumption 6.3.1, we have that the inventory-production system (6.3.1)-(6.3.5) satisfies Assumptions 2.1.2, 2.4.2 and 5.2.1.*

Proof. By Assumption 6.3.1, (6.3.2), (6.3.3), (6.3.5), (6.3.6), (6.3.8) and (6.3.10), then Assumptions 2.1.2(a), 2.4.2(a), 2.4.2(b), 2.4.2(e), and 5.2.1(b)-(e) obviously hold.

From (6.3.9) and (6.3.10), it follows that

$$Q(B|x,a) \geq \delta_0(B)l(x,a) \quad \forall B \in \mathcal{B}(\mathbf{X}), (x,a) \in \mathbb{K},$$

that is, Assumption 2.1.2(b) holds.

From (6.3.9) again, with $u = W$,

$$\begin{aligned}\int_{\mathbf{X}} W(y)Q(dy|x,a) &= W(0)[1-G(x+a)] + \int_0^{x+a} W(x+a-z)g(z)dz \\ &= \nu(W)l(x,a) + W(x)\int_0^{x+a} \exp[\widehat{r}(a-z)]G(dz),\end{aligned} \tag{6.3.11}$$

so that, since $\widehat{r}(a-x) \leq \widehat{r}(\Omega - x)$ for all $a \in A$ and by (6.3.7), we get

$$\int_{\mathbf{X}} W(y)Q(dy|x,a) \leq \nu(W)l(x,a) + \lambda W(x) \quad \forall (x,a) \in \mathbb{K}.$$

This gives Assumption 2.1.2(c).

To prove Assumption 2.1.2(d), with $l(x,a)$ and ν as defined in (6.3.10), note that for each $\varphi \in \Phi_s$

$$l_\varphi(x) \geq 1 - G(x+\Omega).$$

Here, integrating both sides with respect to $\nu = \delta_0$:

$$\nu(l_\varphi) \geq 1 - G(\Omega) \quad \forall \varphi \in \Phi_s. \tag{6.3.12}$$

We only need to show that $G(\Omega) < 1$. If $G(\Omega) = 1$ we obtain

$$\begin{aligned}\bar{z} &= \int_0^\infty zG(dz) \\ &= \int_0^\Omega zG(dz) + \int_\Omega^\infty zG(dz) \\ &\leq \Omega G(\Omega) = \Omega\end{aligned}$$

which contradicts Assumption 6.3.1(c).

Finally, Assumptions 2.4.2(c) and 5.2.1(a) follow from (6.3.9), and Assumption 2.4.2(d) follows from (6.3.11). ■

Proposition 6.3.3 *Suppose that Assumption 6.3.1 holds. Then the inventory-production system (6.3.1)-(6.3.5) has a constrained optimal policy which is*

also a pathwise constrained optimal policy. Moreover, for each $\Lambda \leq 0$ *let* $(\rho(\Lambda), h_\Lambda) \in \mathbb{R} \times B_W(\mathbf{X})$ *be a solution to the AROE*

$$h_\Lambda(x) + \rho(\Lambda) = \sup_{a \in A(x)} \Big[r_\Lambda(x,a) + \int_{\mathbf{X}} h_\Lambda(y) Q(dy|x,a) \Big], \tag{6.3.13}$$

with $r_\Lambda(x,a) := r(x,a) + (c(x,a) - \theta_0) \cdot \Lambda$, *then the constrained optimal value* $V^*(\theta_0)$ *satisfies*

$$V^*(\theta_0) = \min_{\Lambda \leq 0} \rho(\Lambda) \tag{6.3.14}$$

Proof. From Proposition 6.3.2, the assumptions in Theorems 5.3.1, 5.4.1 and 5.5.2 are satisfied. Hence, the stated result follows from these theorems. ■

Chapter 7

Conclusions and open problems

In this thesis we study average reward discrete-time Markov control processes on Borel spaces. Our main results include:

(a) explicit expressions for the invariant measure, the solution of the P.E. and the solution of the AROE,

(b) existence of pathwise average optimal policies, with minimum variance and an asymptotic normality behavior

(c) existence of constrained optimal policies,

(d) existence of constrained pathwise average optimal policies.

To analyze our problems we proceed in three steps:

In the first one, we give explicit expressions for the invariant measures, also for the functions h^*_φ that solve the P.E., and the functions h^* that solve the AROE. This fact will be particularly useful to prove boundedness conditions, necessary for a nice asymptotic behavior (law of large numbers, asymptotic normality) and to prove compactness conditions.

In the second step we prove, under our assumptions, the existence of unconstrained sample-path optimal policies (part (b) above). The main result here is Theorem 3.3.2, which together with Theorem 2.4.3 gives the existence of deterministic stationary sample-path average optimal policies. To this end we applied the law of large numbers for martingales, also known as the martingale stability theorem. Furthermore, we solve a variance-minimization problem. So, we prove the existence among the canonical policies for which

the optimization problem is solved, of those policies with minimal variance. We also show that these canonical policies with minimal variance imply asymptotic normality behavior.

In the third and final step we study part (c) and (d), and we extend the results in the former steps for constrained MCPs. Thus, the unconstrained AROEs in Theorem 2.4.3 is extended to the constrained case in Theorem 5.3.1, which in particular gives us the existence of optimal policies to our problem with expected constraints. Moreover, Theorem 5.4.1 shows that the expected CP can be solved by means of a parametric family of AROEs, which do not depend on unknown parameters. Finally, in Theorem 5.5.2, we extend these results to MCPs with pathwise constraints. In particular, we prove that the constrained optimal policies are also pathwise constrained optimal policies.

There are several standard techniques to analyze discrete-time constrained control problems. For example the so-called *direct method* where the idea is to transform the constrained problem into an equivalent optimization problem in a suitable space of measures, and then one uses the well-known fact that an upper semicontinuous (u.s.c.) function on a compact topological space attains its maximun value. For example, the proof of Theorem 5.3.1 uses in part the direct method in combination with other techniques such as *convex analysis*, *Lagrange multipliers* and *dynamic programming*. On the other hand, to analyse the pathwise constrained problem, we use the strong law of large number for Markov chains and also the martingale stability theorem.

As future work for CMCPs with expected and pathwise constraints, we consider several questions. The first one, is a variance minimization problem, that is, we would like to prove the existence, within the class of stationary optimal policies for the pathwise CP, of one with minimal limiting *average variance*. The second question is to find approximation schemes where the optimum value as well as the optimal policy can be approximated for the problem with pathwise constraints.

References

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

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

[3] Borkar, V.S., *Ergodic control of Markov chains with constraints—the general case*, Siam J. Control Optim. **32**, (1994), pp. 176–186.

[4] Curtain, R., Pritchard, A., *Functional Analysis in Modern Applied Mathematics*, Academic Press, London, 1977.

[5] Dudley, R. M., *Real Analysis and Probability*, Wadsworth & Brooks/Cole Advanced Books and Software, Pacific Grove, CA, 1989.

[6] Ekeland I., Temam, R., *Convex Analysis and Variational Problems*, North-Holland Publishing Company, 1976.

[7] Feinberg, E., Shwartz, A., *Constrained discounted dynamic programming*, Math. Oper. Res. **21** (1996), pp. 922–945.

[8] Föllmer, H., Schied, A., *Stochastic Finance. An Introduction in Discrete Time*, De Gruyter Studies in Mathematics **27**. Walter de Gruyter & Co, Berlin, 2002.

[9] Gordienko, E., Hernández-Lerma, O., *Average cost Markov control processes with weigthed norms: existence of canonical policies*, Appl. Math. (Warsaw) **23** (1995), pp. 199–218.

[10] Gordienko, E., Hernández-Lerma, O., *Average cost Markov control processes with weigthed norms: value iteration*, Appl. Math. (Warsaw) **23** (1995), pp. 219–237.

[11] Guo, X.P., Cao, X.R., *Optimal control of ergodic continuous-time Markov chains with average sample path rewards*, SIAM J. Control Optim. **44** (2005), pp. 29–48.

[12] Haviv, M., *On constrained Markov decision processes,* Oper. Res. Lett. **19** (1996), pp. 25–28.

[13] Hernández-Lerma, O., González-Hernández, J., López-Martínez, R.R., *Constrained average cost Markov control processes in Borel spaces*, SIAM J. Control Optim. **42** (2003), pp. 442–468.

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

[15] Hernández-Lerma, O., Lasserre, J.B., *Further Topics on Discrete-Time Markov Control Processes*, Springer, New York, 1999.

[16] Hernández-Lerma, O., Lasserre, J.B., *Markov Chains and Invariant Probabilities*, Birkhäuser Verlag, Switzerland, 2003.

[17] Hernández-Lerma, O., Vega-Amaya, O., *Infinite-horizon Markov control processes with undiscounted cost criteria: From average to overtaking optimality*, Appl. Math. (Warsaw) **25** (1998), pp. 153–178.

[18] Hernández-Lerma, O., Vega-Amaya, O., Carrasco, G., *Sample-path optimality and variance-minimization of average cost Markov control processes*, SIAM J. Control Optim. **38** (1999), pp. 79–93.

[19] Hilgert, N., Hernández-Lerma, O., *Bias optimality versus strong 0-discount optimality in Markov control processes with unbounded costs*, Acta Appl. Math. **77**, (2003), pp. 215-235.

[20] Luque-Vásquez, F., Hernández-Lerma, O., *Semi-Markov control models with average costs*, Appl. Math. (Warsaw) **26** (1999), pp. 315–331.

[21] Mandl, P. *On the Asymptotic Normality of the Reward in a Controlled Markov Chain*, Colloquia Mathematica Societatis János Bolyai, 9. European Meeting of Statisticians, Budapest (Hungary), 1972.

[22] Meyn, S.P., Tweedie, R.L., *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1993.

[23] Piunovskiy, A.B., *Optimal Control of Random Sequences in Problems with Constraints*, Kluwer, Boston, 1997.

[24] Prieto-Rumeau, T., Hernández-Lerma, O., *Ergodic control of continuous-time Markov chains with pathwise constraints*, SIAM J. Control Optim., *to appear* (2008).

[25] Royden, H.L., *Real Analysis*, 3rd Edition, Macmillan, New York, 1988.

[26] Ross, K.W., Varadarajan, R., *Markov decision processes with sample path constraints*, Oper. Res. **37** (1989), pp. 780–790.

[27] Ross, K.W., Varadarajan, R., (1991). *Multichain Markov decision processes with a sample path constraint*, Math. Oper. Res. **16** (1991), pp. 195–207.

[28] Vega-Amaya, O., *Overtaking optimality for a class of production-inventory systems*. Preprint, Departamento de Matemáticas, Universidad de Sonora, México, 1996.

[29] Vega-Amaya, O., *Markov control processes in Borel spaces: Undiscounted criteria*, Doctoral thesis, UAM-Iztapalapa, México, 1998 (in Spanish).

[30] Vega-Amaya, O., *The average cost optimality equation: a fixed point approach*, Bol. Soc. Mat. Mexicana **9**, (2003), pp. 185-195.

[31] Vega-Amaya, O., *Expected and sample-path constrained average Markov decision processes*, Reporte interno Núm. **35**, Departamento de Matemáticas, Universidad de Sonora, 2007.