Morfismos, Vol. 7, No. 1, 2003, pp. 1–26

The Lagrange approach to constrained Markov control processes: a survey and extension of results *

Raquiel R. López–Martínez [1] Onésimo Hernández–Lerma [2]

Abstract

This paper considers constrained Markov control processes in Borel spaces, with unbounded costs. The criterion to be minimized is the expected total discounted cost and the constraints are imposed on similar criteria. Conditions are given for the constrained problem to be equivalent to a convex program. We present a saddle-point theorem for the Lagrange function associated with the convex program, which is used to obtain the existence of an optimal solution to the constrained problem. In addition, we show that there exists an optimal policy for the constrained problem which is also Pareto optimal for a certain multiobjective Markov control processes.

2000 Mathematics Subject Classification: 90C40, 93E20, 90C25.
Keywords and phrases: Constrained Markov control processes, convex problems, saddle point, Pareto policies.

1 Introduction

This paper gives a unified, self–contained presentation of constrained Markov control processes (MCPs) in *Borel spaces* with *unbounded costs.* The criterion to be minimized is an expected discounted cost and the constraints are imposed on similar discounted cost functionals. The paper has two main objectives. First, it is a survey of several techniques to

[*] **Invited Article.**

[1] Research partially supported by a PROMEP grant.

[2] Research partially supported by CONACyT Grant 37355–E.

analyze constrained MCPs, with emphasis on the Lagrange approach. Second, it extends to constrained MCPs in *general* (i.e. nondenumerable, noncompact) Borel spaces some results on the existence of optimal policies and it also studies the relation between the Lagrange and the Pareto approaches. In particular, we show the existence of an optimal policy which is also Pareto optimal for a certain multiobjective MCP.

The constrained problem (CP) we are concerned with is of the following form: given performance criteria $V_0, V_1, \ldots, V_q$ and constants $k_1, \ldots, k_q$,

$$\text{Minimize } V_0(\pi)$$

over the set of control policies π that satisfy the constraints

$$V_i(\pi) \leq k_i \quad \forall\, i = 1, \ldots, q.$$

Control problems of this form appear in many areas — see, for instance, [1–6, 8, 10–14, 19–25, 28–34]. The easiest way to analyze CP is using the so–called *direct method*. In this method, which of course is also applicable to unconstrained MCPs (e.g. [15], §5.7), the idea is to use occupation measures to transform CP into a "static" optimization problem, say CP'; see [13, 14] and §3 below. If one identifies the set of occupation measures with a convex subset of a suitable *linear space* of (signed) measures, then one can express CP' in an obvious manner as either a linear program or a convex program. The *linear programming* formulation has been done for constrained MCPs in finite [8, 21, 22] or countable [1–3, 20] or even Borel [13, 14] spaces. On the other hand, the *convex programming* approach, which is the one we are interested in this paper, was originally introduced by Beutler and Ross [4] for MCPs with a countable state space and a single constraint, but it has been extended in many directions, for instance, countable state spaces with compact action sets [1, 3, 5, 6, 31, 32] and Borel state spaces [23, 25, 27, 28, 33]. (For the dynamic programming approach, which is not discussed in this paper, see [29].)

As already mentioned above, in this paper we are mainly concerned with the convex programming formulation of constrained MCPs with *general* Borel state space and *unbounded* costs.

We begin in §2 by introducing some basic terminology and notation. In §3 we define the associated discounted occupation measures and state Lemma 3.3, which ensures that we can consider CP as a convex programming problem. In §4 we study the convex problem. In particular, we obtain a saddle-point theorem for the associated Lagrange function, which gives an optimal solution for CP. (A similar result for *average cost* problems appears in [25].) In §5 we establish some connections between the Lagrange approach and the Pareto optimality of a certain multiobjective MCP. Conditions are given under which an optimal policy for CP is Pareto optimal for the multiobjective problem. To illustrate the results in §4 and §5, in §6 we study the so–called stochastic stabilization problem, from [9] and [27]. In particular, we show a saddle point for the Lagrangean associated with this problem. In §7 and §8 we give the proof of Theorems 4.4, 4.5, and 5.4, which require lengthy preliminaries.

2 Constrained MCPs

Constrained MCPs are rather standard and will be introduced only briefly. (If necessary, see for instance [1, 13, 14, 28, 31, 32] for further details.)

The *constrained* Markov control model is of the form

$$(X, A,\ \{A(x)\,|\,x \in X\}, Q, c, \mathbf{d}, \mathbf{k}), \tag{2.1}$$

where X and A are the *state space* and the *control space*, respectively. We shall assume that X and A are Borel spaces, endowed with the corresponding Borel σ-algebras $\mathcal{B}(X)$, $\mathcal{B}(A)$. For each $x \in X$, the nonempty set $A(x)$ in $\mathcal{B}(A)$ consists of the *feasible controls* or *actions* when the system is in state $x \in X$. We suppose that the set

$$\mathbb{K} := \{(x, a)\,|\,x \in X,\ a \in A(x)\} \tag{2.2}$$

of feasible state-action pairs is a Borel subset of $X \times A$. Moreover, Q stands for the *transition law*, and $c\colon \mathbb{K} \to \mathbb{R}$ is a measurable function that denotes the *cost-per-stage* . Finally, $\mathbf{d} = (d_1, \ldots, d_q) : \mathbb{K} \to \mathbb{R}^q$ is a given function and $\mathbf{k} = (k_1, \ldots, k_q)$ is a given vector in $\mathbb{R}^q$, which are used to define the constrained problem (CP) in (2.5) and (2.6), below.

Let Π be the set of all (randomized, history-dependent) admisible control policies. Let Φ be the set of all the stochastic kernels φ on A given X such that $\varphi(A(x)|\, x) = 1$ for all $x \in X$, and let $\mathbb{F}$ be the family of measurable functions $f : X \to A$ for which $f(x) \in A(x)$ for all $x \in X$. As usual, we will identify Φ with the family of *randomized stationary* policies, and $\mathbb{F}$ with the subfamily of *deterministic stationary* policies.

Throughout the following, we consider a fixed *discount factor* $\delta \in$ (0,1), and a fixed *initial distribution* $\gamma_0 \in \mathbb{P}(X)$, where $\mathbb{P}(X)$ denotes the set of probability measures on X. Given the functions c and $\mathbf{d} =$

$(d_1, \ldots, d_q)$ as in (2.1), for each policy $\pi \in \Pi$, consider the expected δ-discounted cost functions

$$V_0(\pi, \gamma_0) := (1-\delta) E^{\pi}_{\gamma_0} \left[\sum_{t=0}^{\infty} \delta^t c(x_t, a_t) \right], \tag{2.3}$$

$$V_i(\pi, \gamma_0) := (1-\delta) E^{\pi}_{\gamma_0} \left[\sum_{t=0}^{\infty} \delta^t d_i(x_t, a_t) \right] \text{ for } i = 1, \ldots, q. \tag{2.4}$$

Furthermore, letting $\mathbf{k} = (k_1, \ldots, k_q)$ be the q-vector in (2.1), define a subset Δ of Π as

$$\Delta := \{\pi \mid V_0(\pi, \gamma_0) < \infty \;\text{ and } V_i(\pi, \gamma_0) \leq k_i \; (i = 1, \ldots, q)\}. \tag{2.5}$$

With this notation, we may then define the *constrained problem* (CP) we are concerned with as follows:

$$\mathbf{CP}: \quad \text{Minimize } V_0(\pi, \gamma_0) \quad \text{subject to } \pi \in \Delta. \tag{2.6}$$

If there exists a policy π^* in Δ that solves CP, that is,

$$V_0(\pi^*, \gamma_0) = \inf\{V_0(\pi, \gamma_0) \,|\, \pi \in \Delta\} =: V^*(\gamma_0), \tag{2.7}$$

then π^* is said to be an *optimal policy* for CP, and $V^*(\gamma_0)$ is called the *optimal value* of CP.

3 CP as a "static" optimization problem

The following conditions are used, in particular, to express CP as an optimization problem on a certain set of occupation measures —see Lemma 3.3.

Assumption 3.1

(a) The set $\mathbb{K}$ (defined in (2.2)) is closed.

(b) $c(x, a)$ is *nonnegative and inf-compact*, which means that for each $r \in \mathbb{R}$ the set $\{(x, a) \in \mathbb{K} \,|\, c(x, a) \leq r\}$ is compact.

(c) $d_i(x,a)$ is nonnegative and lower semicontinuous (l.s.c.) for $i = 1,\ldots,q$.

(d) The transition law Q is *weakly continuous*, that is (denoting by $C_b(S)$ the space of continuous bounded functions on a topological spaces S), Q is such that $\int_X u(y)Q(dy|\cdot)$ belongs to $C_b(\mathbb{K})$ for each function u in $C_b(X)$.

(e) CP is *consistent*, that is, the set Δ in (2.5) is nonempty.

Observe that Assumption 3.1(b) yields, in particular, that c is l.s.c.

Assumptions 3.1(b) and (c) can be replaced with the following: *The cost functions c and $d_1,\ldots,d_q$ are nonnegative and l.s.c., and at least one of them is inf-compact.* On the other hand, the "nonnegativity" condition on c and d_i may be replaced with "boundedness from below".

Occupation measures. For each policy $\pi \in \Pi$, we define the occupation measure $\mu^\pi = \mu^\pi_{\gamma_0}$ as

$$(3.1) \qquad \mu^\pi(\Gamma) := (1-\delta)\sum_{t=0}^{\infty} \delta^t P^\pi_{\gamma_0}\left[(x_t,a_t)\in\Gamma\right] \ \forall \Gamma \in \mathcal{B}(X\times A).$$

Then μ^π is a probability measure (p.m.) on $X\times A$, which is concentrated on $\mathbb{K}$, that is, $\mu^\pi(\mathbb{K}^c)=0$, where $\mathbb{K}^c$ stands for the complement of $\mathbb{K}$. Moreover, using the notation

$$\langle \mu, h\rangle := \int h d\mu,$$

we can write (2.3) and (2.4) as

$$(3.2) \qquad V_0(\pi,\gamma_0) = \langle \mu^\pi, c\rangle \text{ and } V_i(\pi,\gamma_0) = \langle \mu^\pi, d_i\rangle \ (i=1,\ldots,q),$$

respectively.

We shall denote by $\mathbb{P}(\mathbb{K})$ the set of p.m.'s on $X\times A$ that are concentrated on $\mathbb{K}$, and by $\mathbb{P}_{O\delta}(\mathbb{K})$ the subset of occupation measures. Further, for a p.m. μ in $\mathbb{P}(\mathbb{K})$, we denote by $\widehat{\mu}$ its *marginal* on X, that is, $\widehat{\mu}(B) := \mu(B\times A)$ for all B in $\mathcal{B}(X)$.

Remark 3.2 (See Remark 6.3.1 and Theorem 6.3.7 in [15].) For each policy $\pi\in\Pi$, the occupation measure $\mu^\pi \in \mathbb{P}_{O\delta}(\mathbb{K})$ satisfies the following:

$$(3.3) \quad \widehat{\mu^\pi}(B) = (1-\delta)\gamma_0(B) + \delta\int Q(B|x,a)\mu^\pi(d(x,a)) \ \forall B\in\mathcal{B}(X).$$

Conversely, if μ is a p.m. in $\mathbb{P}(\mathbb{K})$ that satisfies (3.3), i.e.,

$$\widehat{\mu}(B) = (1-\delta)\gamma_0(B) + \delta \int Q(B|x,a)\mu(d(x,a)) \ \forall B \in \mathcal{B}(X), \tag{3.4}$$

then μ is in $\mathbb{P}_{O\delta}(\mathbb{K})$. In other words, there is a policy π for which μ is the associated occupation measure, that is, $\mu = \mu^\pi$. Therefore,

$$\mathbb{P}_{O\delta}(\mathbb{K}) = \{\mu \in \mathbb{P}(\mathbb{K}) \mid \mu \text{ satisfies } (3.4)\}.$$

We define the following subsets of $\mathbb{P}_{O\delta}(\mathbb{K})$:

(3.5)

$$\mathbb{P}_\delta(\mathbb{K}) := \{\mu \in \mathbb{P}_{O\delta}(\mathbb{K}) | \langle \mu, c\rangle < \infty, \text{ and } \langle \mu, d_i\rangle < \infty, \ i = 1, \dots q\},$$

and

$$\Delta_\delta := \{\mu \in \mathbb{P}_\delta(\mathbb{K}) | \ \langle \mu, d_i\rangle \leq k_i, \ i = 1, \dots q\}. \tag{3.6}$$

With this notation we can then state the following key fact.

Lemma 3.3 *CP is equivalent to the problem:*

$$\begin{aligned} \mathbf{CP}' : \quad & \text{Minimize } \langle \mu, c\rangle \\ & \text{subject to} : \mu \in \Delta_\delta. \end{aligned}$$

Proof: The lemma is a consequence of (3.2) and Remark 3.2. □

4 CP as a convex program

In Lemma 3.3 we already transformed CP into the "static" optimization problem CP′. We next use CP′ to restate CP as a *convex program.*

Let f and G be the functions on $\mathbb{P}_\delta(\mathbb{K})$ defined as

$$f(\mu) := \langle \mu, c\rangle \quad \text{and} \quad G(\mu) := (G_1(\mu), \dots, G_q(\mu)),$$

with $G_i(\mu) := \langle \mu, d_i\rangle - k_i$ for $i = 1, \dots, q$. Obviously, f and G are convex functions. It is just as obvious that $\mathbb{P}_\delta(\mathbb{K})$ is a convex set. Thus, by Lemma 3.3 we can represent CP as the convex problem

$$\begin{aligned} \text{Minimize} \quad & f(\mu) \\ \text{subject to} : \quad & \mu \in \mathbb{P}_\delta(\mathbb{K}) \quad \text{and} \quad G(\mu) \leq \theta, \end{aligned} \tag{4.1}$$

where θ is the vector zero in $\mathbb{R}^q$, and $G(\mu) \leq \theta$ means that $G_i(\mu) \leq 0$ for all $i = 1, \ldots, q$. Observe that the constraint in (4.1) can also be written as $\mu \in \Delta_\delta$.

The *Lagrangean* $L : \mathbb{P}_\delta(\mathbb{K}) \times \mathbb{R}^q_+ \to \mathbb{R}$ associated with problem (4.1) is given by

$$L(\mu, \boldsymbol{\alpha}) := f(\mu) + G(\mu) \cdot \boldsymbol{\alpha}, \tag{4.2}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_q)$ is in $\mathbb{R}^q_+$, and "$\cdot$" denotes the inner product in $\mathbb{R}^q$.

Remark 4.1 (a) (See, for instance, [9, p. 88,89] or [18, p. 89]). If μ is in $\mathbb{P}(\mathbb{K})$, then there exists $\varphi \in \Phi$ such that μ can be "disintegrated" as

$$\mu(B \times C) = \int_B \varphi(C|x)\widehat{\mu}(dx) \quad \forall\, B \in \mathcal{B}(X), C \in \mathcal{B}(A), \tag{4.3}$$

where $\widehat{\mu}$ is the marginal of μ on X. In abbreviated form we write (4.3) as $\mu = \widehat{\mu}\cdot \varphi$.
(b) If $\mu = \widehat{\mu}\cdot \varphi$ is in $\mathbb{P}_{O\delta}(\mathbb{K})$, then it follows from (3.4) that μ is the occupation measure of the policy $\varphi \in \Phi$, that is, $\mu = \mu^\varphi$.

The following saddle-point result gives conditions for problem (4.1) to have a solution.

Theorem 4.2 *Suppose that there exists $(\mu^*, \boldsymbol{\alpha}^*) \in \mathbb{P}_\delta(\mathbb{K}) \times \mathbb{R}^q_+$ such that the Lagrangean L has a saddle point at $(\mu^*, \boldsymbol{\alpha}^*)$, i.e.,*

$$L(\mu^*, \boldsymbol{\alpha}) \leq L(\mu^*, \boldsymbol{\alpha}^*) \leq L(\mu, \boldsymbol{\alpha}^*) \tag{4.4}$$

for all $(\mu, \boldsymbol{\alpha})$ in $\mathbb{P}_\delta(\mathbb{K}) \times \mathbb{R}^q_+$. Then
(a) μ^* solves problem (4.1), and
(b) the disintegration $\mu^* = \widehat{\mu^*} \cdot \varphi^*$ of μ^* satisfies that φ^* is an optimal policy for CP.

Proof: The proof of part (a) is similar to that of Theorem 2 in [26, p. 221], and, therefore, is omitted. Part (b) follows from (a), the Remark 4.1(b), and the equivalence of CP and problem (4.1). □

In view of Theorem 4.2, to prove that the problem (4.1) is solvable it suffices to show the existence of a saddle point for L. This is true, in particular, if the following condition holds.

Assumption 4.3 (Slater condition) There exists $\mu_1 \in \mathbb{P}_\delta(\mathbb{K})$ such that $G(\mu_1) < \theta$, that is, $G_i(\mu_1) < 0$ for $i = 1, \ldots, q$.

Theorem 4.4 *Under Assumptions 3.1 and 4.3 , there exists a saddle point* $(\mu^*, \boldsymbol{\alpha}^*)$ *for the Lagrangean L, and, therefore, CP is solvable.*

Proof: See §7. □

To summarize, Theorem 4.4 gives the existence of a saddle point $(\mu^*, \boldsymbol{\alpha}^*)$ for L, which, by Theorem 4.2 yields an optimal policy φ^* for CP. It turns out that *the converse is also true*, as shown in the following result.

Theorem 4.5 *Suppose that Assumptions 3.1 and 4.3 hold. If* $\mu^* = \widehat{\mu^*} \cdot \varphi^* \in \Delta_\delta$ *is such that* φ^* *is an optimal policy for CP, then the Lagrangean L has a saddle point.*

Proof: See §7. □

Remark 4.6 (See Remark 4.2.5, p. 51 in [7].) In our present context, Assumption 4.3 is equivalent to the so-called *Karlin condition* (or constraint qualification), according to which there is no nonzero vector $\boldsymbol{\alpha} \in \mathbb{R}^{\mathrm{q}}_{+}$ for which $G(\mu) \cdot \boldsymbol{\alpha} \geq 0$ for all $\mu \in \mathbb{P}_\delta(\mathbb{K})$.

5 The Lagrange approach vs Pareto optimality

In this section we compare the Lagrange approach to CP with the Pareto optimality of a certain *multiobjective* MCP. With this in mind, we first briefly introduce multiobjective MCPs (for more information see, for instance [17] or [28]).

Let $V_0(\pi, \gamma_0)$ and $V_i(\pi, \gamma_0)$ be as in (2.3) and (2.4), and let $V(\pi, \gamma_0) \in \mathbb{R}^{q+1}$ be the cost vector

$$V(\pi, \gamma_0) := (V_0(\pi, \gamma_0), \ldots, V_q(\pi, \gamma_0)). \tag{5.1}$$

The *multiobjective control problem* we are concerned with is to find a policy π^* that “minimizes” $V(\cdot, \gamma_0)$ in the sense of Pareto. To state this in precise form, we first simplify the notation by writing $V(\pi, \gamma_0)$ simply as $V(\pi)$.

Definition 5.1 Let $\Gamma(\Pi) \subset \mathbb{R}^{q+1}$ be the set of cost vectors in (5.1), i.e.,

$$\Gamma(\Pi) := \{V(\pi) \mid \pi \in \Pi\},$$

which is sometimes called the *performance set* of the multiobjective MCP. Then a policy π^* is said to be *Pareto optimal* (or a *Pareto policy*) if there is no $\pi \in \Pi$ such that $V(\pi) \neq V(\pi^*)$ and $V_i(\pi) \leq V_i(\pi^*)$ for all $i = 0, \ldots, q$. The set of cost vectors in $\Gamma(\Pi)$ corresponding to Pareto policies is called the *Pareto set* of $\Gamma(\Pi)$, and it is denoted by $\mathrm{Par}(\Gamma(\Pi))$.

Let $\mathbb{R}^{q+1}_{++}$ be set of vectors in $\mathbb{R}^{q+1}$ with strictly positive components. Let $\boldsymbol{\beta} \in \mathbb{R}^{q+1}_{++}$, and consider the scalar (or real-valued) cost-per -stage function

$$C^{\boldsymbol{\beta}}(x,a) := \beta_0 c(x,a) + \sum_{i=1}^{q} \beta_i d_i(x,a), \tag{5.2}$$

and the δ-discounted cost $V^{\boldsymbol{\beta}}(\pi) = V^{\boldsymbol{\beta}}(\pi, \gamma_0)$ with

$$V^{\boldsymbol{\beta}}(\pi) := (1-\delta) E^{\pi}_{\gamma_0} \left[\sum_{t=0}^{\infty} \delta^t C^{\boldsymbol{\beta}}(x_t, a_t) \right]. \tag{5.3}$$

Using (5.1) and (5.2) we may write $V^{\boldsymbol{\beta}}(\pi)$ as

$$V^{\boldsymbol{\beta}}(\pi) = \boldsymbol{\beta} \cdot V(\pi) = \sum_{i=0}^{q} \beta_i V_i(\pi). \tag{5.4}$$

Let

$$\Lambda := \{\boldsymbol{\beta} \in \mathbb{R}^{q+1}_{++} \mid \sum_{i=0}^{q} \beta_i = 1\}. \tag{5.5}$$

We may then obtain the existence of Pareto policies by the standard "scalarization" approach, as follows.

Theorem 5.2 *Choose an arbitrary vector* $\boldsymbol{\beta} \in \Lambda$*. If* $\pi^* \in \Pi$ *is an optimal policy for the scalar criterion (5.3), that is,*

$$V^{\boldsymbol{\beta}}(\pi^*) \leq V^{\boldsymbol{\beta}}(\pi) \;\; \forall \, \pi \in \Pi, \tag{5.6}$$

then π^* *is Pareto optimal.*

For a proof of Theorem 5.2 see, for instance, Theorem 3.2(a) in [17].

In general, the constrained problem CP in (2.6) can have optimal policies that are *not* Pareto optimal. On the other hand, if CP has a *unique* optimal policy π^*, then it is easily seen (directly from the Definition 5.1) that π^* is a Pareto policy. The following two theorems give other cases in which an optimal policy for CP is in fact a Pareto policy.

Theorem 5.3 *Let* $(\mu^*, \boldsymbol{\alpha}^*) \in \mathbb{P}_\delta(\mathbb{K}) \times \mathbb{R}^q_{++}$ *be a saddle point for the Lagrangean* L*, and disintegrate* μ^* *as* $\widehat{\mu^*} \cdot \varphi^*$*. Then* φ^* *is Pareto optimal.*

Proof: From the definition (4.4) of a saddle point, we have that

$$L(\mu^*, \boldsymbol{\alpha}^*) \leq L(\mu, \boldsymbol{\alpha}^*) \ \forall \, \mu \in \mathbb{P}_\delta(\mathbb{K}). \tag{5.7}$$

On the other hand, from (3.2) and the definition (4.2) of L it follows that

$$L(\mu, \boldsymbol{\alpha}) = V_0(\pi) + \sum_{i=1}^{q} \alpha_i (V_i(\pi) - k_i), \tag{5.8}$$

where π is a policy associated to the occupation measure μ. Hence, from (5.7) and (5.8) we have that

$$V_0(\varphi^*) + \sum_{i=1}^{q} \alpha_i^* (V_i(\varphi^*) - k_i) \leq V_0(\pi) + \sum_{i=1}^{q} \alpha_i^* (V_i(\pi) - k_i) \quad \forall \, \pi \in \Pi.$$

Equivalently, defining $\boldsymbol{\beta}^* := (1, \boldsymbol{\alpha}^*) \in \mathbb{R}^{q+1}_{++}$, we have

$$\boldsymbol{\beta}^* \cdot V(\varphi^*) - \boldsymbol{\alpha}^* \cdot \mathbf{k} \leq \boldsymbol{\beta}^* \cdot V(\pi) - \boldsymbol{\alpha}^* \cdot \mathbf{k} \quad \forall \, \pi \in \Pi,$$

and so

$$\boldsymbol{\beta}^* \cdot V(\varphi^*) \leq \boldsymbol{\beta}^* \cdot V(\pi) \quad \forall \, \pi \in \Pi. \tag{5.9}$$

Finally, let $P = 1 + \sum_{i=1}^{q} \alpha_i^*$. Then, multiplying both sides of (5.9) by $1/P$, it follows from Theorem 5.2 that φ^* is Pareto optimal. □

Now consider the following subset of $\Gamma(\Pi)$

$$\Gamma^*(\Pi) := \{V(\pi) \mid \pi \text{ an optimal policy for CP}\}. \tag{5.10}$$

Let $\mathrm{Par}(\Gamma^*(\Pi))$ be the Pareto set of $\Gamma^*(\Pi)$.

Theorem 5.4 *Under Assumption 3.1, the Pareto set* Par $(\Gamma^*(\Pi))$ *of* $\Gamma^*(\Pi)$ *is nonempty.*

Proof: See §8. □

It turns out that the nonemptiness of Par $(\Gamma^*(\Pi))$ in Theorem 5.4 ensures the existence of a Pareto policy that is optimal for CP.

Theorem 5.5 *Under Assumptions 3.1 and 4.3, there exists an optimal policy π^* for CP, which is also Pareto optimal.*

Proof: From Theorem 5.4 there exists a policy π^* such that $V(\pi^*)$ is in Par$(\Gamma^*(\Pi))$. By (5.10), π^* is an optimal policy for CP. We now claim that π^* is Pareto optimal, that is, $V(\pi^*)$ is in Par$(\Gamma(\Pi))$. Indeed, if π^* is not Pareto optimal, then there exists a policy $\pi_1 \in \Pi$ such that $V(\pi_1) \neq V(\pi^*)$ and $V_i(\pi_1) \leq V_i(\pi^*)$ for $i = 0, \ldots, q$. Hence, $V(\pi_1) \in \Gamma^*(\Pi)$, which contradicts our assumption on π^*. Therefore, π^* is Pareto optimal. □

6 Example

To illustrate the results in Sections 4 and 5, we next consider the following problem, which is similar to the *stochastic stabilization problem* in [9, 27]. First, we show that Assumptions 3.1 and 4.3 hold. Then, we prove that this problem is solvable using the Lagrange approach, that is, we shall obtain a saddle point for the Lagrange function. Finally, we construct the corresponding Pareto set. For notational ease, we shall write the δ-discounted costs in (2.3) and (2.4) without the factor $(1-\delta)$.

Consider the scalar linear system

$$x_{t+1} = x_t - a_t + \xi_t \quad \text{for } t = 0, 1, \ldots, \tag{6.1}$$

with state and control spaces $X = A = \mathbb{R}$. The disturbances ξ_t are i.i.d. random variables, independent of the initial state x_0, and such that

$$E(\xi_0) = 0 \quad \text{and} \quad E(\xi_0^2) =: \sigma^2 < \infty. \tag{6.2}$$

Let $c(x, a)$ and $d(x, a)$ be the quadratic costs defined as

$$c(x, a) = x^2 + a^2, \;\; d(x, a) = (x - a)^2, \tag{6.3}$$

and consider the following constrained problem in which k is a given positive constant.

$$\begin{aligned} \text{Minimize} \qquad & V_0(\pi,\gamma_0) := E^{\pi}_{\gamma_0}\left[\sum_{t=0}^{\infty}\delta^t(x_t^2+a_t^2)\right] \\ \text{subject to :} \qquad & V_1(\pi,\gamma_0) := E^{\pi}_{\gamma_0}\left[\sum_{t=0}^{\infty}\delta^t(x_t-a_t)^2\right] \le k. \end{aligned}$$

It is clear that the Assumptions 3.1(a), (b), (c) are satisfied in this example. Moreover, by the continuity of the right-hand side of (6.1) with respect to x_t and a_t for every ξ_t it follows that also Assumption 3.1(d) holds. On the other hand, if we take $\pi = f_0 \in \mathbb{F}$ as the "identity"policy $f_0(x) := x$ for all $x \in X$, we see that $V_1(f_0,x) = 0$, and, therefore, Assumptions 3.1(e) and 4.3 are both satisfied. Summarizing, Assumptions 3.1 and 4.3 hold for this problem.

Now, from (3.2) and (4.2) the corresponding Lagrange function is

$$L(\pi,\alpha) = V_0(\pi,\gamma_0) + (V_1(\pi,\gamma_0) - k)\cdot\alpha \tag{6.4}$$

with $\alpha \ge 0$. Let

$$L_1(\alpha) := \inf_{\pi\in\Pi} L(\pi,\alpha). \tag{6.5}$$

Note that defining the new cost per-stage function

$$C^{\alpha}(x,a) := c(x,a) + \alpha\cdot d(x,a) = x^2 + a^2 + \alpha(x-a)^2$$

and denoting by $V^{\alpha}(\pi,\gamma_0)$ the corresponding δ-discounted cost, we may express (6.4) as

$$L(\pi,\alpha) = V^{\alpha}(\pi,\gamma_0) - \alpha\cdot k.$$

Therefore, finding a policy that attains the minimun in (6.5) becomes a linear-quadratic problem; see, for instance, p. 162 in [9], p. 70 in [15], or p. 253 in [28]. From any of these references we have

$$\inf_{\pi\in\Pi} V^{\alpha}(\pi,x) - k\cdot\alpha = z(\alpha)v(x) - k\cdot\alpha \quad \forall\, x \in X,$$

with $v(x) := x^2 + (1-\delta)^{-1}\delta\sigma^2$, and $z(\alpha)$ is the maximal solution of the quadratic equation

$$\delta z^2 + (1+\alpha-2\delta)z - 1 - 2\alpha = 0. \tag{6.6}$$

Therefore, assuming that the initial distribution γ_0 satisfies that

$$\bar{\gamma_0} := \int v(x)\gamma_0(dx) < \infty, \tag{6.7}$$

we can express (6.5) as

$$L_1(\alpha) = z(\alpha)\bar{\gamma_0} - k \cdot \alpha. \tag{6.8}$$

Moreover, the deterministic stationary policy $f_\alpha \in \mathbb{F}$ given by

$$f_\alpha(x) = \frac{\alpha + \delta z(\alpha)}{1 + \alpha + \delta z(\alpha)} x \tag{6.9}$$

is optimal for $V^\alpha(\pi, x)$ for all $x \in \mathbb{R}$, and so we also have $L_1(\alpha) = L(f_\alpha, \alpha)$ for each $\alpha \geq 0$. Now, to obtain a saddle point for the Lagrangean in (6.4) we first prove the following, which can be seen as an "explicit" form of Lemma 7.2, below.

Proposition 6.1 *If the constraint constant k satisfies the inequality*

$$0 < k < K, \tag{6.10}$$

where $K := \bar{\gamma_0}(1 + 2\delta - \sqrt{1 + 4\delta^2})/2\delta\sqrt{1 + 4\delta^2}$, then there exists a unique $\alpha^* > 0$ such that

$$L_1(\alpha^*) = \max_{\alpha \geq 0} L_1(\alpha).$$

Proof: We differentiate the function L_1 in (6.8) with respect to α, to get $L_1'(\alpha) = z'(\alpha)\bar{\gamma_0} - k$.

Let us now show that $L_1'(\alpha) = 0$ has a unique positive solution. With this in mind, first note that the positive solution of (6.6) is

$$z(\alpha) = \frac{-(1 + \alpha - 2\delta) + \sqrt{(1 + \alpha - 2\delta)^2 + 4\delta(1 + 2\alpha)}}{2\delta}.$$

Hence

$$z'(\alpha) = -\frac{1}{2\delta} + \frac{1 + \alpha + 2\delta}{2\delta\sqrt{(1 + \alpha - 2\delta)^2 + 4\delta(1 + 2\alpha)}},$$

and so

$$L_1'(\alpha) = \left[-\frac{1}{2\delta} + \frac{1 + \alpha + 2\delta}{2\delta\sqrt{(1 + \alpha - 2\delta)^2 + 4\delta(1 + 2\alpha)}} \right] \bar{\gamma_0} - k. \tag{6.11}$$

According to (6.10) and (6.11) we have

$$L_1'(0) = \frac{\bar{\gamma}_0(1+2\delta-\sqrt{1+4\delta^2})}{2\delta\sqrt{1+4\delta^2}} - k > 0.$$

On the other hand,

$$\lim_{\alpha\to\infty} L_1'(\alpha) = -k < 0.$$

Hence the equation $L_1'(\alpha) = 0$ has a positive solution. Moreover, from (6.11), $L_1'(\alpha) = 0$ becomes

$$(1+\alpha+2\delta)^2 = 4\delta^2(k(\bar{\gamma}_0)^{-1} + (2\delta)^{-1})^2((1+\alpha-2\delta)^2 + 4\delta(1+2\alpha)).$$

As this equation is quadratic in α, it has a unique positive solution. □

Let α^* be as in Proposition 6.1 and define $z^* = z(\alpha^*)$ and $f^* := f_{\alpha^*}$ as in (6.9), that is,

$$f^*(x) := f_{\alpha^*}(x) = (\alpha^* + \delta z^*)(1 + \alpha^* + \delta z^*)^{-1}x.$$

Then (f^*, α^*) *is a saddle point* for L, and, therefore, from Theorem 4.2 it follows f^* *is an optimal policy for* CP. Moreover, as α^* is positive, from Theorem 5.3 we have that f^* *is Pareto optimal.*

Remark 6.2 If $\alpha = 0$, then $f_0^*(x) = \delta z_0 x(1+\delta z_0)^{-1}$ is optimal for V_0, that is,

$$V_0(f_0^*, \gamma_0) = \inf_{\pi\in\Pi} V_0(\pi, \gamma_0)$$

where z_0 is the positive solution of the quadratic equation

$$\delta z^2 + (1-2\delta)z - 1 = 0. \tag{6.12}$$

On the other hand, we can see that the "identity" policy $f_0(x) = x$ is optimal for V_1, and obviously, $V_1(f_0, \gamma_0) = 0$, that is,

$$\inf_{\pi\in\Pi} V_1(\pi, \gamma_0) = 0.$$

Proposition 6.3 *Let $\widehat{f}$ be a constant, and $f \in \mathbb{F}$ a stationary policy given by $f(x) := \widehat{f}x$ for all $x \in X$. Let $\theta := 1 - \widehat{f}$. If $|\theta| < 1$, then*

$$V_0(f, \gamma_0) = \frac{1+\widehat{f}^2}{1-\delta\theta^2}\bar{\gamma}_0, \tag{6.13}$$

$$V_1(f, \gamma_0) = \frac{(1-\widehat{f})^2}{1-\delta\theta^2}\bar{\gamma}_0. \tag{6.14}$$

In particular, for K and f_0^* as in (6.10) and Remark 6.2,

$$V_1(f_0^*, \gamma_0) = K. \tag{6.15}$$

Proof: Replacing a_t in (6.1) with $a_t = f(x_t) = \widehat{f} x_t$, we obtain

$$x_t = (1 - \widehat{f})x_{t-1} + \xi_{t-1} = \theta x_{t-1} + \xi_{t-1} \quad \forall\, t = 1, 2, \ldots.$$

Hence, for all $t = 1, 2, \ldots$

$$x_t = \theta^t x_0 + \sum_{j=0}^{t-1} \theta^j \xi_{t-1-j},$$

and so

$$E_x^f(x_t^2) = \theta^{2t} x^2 + \frac{\sigma^2(1 - \theta^{2t})}{1 - \theta^2}.$$

This yields that

$$E_x^f \left(\sum_{t=0}^{\infty} \delta^t x_t^2 \right) = \frac{1}{1 - \delta\theta^2} \left(x^2 + \frac{\sigma^2 \delta}{1 - \delta} \right) = \frac{v(x)}{1 - \delta\theta^2}. \tag{6.16}$$

Hence, from (6.7),

$$E_{\gamma_0}^f \left(\sum_{t=0}^{\infty} \delta^t x_t^2 \right) = \frac{\bar{\gamma}_0}{1 - \delta\theta^2}. \tag{6.17}$$

Now note that using $a = f(x) = \widehat{f} x$ in (6.3) we get

$$c(x, a) = (1 + \widehat{f}^{2})x^2 \quad \text{and} \quad d(x, a) = (1 - \widehat{f})^2 x^2 \tag{6.18}$$

for all x. Thus, inserting (6.17) and (6.18) in V_0 and V_1 we obtain (6.13) and (6.14). Finally, from (6.14) and Remark 6.2 we have

$$V_1(f_0^*, \gamma_0) = \frac{\bar{\gamma}_0}{(1 + \delta z_0)^2 - \delta}. \tag{6.19}$$

On the other hand, from (6.12) we get

$$z_0 = \frac{2\delta - 1 + \sqrt{1 + 4\delta^2}}{2\delta}. \tag{6.20}$$

Hence, substituting (6.20) in (6.19) we obtain (6.15). □

Remark 6.4 Suppose that instead of (6.10) we have

$$k \geq K,$$

and let $f_0^*(x) = \delta z_0 x(1+\delta z_0)^{-1}$ be the optimal policy for V_0 (see Remark 6.2). Then, from (6.15) it follows that

$$V_1(f_0^*, \gamma_0) = K \leq k$$

and, therefore, f_0^* is an optimal policy for the constrained problem. Moreover, f_0^* is the unique optimal policy for CP, and so it is Pareto optimal, that is, $(V_0(f_0^*, \gamma_0), V_1(f_0^*, \gamma_0))$ belongs to the Pareto set. (See Figure 6.1.)

The Pareto set. We next construct the Pareto set in an explicit form. As seen above, f^* is an optimal policy for CP which is also Pareto optimal, that is, $(V_0(f^*, \gamma_0), V_1(f^*, \gamma_0))$ is in the Pareto set. When the constraint constant k varies in the interval $(0, K)$, with K as in (6.10), then $(V_0(f^*, \gamma_0), V_1(f^*, \gamma_0))$ describes the Pareto set. Obviously, $V_0(f^*, \gamma_0)$ is the optimal value for the constrained problem, that is, $V_0(f^*, \gamma_0) = V^*(\gamma_0)$. Now, we wish to find the value of $V_1(f^*, \gamma_0)$.

Proposition 6.5 *For each k as in (6.10),*

$$V_1(f^*, \gamma_0) := E_{\gamma_0}^{f^*}\left[\sum_{t=0}^{\infty} \delta^t (x_t - a_t)^2\right] = k$$

and so $(V_0(f^*, \gamma_0), V_1(f^*, \gamma_0)) = (V^*(\gamma_0), k)$ belongs to the Pareto set.

Proof: Since $(f^*, \boldsymbol{\alpha}^*)$ is a saddle point and f^* is an optimal policy for CP we have

$$V^*(\gamma_0) \leq L(f^*, \boldsymbol{\alpha}^*) = V^*(\gamma_0) + (V_1(f^*, \gamma_0) - k)\boldsymbol{\alpha}^*.$$

On the other hand, as $(V_1(f^*, \gamma_0) - k)\boldsymbol{\alpha}^* \leq 0$, it follows that

$$V^*(\gamma_0) + (V_1(f^*, \gamma_0) - k)\boldsymbol{\alpha}^* \leq V^*(\gamma_0)$$

and so we have $(V_1(f^*, \gamma_0) - k)\boldsymbol{\alpha}^* = 0$. This equality together with Proposition 6.1 yields that $V_1(f^*, \gamma_0) = k$. □

Proposition 6.5 ensures that $(V^*(\gamma_0), k)$ belongs to the Pareto set when k varies in $(0, K)$. Furthermore, if α^* is as in Proposition 6.1, it is clear then that $V^*(\gamma_0) = L_1(\alpha^*)$. Now, in connection with the Figure 6.1, let us fix $w = k$ and calculate $y = L_1(\alpha^*)$. First, we note the following facts. Proposition 6.3 yields that

$$V_1(f^*, \gamma_0) = \frac{\bar{\gamma}_0}{(1 + \boldsymbol{\alpha}^* + \delta z^*)^2 - \delta}. \tag{6.21}$$

Further, from (6.6) with $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ we have

$$\boldsymbol{\alpha}^* = \frac{\delta(z^*)^2 + (1 - 2\delta)z^* - 1}{2 - z^*} \tag{6.22}$$

and subtituting this value of $\boldsymbol{\alpha}^*$ in (6.21) it follows that

$$V_1(f^*, \gamma_0) = \frac{(2 - z^*)^2\bar{\gamma}_0}{1 - \delta(2 - z^*)^2}. \tag{6.23}$$

Hence, from Proposition 6.5 we get

$$\frac{(2 - z^*)^2\bar{\gamma}_0}{1 - \delta(2 - z^*)^2} = k. \tag{6.24}$$

Now, substituting 6.22 in $L(\boldsymbol{\alpha}^*)$ we obtain

$$\begin{aligned} y &= z^*\bar{\gamma}_0 - \frac{\delta(z^*)^2 + (1 - 2\delta)z^* - 1}{2 - z^*}k \\ &= \frac{-(z^*)^2(\bar{\gamma}_0 + \delta k) + 2z^*(\bar{\gamma}_0 + \delta k) - kz^* + k}{2 - z^*}. \end{aligned} \tag{6.25}$$

From (6.24) we have that

$$-(z^*)^2(\bar{\gamma}_0 + \delta k) = 4(\bar{\gamma}_0 + \delta k)(1 - z^*) - k. \tag{6.26}$$

The latter equality together with (6.25) gives

$$y = 2(\bar{\gamma}_0 + \delta k) - \frac{z^*}{2 - z^*}k. \tag{6.27}$$

Now let

$$s := \frac{z^*}{2 - z^*}.$$

Solving this equation for z^* and substituting the solution in (6.24) we get

$$\frac{4\bar{\gamma}_0}{(1 + s)^2 - 4\delta} = k,$$

which yields

$$w = k = \bar{\gamma_0}\frac{4}{(s+1)^2 - 4\delta} \tag{6.28}$$

and so, from (6.25),

$$y = 2(\bar{\gamma_0} + k\delta) - \frac{4\delta\bar{\gamma_0}}{(s+1)^2 - 4\delta}. \tag{6.29}$$

In (6.28) and (6.29) s is the parameter which varies as k is in $(0, K)$. The graph of (6.28)-(4.29) is the Pareto set, which is represented in Figure 6.1.

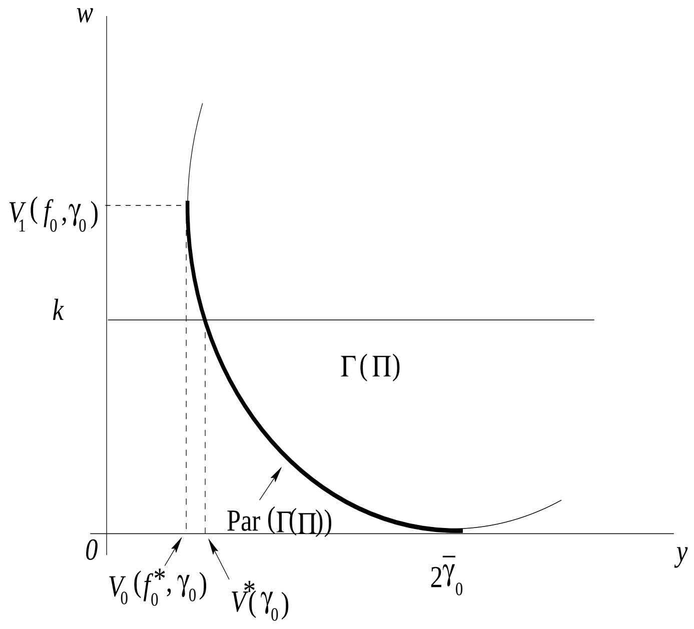

Figure 6.1

7 Proof of Theorems 4.4 and 4.5

The proof of Theorems 4.4 and 4.5 is based on the following preliminary facts. Consider the functions

$$L_1(\boldsymbol{\alpha}) := \inf_{\mu \in \mathbb{P}_\delta(\mathbb{K})} L(\mu, \boldsymbol{\alpha}), \tag{7.1}$$

$$L_2(\mu) := \sup_{\boldsymbol{\alpha} \geq \theta} L(\mu, \boldsymbol{\alpha}), \tag{7.2}$$

and let $V^*(\gamma_0)$ be as in (2.7). Note that, by Lemma 3.3,

$$V^*(\gamma_0) = \inf\{\langle \mu, c\rangle \mid \mu \in \Delta_\delta\}.$$

Remark 7.1 As $\Delta_\delta \subset \mathbb{P}_\delta(\mathbb{K})$, for each $\boldsymbol{\alpha} \in \mathbb{R}^q_+$ we have

$$L_1(\boldsymbol{\alpha}) \leq \inf_{\mu \in \Delta_\delta} L(\mu, \boldsymbol{\alpha}) \leq \inf_{\mu \in \Delta_\delta} \langle \mu, c\rangle = V^*(\gamma_0),$$

that is, $L_1(\boldsymbol{\alpha}) \leq V^*(\gamma_0)$ for all $\boldsymbol{\alpha} \in \mathbb{R}^q_+$. Similarly, $V^*(\gamma_0) \leq L_2(\mu)$ for all $\mu \in \Delta_\delta$. Hence

$$\sup_{\boldsymbol{\alpha} \geq \theta} L_1(\boldsymbol{\alpha}) \leq V^*(\gamma_0) \leq \inf_{\mu \in \mathbb{P}_\delta(\mathbb{K})} L_2(\mu). \tag{7.3}$$

The following lemmas show that equality holds throughout (7.3).

Lemma 7.2 *Under Assumptions 3.1 and 4.3, there exists* $\boldsymbol{\alpha}^*$ *in* $\mathbb{R}^q_+$ *such that*

$$L_1(\boldsymbol{\alpha}^*) = \sup_{\boldsymbol{\alpha} \geq \theta} L_1(\boldsymbol{\alpha}) = V^*(\gamma_0).$$

Proof: In the space $\mathbb{R} \times \mathbb{R}^{\mathrm{q}}$ define the sets

$$B_1 := \{(r, \boldsymbol{\alpha}) | \ r \geq f(\mu), \boldsymbol{\alpha} \geq G(\mu) \text{ for some } \mu \in \mathbb{P}_\delta(\mathbb{K}) \ \},$$
$$B_2 := \{(r, \boldsymbol{\alpha}) | \ r \leq V^*(\gamma_0), \boldsymbol{\alpha} \leq \theta \ \}.$$

The set B_2 is obviously convex, and so is B_1 because f and G are convex. By definition of $V^*(\gamma_0)$, the set B_1 contains no interior points of B_2. On the other hand, it is clear that B_2 contains an interior point. Thus, by the Separating Hyperplane Theorem (see, for example, [26], p. 133, Theorem 3), there is a vector $(r^*, \boldsymbol{\alpha}^*) \in \mathbb{R} \times \mathbb{R}^q$ such that

$$r^* r_1 + \boldsymbol{\alpha}_1 \cdot \boldsymbol{\alpha}^* \geq r^* r_2 + \boldsymbol{\alpha}_2 \cdot \boldsymbol{\alpha}^*$$

for all $(r_1, \boldsymbol{\alpha}_1) \in B_1$ and all $(r_2, \boldsymbol{\alpha}_2) \in B_2$. By the definition of B_2 it follows that $r^* \geq 0, \boldsymbol{\alpha}^* \geq \theta$. We next show that in fact $r^* > 0$. Indeed, as the vector $(V^*(\gamma_0), \theta)$ is in B_2, we have

$$(7.4) \qquad r^* r + \boldsymbol{\alpha} \cdot \boldsymbol{\alpha}^* \geq r^* V^*(\gamma_0)$$

for all $(r, \boldsymbol{\alpha}) \in B_1$. Thus, if $r^* = 0$, then $\boldsymbol{\alpha} \cdot \boldsymbol{\alpha}^* \geq 0$ for all $\boldsymbol{\alpha} \in \mathbb{R}^q$ such that $(r, \boldsymbol{\alpha}) \in B_1$. In particular, taking $\boldsymbol{\alpha} = G(\mu_1)$ with μ_1 as in Assumption 4.3, we obtain $G(\mu_1) \cdot \boldsymbol{\alpha}^* \geq 0$, which implies that $G_i(\mu_1) \geq 0$ for some $i = 1, \ldots, q$. As this contradicts Assumption 4.3, it follows that $r^* > 0$ and, without loss of generality, we may assume $r^* = 1$.

Now, since the point $(V^*(\gamma_0), \theta)$ is in the closure of both B_1 and B_2, we have (with $r^* = 1$ in (7.4))

$$\begin{aligned} V^*(\gamma_0) &= \inf_{(r, \boldsymbol{\alpha}) \in B_1} [r + \boldsymbol{\alpha} \cdot \boldsymbol{\alpha}^*] \leq \inf_{\mu \in \mathbb{P}_\delta(\mathbb{K})} [f(\mu) + G(\mu) \cdot \boldsymbol{\alpha}^*] \\ &= \inf_{\mu \in \mathbb{P}_\delta(\mathbb{K})} L(\mu, \boldsymbol{\alpha}^*) \leq \inf_{\mu \in \Delta_\delta} f(u) = V^*(\gamma_0). \end{aligned}$$

Hence, recalling (7.3) the lemma is proved. □

By (7.1) and (7.2) the following lemma is a “minimax” result.

Lemma 7.3 *Under Assumptions 3.1 and 4.3, we have*

$$(7.5) \qquad \max_{\boldsymbol{\alpha} \geq \theta} L_1(\boldsymbol{\alpha}) = \inf_{\mu \in \mathbb{P}_\delta(\mathbb{K})} L_2(\mu) = V^*(\gamma_0).$$

Proof: Since $G(\mu) \cdot \boldsymbol{\alpha} \leq \theta$ for all $\mu \in \Delta_\delta$ and $\boldsymbol{\alpha} \geq 0$, we see that

$$L_2(\mu) = \sup_{\boldsymbol{\alpha} \geq \theta} L(\mu, \boldsymbol{\alpha}) = \langle \mu, c \rangle \quad \text{for all } \mu \in \Delta_\delta.$$

Hence

$$\inf_{\mu \in \Delta_\delta} L_2(\mu) = V^*(\gamma_0).$$

It follows that

$$\inf_{\mu \in \mathbb{P}_\delta(\mathbb{K})} L_2(\mu) \leq V^*(\gamma_0),$$

and so, by (7.3) and Lemma 7.2, the equality (7.5) holds. □

Lemma 7.4 *Under Assumption 3.1, there exists a p.m.* μ^* *in* $\mathbb{P}_\delta(\mathbb{K})$ *such that*

$$L_2(\mu^*) = \inf_{\mu \in \mathbb{P}_\delta(\mathbb{K})} L_2(\mu) = V^*(\gamma_0).$$

Proof: If μ is in $\mathbb{P}_\delta(\mathbb{K})$ but not in Δ_δ, then there exists i_0 in $\{1,\ldots,q\}$ such that $G_{i_0}(\mu) > 0$, which implies that $L_2(\mu) = +\infty$. Therefore,

$$\inf_{\mu\in\mathbb{P}_\delta(\mathbb{K})} L_2(\mu) = \inf_{\mu\in\Delta_\delta} L_2(\mu) = V^*(\gamma_0). \tag{7.6}$$

On the other hand, for all $\mu \in \Delta_\delta$ and $\boldsymbol{\alpha} \geq \theta$, we have $G(\mu)\cdot\boldsymbol{\alpha} \leq 0$, and so it follows that

$$L_2(\mu) = \sup_{\boldsymbol{\alpha}\geq\theta} L(\mu,\boldsymbol{\alpha}) = \langle\mu, c\rangle \ \forall\ \mu \in \Delta_\delta. \tag{7.7}$$

Therefore, from the (7.6), (7.7), together with Lemma 3.3 and Theorem 3.2 in [13], the desired conclusion follows. □

We are now ready for the proof of Theorems 4.4 and 4.5.

Proof of Theorem 4.4. Let $\boldsymbol{\alpha}^*$ and μ^* be as in Lemma 7.2 and Lemma 7.4, respectively. From lemma 7.3 we have that

$$L(\mu^*,\boldsymbol{\alpha}^*) = V^*(\gamma_0).$$

Now, by the latter equality together with Lemmas 7.2, 7.4 and the definition of L_1 and L_2 it follows that

$$L(\mu^*,\boldsymbol{\alpha}^*) = L_1(\boldsymbol{\alpha}^*) \leq L(\mu,\boldsymbol{\alpha}^*) \text{ for all } \mu \in \mathbb{P}_\delta(\mathbb{K}),$$

and, similarly,

$$L(\mu^*,\boldsymbol{\alpha}^*) = L_2(\mu^*) \geq L(\mu^*,\boldsymbol{\alpha}) \text{ for all } \boldsymbol{\alpha} \geq \theta.$$

Therefore, the pair $(\mu^*,\boldsymbol{\alpha}^*)$ is a saddle point. □

Proof of Theorem 4.5. Let $\boldsymbol{\alpha}^*$ be as in Lemma 7.2. As $G(\mu^*) \leq 0$ and $f(\mu^*) = V^*(\gamma_0)$, it follows that

$$L(\mu^*,\boldsymbol{\alpha}^*) \leq L(\mu,\boldsymbol{\alpha}^*) \text{ for all } \mu \in \mathbb{P}_\delta(\mathbb{K}),$$

which gives the second inequality in (4.4). On other hand, since

$$V^*(\gamma_0) \leq f(\mu^*) + G(\mu^*)\cdot\boldsymbol{\alpha}^* \leq f(\mu^*) = V^*(\gamma_0),$$

we have $G(\mu^*)\cdot\boldsymbol{\alpha}^* = 0$. Therefore,

$$L(\mu^*,\boldsymbol{\alpha}) - L(\mu^*,\boldsymbol{\alpha}^*) = G(\mu^*)\cdot\boldsymbol{\alpha} - G(\mu^*)\cdot\boldsymbol{\alpha}^* = G(\mu^*)\cdot\boldsymbol{\alpha} \leq 0,$$

and the first inequality in (4.4) follows. □

8 Proof of Theorem 5.4

For completeness, we first state some well-known results that are needed to prove Theorem 5.4.

Lemma 8.1 *Let Y be a metric space and M a family of probability measures on Y. If there exists a nonnegative and inf-compact function v on Y such that*

$$\sup\{\langle\mu, v\rangle,\ \mu \in M\} < \infty,$$

then M is relatively compact, that is, for each sequence $\{\mu_n\}$ in M there is a probability measure μ on Y and a subsequence $\{\mu_m\}$ of $\{\mu_n\}$ such that μ_m converges weakly to μ in the sense that

$$\langle\mu_m, v\rangle \to \langle\mu, v\rangle \quad \forall v \in C_b(Y). \tag{8.1}$$

To prove Lemma 8.1, one first shows that the hypothesis implies that M is *tight*, and then the relative compactness of M follows from Prohorov's Theorem (see [16]).

Lemma 8.2 *Let Y a metric space, and $v : Y \to \mathbb{R}$ lower semicontinuous and bounded below. If $\{\mu_n\}$ and μ are probability measures on Y and μ_n converges weakly to μ (that is, as in (8.1)), then*

$$\liminf_{n\to\infty}\langle\mu_n, v\rangle \geq \langle\mu, v\rangle.$$

Lemma 8.2 is well known (and easy to prove): see, for instance, statement (12.3.37) in [16, p. 243]

Lemma 8.3 *The set $\mathbb{P}_\delta(\mathbb{K})$ is closed with respect to the topology of weak convergence.*

For a proof of Lemma 8.3 see Lemma 5.5 in [17], for instance.

Let

$$\Delta'_\delta := \{\mu \in \Delta_\delta \mid \mu \text{ is an optimal solution for (4.1)}\}.$$

Lemma 8.4 *Let $V^{\boldsymbol{\beta}}(\pi)$ and $\Gamma^*(\Pi)$ be as in (5.3) and (5.10), respectively. Let Π^* be set of policies π such that $V(\pi)$ in $\Gamma^*(\Pi)$. Then there exists a policy π^* such that*

$$V^{\boldsymbol{\beta}}(\pi^*) = \min_{\pi\in\Pi^*} V^{\boldsymbol{\beta}}(\pi). \tag{8.2}$$

Proof: It is clear that minimizing $V^{\boldsymbol{\beta}}(\cdot)$ on $\Gamma^*(\Pi)$ is equivalent to minimizing $\langle \cdot, C^{\boldsymbol{\beta}} \rangle$ on Δ'_δ, with $C^{\boldsymbol{\beta}}$ as in (5.2). Let $\rho^* := \inf\{\langle \mu, C^{\boldsymbol{\beta}} \rangle \mid \mu \in \Delta'_\delta\}$ and take a sequence $\{\mu_n\}$ in Δ'_δ such that

$$\langle \mu_n, C^{\boldsymbol{\beta}} \rangle \downarrow \rho^*.$$

Therefore, given $\epsilon > 0$, there exists an integer N such that

$$\rho^* \leq \langle \mu_n, C^{\boldsymbol{\beta}} \rangle \leq \rho^* + \epsilon \quad \forall n \geq N. \tag{8.3}$$

On the other hand, by definition of Δ'_δ, it follows that

$$\langle \mu_n, c \rangle = V^*(\gamma_0) \quad \text{for all} \quad n \geq 0 \tag{8.4}$$

with $V^*(\gamma_0)$ as in (2.7), which implies that

$$\sup_n \langle \mu_n, c \rangle = V^*(\gamma_0).$$

Since c is inf-compact (Assumption 3.1(b)), from Lemma 8.1 it follows that $\{\mu_n\}$ is relatively compact, that is, there exists a probability measure μ^* on $\mathbb{K}$ and a subsequence $\{\mu_m\}$ of $\{\mu_n\}$ that converges weakly to μ^*. The latter convergence, together with (8.3) and Lemma 8.2, yields that $\langle \mu^*, C^{\boldsymbol{\beta}} \rangle = \rho^*$. Finally, from Lemma 8.3 we conclude that μ^* is indeed a p.m. in Δ'_δ, and so the disintegration $\mu^* = \widehat{\mu^*} \cdot \varphi^*$ of μ^* is such that $\pi^* := \varphi^*$ satisfies (8.2). □

Proof of Theorem 5.4 From Lemma 8.4 and Theorem 5.2, it follows that $\mathrm{Par}(\Gamma^*(\Pi)) \neq \varnothing$. □

Raquiel R. López–Martínez
Facultad de Matemáticas
Universidad Veracruzana
A. P. 270
Xalapa, Ver., 91090
México
ralopez@uv.mx

Onésimo Hernández–Lerma
Departamento de Matemáticas
CINVESTAV–IPN
A. P. 14–740
México D.F., 07000
México
ohernand@math.cinvestav.mx

References

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

[2] Altman E., *Constrained Markov decision processes with total cost criteria: occupation measures and primal LP,* Math. Meth. Oper. Res., **43** (1996), 45-72.

[3] Altman E., *Constrained Markov decision processes with total cost criteria: Lagrange approach and dual LP,* Math. Meth. Oper. Res., **48** (1998), 387-417.

[4] Beutler F. J.; Ross K. W., *Optimal policies for controlled Markov chains with a constraint,* J. Math. Anal. Appl., **112** (1983), 236-252.

[5] Borkar V.S., *A convex analytic approach to Markov decision processes,* Prob. Theory Related Fields, **78** (1988), 583-602.

[6] Borkar V.S., *Ergodic control of Markov chains with constraints—the general case,* SIAM J. Control Optim., **32** (1994), 176-186.

[7] Craven B. D., Mathematical Programming and Control Theory, Chapman and Hall, London, 1978.

[8] Derman B. D.; Veinott A. F. Jr., *Constrained Markov decision chains,* Management Science, **19** (1972), 389-390.

[9] Dynkin E.B.; Yushkevich A. A., Controlled Markov Processes, Springer-Verlag, Berlin, 1979.

[10] Feinberg E.; Shwartz A., *Constrained discounted dynamic programming,* Math. Oper. Res., **21** (1996), 922-945.

[11] Feinberg E.; Shwartz A., *Constrained dynamic programming with two discount factors: applications and an algorithm,* IEEE Trans. Autom. Control, **44** (1999), 628-631.

[12] Golabi K.; Kulkarni R.B.; Way G.B., *A statewide pavement management system,* Interfaces, **12** (1982), 5-21.

[13] Hernández–Lerma O.; González-Hernández J., *Constrained Markov control processes in Borel spaces: the discounted case*, Math. Meth. Oper. Res., **52** (2000), 271-285.

[14] Hernández–Lerma O.; González-Hernández J.; López-Martínez R.R., *Constrained average cost Markov control processes in Borel spaces*, SIAM J. Control Optim., (to appear).

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

[16] Hernández–Lerma O.; Lasserre J.B., Further Topics on Discrete-Time Markov Control Processes, Springer-Verlag, New York, 1999.

[17] Hernández–Lerma O.; Romera R. , *Multiobjective Markov control processes: a linear programing approach*, preprint, Departamento de Matemáticas, CINVESTAV-IPN, México, 2003. (Submitted).

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

[19] Hordijk A.; Spieksma F., *Constrained admission control to a queueing system,* Adv. Appl. Prob., **21** (1989), 409-431.

[20] Huang Y.; Kurano M., *The LP approach in average rewards MDPs with multiple cost constraints: The countable state case,* J. Inform. Optim. Sci., **18** (1997), 33-47.

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

[22] Kallenberg L. C. M., *Survey of linear programming for standard and nonstandard Markovian control problems, Part I: Theory,* ZOR–Math. Methods Oper. Res., **40** (1994), 1-42.

[23] Kurano M., Nakagami J.; Y. Huang, *Constrained Markov decision processes with compact state and action spaces: the average case,* Optimization, **48** (2000), 255-269.

[24] A. Lazar, *Optimal flow control of a class of queueing networks in equilibrium,* IEEE Trans. Autom. Control, **28** (1983), 1001-1007.

[25] López–Martínez R.R., *A saddle–point theorem for constrained Markov control processes*, Morfismos, **3** (1999), 69–79. (Available at http://chucha.math.cinvestav.mx/morfismos/v3n2/index.html)

[26] Luenberger D. G., Optimization by Vector Space Methods, Wiley, New York, 1969.

[27] Mao X.; Piunovskiy A.B., *Strategic measure in optimal control problems for stochastic sequences,* Stoch. Anal. Appl., **18** (2000), 755-776.

[28] Piunovskiy A.B., Optimal Control of Random Sequences in Problems with Constraints, Kluwer, Boston, 1997.

[29] Piunovskiy A.B.; Mao X., *Constrained Markovian decision processes: the dynamic programming approach,* Oper. Res. Letters, **27** (2000), 119-126.

[30] Ross K.; Varadarajan R., *Multichain Markov decision processes with a sample path constraint: a decomposition approach,* Math. Oper. Res., **16** (1991), 195-207.

[31] Sennott L.I., *Constrained discounted Markov decision chains,* Prob. Eng. Inform. Sci., **5** (1991), 463-475.

[32] Sennott L.I., *Constrained average cost Markov decision chains,* Prob. Eng. Inform. Sci., **7** (1993), 69-83.

[33] Tanaka K., *On discounted dynamic programming with constraints,* J. Math. Anal. Appl., **155** (1991), 264-277.

[34] Vakil F.; Lazar A. A., *Flow control protocols for integrated networks with partially observed voice traffic,* IEEE Trans. Autom. Control, **32** (1987), 2-14.