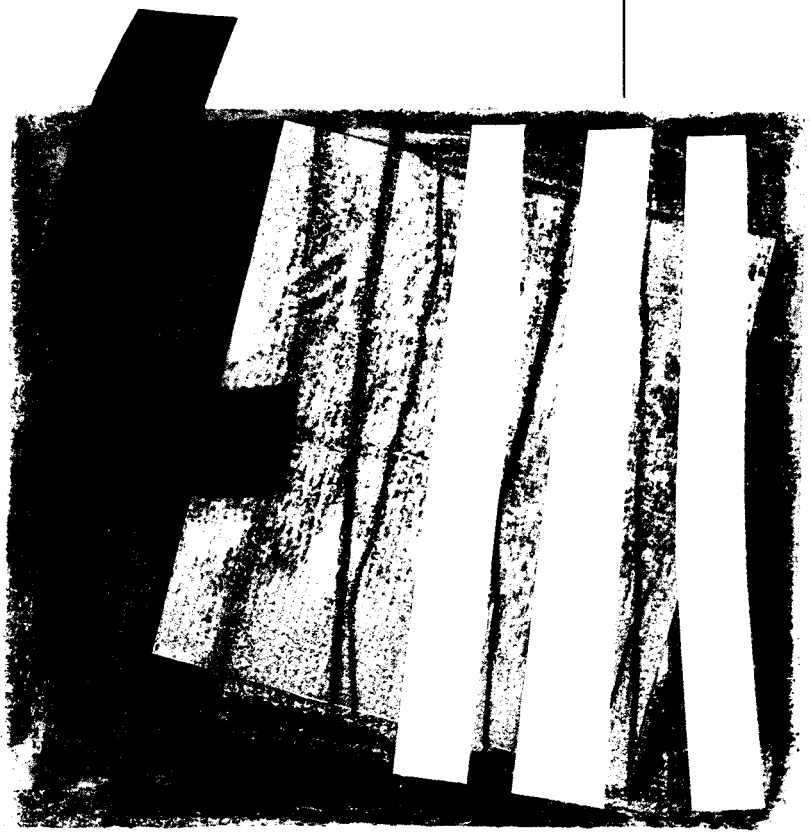
PARETO OPTIMALITY IN MULTIOBJECTIVE MARKOV CONTROL PROCESSES

Onésimo Hernández-Lerma
and Rosario Romera

00-28



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 00-28
Statistics and Econometrics Series (11)
April 2000

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 624-9849

PARETO OPTIMALITY IN MULTIOBJECTIVE MARKOV CONTROL PROCESSES.

Onesimo Hernandez-Lerma* and Rosario Romera**

Abstract

This paper studies discrete-time multiobjective Markov control processes (MCPs) on Borel spaces and with unbounded costs. Under mild assumptions, it shows the existence of Pareto optimal control policies, which are also characterized as optimal policies for a certain class of single-objective (or "scalar") MCPs. A similar result is obtained for *strong* Pareto optimal policies, which are Pareto optimal policies whose cost vector is the closest, in the Euclidean norm, to the virtual minimum. To obtain these results, the basic idea is to transform the multiobjective MCP into an equivalent *multiobjective measure problem* (MMP). In addition, MMP is restated as a primal multiobjective linear program and it is shown that solving the scalarized MCPs is in fact the same as solving the dual of MMP. A multiobjective LQ example illustrates the main results.

Keywords: Markov control processes; multiobjective control; Pareto optimality.

*Corresponding author. Departamento de Matemáticas, CINVESTAV-IPN, E-mail: ohernand@math.cinvestav.mx;** Universidad Carlos III de Madrid, Departamento de Estadística y Econometría, E-mail: mrromera@est-econ. AMS subject classification 2000: 93E20,90C40,90C27.

1 Introduction

In a standard optimal control problem there is a controller that wishes to optimize a *single* objective function. Thus, for instance, in a production control problem it is tacitly assumed that the given objective function somehow aggregates several different costs (manufacturing costs, holding costs, distribution costs, etc.) and possibly several income sources (for example, sales, investments, and so on). However, there are situations in which it is convenient, or perhaps even necessary, to optimize separately these functions and the controller is then led to consider a *multiobjective* problem of the form (say): "minimize" the cost vector

$$V(\pi) := (V_1(\pi), \ldots, V_q(\pi)) \in \mathbb{R}^q$$

over the class of all admissible policies π (see Section 2 for details.) In particular, if π^* minimizes $V(\pi)$ in the sense of Pareto, then π^* is said to be *Pareto optimal*, or simply a *Pareto* policy. On the other hand, letting

$$V_i^* := \inf_{\pi} V_i(\pi) \text{ for } i = 1, \ldots, q,$$

and defining the *virtual minimum* $V^* := (V_1^*, \ldots, V_q^*)$, an important issue is to find *strong* Pareto policies, namely, Pareto policies π^* whose cost vector $V(\pi^*)$ is the "closest" (e.g. in the usual Euclidean norm) to V^*.

In this paper we study discrete-time multiobjective Markov control processes (MCPs) on *Borel spaces* and with *unbounded costs*. The main problems we are concerned with are the existence and characterization of both Pareto and strong Pareto control policies. To do this the key idea is to use occupation measures to transform the multiobjective MCP into an equivalent *multiobjective measure problem* (MMP) on a suitable linear space of measures. This greatly simplifies the original problem and it also has important consequences. First, it gives, in our general setting, the usual characterization of Pareto optimal policies via the "scalarization" approach. Thus, the multiobjective MCP can be reduced to single-objective (or scalar) MCPs with a "weighted" objective function of the form

$$\lambda \cdot V(\pi) := \lambda_1 V_1(\pi) + \cdots + \lambda_q V_q(\pi)$$

for some vectors λ in the nonnegative orthant $\mathbb{R}_+^q$. Second, restating the MMP as a primal *multiobjective linear program* it is shown that the scalarization approach is in fact the same as solving the corresponding *dual* linear program. Third, using

the MMP it is trivially deduced that the *performance set*, that is, the set of all cost vectors $V(\pi)$, is *convex*. Therefore, the MMP essentially reduces the question of existence of strong Pareto policies to the problem of finding the distance from the virtual minimum V^* to a convex set, which is a standard optimization problem (see [4] or [21], for instance).

The existence and characterization of Pareto minima (also known as *efficient points*) are standard topics in multiobjective optimization (see [3-5, 23] and their references). In control theory, however, these topics are not as well developed and all of the literature is restricted to some classes of MCPs, for example, with a countable state space [8-11, 17-19, 26-30] or in Borel spaces but with bounded costs [20, 24, 25]. On the other hand, some papers [10, 17, 18, 20, 24, 25] deal with a vector-minimization problem more general than ours, in the sense that, instead of $\mathbb{R}^q_+$, they work with the partial order induced by an arbitrary pointed convex cone K in $\mathbb{R}^q$. But it turns out that they restrict the control problem to some *subclass* of policies (for instance, deterministic stationary) and, moreover, in the case of Borel state spaces, they *assume* that the performance set is convex. Here, we work with the set of all policies and, as already noted, the convexity of the performance set is an straightforward consequence of the MMP. At any rate, extending our results to a general pointed convex cone K seems to be a purely notational problem.

It is worth noting that, in addition to the scalarization approach (*cum* MMP) used here, there are other methods to study multiobjective MCPs. For example, there are multiobjective versions of value iteration [17, 18, 30] and of policy iteration [10, 27, 28]. Still another method, introduced in [29], is to transfer the multiobjective MCP into a MCP with partial state observations. All of these methods, however, have been studied only for problems with a countable state space, and, as they are computationally appealing, it would be interesting to see if they can be extended to more general spaces.

The remainder of the paper is organized as follows. In Section 2 we introduce the multiobjective MCP we are concerned with, as well as the precise notion of Pareto optimality. We consider a vector of discounted cost criteria but in Section 8 we briefly explain, among other things, how our results can be translated to average costs. In Section 3 we state our hypotheses (Assumption 3.1) and the so-called "theorem of equivalence" in Pareto optimality [3]. In fact, we state this theorem in two parts, Theorem 3.2(a) and (the converse) Theorem 3.3, because the proof of the latter requires

the MPP, which is not introduced until Section 4. On the other hand, Theorem 3.2(a) is the easy part of the "theorem of equivalence" and it directly yields the existence of Pareto optimal policies. Section 3 also includes Example 3.4 on the multiobjective LQ (Linear system with Quadratic costs) MCP in which explicit Pareto optimal policies can be calculated. In Section 5 we introduce the virtual minimum V^* for our multiobjective MCP, and show the existence of *strong Pareto* policies. We also extend a result of Tanaka [24] that can be very useful to compute strong Pareto policies; see Theorem 5.2(b). This fact is illustrated in Example 5.7, which is a continuation of the LQ Example 3.4. Section 6 presents the *multiobjective Linear Programming* (LP) formulation of the multiobjective MCP. The idea (as for scalar and constrained MCPs [1, 13-16]) is to introduce suitable dual pairs of vector spaces in which the MPP (4.7) can be formulated as a multiobjective linear program. The multiobjective LP formulation is borrowed from Balbás and Heras [5]. Section 7 contains the proof of Theorem 3.3, and, finally, in Section 8 we briefly mention some connections between our multiobjective MCP and constrained MCPs, multiobjective problems with average cost criteria, and multiobjective problems with "mixed" average and discounted criteria.

Remark 1.1. *(***Notation***.) If S is a Borel space (that is, a Borel subset of a complete and separable metric space), we denote its Borel σ-algebra by $\mathcal{B}(S)$. If S and T are Borel spaces, then a stochastic kernel on S given T is a function $(t, B) \mapsto q(B|t)$ from $T \times \mathcal{B}(S)$ to the interval $[0, 1]$ such that $q(B|\cdot)$ is a measurable function on T for each fixed $B \in \mathcal{B}(S)$, and $q(\cdot\,|t)$ is a probability measure on $\mathcal{B}(S)$ for each fixed t.*

2 Multiobjective MCPs

The material in this section is quite standard —see, for instance [1, 7, 15, 16, 22] for additional details, if necessary.

The *multiobjective Markov control model* can be represented as

$$(\mathrm{X}, A, \mathbb{K}, Q, (c_1, \ldots, c_q), \delta, \gamma_0), \tag{2.1}$$

where X and A are Borel spaces that stand for the *state space* and the *control* (or *action*) *set*, respectively. We also have the *constraint set* $\mathbb{K}$, a Borel subset of $\mathrm{X} \times A$, and which is assumed to contain the graph of a measurable map from X to A (this

ensures that the set $\mathbb{F}$ in Definition 2.1, below, is nonempty). For each $x \in \mathrm{X}$, the x-section in $\mathbb{K}$, namely

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

is a (nonempty) Borel subset of A whose elements are the admissible control actions in the state x. The *transition law* Q is a stochastic kernel on X given $\mathbb{K}$, whereas

$$c := (c_1, \ldots, c_q) : \mathbb{K} \to \mathbb{R}^q \tag{2.2}$$

is a vector function whose components are used to define the different cost criteria. Finally, $\delta \in (0, 1)$ is a given *discount factor*, and γ_0 is the *initial distribution*, a probability measure on X.

If $q = 1$, then (2.1) will be referred to as a "scalar" (or "standard") Markov control model.

Definition 2.1. *Φ denotes the family of stochastic kernels φ on A given* X *that satisfy the constraint $\varphi(A(x)|x) = 1$ for all $x \in \mathrm{X}$, and $\mathbb{F}$ stands for the class of measurable functions f from* X *to A such that $f(x) \in A(x)$ for all $x \in \mathrm{X}$.*

Let $H_0 := \mathrm{X}$, and $H_n := \mathbb{K}^n \times \mathrm{X}$ for $n = 1, 2, \ldots$. A *control policy* is a sequence $\pi = \{\pi_n\}$ of stochastic kernels π_n on A given H_n that satisfy the condition

$$\pi_n(A(x_n)|h_n) = 1 \tag{2.3}$$

for each "history" $h_n = (x_0, a_0, \ldots, x_{n-1}, a_{n-1}, x_n)$ in H_n and $n = 0, 1, \ldots$. We denote by Π the set of policies. A control policy $\pi = \{\pi_n\}$ is said to be *randomized stationary* if there exists $\varphi \in \Phi$ such that $\pi_n(\cdot\,|h_n) = \varphi(\cdot\,|x_n)$ for every history $h_n \in H_n$ and $n = 0, 1, \ldots$. The set of such policies will be identified with the family Φ in Definition 2.1. On the other hand, $\pi = \{\pi_n\}$ is called *deterministic stationary* if there exists $f \in \mathbb{F}$ such that $\pi_n(\cdot\,|h_n)$ is the Dirac measure concentrated at $f(x_n)$ for all $h_n \in H_n$ and $n = 0, 1, \ldots$. We shall identify $\mathbb{F}$ with the collection of deterministic stationary policies.

The multiobjective MCP. Consider the control model (2.1), and let $(\Omega, \mathcal{F})$ be the (canonical) measurable space consisting of the sample space $\Omega := (\mathrm{X} \times A)^\infty$, and the corresponding product σ-algebra $\mathcal{F}$. Then, for each policy $\pi \in \Pi$, there is a probability measure $P^\pi_{\gamma_0}$ and a stochastic process $\{(x_t, a_t), t = 0, 1, \ldots\}$ defined on Ω in a canonical way, where x_t and a_t represent the state and the control variables

at the time t ($t = 0, 1, \dots$) when using the policy π. The expectation operator with respect to $P^{\pi}_{\gamma_0}$ is denoted by $E^{\pi}_{\gamma_0}$.

For each $i = 1, \dots, q$ and $\pi \in \Pi$, consider the δ-discounted cost

$$V_i(\pi, \gamma_0) := (1 - \delta) E^{\pi}_{\gamma_0} \Big[\sum_{t=0}^{\infty} \delta^t c_i(x_t, a_t) \Big]. \tag{2.4}$$

and let $V(\pi, \gamma_0) \in \mathbb{R}^q$ be the cost vector

$$V(\pi, \gamma_0) := (V_1(\pi, \gamma_0), \dots, V_q(\pi, \gamma_0)). \tag{2.5}$$

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 a precise form we first introduce some notation and terminology.

We consider $\mathbb{R}^q$ with the usual partial order; that is, for q-vectors u and v, the inequality $u \leq v$ means that $u_i \leq v_i$ for all $i = 1, \dots, q$. We also have

$$u < v \Leftrightarrow u \leq v \text{ and } u \neq v;$$

$$u \ll v \Leftrightarrow u_i < v_i \text{ for all } i = 1, \dots, q.$$

Definition 2.2. *Let Γ be a subset of $\mathbb{R}^q$. A vector $u^* \in \Gamma$ is said to be Pareto optimal (or an efficient point) on Γ if there is no $u \in \Gamma$ such that $u < u^*$. The set of Pareto optimal vectors on Γ is called the Pareto set of Γ.*

These concepts can be extended to multiobjective MCPs as follows.

Definition 2.3. *Let $\Gamma(\Pi) \subset \mathbb{R}^q$ be the set of cost vectors in (2.5), i.e.,*

$$\Gamma(\Pi) := \{V(\pi, \gamma_0) | \pi \in \Pi\}. \tag{2.6}$$

Then a policy π^* is said to be Pareto optimal (or a Pareto policy) if its corresponding cost vector $V(\pi^*, \gamma_0)$ is in the Pareto set of $\Gamma(\Pi)$.

In other words, π^* is a Pareto policy if there is no other policy π such that $V(\pi, \gamma_0) < V(\pi^*, \gamma_0)$. The family of Pareto optimal policies is denoted by $\mathrm{Par}(\Pi)$. In the following section we give conditions that, in particular, ensure that $\mathrm{Par}(\Pi)$ is nonempty.

Remark 2.4. *As the initial distribution γ_0 is fixed, to simplify the notation we shall drop γ_0 from expressions such as (2.4)-(2.6). Thus, for instance, we shall write $V(\pi, \gamma_0)$ simply as $V(\pi)$. However, it is important to keep in mind that, in general, Pareto optimal policies depend on the initial distribution.*

Remark 2.5. *Usually, the definition (2.4) of the δ-discounted cost does not include the factor $(1-\delta)$. However, as in [1], the present definition is convenient because it is easily related to the average cost criterion; see (8.2) and (8.3).*

3 Pareto optimal policies

To study the existence and characterization of Pareto policies, *in the remainder of the paper we impose the following assumption.*

Assumption 3.1. *The multiobjective Markov control model (2.1) satisfies that:*

(a) *The constraint set $\mathbb{K} \subset \mathrm{X} \times A$ is closed.*

(b) *The functions c_i are nonnegative and lower semicontinuous and, moreover, at least one of them, say c_1, is inf-compact, which means that for each $r \in \mathbb{R}$, the level set*

$$K_r := \{(x,a) \in \mathbb{K} | c_1(x,a) \leq r\} \tag{3.1}$$

is compact.

(c) *The transition law Q is weakly continuous; that is, denoting by $C_b(S)$ the space of continuous bounded functions on a topological space S, the map*

$$(x,a) \mapsto \int_{\mathrm{X}} h(y) Q(dy|x,a) \text{ is in } C_b(\mathbb{K}) \text{ for each } h \in C_b(\mathrm{X}). \tag{3.2}$$

(d) *There exists a policy $\pi \in \Pi$ such that $V_i(\pi) < \infty$ for all $i = 1, \ldots, q$. (Recall that $V_i(\pi, \gamma_0) = V_i(\pi)$; see Remark 2.4.)*

Observe that Assumption 3.1 is not restrictive at all. In fact, it holds in most applications to queueing systems, productions models, etc. In particular, Assumption 3.1(c) holds if the state process $\{x_t\}$ evolves according to a discrete-time equation of the form

$$x_{t+1} = G(x_t, a_t, \xi_t), \ t = 0, 1, \ldots,$$

where the $\{\xi_t\}$ are i.i.d. disturbances independent of the initial state x_0, and $G(x,a,s)$ is a given measurable function, continuous in $(x,a) \in \mathbb{K}$ for each s. This class of systems includes the LQ problem in Examples 3.4 and 5.7.

The existence problem. Let $\mathbb{R}^q_{++}$ be the set of strictly positive q-vectors (that is, $\lambda \gg 0$). To study the existence of Pareto policies we shall first follow the well-known "scalarization" approach. Thus, given $\lambda \in \mathbb{R}^q_{++}$ we consider the scalar (or real-valued) cost-per-stage function

$$c^\lambda(x,a) := \lambda \cdot c(x,a) = \sum_{i=1}^{q} \lambda_i c_i(x,a), \tag{3.3}$$

and, as in (2.4), we consider a δ-discounted cost $V^\lambda(\pi) \equiv V^\lambda(\pi, \gamma_0)$, with

$$V^\lambda(\pi) := (1-\delta) E^\pi_{\gamma_0}\Big[\sum_{t=0}^{\infty} \delta^t c^\lambda(x_t, a_t)\Big]. \tag{3.4}$$

Using (3.3) and (2.5) we may write $V^\lambda(\pi)$ as

$$V^\lambda(\pi) = \lambda \cdot V(\pi) = \sum_{i=1}^{q} \lambda_i V_i(\pi). \tag{3.5}$$

It is clear that minimizing $V^\lambda(\cdot)$ over Π is equivalent to minimize $V^\lambda(\cdot)$ multiplied by a positive constant. Hence, occasionally we shall assume that the vector λ in (3.3)-(3.5) belongs to the set

$$\Lambda := \{\lambda \in \mathbb{R}^q_{++} \,|\, \sum_{i=1}^{q} \lambda_i = 1\}. \tag{3.6}$$

We may then state an existence result as follows.

Theorem 3.2. *Choose an arbitrary vector $\lambda \in \Lambda$.*

(a) *If $\pi^* \in \Pi$ is an optimal policy for the scalar criterion (3.4), that is,*

$$V^\lambda(\pi^*) \le V^\lambda(\pi) \quad \forall \pi \in \Pi. \tag{3.7}$$

then π^* is Pareto optimal.

(b) *There exists a deterministic stationary Pareto policy f_λ; that is, f_λ is in* $\mathrm{Par}(\Pi) \cap \mathbb{F}$.

Proof. (a) Suppose that π^* satisfies (3.7) but is *not* Pareto optimal. It follows that $V(\pi) < V(\pi^*)$ for some policy π, which in turn gives that $V^\lambda(\pi) < V^\lambda(\pi^*)$. This contradicts (3.7) and so (a) follows.

(b) By (3.3) and Assumption 3.1(b), the function c^λ is nonnegative and inf-compact. This fact, together with Assumption 3.1(a), (c), (d), implies the existence of a policy $\pi^* \in \mathbb{F}$ that satisfies (3.7); see for instance [12] or Theorem 4.2.3 in [15]. Hence, by part (a), the policy $f_\lambda := \pi^*$ satisfies (b). ∎

The converse of Theorem 3.2(a) is sometimes called the "theorem of equivalence" in Pareto optimality [3]. As our proof of that converse requires a special formulation (introduced in Section 4) we next state the result but its proof is postponed until Section 7.

Theorem 3.3. *(The "theorem of equivalence".) Let Π_0 be the family of policies that satisfy Assumption 3.1(d). If $\pi^* \in \Pi_0$ is Pareto optimal, then there exists a vector $\lambda^* \in \Lambda$ such that*

$$V^{\lambda^*}(\pi^*) \leq V^{\lambda^*}(\pi) \quad \forall \pi \in \Pi. \tag{3.8}$$

Proof. See Section 7.

The following example illustrates Theorem 3.2.

Example 3.4. Let α and β be nonzero real numbers and consider the scalar linear system

$$x_{t+1} = \alpha x_t + \beta a_t + \xi_t \text{ for } t = 0, 1, \ldots, \tag{3.9}$$

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

$$E(\xi_0) = 0 \quad \text{and} \quad E(\xi_0^2) =: \sigma^2 < \infty. \tag{3.10}$$

For $i = 1, \ldots, q$, let s_i and r_i be strictly positive numbers, and let $c_i(x, a)$ be the quadratic cost

$$c_i(x, a) := s_i x^2 + r_i a^2. \tag{3.11}$$

Then, for each vector $\lambda \in \mathbb{R}^q_{++}$, the scalar problem (3.3)-(3.5) corresponds to the linear system (3.9) with quadratic cost

$$c^\lambda(x,a) = (\lambda \cdot s)x^2 + (\lambda \cdot r)a^2 \tag{3.12}$$

with $s := (s_1, \ldots, s_q)$ and $r := (r_1, \ldots, r_q)$. Moreover, for each $i = 1, \ldots, q$, let z_i be the unique positive solution of the Riccati equation

$$\delta\beta^2 z^2 + (r_i - r_i\alpha^2\delta - s_i\beta^2\delta)z - s_i r_i = 0. \tag{3.13}$$

Replace s_i and r_i with the coefficients $\lambda \cdot s$ and $\lambda \cdot r$ in (3.12), respectively, and let $z(\lambda)$ be the corresponding unique positive solution of (3.13). Then, as is well-known (see, for instance, p. 72 in [15]), the optimal control policy $f_\lambda \in \mathbb{F}$ for the scalar problem is

$$f_\lambda(x) = -\big[\lambda \cdot r + \delta\beta^2 z(\lambda)\big]^{-1}\alpha\beta\delta z(\lambda)x \quad \forall x \in \mathrm{X}, \tag{3.14}$$

and, moreover, for each initial state $x_0 = x$, the optimal cost function is

$$V^\lambda(f_\lambda, x) = z(\lambda)\big[(1-\delta)x^2 + \delta\sigma^2\big] \quad \forall x \in \mathrm{X}, \tag{3.15}$$

with σ^2 as in (3.10). Therefore, assuming that the initial distribution γ_0 satisfies that

$$\overline{\gamma}_0 := \int x^2 \gamma_0(dx) < \infty, \tag{3.16}$$

the optimal cost $V^\lambda(f_\lambda) \equiv V^\lambda(f_\lambda, \gamma_0)$ in the left-hand side of (3.7) is obtained by integrating both sides of (3.15) with respect to γ_0. This yields

$$V^\lambda(f_\lambda) = z(\lambda)\big[(1-\delta)\overline{\gamma}_0 + \delta\sigma^2\big]. \tag{3.17}$$

4 A multiobjective measure problem

In this section we reformulate the multiobjective MCP as an equivalent *multiobjective measure problem* (MMP) on a suitable vector space of measures. this reformulation greatly simplifies the proofs of some results and, in addition, it can be used to write the multiobjective MCP as a *multiobjective linear program* (see Section 6).

Occupation measures. For each policy $\pi \in \Pi$, let $\mu^\pi \equiv \mu^\pi_{\gamma_0}$ be the corresponding δ-discount expected *occupation measure*, which is defined as

$$\mu^\pi(D) := (1-\delta)\sum_{t=0}^{\infty} \delta^t P^\pi_{\gamma_0}\big[(x_t, a_t) \in D\big] \quad \forall D \in \mathcal{B}(\mathrm{X} \times A). \tag{4.1}$$

This is a probability measure on $\mathrm{X} \times A$ that, by (2.3), is concentrated on $\mathbb{K}$. Moreover, if π is in Π_0, where Π_0 is as in Theorem 3.3, then a standard argument (see, for instance, Remark 9.4.2(b) in [16, p. 85]) yields that $V_i(\pi)$ in (2.4) can be written as

$$V_i(\pi) = \langle \mu^\pi, c_i \rangle := \int_{\mathbb{K}} c_i d\mu^\pi \quad (i = 1, \ldots, q). \tag{4.2}$$

To state other properties of occupation measures we shall use the following notation: if μ is a finite signed measure on $\mathrm{X} \times A$, we denote its *variation* by $|\mu| = \mu^+ + \mu^-$, and its *marginal* (or projection) on X by $\widehat{\mu}$, that is,

$$\widehat{\mu}(B) := \mu(B \times A) \quad \forall B \in \mathcal{B}(\mathrm{X}).$$

We also introduce the following sets of measures.

Definition 4.1. *$M(\mathbb{K})$ denotes the vector space of finite signed measures on $\mathrm{X} \times A$, concentrated on $\mathbb{K}$, and such that*

$$\langle |\mu|, c_i \rangle = \int c_i d|\mu| < \infty \quad \forall i = 1, \ldots, q. \tag{4.3}$$

Further, $M_+(\mathbb{K}) \subset M(\mathbb{K})$ stands for the convex cone of nonnegative measures in $M(\mathbb{K})$, and $M_\delta(\mathbb{K}) \subset M_+(\mathbb{K})$ is the subfamily of nonnegative measures for which

$$\widehat{\mu}(B) = (1-\delta)\gamma_0(B) + \delta \int_{\mathbb{K}} Q(B|x,a)\mu(d(x,a)) \ \ \forall B \in \mathcal{B}(\mathrm{X}). \tag{4.4}$$

As $\widehat{\mu}(\mathrm{X}) = \mu(\mathrm{X} \times A)$, it is evident from (4.4) that

$$M_\delta(\mathbb{K}) \text{ is a convex set of probability measures.} \tag{4.5}$$

It also turns out that $M_\delta(\mathbb{K})$ coincides with the family of occupation measures in (4.1). More precisely (as in [12, pp. 386-387] or [15, Theorem 6.3.7], for instance), we have:

Lemma 4.2. *If π is a policy in Π_0, then its occupation measure μ^π is in $M_\delta(\mathbb{K})$. Conversely, if μ is in $M_\delta(\mathbb{K})$, then μ is the occupation measure of a policy in Π_0 (that is, there exists $\pi \in \Pi_0$ such that $\mu^\pi = \mu$).*

For $\mu \in M_\delta(\mathbb{K})$ and c as in (2.2), let

$$\langle \mu, c \rangle := (\langle \mu, c_1 \rangle, \ldots, \langle \mu, c_q \rangle). \tag{4.6}$$

Then by (4.2) and Lemma 4.2 our multiobjective MCP can be expressed, equivalently, as the following *multiobjective measure problem* (MMP):

$$\text{minimize } \{\langle \mu, c \rangle | \mu \in M_\delta(\mathbb{K})\}. \tag{4.7}$$

This is indeed the case because, by Assumption 3.1(d), in our original multiobjective MCP we may restrict ourselves to the set

$$\Gamma(\Pi_0) := \{V(\pi) | \pi \in \Pi_0\} \tag{4.8}$$

in lieu of the set $\Gamma(\Pi)$ in (2.6).

In the following section we use the MMP (4.7) to show the existence of "strong" Pareto policies.

5 Strong Pareto optimality

For each $i = 1, \ldots, q$, let $V_i^* \equiv V_i^*(\gamma_0)$ be the optimal δ-discounted cost of the scalar MCP with cost-per-stage $c_i(x, a)$, that is,

$$V_i^* := \inf_\pi V_i(\pi) \quad (\text{with } V_i(\pi) \text{ as in}(2.4)).$$

The q-vector $V^* := (V_1^*, \ldots, V_q^*)$ is called the *virtual minimum* for the multiobjective MCP. (V^* is also known as the *utopian* or the *ideal* or the *shadow minimum.*) Let $\|\cdot\|$ be the Euclidean norm in $\mathbb{R}^q$, and let $\rho : \Pi_0 \to \mathbb{R}_+$ be the map defined as

$$\rho(\pi) := \|V(\pi) - V^*\| \quad \text{for} \quad \pi \in \Pi_0. \tag{5.1}$$

This is a *utility function* for the multiobjective MCP in the sense that if π and π' are such that $V(\pi) < V(\pi')$, then $\rho(\pi) < \rho(\pi')$. (In (5.1) we took the Euclidean norm to fix ideas, but in fact we may take *any norm* in $\mathbb{R}^q$. See Remark 5.6(a).)

Definition 5.1. *A policy $\pi^* \in \Pi_0$ is said to be strong Pareto optimal (or a strong Pareto policy) if it minimizes the function ρ, that is,*

$$\rho(\pi^*) = \inf\{\rho(\pi) | \pi \in \Pi_0\} =: \rho^*. \tag{5.2}$$

As ρ is a utility function, it is clear that a strong Pareto optimal policy is Pareto optimal, but of course the converse is not true.

Let $\Gamma(\Pi_0)$ be as in (4.8). For each $\lambda \in \mathbb{R}^q$, let

$$\Delta(\lambda) := \inf\{\lambda \cdot (V(\pi) - V^*) | \pi \in \Pi_0\} \tag{5.3}$$

be the so-called *support function* of $\Gamma(\Pi_0) - V^*$ at λ. Moreover, let $S \subset \mathbb{R}^q$ be the closed unit sphere centered at the origin, and let S_1 be its boundary, i.e.,

$$S := \{\lambda \mid \|\lambda\| \leq 1\} \text{ and } S_1 := \{\lambda \mid \|\lambda\| = 1\}$$

Theorem 5.2. *Suppose that $\rho^* > 0$. Then:*

(a) *There exists a strong Pareto policy;*

(b) *There exists a vector $\lambda^* \in S_1 \cap \mathbb{R}^q_{++}$ such that*

$$\rho^* = \Delta(\lambda^*) = \max_{\lambda \in S} \Delta(\lambda) \tag{5.4}$$

and, moreover, for any strong Pareto policy π^*, the vector λ^* is "aligned" with $V(\pi^*) - V^*$, i.e.,

$$\lambda^* \cdot (V(\pi^*) - V^*) = \|\lambda^*\| \, \|V(\pi^*) - V^*\| = \rho^*. \tag{5.5}$$

For completeness and ease of reference, before proving Theorem 5.2 we state some well-known technical facts.

Lemma 5.3. *Let Y be a metric space and M a family of probability measures on Y.*

(a) *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 tight, that is, for each $\epsilon > 0$ there exists a compact set $K \subset Y$ for which

$$\mu(K) \geq 1 - \epsilon \quad \forall \mu \in M.$$

(b) *If M is tight, 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*

$$\langle \mu_m, u \rangle \to \langle \mu, u \rangle \quad \forall u \in C_b(Y). \tag{5.6}$$

Part (a) in Lemma 5.3 follows directly from the definition of inf-compactness (see Assumption 3.1(b)) and the definition of tightness. On the other hand, (b) is (a part of) *Prohorov's Theorem* —see [6], for instance.

Lemma 5.4. *Let Y be a metric space, and $v : Y \to \mathbb{R}$ lower semicontinuous and bounded below. If μ_m and μ are probability measures on Y and μ_m converges weakly to μ (that is, as in (5.6)), then*

$$\liminf_{m\to\infty} \langle \mu_m, v \rangle \geq \langle \mu, v \rangle. \tag{5.7}$$

Lemma 5.4 is well known (and easy to prove): see, for instance, statement (12.3.37) in [16, p. 225].

Lemma 5.5. *The set $M_\delta(\mathbb{K})$ (in Definition 4.1) is closed with respect to the topology of weak convergence.*

Proof. Let $\{\mu_m\}$ be a sequence in $M_\delta(\mathbb{K})$ such that μ_m converges weakly to μ. Choose an arbitrary function h in $C_b(\mathrm{X})$. By (3.2), $\int h(y)Q(dy|\cdot\,)$ is in $C_b(\mathbb{K})$, and, therefore, by the weak convergence of μ_m to μ, we get

$$\int\int h(y)Q(dy|x,a)\mu_m(d(x,a)) \to \int\int h(y)Q(dy|x,a)\mu(d(x,a)).$$

Similarly, the marginals $\widehat{\mu}_m$ converge weakly to the marginal $\widehat{\mu}$. Hence, as each μ_m satisfies (4.4), so does the limiting probability measure μ. Thus, to complete the proof that μ is in $M_\delta(\mathbb{K})$, it only remains to show that (4.3) holds for μ. This, however, follows from Assumption 3.1(b) and Lemma 5.4, which together yield

$$\liminf_{m\to\infty} \langle \mu_m, c_i \rangle \geq \langle \mu, c_i \rangle \quad \forall i = 1, \ldots, q.$$

This implies that μ satisfies (4.3). ∎

We are finally ready for the proof of Theorem 5.2.

Proof of Theorem 5.2. (a) To simplify the proof we may assume that $V^* = 0$, and then the general case is obtained by translation. Thus, instead of (5.1) we now have

$$\rho(\pi) = \|V(\pi)\| \quad \text{for} \quad \pi \in \Pi_0.$$

Moreover, by Lemma 4.2 and using (4.6), we may express ρ^* in (5.2) as

$$\rho^* = \inf\{\|\langle \mu, c\rangle\| \,|\, \mu \in M_\delta(\mathbb{K})\}.$$

Now let $\{\mu_n\}$ be a sequence in $M_\delta(\mathbb{K})$ such that, as $n \to \infty$,

$$\|\langle \mu_n, c\rangle\| \downarrow \rho^*. \tag{5.8}$$

Choose an arbitrary $\epsilon > 0$ and let $n(\epsilon)$ be such that

$$\rho^* \le \|\langle \mu_n, c\rangle\| \le \rho^* + \epsilon \quad \forall n \ge n(\epsilon).$$

This implies the existence of a constant k such that $\langle \mu_n, c_i\rangle \le k$ for all $n \ge n(\epsilon)$ and $i = 1, \ldots, q$. In particular,

$$\langle \mu_n, c_1\rangle \le k \quad \forall n \ge n(\epsilon). \tag{5.9}$$

Thus, as c_1 is inf-compact (Assumption 3.1(b)), (5.9) and Lemma 5.3 imply the existence of a subsequence $\{\mu_m\}$ of $\{\mu_n\}$ and a probability measure μ^* on $\mathrm{X} \times A$, concentrated on $\mathbb{K}$ (by Assumption 3.1(a)), such that μ_m converges weakly to μ^*. By Lemma 5.5, μ^* is in $M_\delta(\mathbb{K})$, and, by (5.7) and (5.8),

$$\|\langle \mu^*, c\rangle\| = \rho^*. \tag{5.10}$$

Finally, let $\pi^* \in \Pi_0$ be the policy associated to μ^*, and use (4.2) to rewrite (5.10) as $\|V(\pi^*)\| = \rho^*$. This completes the proof of part (a).

(b) If $\pi^* \in \Pi_0$ is strong Pareto optimal, then the support function in (5.3) becomes

$$\Delta(\lambda) = \lambda \cdot (V(\pi^*) - V^*),$$

and the vector $\lambda^* := (V(\pi^*) - V^*)/\|V(\pi^*) - V^*\|$ satisfies (5.4) and (5.5). ∎

Remark 5.6. *Part (b) in Theorem 5.2 can be obtained in other ways. For instance, let us write the performance set* $\Gamma(\Pi_0)$ *in (4.8) as*

$$\Gamma(\Pi_0) = \{\langle \mu, c\rangle | \mu \in M_\delta(\mathbb{K})\}, \tag{5.11}$$

Hence, by (4.5),

$$\Gamma(\Pi_0) \text{ is a convex subset of } \mathbb{R}^q_+, \tag{5.12}$$

and so the problem of finding a strong Pareto policy reduces to the problem of finding the distance from the virtual minimum V^* *to the convex set* $\Gamma(\Pi_0)$. *Therefore, Theorem 5.2(b) turns out to be a special case of the "Minimum Norm Duality" in Luenberger [21, p. 136, Theorem 1]. This result from [21] is true for an arbitrary normed linear space (not necessarily* $\mathbb{R}^q$*). Hence, in (5.1) we may take any norm instead of the Euclidean one.*

(b) Alternatively, by the Minimax Theorem (see, for instance, [1, p.129] or [4, p. 126]), there exists a vector λ^* *in* $S_1 \cap \mathbb{R}^q_{++}$ *such that*

$$\max_{\lambda \in S} \min_{\pi \in \Pi_0} \lambda \cdot (V(\pi) - V^*) = \min_{\pi \in \Pi_0} \max_{\lambda \in S} \lambda \cdot (V(\pi) - V^*) \tag{5.13}$$

$$= \min_{\pi \in \Pi_0} \lambda^* \cdot (V(\pi) - V^*). \tag{5.14}$$

Finally, as in the proof of Theorem 3.2(b), we can use standard dynamic programming results to obtain $\pi^* \in \Pi_0$ *that attains the minimum in (5.14), and, therefore, satisfies (5.5). This approach would in fact give a different proof of Theorem 5.2, but of course to use the minimax result (5.13) we still need to verify conditions such as (5.12). We chose the slightly longer approach via Lemmas 5.3, 5.4 and 5.5 because these results are also needed below.*

Observe that (5.13) gives a "curious" interpretation of Theorem 5.2(b) as a zero-sum game: the controller (or "player") wishes to minimize $g(\lambda, \pi) := \lambda \cdot (V(\pi) - V^*)$ *over* $\pi \in \Pi_0$, *whereas "nature" tries to maximize* $g(\lambda, \pi)$ *over* $\lambda \in S$.

Example 5.7. (Example 3.4 continued). Consider again the LQ problem (3.9)-(3.11). For each $i = 1, \ldots, q$, let $V_i^* = V_i^*(\gamma_0)$ be the corresponding optimal δ-discounted cost; that is (as in (3.17)), V_i^* is given by

$$V_i^* = k(\gamma_0) z_i, \text{ with } k(\gamma_0) := (1-\delta)\overline{\gamma}_0 + \delta\sigma^2,$$

where z_i is the unique positive solution of (3.13). Thus, letting $z^* := (z_1, \ldots, z_q)$, the LQ problem's virtual minimum $V^* = (V_1^*, \ldots, V_q^*)$ becomes

$$V^* = k(\gamma_0)z^*. \tag{5.15}$$

Moreover, to find a strong Pareto policy we may proceed as follows. From (5.15) and (3.17), the support function in (5.3) is given by

$$\Delta(\lambda) = k(\gamma_0)[z(\lambda) - \lambda \cdot z^*] \quad \forall \lambda \in \mathbb{R}^q.$$

Now let $\lambda^* \in S_1 \cap \mathbb{R}^q_{++}$ be as in Theorem 5.2(b). Then a strong Pareto optimal policy is obtained from (3.14) taking $\lambda = \lambda^*$, and the cost vector "closest" to V^* is given by (3.17) with $\lambda = \lambda^*$.

6 The multiobjective LP approach

In this section we follow Balbás and Heras [5] to formulate our multiobjective MCP as a *multiobjective linear program*. This requires to introduce two dual pairs $(M(\mathbb{K}), F(\mathbb{K}))$ and $(M(\mathrm{X}), F(\mathrm{X}))$ of vector spaces, which are essentially the same as those defined in [15, §6.3] or [16, §12.3]. (The reader may consult the latter references or [2] for general facts on infinite-dimensional scalar linear programming (LP).)

Define $w : \mathbb{K} \to \mathbb{R}_{++}$ as

$$w(x, a) := 1 + c_1(x, a) + \cdots + c_q(x, a). \tag{6.1}$$

(More generally, our approach may use any nonnegative "weight" function $w(x, a)$ provided that it is bounded away from zero and that it majorizes all of the functions $c_i(x, a)$. Thus, instead of w in (6.1) we could use, for instance, $w := \epsilon + \max(c_1, \ldots, c_q)$ for any $\epsilon > 0$.) Observe that (4.3) is equivalent to

$$\int w d|\mu| < \infty. \tag{6.2}$$

Therefore, the vector space $M(\mathbb{K})$ can be described as the space of finite signed measures μ on $\mathrm{X} \times A$, concentrated on $\mathbb{K}$, and for which (6.2) holds.

Now let $F(\mathbb{K})$ be the vector space of real-valued measurable functions v on $\mathbb{K}$ such that

$$\|v\|_w := \sup_{(x,a)} |v(x,a)|/w(x,a) < \infty. \tag{6.3}$$

From (6.1), it follows that each of the cost functions c_i belongs to $F(\mathbb{K})$, and, on the other hand, $(M(\mathbb{K}), F(\mathbb{K}))$ is a dual pair of vector spaces with respect to the bilinear form

$$\langle \mu, v \rangle := \int v d\mu \quad \text{for} \quad \mu \in M(\mathbb{K}), v \in F(\mathbb{K}). \tag{6.4}$$

We also consider another dual pair $(M(\mathrm{X}), F(\mathrm{X}))$ defined exactly as above but replacing $\mathbb{K}$ and w with X and

$$w_0(x) := \inf_{a \in A(x)} w(x,a) \quad \forall x \in \mathrm{X},$$

respectively.

Weak topologies. In the remainder of this section we consider $M(\mathbb{K})$ to be endowed with the weak toplogy $\sigma(M(\mathbb{K}), F(\mathbb{K}))$, which will be referred to as the *σ-weak topology*. Thus a sequence (or a net) $\{\mu_n\}$ *σ-converges* to μ if

$$\langle \mu_n, v \rangle \to \langle \mu, v \rangle \quad \forall v \in F(\mathbb{K}). \tag{6.5}$$

This should not be confused with the "weak convergence" (5.6), which is restricted to *continuous and bounded* functions. (Note that, of course, $C_b(\mathbb{K}) \subset F(\mathbb{K})$.) The vector spaces $F(\mathbb{K}), M(\mathbb{K})$, and $F(\mathrm{X})$ are also endowed with the corresponding σ-weak topologies.

Let $L : M(\mathbb{K}) \to M(\mathrm{X})$ be the linear map $\mu \mapsto L\mu$ defined as

$$(L\mu)(B) := \widehat{\mu}(B) - \delta \int_{\mathbb{K}} Q(B|x,a)\mu(d(x,a)). \tag{6.6}$$

The *adjoint* $L^* : F(\mathrm{X}) \to F(\mathbb{K})$ of L, that is, the linear map L^* for which

$$\langle L\mu, u \rangle = \langle \mu, L^* u \rangle \quad \forall \mu \in M(\mathbb{K}), u \in F(\mathrm{X}), \tag{6.7}$$

is given by

$$(L^*u)(x,a) = u(x) - \delta \int_{\mathrm{X}} u(y)Q(dy|x,a) \quad \forall (x,a) \in \mathbb{K}. \tag{6.8}$$

To ensure that L^* indeed maps $F(\mathrm{X})$ into $F(\mathbb{K})$, or, equivalently, that

$$L \text{ is } \sigma - \textit{weakly continuous},$$

we suppose the following.

Assumption 6.1. $\int_{\mathrm{X}} w_0(y)Q(dy|\cdot)$ *is in* $F(\mathbb{K})$*; that is, for some constant* k*,*

$$\int_{\mathrm{X}} w_0(y)Q(dy|x,a) \leq kw(x,a) \quad \forall (x,a) \in \mathbb{K}.$$

Note that Assumptions 6.1 and 3.1(d) ensure that the initial distribution γ_0 is in $M(\mathrm{X})$.

In the remainder of this section we suppose that Assumptions 3.1 and 6.1 are satisfied.

Multiobjective LP. For each μ in $M(\mathbb{K})$, let $\langle \mu, c \rangle$ be as in (4.6) and consider the *primal program* (PP):

$$\begin{aligned} &\text{minimize } \langle \mu, c \rangle \\ &\text{subject to: } L\mu = (1-\delta)\gamma_0, \quad \mu \in M_+(\mathbb{K}). \end{aligned} \tag{6.9}$$

Comparing (PP) with the MMP (4.7) we can see that they are essentially the same but the former has a little more "structure": the constraint (4.4) has been rewritten in (6.9) using the σ-weakly continuous map L.

A feasible solution μ^* for (PP) is said to be *optimal* if there is no feasible μ such that $\langle \mu, c \rangle < \langle \mu^*, c \rangle$. If such an optimal solution exists, then (PP) is said to be *solvable*. Thus, from Theorem 3.2(b) and the equivalence of (4.7) and the multiobjective MCP, we conclude the following.

Corollary 6.2. (PP) *is solvable.*

To state the *dual program* we need some notation. Let $F(\mathrm{X})^q$ be the vector space of $\mathbb{R}^q$-valued functions $u = (u_1, \ldots, u_q)$ with $u_i \in F(\mathrm{X})$ for all $i = 1, \ldots, q$. For $u \in F(\mathrm{X})^q$ and $\lambda \in \mathbb{R}^q$, let $u^\lambda \in F(\mathrm{X})$ and $L^*u \in F(\mathbb{K})^q$ be the functions given by

$$u^\lambda := \lambda \cdot u = \sum_{i=1}^{q} \lambda_i u_i \quad \text{and} \quad L^*u := (L^*u_1, \ldots, L^*u_q), \tag{6.10}$$

respectively. Moreover, if ν is in $M(\mathrm{X})$, we write

$$\langle \nu, u \rangle := (\langle \nu, u_1 \rangle, \ldots, \langle \nu, u_q \rangle).$$

Then, from [5, p. 380], we can see that the *dual program* (DP) of (PP) is as follows:

$$\begin{aligned} &\text{(DP) maximize } \langle (1-\delta)\gamma_0, u \rangle \\ &\qquad \text{subject to: } \lambda \cdot L^*u \le \lambda \cdot c \text{ with } u \in F(\mathrm{X})^q, \text{ for some } \lambda \in \mathbb{R}^q_{++}. \end{aligned} \tag{6.11}$$

In fact, if we let

$$F_\lambda := \{u \in F(\mathrm{X})^q | \lambda \cdot \langle L\mu, u \rangle \le \lambda \cdot \langle \mu, c \rangle \ \ \forall \mu \in M_+(\mathrm{X})\}$$

and use (6.7), it then follows that the dual constraint (6.11) can be expressed as in [5], namely:

$$u \textit{ is in } F_\lambda \textit{ for some } \lambda \in \mathbb{R}^q_{++}.$$

On the other hand, using (6.10) and (6.8) we can write (6.11) in the more explicit form

$$u^\lambda(x) \le c^\lambda(x,a) + \delta \int_{\mathrm{X}} u^\lambda(y) Q(dy|x,a) \quad \forall (x,a) \in \mathbb{K}, \tag{6.12}$$

for some $\lambda \in \mathbb{R}^q_{++}$. The latter inequality yields

$$u^\lambda(x) \le \min_{a \in A(x)} \left[c^\lambda(x,a) + \delta \int_{\mathrm{X}} u^\lambda(y) Q(dy|x,a) \right] \ \ \forall x \in \mathrm{X}, \tag{6.13}$$

which, when the *equality* holds, that is,

$$u^\lambda(x) = \min_{a \in A(x)} \left[c^\lambda(x,a) + \delta \int_{\mathrm{X}} u^\lambda(y) Q(dy|x,a) \right] \ \ \forall x \in \mathrm{X}. \tag{6.14}$$

becomes the *dynamic programming equation* (d.p.e) for the scalar MCP with cost function $(1-\delta)^{-1} V^\lambda(\pi, x)$, where $V^\lambda(\pi, x)$ is the function in (3.5) when the initial state is $x_0 = x$.

Remark 6.3. *Let* $V_*^\lambda(x) := \inf_\pi V^\lambda(\pi, x)$ *for all* $x \in \mathrm{X}$. *Then* $(1-\delta)^{-1}V_*^\lambda(x)$ *is the (pointwise) minimal solution of the d.p.e. (6.14). Moreover, if* V_*^λ *is in* $F(\mathrm{X})$ *and* u^λ *satisfies (6.12)-(6.13), then well-known arguments (see [15, Lemma 4.2.7], for instance) give that*

$$u^\lambda(x) \leq (1-\delta)^{-1}V_*^\lambda(x) \quad \forall x \in \mathrm{X}, \tag{6.15}$$

and for this reason u^λ *is said to be a subsolution of the d.p.e. (6.14). Note that (6.15) yields*

$$\langle (1-\delta)\gamma_0, u^\lambda \rangle \leq \langle \gamma_0, V_*^\lambda \rangle. \tag{6.16}$$

Therefore (by the equivalence of (6.11) and (6.12)), we can see the dual program (DP) *as the problem of maximizing integrals as in the left-hand side of (6.16) over the family of subsolutions* u^λ *of the d.p.e. for a class of scalar MCPs parameterized by* $\lambda \in \mathbb{R}^q_{++}$. *Thus, the multiobjective LP formulation gives us a "primal-dual" interpretation of the relation between our original multiobjective MCP and the scalar MCPs in (3.3)-(3.5). This interpretation can also be obtained from the "complementary slackness" property in the following proposition from [5] adapted to our current situation.*

Proposition 6.4. *Let* μ *be a feasible solution for* (PP) *and* u *a feasible solution for* (DP). *Then*

(a) *(Weak duality.) We never have* $\langle (1-\delta)\gamma_0, u \rangle > \langle \mu, c \rangle$.

(b) *(Complementary slackness.) If in addition*

$$\langle \mu, c - L^* u \rangle = 0, \tag{6.17}$$

then μ *is optimal for* (PP) *and* u *is optimal for* (DP).

Proof. Part (a) is straightforward, and in turn (a) implies (b) because, by (6.7) and (6.9), we can write (6.17) as

$$\langle (1-\delta)\gamma_0, u \rangle = \langle \mu, c \rangle. \quad \blacksquare$$

Now, to obtain the primal-dual interpretation mentioned in the last sentence of Remark 6.4, it suffices to note that (6.17) is equivalent to

$$\langle \mu, c^\lambda - L^* u^\lambda \rangle = 0 \quad \forall \lambda \in \mathbb{R}^q_{++}. \tag{6.18}$$

In fact, by (6.8), we can recognize the integrand $c^\lambda - L^*u^\lambda$ in (6.18) as the difference between the two sides of (6.12). Therefore, we can obtain a solution (μ, u^λ) for (6.18) in the obvious manner: choose an arbitrary $\lambda \in \mathbb{R}^q_{++}$ and let V^λ_* be as in Remark 6.4. Let

$$u^\lambda_*(x) := (1-\delta)^{-1}V^\lambda_*(x) \quad \forall x \in \mathrm{X}.$$

Furthermore, (as in the proof of Theorem 3.2(b)) let $f_* \in \mathbb{F}$ be a stationary policy such that $f_*(x) \in A(x)$ attains the minimum in the d.p.e. (6.14) for all $x \in \mathrm{X}$, and, finally, let μ_* be the occupation measure associated with f_*. Then, by their very definitions, it follows that μ_* is feasible for (PP), u^λ_* is feasible for (DP), and

$$\langle \mu_*, c^\lambda - L^*u^\lambda_* \rangle = 0.$$

7 Proof of Theorem 3.3

Suppose that $\pi^* \in \Pi_0$ is Pareto optimal and let μ^* be its occupation measure (see (4.1)). By (4.2), (4.6) and Lemma 4.2, to prove Theorem 3.3 it suffices to show the existence of a q-vector $\lambda^* \in \Lambda$ (the set defined in (3.6)) such that

$$\lambda^* \cdot \langle \mu^*, c \rangle \leq \lambda^* \cdot \langle \mu, c \rangle \quad \forall \mu \in M_\delta(\mathbb{K}),$$

that is

$$\lambda^* \cdot \langle \mu - \mu^*, c \rangle \geq 0 \quad \forall \mu \in M_\delta(\mathbb{K}). \tag{7.1}$$

With this in mind, consider the set $\Gamma(\Pi_0)$ in (5.11) and let $Y \subset \mathbb{R}^q$ be the set given by

$$Y := \Gamma(\Pi_0) + \mathbb{R}^q_+ - \langle \mu^*, c \rangle.$$

As already noted in (5.12), $\Gamma(\Pi_0) \subset \mathbb{R}^q$ is a convex set and, therefore, so is Y. Let Y^+ and $c(Y)$ be the *polar* (or positive conjugate) *cone* of Y and the *cone generated* by Y, respectively; that is

$$Y^+ := \{z \in \mathbb{R}^q | z \cdot y \geq 0 \ \ \forall y \in Y\},$$
$$c(Y) := \{z \in \mathbb{R}^q | z = ry \text{ for some } y \in Y \text{ and } r \geq 0\}.$$

Moreover, let $\mathbb{R}^q_- := -\mathbb{R}^q_+$ be the nonpositive orthant in $\mathbb{R}^q$.

Now observe that, as π^* is in $\Pi_0 \cap \mathrm{Par}(\Pi)$, μ^* is an optimal solution for the MMP (4.7). Hence, there is no $\mu \in M_\delta(\mathbb{K})$ for which

$$\langle \mu, c \rangle < \langle \mu^*, c \rangle,$$

which implies that

$$c(Y) \cap \mathbb{R}^q_- = \{0\}. \tag{7.2}$$

Therefore, to prove (7.1) it suffices to show that

$$\mathbb{R}^q_{++} \cap Y^+ \neq \emptyset, \tag{7.3}$$

because if z is a vector in $\mathbb{R}^q_{++} \cap Y^+$, then

$$z \cdot (\langle \mu, c \rangle + y - \langle \mu^*, c \rangle) \geq 0 \quad \forall \mu \in M_\delta(\mathbb{K}), y \in \mathbb{R}^q_+,$$

which taking $y = 0$ yields

$$z \cdot (\langle \mu, c \rangle - \langle \mu^*, c \rangle) \geq 0 \quad \forall \mu \in M_\delta(\mathbb{K}).$$

Thus, letting $\overline{z} := \sum z_i$, we get (7.1) with $\lambda^* := z/\overline{z}$.

We will prove (7.3) by contradiction. Suppose that (7.3) does not hold, i.e.,

$$\mathbb{R}^q_{++} \cap Y^+ = \emptyset.$$

Then, as $\mathbb{R}^q_{++}$ and Y^+ are both convex sets and the interior of $\mathbb{R}^q_{++}$ is of course nonempty, the "separating hyperplane theorem" [4, 21] yields the existence of a q-vector $z \neq 0$ such that

$$z \cdot \lambda \leq z \cdot y \quad \forall \lambda \in \mathbb{R}^q_{++}, y \in Y^+. \tag{7.4}$$

In particular, as $y = 0$ is in Y^+, we get $z \cdot \lambda \leq 0 \ \ \forall \lambda \in \mathbb{R}^q_{++}$, which implies that z is in $\mathbb{R}^q_-$. Similarly, as $\lambda \in \mathbb{R}^q_{++}$ in (7.4) can be chosen arbitrarily close to the vector $0 \in \mathbb{R}^q$, (7.4) gives $z \cdot y \geq 0 \ \ \forall y \in Y^+$. Therefore, z is in the bipolar cone $(Y^+)^+$ of Y; that is, z is in $(Y^+)^+ = c(Y)$. To conclude, we have proved that z is a nonzero vector in $c(Y) \cap \mathbb{R}^q_-$, which contradicts (7.2). This completes the proof of (7.3), which, as was already noted, gives (7.1). ■

8 Further remarks

In this final section we briefly discuss some connections between our results and other problems for MCPs.

Constrained MCPs. For each $i = 1, \ldots, q$, let $V_i(\pi) = V_i(\pi, \gamma_0)$ be as in (2.4), and let $k_2, \ldots, k_q$ be $q-1$ nonnegative given numbers. Then the problem

$$\begin{aligned} &\text{minimize } V_1(\pi) \\ &\text{subject to: } V_i(\pi) \le k_i \text{ for } i = 2, \ldots, q;\ \pi \in \Pi, \end{aligned} \tag{8.1}$$

is called a *constrained MCP*. In this case, a policy π for which (8.1) holds and, in addition, $V_q(\pi) < \infty$ is said to be *feasible* for the constrained MCP. Let us suppose that the set $\Pi_{co} \subset \Pi$ of feasible policies is nonempty. Then, under Assumption 3.1, there is an optimal policy $\pi^* \in \Pi_{co}$ for the constrained MCP (see [13]), and if, in addition, π^* is the *unique* optimal policy, then π^* is easily seen to be Pareto optimal for the multiobjective MCP in Section 2 above. Moreover, Theorem 3.3 yields that π^* is optimal for the scalar, or "weighted", cost criterion (3.4)-(3.5) for some q-vector $\lambda = \lambda^*$ in Λ.

For additional results on constrained MCPs or for MCPs with weighted criteria, see, for instance, [1, 8, 9, 13, 14, 19, 22, 26].

Average cost. Let us rewrite (2.4) as

$$V_i(\pi, \gamma_0) = \limsup_{n\to\infty} E^{\pi}_{\gamma_0}\Big[\sum_{t=0}^{n-1} \delta^t c_i(x_t, a_t)\Big] \Big/ \sum_{t=0}^{n-1} \delta^t. \tag{8.2}$$

This is, of course, the same as (2.4) if $0 < \delta < 1$, whereas if $\delta = 1$ we get the *average cost* (AC) criterion

$$J_i(\pi, \gamma_0) = \limsup_{n\to\infty} \frac{1}{n} E^{\pi}_{\gamma_0}\Big[\sum_{t=0}^{n-1} c_i(x_t, a_t)\Big]. \tag{8.3}$$

A key difference with respect to the discounted cost problem (in which the initial distribution γ_0 is *fixed*) is that in the AC case γ_0 is also a decision variable, so that in fact we have a "minimum pair" problem as in [14, 15, 16]. In other words, let $\mathbb{P}(X)$

be the set of probability measures on X, and let $J_i(\pi, \gamma_0)$ be as in (8.3) *for each pair* $(\pi, \gamma_0) \in \Pi \times \mathbb{P}(X)$. Moreover, let Δ be the set of all those *pairs* (π, γ_0) such that

$$J_i(\pi, \gamma_0) < \infty \quad \forall i = 1, \ldots, q, \tag{8.4}$$

and replace Assumption 3.1(d) with

$$\Delta \textit{ is nonempty.} \tag{8.5}$$

This set Δ plays the role of Π_0 in Theorem 3.3. Further, let $J(\pi, \gamma_0) := (J_1(\pi, \gamma_0), \ldots, J_q(\pi, \gamma_0))$ and

$$\Gamma(\Delta) := \{J(\pi, \gamma_0) | (\pi, \gamma_0) \in \Delta\}. \tag{8.6}$$

Then, under (8.5) and Assumption 3.1(a), (b), (c), all of the results in Sections 3, 4 and 5 remain valid when $\delta = 1$, with some obvious changes. For example, the set $M_1(\mathbb{K})$ in Definition 4.1 (and (4.5)) is the set of probability measures μ on $X \times A$, concentrated on $\mathbb{K}$, and such that (as in (4.4))

$$\widehat{\mu}(B) = \int_{\mathbb{K}} Q(B|x, a)\mu(d(x, a)). \tag{8.7}$$

Similarly, the virtual minimum in Section 5 is now given by $J^* = (J_1^*, \ldots, J_q^*)$ with

$$J_i^* := \inf\{J_i(\pi, \gamma_0) | (\pi, \gamma_0) \in \Delta\},$$

and, by (8.7), the constraint equation (6.9) in the multiobjective LP formulation becomes

$$L_1\mu = 0, \quad \mu \in M_+(\mathbb{K}), \tag{8.8}$$

where L_1 is given by (6.6) with $\delta = 1$. Finally, as in the discounted case (8.1), we can also consider constrained MCPs with the AC criterion and if there is a *unique* optimal policy for the constrained problem, then it is Pareto optimal for the multiobjective MCP.

Remark 8.1. *In [14-16] a probability measure μ for which (8.8) holds is called stable.*

Mixed average-discounted criteria. The "minimum pair" approach in (8.4)-(8.6) can be used to study multiobjective MCPs with cost vectors of the form

$$(J_1(\pi, \gamma_0), \ldots, J_r(\pi, \gamma_0), V_{r+1}(\pi, \gamma_0), \ldots, V_q(\pi, \gamma_0))$$

in which the $J_i(\pi, \gamma_0)$ are ACs as in (8.3), and the $V_j(\pi, \gamma_0)$ are discounted costs as in (8.2) *with possibly different discount factors* δ_j $(j = r+1, \ldots, q)$. The key fact that allows to do this is that using (8.5) and Assumption 3.1(a), (b), (c) the original multiobjective MCP is reduced to solving a Pareto problem of the form (4.7) but on the set $M_1(\mathbb{K})$ of *stable probability measures*. The corresponding technical details are essentially the same as in Remarks 2.2(c) and 3.8(b) of [14].

References

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

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

[3] J.-P. Aubin, *A Pareto minimum principle*, in Differential Games and Related Topics, ed. by H.W. Kuhn and G.P. Szegö, North-Holland, Amsterdam, 1971, pp. 147-175.

[4] J.-P. Aubin, *Optima and Equilibria*, Springer-Verlag, Berlin, 1993.

[5] A. Balbás and A. Heras, *Duality theory for infinite dimensional multiobjective linear programming*, Euro. J. Oper. Res., 68(1993), pp. 379-388.

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

[7] E.B. Dynkin and A.A. Yushkevich, *Controlled Markov Processes*, Springer-Verlag, Berlin, 1979.

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

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

[10] N. Furukawa, *Characterization of optimal policies in vector-valued Markov decision processes*, Math. Oper. Res., 5(1980), pp. 271-279.

[11] M.K. Ghosh, *Markov decision processes with multiple costs*, Oper. Res. Lett., 9(1990), pp. 257-260.

[12] J. González-Hernández and O. Hernández-Lerma, *Envelopes of sets of measures, tightness, and Markov control processes*, Appl. Math. Optim., 40(1999), pp. 377-392.

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

[14] O. Hernández-Lerma, J. González-Hernández and R.R. López-Martínez, *Constrained average cost Markov control processes in Borel spaces*, Internal Report, CINVESTAV-IPN, México, 1999. (Submitted.)

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

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

[17] M.I. Henig, *Vector-valued dynamic programming*, SIAM J. Control Optim., 21(1983), pp. 490-499.

[18] M.I. Henig, *The principle of optimality in dynamic programming with returns in partially ordered sets*, Math. Oper. Res., 10(1985), pp. 462-471.

[19] D. Krass, J. Filar and S.S. Sinha, *A weighted Markov decision process*, Oper. Res., 40(1992), pp. 1180-1187.

[20] H.-C. Lai and K. Tanaka, *Average-time criterion for vector-valued Markovian decision systems*, Nihonkai Math. J., 2(1991), pp. 71-91.

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

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

[23] Y. Sawaragi, H. Nakayama and T. Tanino, *Theory of Multiobjective Optimization*, Academic Press, New York, 1985.

[24] K. Tanaka, *The closest solution to the shadow minimum of a cooperative dynamic game*, Computers Math. Appl., 18(1989), pp. 181-188.

[25] K. Tanaka and C. Matsuda, *On continuously discounted vector valued Markov decision process*, J. Inform. Optim. Sci., 11(1990), pp. 33-48.

[26] L.C. Thomas, *Constrained Markov decision processes as multi-objective problems*, in Multi-Objective Decision Making, ed. by D.J. White, S. French and R. Hartley, Academic Press, London, 1983, pp. 77-94.

[27] K. Wakuta, *Optimal stationary policies in the vector-valued Markov decision process*, Stoch. Proc. Appl., 42(1992), pp. 149-156.

[28] K. Wakuta, *Vector-valued Markov decision process and the systems of linear inequalities*, Stoch. Proc. Appl., 56(1995), pp. 159-169.

[29] C.C. White, III, and W.K. Kwang, *Solution procedures for vector criterion Markov decision processes*, Large Scale Systems, 1(1980), pp. 129-140.

[30] D.J. White, *Multi-objective infinite-horizon discounted Markov decision processes*, J. Math. Anal. Appl., 89(1982), pp. 639-647.