Morfismos, Vol. 8, No. 1, 2004, pp. 1–33

Multiobjective Markov Control Processes: a Linear Programming Approach *

Onésimo Hernández-Lerma Rosario Romera

Abstract

This paper studies discrete-time multiobjective Markov control processes (MCPs) on Borel spaces and unbounded costs. Under mild assumptions, it shows the existence of Pareto policies, which, as in multiobjective optimization problems, are also characterized as optimal policies for a certain class of single-objective (or "scalar") MCPs. A similar result is obtained for *strong* Pareto policies, which are Pareto 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 *dual program* is in fact the same as solving the scalarized MCPs. A multiobjective LQ example illustrates the main results.

2000 Mathematics Subject Classification: 93E20, 90C40, 90C29.
Keywords and phrases: Markov control processes, multiobjective control problems, Pareto optimality, (infinite–dimensional) multiobjective linear programming.

1 Introduction

In a standard optimal control problem there is a decision–maker or 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

*Invited Article. Research partially supported by CONACYT (México) Grant 37355-E for OHL, and DGES (Spain) Grant PB96-0111 for RR.

(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 a *Pareto policy.* On the other hand, letting

$$V_i^* := \inf_\pi V_i(\pi) \quad \text{for} \quad i = 1, \ldots, q, \tag{1.1}$$

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^*. This is the control–theoretic analogue of a *goal programming* problem [36] in which the *goal* or *target* is V^*. (We might of course consider other "goals", but V^* is the most common.) Still another key problem occurs when the individual costs $V_1(\pi), \ldots, V_q(\pi)$ are ranked in order of "importance". In this case, a *lexicographically* (or *hierarchically*) *optimal policy* turns out to be a particular Pareto policy.

Contributions of this paper. In this paper we study discrete-time multiobjective Markov control processes (MCPs) on *Borel spaces* and *unbounded costs.* The main problems we are concerned with are the existence and characterization of both Pareto and strong Pareto policies, and also of *weak* and *proper* Pareto policies (Definition 2.5). Actually, the *existence* of Pareto, weak Pareto, and proper Pareto policies is very easy because it can be obtained via the usual *scalarization approach*, in which the multiobjective MCP is reduced to a single–objective (or standard or scalar) MCP with a "weighted" objective function of the form

$$\lambda\cdot V(\pi) := \lambda_1 V_1(\pi) + \cdots + \lambda_q V_q(\pi) \tag{1.2}$$

for some vectors λ in the nonnegative orthant $\mathbb{R}^q_+$. However, the existence of *strong* Pareto policies as well as the *characterization* problem are more complicated, and, to the best of our knowledge, this is the first paper dealing with these issues for *general* MCPs. (See below for related literature.)

To study the latter problems we propose here to use so–called occupation measures to transform the multiobjective MCP into an equivalent *multiobjective measure problem* (MMP) on a suitable space of measures. The original multiobjective control problem is thus greatly simplified because the MMP turns out to have a *linear* objective (vector) function defined on a *convex* set of measures. This implies that, for instance, the existence of strong Pareto policies essentially reduces to find the distance from the virtual minimum V^* to a convex set. Similarly, the characterization of Pareto policies (known as the "theorem of equivalence" in Pareto optimality [4]) can be obtained by standard convex–analytic arguments. Moreover, introducing suitable vector spaces, we reformulate the MMP as a primal *multiobjective linear program* and this allows us to show that solving the *dual* linear program is in fact the same as solving the scalar problem (1.2) using *dynamic programming*. As far as we can tell, this interpretation of the scalarization approach for *multiobjective* control problems as the *dual* of a multiobjective linear program has never been reported before in the literature. We should also note that to obtain the latter duality result, as well as the characterization of Pareto policies (Theorem 3.4, below) without our MMP approach would be extremely difficult — perhaps impossible — to obtain.

Related literature. Vector optimization problems can be traced back to (at least) the late 19th century; see e.g. [4, 31] for earlier references. However, according to the excellent survey by Salukvadze [35, Chapter 1], in control theory they were first introduced by Zadeh [46] in 1963. The scalarization and the hierarchical (or lexicographical) approaches were introduced by Reid and Citron [34] and Waltz [43], respectively.

Concerning multiobjective MCPs, the existence and characterization of Pareto policies have been studied by many authors but for particular classes of MCPs, for instance with a *countable* state space [10-12,14,23,24,28,40-42,44,45], or in Borel spaces but with *bounded* costs [29, 37, 39]. It should be noted that for some of these classes of MCPs one can obtain very interesting results. For example, if the state and action spaces are both *finite*, the set of Pareto policies can be completely characterized using Theorem 1 of Arrow et al. [3], as in [14]. Moreover, for *finite* state spaces, there are multiobjective versions of *value iteration* [23, 24, 45] and of *policy iteration* [12, 41, 42], which, as they are computationally appealing, it would be interesting to investigate if they can be extended to MCPs in uncountable spaces. On the other hand, some papers [12, 23, 24, 29] deal with a vector–minimization problem more general than ours in the sense that instead of the convex cone $\mathbb{R}^q_+$, they work with the partial order induced by an arbitrary pointed convex cone in $\mathbb{R}^q$. But it turns out that they restrict the control problem to some *subclass* of policies (e.g. deterministic stationary), whereas here we work with the set of all (randomized, history–dependent) policies. At any rate, extending our MMP approach to the case of a general pointed convex cone seems to be a purely notational problem.

Organization of the paper. 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 notions 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 [4]. In fact, we state this theorem in two parts, Theorem 3.2(a) and (the converse) Theorem 3.4, because the proof of the latter requires the MMP, which is 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 policies. Section 3 also includes Example 3.5 on a multiobjective LQ (Linear system with Quadratic costs) MCP in which explicit Pareto 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 [37] 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 3.5. Section 6 presents the *multiobjective Linear Programming* (LP) formulation of the multiobjective MCP. The idea (as for scalar and constrained MCPs [16,17,20,21]) is to introduce suitable dual pairs of vector spaces in which the MMP (4.7) can be formulated as a multiobjective linear program. The multiobjective LP formulation is borrowed from Balbás and Heras [7]. Section 7 contains the proof of Theorem 3.4, and, finally, in Section 8 we briefly mention some connections between our multiobjective MCPand 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

A *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 $\pi = \{\pi_n, n = 0, 1, \ldots\}$ 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 control 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 time t $(t = 0, 1, \ldots)$ when using the policy π. The expectation operator with respect to $P^\pi_{\gamma_0}$ is denoted by $E^\pi_{\gamma_0}$.

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

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

which will be well defined under our Assumption 3.1 below. Now let $V(\pi, \gamma_0) \in \mathbb{R}^q$ be the cost vector

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

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.

Pareto optimality. 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 $i = 1, \ldots, q$. We also have

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

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

A sequence $\{u^k\} \subset \mathbb{R}^q$ converging to u is said to *converge in the direction* $v \in \mathbb{R}^q$ if there is a sequence of positive numbers t_k such that $t_k \to 0$ and

$$\lim_{k\to\infty} (u^k - u)/t_k = v. \tag{2.6}$$

Let Γ be a subset of $\mathbb{R}^q$. The *tangent cone to* Γ *at* $u \in \Gamma$, denoted $\mathrm{T}(\Gamma, u)$, is the set of all the directions $v \in \mathbb{R}^q$ in which some sequence in Γ converges to u. There are several equivalent definitions of tangent cone; see e.g. [5]. In particular, if Γ is a *convex* set, then ([5],p. 64)

$$\mathrm{T}(\Gamma, u) = \text{ closure } [\bigcup_{t>0} \frac{1}{t}(\Gamma - u)\,]. \tag{2.7}$$

Note that $\Gamma - u$ *is contained in* $\mathrm{T}(\Gamma, u)$.

Definition 2.2. Let Γ be a subset of $\mathbb{R}^q$. A vector u^* in Γ is said to be

(a) a *Pareto point* of Γ if there is no $u \in \Gamma$ such that $u < u^*$;

(b) a *weak Pareto point* of Γ if there is no $u \in \Gamma$ such that $u << u^*$;

(c) a *proper Pareto point* of Γ if u^* is a Pareto point and, in addition, the tangent cone to Γ at u^* does not contain vectors $v < 0$.

Let $\mathrm{Par}(\Gamma)$, $\mathrm{WPar}(\Gamma)$ and $\mathrm{PPar}(\Gamma)$ denote, respectively, the set of Pareto points of Γ, the set of weak Pareto points, and the set of proper Pareto points. Then

$$\mathrm{PPar}(\Gamma) \subset \mathrm{Par}(\Gamma) \subset \mathrm{WPar}(\Gamma). \tag{2.8}$$

Moreover, if Γ is a *closed convex set*, then (by Theorem 1 in [3]) $\mathrm{Par}(\Gamma)$ is contained in the closure of $\mathrm{PPar}(\Gamma)$.

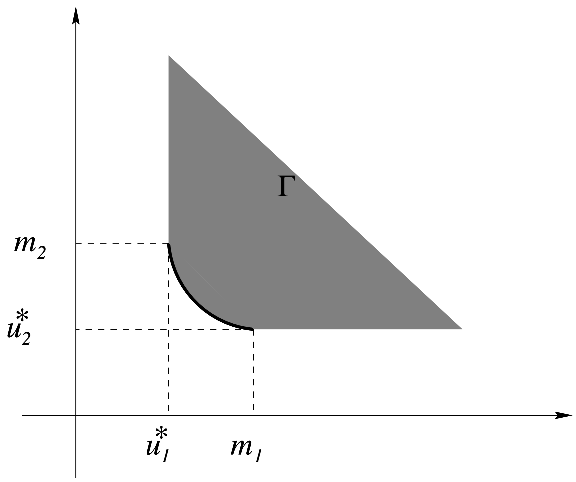

Figure 1.

Example 2.3.Let $\Gamma \subset \mathbb{R}^2$ be as in Figure 1. Then WPar(Γ) coincides with the boundary of Γ, whereas Par(Γ) is the subset of the boundary consisting of the vector (u_1^*, m_2) and the vectors whose first coordinate is in the half–closed interval $(u_1^*, m_1]$. Finally, the proper Pareto points of Γ are the vectors in Par(Γ) with first coordinate in the open interval (u_1^*, m_1). Also note that the vector (u_1^*, m_2) is the *lexicographical minimum* of Γ in the sense of the following definition.

Definition 2.4. If u and v are vectors in $\mathbb{R}^q$, u is said to be *lexicographically smaller than* v (in symbols: $u \leq_L v$) if the first nonzero term of the sequence $v_1 - u_1, \ldots, v_q - u_q$ is positive. Moreover, a vector $\widehat{u}$ in $\Gamma \subset \mathbb{R}^q$ is called the *lexicographical minimum* of Γ if $\widehat{u} \leq_L u$ for all $u \in \Gamma$.

A direct application of Definitions 2.2 and 2.4 shows that the lexicographical minimum is a Pareto point.

Pareto policies. The above concepts can be extended to multiobjective MCPs in the same way as it is done for vector optimization problems [27, 31, 36]. First, as the initial distribution γ_0 is *fixed*, we shall simplify the notation by dropping γ_0 from expressions such as (2.4) and (2.5). For instance, we shall write $V_i(\pi, \gamma_0)$ simply as $V_i(\pi)$.

Definition 2.5.Let $\Gamma(\Pi)$ be the set of cost vectors in (2.5), i.e.

$$\Gamma(\Pi) := \{V(\pi) | \pi \in \Pi\}. \tag{2.9}$$

A policy $\pi^* \in \Pi$ is said to be

(a) a *Pareto policy* (respectively, a *weak Pareto policy* or a *proper Pareto policy*) if its corresponding cost vector $V(\pi^*)$ is in $\mathrm{Par}(\Gamma(\Pi))$ (respectively, in $\mathrm{WPar}(\Gamma(\Pi))$ or in $\mathrm{PPar}(\Gamma(\Pi))$);

(b) *lexicographically optimal* if $V(\pi^*)$ is the lexicographical minimum of $\Gamma(\Pi)$.

In other words, $\pi^* \in \Pi$ is a Pareto policy (or Pareto optimal) if there is no policy π such that $V(\pi) < V(\pi^*)$, and similarly for weak or proper Pareto policies.

The set $\Gamma(\Pi)$ in (2.9) is called the *performance set* (also known as the *objective* or *achievable* set) of the multiobjective MCP. An example in which $\Gamma(\Pi)$ is similar to the set Γ in Figure 1 is given in [18], where it is shown that the so–called *cμ–rule* for priority queues is lexicographically optimal — hence a nonproper Pareto policy. In fact, there are many examples of lexicographically optimal policies, including Blackwell optimal policies [26], bias optimal policies [21, 25], and average cost optimal policies that in addition minimize the cost variance [22].

Remark 2.6.(a) To find lexicographically optimal policies we may proceed as follows. Let $\Pi_0 := \Pi$, and for $i = 1, \dots, q$ let

$$\widehat{V}_i := \inf\{V_i(\pi) | \pi \in \Pi_{i-1}\}, \tag{2.10}$$

and, finally, let Π_i be the set of policies in Π_{i-1} that attain the minimum in (2.10). Then, assuming that the sets Π_i are nonempty, Π_q consists of the lexicographically optimal policies. Moreover, if $\widehat{\pi}$ is in Π_q, then its cost vector $V(\widehat{\pi}) = (\widehat{V}_1, \dots, \widehat{V}_q)$ is of course the lexicographical minimum of $\Gamma(\Pi)$.

(b) The procedure in (2.10) is also valid for $q = \infty$, that is, for *infinite* cost vectors $V(\pi)$, as in Blackwell optimality [26], for instance.

(c) If for some $i = 1, \dots, q$ the set Π_i in (a) consists of a *single* policy $\widehat{\pi}_i$, then $\widehat{\pi}_i$ is the unique lexicographically optimal policy.

(d) As in (2.8), $\mathrm{PPar}(\Gamma(\Pi)) \subset \mathrm{Par}(\Gamma(\Pi)) \subset \mathrm{WPar}(\Gamma(\Pi))$.

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 $V_i(\pi, \gamma_0) \equiv V_i(\pi)$.)

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 ξ_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 3.5 and 5.7, below.

The existence problem. To study the existence of Pareto policies we shall first follow the well-known “scalarization” approach. Thus, given a q–vector $\lambda > 0$ 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

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

where $\mathbb{R}^q_{++}$ is the set of vectors $\lambda \gg 0$. We may then state an existence result as follows. (Observe that part (d) in Theorem 3.2 gives a little more than the existence of Pareto policies because, in fact, it ensures the existence of *deterministic stationary* Pareto policies.)

Theorem 3.2. *Suppose that for some q–vector $\lambda = (\lambda_1, \ldots, \lambda_q) > 0$ there is a policy $\pi^* \in \Pi$ that is optimal for the scalar criterion (3.4), i.e.*

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

Then:

(a) π^* is a weak Pareto policy.

(b) If in addition

$$V^\lambda(\pi^*) < V^\lambda(\pi) \quad \forall\, \pi \in \Pi \quad \text{with} \quad V(\pi) \ne V(\pi^*), \tag{3.8}$$

then π^* is a Pareto policy.

(c) If $\lambda \gg 0$ (in particular if λ is in Λ), then π^* is a proper Pareto policy.

(d) If $\lambda_1 > 0$, then there exists a deterministic stationary policy $f_\lambda \in \mathbb{F}$ that is a weak Pareto policy; if, moreover, $\pi^* \equiv f_\lambda$ satisfies (3.8), then f_λ is a Pareto policy. Finally, if $\lambda \gg 0$, then f_λ is a proper Pareto policy.

Proof: (a) Suppose that π^* is *not* a weak Pareto policy. Then there exists a policy $\pi \in \Pi$ such that $V(\pi) \ll V(\pi^*)$ and, therefore, as $\lambda > 0$, we get $V^\lambda(\pi) < V^\lambda(\pi^*)$, which contradicts (3.7).

(b) Similarly, if π^* is *not* a Pareto policy, there exists $\pi \in \Pi$ such that $V(\pi) < V(\pi^*)$. Hence $V^\lambda(\pi) \leq V^\lambda(\pi^*)$, which contradicts (3.8).

(c) If π^* is *not* a proper Pareto policy, then the tangent cone to $\Gamma(\Pi)$ at $V(\pi^*)$, i.e. $T(\Gamma(\Pi), V(\pi^*))$, contains a vector $v < 0$. Therefore, there exists a sequence $\{\pi^k\}$ in Π and a sequence $\{t_k\}$ of positive numbers such that

$$\lim_{k\to\infty} (V(\pi^k) - V(\pi^*))/t_k = v.$$

As $\lambda \gg 0$, we have $\lambda \cdot v < 0$. It follows that for all k sufficiently large

$$\lambda \cdot (V(\pi^k) - V(\pi^*)) = V^\lambda(\pi^k) - V^\lambda(\pi^*) < 0,$$

which contradicts (3.7).

(d) Suppose that $\lambda_1 > 0$. Then, by Assumption 3.1(b), $\lambda_1 \cdot c_1(x, a)$ is nonnegative and inf–compact, and, therefore (by (3.3) and the first part of Assumption 3.1(b)), so is c^λ. The latter fact together with Assumption 3.1(a),(c),(d) implies the existence of a deterministic stationary policy $\pi^* \equiv f^\lambda$ that satisfies (3.7); see e.g. [15] or Theorem 4.2.3 in [20]. Hence, by part (a), f^λ is a weak Pareto policy. The remaining statements in (d) are proved similarly. □

To obtain the converse of parts (a),(b),(c) in Theorem 3.2 we will use a special reformulation (introduced in Section 4) of the original multiobjective MCP. This requires to restrict the "admissible" policies to the following set.

Definition 3.3.Π_0 denotes the set of policies $\pi \in \Pi$ for which $V_i(\pi) < \infty$ for all $i = 1, \ldots, q$.

By our Assumption 3.1(d), the set Π_0 is nonempty. The following theorem is proved in Section 7.

Theorem 3.4. *Let π^* be a policy in Π_0. If π^* is a weak Pareto policy, then there exists a q–vector $\lambda > 0$ for which (3.7) holds. If π^* is a proper Pareto policy, then $\lambda \gg 0$.*

As a Pareto policy is weak Pareto (recall (2.8)), Theorem 3.4 tacitly includes the case in which π^* is a nonproper Pareto policy. Thus Theorem 3.4 is a (slight) extension of the so–called "theorem of equivalence" in Pareto optimality [4]. Finally, observe that Theorems 3.4 and 3.2 indeed characterize weak and proper Pareto policies because they yield that, for instance, $\pi^* \in \Pi_0$ is a proper Pareto policy *if and only if* π^* minimizes the scalar criterion (3.4) for some q–vector $\lambda \gg 0$.

The following example illustrates Theorem 3.2.

Example 3.5. Let α and β be nonzero real numbers and consider the scalar linear system

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

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 q–vector $\lambda > 0$, 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

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

Now 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 [20]), 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

$$V^\lambda(f_\lambda) = k(\gamma_0)z(\lambda), \quad \text{with} \quad k(\gamma_0) := (1-\delta)\overline{\gamma}_0 + \delta\sigma^2. \tag{3.17}$$

By Theorem 3.2, f_λ is a proper Pareto policy if $\lambda \gg 0$, and a weak Pareto policy if $\lambda > 0$. In particular, let $e(i)$ be the unit vector with coordinates $e_i(i) = 1$ and $e_j(i) = 0$ for $j \neq i$. Then replacing λ in (3.17) with $e(i)$ we obtain the "partial" minimum cost in (1.1), i.e.

$$V_i^* := \inf_\pi V_i(\pi) = V_i(f_{e(i)}) = k(\gamma_0)z_i \quad \forall\ i = 1, \dots, q. \tag{3.18}$$

This gives the *virtual minimum* $V^* = (V_1^*, \dots, V_q^*)$, which is illustrated in Figure 2 for the case $q = 2$. In that figure, the Pareto set $\mathrm{Par}(\Gamma(\Pi))$ is the part of the boundary of $\Gamma(\Pi)$ with first coordinate in $[V_1^*, V_1(f_{e(2)})]$. On the other hand, by the uniqueness of optimal policies for LQ (linear–quadratic) systems, it follows from Remark 2.6(a),(c) that $f_{e(1)}$ is the *lexicographically optimal policy*, whose corresponding cost vector $\widehat{V} := V(f_{e(1)})$ has coordinates $\widehat{V}_i = V_i(f_{e(1)})$ for all $i = 1, \dots, q$, i.e.

$$\widehat{V} = (V_1^*, V_2(f_{e(1)}), \dots, V_q(f_{e(1)})). \tag{3.19}$$

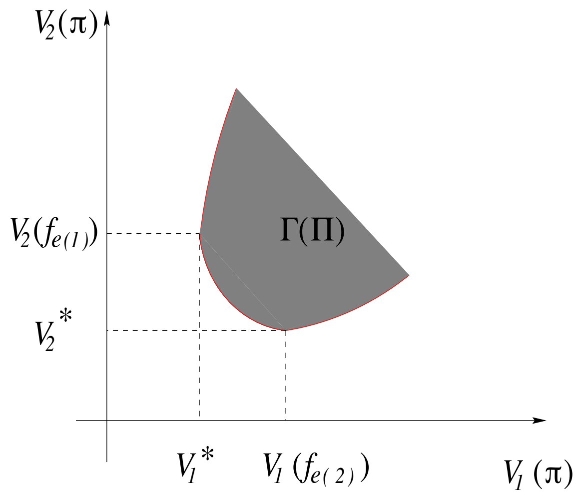

Figure 2. See (3.18), (3.19).

Remark 3.6.Consider a *single*, or scalar, LQ system with cost $c(x,a) = sx^2 + ra^2$; see (3.11). If the coefficients s and r are both positive, then an optimal policy for this problem can be interpreted as a *proper Pareto policy* for a two–dimensional *multiobjective* control problem with individual costs $c_1(x,a) := x^2$ and $c_2(x,a) := a^2$. In fact, a similar interpretation is valid for any scalar control problem with *additive costs*, say of the form

$$c(x,a) = r_1 c_1(x,a) + \cdots + r_q c_q(x,a)$$

with positive coefficients $r_1, \ldots, r_q$. See [19] for details.

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

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

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 (see Definition 3.3), then a standard argument (see, for instance, Remark 9.4.2(b) in [21, 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

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

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

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

It also turns out that $M_\delta(\mathbb{K})$ coincides with the family of occupation measures in (4.1). More precisely (as in [15, pp. 386-387] or [20, Theorem 6.3.7], for instance), we have the following result in which Π_0 is as in Definition 3.3.

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

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

Now consider the following *multiobjective measure problem* (MMP):

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

By (4.2) and Lemma 4.2, MMP is equivalent to our original multiobjective MCP if we restrict ourselves — which we do in the rest of this paper — to the set

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

in lieu of the set $\Gamma(\Pi)$ in (2.9). On the other hand, from (4.2), (4.5) and Lemma 4.2 we may immediately conclude the following.

Lemma 4.3. $\Gamma(\Pi_0)$ *can be expressed as*

$$\Gamma(\Pi_0) = \{\langle \mu, c \rangle | \mu \in M_\delta(\mathbb{K})\}, \tag{4.9}$$

which is a convex subset of $\mathbb{R}^q_+$.

Actually, the convexity of $\Gamma(\Pi_0)$ is a well–known fact (see e.g. [10, 33, 38]). However, we wish to emphasize here that this convexity is a straightforward, *trivial*, consequence of the MMP formulation: see (4.6) and (4.5). This illustrates the advantage of using the MMP instead of the original multiobjective MCP.

In the following section we use the MMP (4.7) to show the existence of "strong" Pareto policies, and in Section 7 we use it to prove Theorem 3.4.

5 Strong Pareto optimality

For each $i = 1, \dots, q$, 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 $V^* := (V_1^*, \dots, V_q^*)$ 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* (or a *strongly monotonically increasing* function [27]) 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.)

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 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.,

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

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. The following lemma can be obtained from the definition of inf–compactness and Prohorov's Theorem [8].

Lemma 5.3. *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

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

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: see, for instance, statement (12.3.37) in [21, 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

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

This implies that μ satisfies (4.3). □

Proof of Theorem 5.2. (a) By (4.6) and Lemma 4.2, we may express ρ^* in (5.2) as

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

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

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

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

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

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),

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

Finally, let $\pi^* \in \Pi_0$ be the policy associated to μ^*, and use (4.2) to rewrite (5.10) as $\|V(\pi^*) - V^*\| = \rho^*$. 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.By the convexity of $\Gamma(\Pi_0)$ (Lemma 4.3), finding a strong Pareto policy essentially reduces to the problem of finding the distance from the virtual minimum V^* to the convex set $\Gamma(\Pi_0)$. This yields, in particular, that part (b) in Theorem 5.2 can be seen as a special case of the "Minimum Norm Duality" in Luenberger [32, p. 136, Theorem 1]. Hence, as the latter result is true for an arbitrary *normed linear space*, in (5.1) we may take any norm instead of the Euclidean one. For instance, one could take a weighted ℓ_p–norm, with $1 \leq p \leq \infty$, which is very common in vector optimization [27, 31, 36].

Example 5.7. (Example 3.5 continued). Consider again the LQ problem (3.9)–(3.11). For each $i = 1, \ldots, q$, let $V_i^* = k(\gamma_0)z_i$ be the partial minimum in (3.18), 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

$$V^* = k(\gamma_0)z^*. \tag{5.11}$$

Moreover, to find a strong Pareto policy we may proceed as follows. From (5.11) 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 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 [7] 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 [20, §6.3] or [21, §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. Henceforth 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.

In the remainder of this section we suppose that Assumption 3.1 and the following Assumption 6.1 are both satisfied.

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

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

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

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

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

By Assumption 6.1, L^* indeed maps $F(\mathrm{X})$ into $F(\mathbb{K})$, which is equivalent to say that L *is σ–weakly continuous.*

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(d) 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 $u = (u_1, \dots, u_q)$ with $u_i \in F(\mathrm{X})$

for all $i = 1, \dots, q$. 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

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

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

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

Then, from [7, p. 380], we can see that the *dual program* (DP) of (PP) is as follows:

(DP) maximize $\langle (1-\delta)\gamma_0, u \rangle$

(6.11) subject to: $\lambda \cdot L^* u \leq \lambda \cdot c$ with $u \in F(\mathrm{X})^q$, for some $\lambda \in \mathbb{R}^q_{++}$.

In fact, if we let

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

and use (6.7), it then follows that the dual constraint (6.11) can be expressed as in [7], namely:

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

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

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

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

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

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 [20, Lemma 4.2.7], for instance) give that

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

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

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

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 [7] 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

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

Now, to obtain the primal-dual interpretation mentioned in the last part of Remark 6.3, 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 6.3. 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(d)), 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 3.4

Let us first suppose that $\pi^* \in \Pi_0$ is a *proper* Pareto policy. Let $\mu^* \in M_\delta(\mathbb{K})$ be the occupation measure corresponding to π^* (see (4.1)). By (4.2), (4.6) and Lemma 4.2, to prove Theorem 3.4 it suffices to show the existence of a q–vector $\lambda \gg 0$ such that

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

(cf. (3.7)) or, equivalently,

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

with c^λ as in (3.3). With this in mind, consider the set $\Gamma(\Pi_0)$ in (4.9), and let $T_0 := T(\Gamma(\Pi_0), \langle \mu^*, c \rangle)$ be the tangent cone to $\Gamma(\Pi_0)$ at $\langle \mu^*, c \rangle$. As $\Gamma(\Pi_0)$ is convex (Lemma 4.3), we have

$$\Gamma(\Pi_0) - \langle \mu^*, c \rangle \subset T_0. \tag{7.2}$$

(Recall (2.7).) Let B be the set of q–vectors $u < 0$ such that

$$\sum_{i=1}^{q} u_i = -1, \tag{7.3}$$

and note that $T_0 - B$ is a convex set that does not contain the vector zero. Therefore, by a well–known separation theorem (e.g. [5, p. 30], [32, p. 133], there exists a vector $\lambda \neq 0$ such that

$$\lambda \cdot (v - u) > 0 \quad \forall\, v \in T_0,\ u \in B.$$

In particular, by (7.2),

$$\lambda \cdot (\langle \mu, c\rangle - \langle \mu^*, c\rangle) > \lambda \cdot u \quad \forall\, \mu \in M_\delta(\mathbb{K}),\ u \in B, \tag{7.4}$$

and taking $\mu = \mu^*$ we obtain that $\lambda \cdot u < 0$ for all $u \in B$. Therefore, choosing an arbitrary $i \in \{1, \dots, q\}$ and letting $u \in B$ be the vector with components $u_i = -1$ and $u_j = 0$ for $j \neq i$, we conclude that $\lambda_i > 0$; hence, as $i \in \{1, \dots, q\}$ was arbitrary, $\lambda \gg 0$. Thus to complete the proof it only remains to verify that μ^* and λ satisfy (7.1), so that μ^* indeed minimizes $\lambda \cdot \langle \mu, c\rangle = \langle \mu, c^\lambda\rangle$. Suppose that this is not the case and let $\mu \in M_\delta(\mathbb{K})$ be such that $\langle \mu, c^\lambda\rangle < \langle \mu^*, c^\lambda\rangle$, i.e.

$$\langle \mu - \mu^*, c^\lambda\rangle < 0. \tag{7.5}$$

For each $r \geq 0$, let v_r be the vector in T_0 defined as

$$v_r := r(\langle \mu, c\rangle - \langle \mu^*, c\rangle) = r\langle \mu - \mu^*, c\rangle.$$

Then, by (7.5), $\lambda \cdot v_r = r\langle \mu - \mu^*, c^\lambda\rangle \to -\infty$ as $r \to \infty$, which contradicts (7.4). This completes the proof of Theorem 3.4 when π^* is a proper Pareto policy.

Let us now suppose that π^* is a *weak* Pareto policy and let μ^* be the corresponding occupation measure. Then $\langle \mu^*, c\rangle$ is a weak Pareto point of $\Gamma(\Pi_0)$, i.e. there is no $\mu \in M_\delta(\mathbb{K})$ such that $\langle \mu, c\rangle \ll \langle \mu^*, c\rangle$. Let

$$C_1 := \{u \in \mathbb{R}^q | u \ll \langle \mu^*, c\rangle\},$$

$$C_2 := \{u \in \mathbb{R}^q | u \geq \langle \mu, c\rangle \text{ for some } \mu \in M_\delta(\mathbb{K})\}.$$

Then C_1 and C_2 are disjoint convex sets, and in addition C_1 is open. Therefore, by the separation theorem in [32, p. 133, Theorem 3], there is a q–vector $\lambda \neq 0$ and a real number α such that

$$\lambda \cdot u < \alpha \leq \lambda \cdot v \quad \forall\, u \in C_1,\ v \in C_2. \tag{7.6}$$

Moreover, the vector $\langle \mu^*, c\rangle$ is in the intersection of C_2 and the closure of C_1, which yields that $\alpha = \lambda \cdot \langle \mu^*, c\rangle$. Hence the first inequality in (7.6) gives

$$\lambda \cdot (\langle \mu^*, c\rangle - w) \leq \lambda \cdot \langle \mu^*, c\rangle \quad \forall\, w \in \mathbb{R}^q_+,$$

which implies that $\lambda \cdot w \geq 0$ for all $w \geq 0$, and so $\lambda \geq 0$. Thus $\lambda > 0$ because $\lambda \neq 0$. Finally, by the definition of C_2 and the second inequality in (7.6), we obtain that $\lambda \cdot \langle \mu^*, c\rangle \leq \lambda \cdot \langle \mu, c\rangle$ for all $\mu \in M_\delta(\mathbb{K})$, which concludes the proof of Theorem 3.4. □

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

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

is called a *constrained MCP*. In this case, a policy π for which (8.1) holds and, in addition, $V_1(\pi) < \infty$ 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 e.g. [16]), and under an additional Slater–like condition, π^* is also a *Pareto policy* for the multiobjective MCP in Section 2 above; see [30], for instance.

For additional results on constrained MCPs or for MCPs with weighted criteria, see, for instance, [1, 10, 11, 14, 16, 17, 28, 30, 33, 38].

Average cost. Let us rewrite (2.4) as

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

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

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

It is easily verified that 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 $\mathrm{X} \times A$, concentrated on $\mathbb{K}$, and such that (as in (4.4))

$$\widehat{\mu}(B) = \int_{\mathbb{K}} Q(B|x, a)\mu(d(x, a)). \tag{8.4}$$

Similarly, by (8.4), the constraint equation (6.9) in the multiobjective LP formulation becomes

$$L_1\mu = 0, \quad \mu \in M_+(\mathbb{K}), \tag{8.5}$$

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 obtain an optimal policy for the constrained problem, which is a Pareto policy for the multiobjective MCP. For details see [17], where a probability measure μ for which (8.5) holds is called *stable*.

Mixed average-discounted criteria. The average cost case in (8.3)-(8.5) 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 us to do this is that 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 3.7(b) of [17].

Further research: the balance space approach. In this paper we used two main approaches to analyze a multiobjective MCP: the scalarization approach (to study the problem of existence of Pareto policies) and the MMP approach (to study the characterization of Pareto policies). In fact, the former approach is the "dual" of the latter in a precise sense (see Section 6). On the other hand, there is a *nonscalarized* approach called the *balance space approach* introduced by Galperin [13] for vector optimization problems. This approach, in addition to allowing an interesting economic interpretation of the so–called "balance points", has proved to be very effective from the computational viewpoint and also to study key issues, such as the sensitivity of vector minimization problems [6]. It might be worth investigating if this effectiveness also holds for multiobjective *control* problems.

Onésimo Hernández-Lerma
Departamento de Matemáticas,
CINVESTAV-IPN,
A.P. 14-470,
México D.F. 07000,
México.
ohernand@math.cinvestav.mx

Rosario Romera
Departamento de Estadística,
Universidad Carlos III de Madrid
Calle Madrid 126,
Getafe 28903,
Madrid España,
rosario.romera@uc3m.es

References

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

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

[3] Arrow K.J.; Barankin E.W.; Blackwell D.,*Admissible points of convex sets*, Annals of Mathematics Studies, **2** (1950), pp. 87–91.

[4] Aubin J.-P.,*A Pareto minimum principle*, in Differential Games and Related Topics (Proc. Internat. Summer School, Varenna), North-Holland, Amsterdam, 1970, pp. 147-175.

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

[6] Balbás A.; Guerra P.J., *Measuring the balance space sensitivity in vector optimization*, Lecture Notes Economics and Math. Syst., to appear.

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

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

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

[13] Galperin E.A., *Pareto analysis vis–à–vis balance space approach in multiobjective global optimization*, J. Optim. Theory Appl., **93** (1997), 533-545.

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

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

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

[17] Hernández-Lerma O.; González-Hernández J.; López-Martínez R.R., *Constrained average cost Markov control processes in Borel spaces*, SIAM J. Control Optim., **42** (2003), 442-468.

[18] Hernández-Lerma O.; Hoyos-Reyes L.F., *A multiobjective control approach to priority queues*, Math. Meth. Oper. Res., **53** (2001), 265–277.

[19] Hernández-Lerma O; Hoyos-Reyes L.F., *A multiobjective formulation of optimal control problems with additive costs*, Internal Report 286, Departamento de Matemáticas, CINVESTAV–IPN, 2000.

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

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

[22] Hernández-Lerma O.; Vega-Amaya O.; Carrasco G., *Sample-path optimality and variance-minimization of average cost Markov control processes*, SIAM J. Control Optim., **38** (2000), 79-93.

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

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

[25] Hilgert N.; Hernández-Lerma O., *Bias optimality versus strong 0-discount optimality in Markov control processes with unbounded costs*, Acta Appl. Math., **77** (2003), 215-235.

[26] Hordijk A.; Yushkevich A.A., *Blackwell optimality in the class of all policies in Markov decision chains with a Borel state space and unbounded rewards*, Math. Meth. Oper. Res., **50** (1999), 421-448.

[27] Jahn J., *Theory of vector maximization: various concepts of efficient solutions*, Internat. Ser. Oper. Res. Management Sci., **21** (1999). Chapter 2.

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

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

[30] López-Martínez R.R.; Hernández-Lerma O., *The Lagrange and Pareto approaches to constrained Markov control processes*, Morfismos, **7** (2003), 1-26.

[31] Luc D.T., Theory of Vector Optimization, Lecture Notes in Economics and Math. Systs. **319**, Springer–Verlag, Berlin, 1989.

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

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

[34] Reid R.W.; Citron S.J., *On noninferior performance index vectors*, J. Optim. Theory Appl., **7** (1971), 11-28.

[35] Salukvadze M.E., Vector-Valued Optimization Problems in Control Theory, Academic Press, New York, 1979.

[36] Sawaragi Y.; Nakayama H.; Tanino T., Theory of Multiobjective Optimization, Academic Press, New York, 1985.

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

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

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

[40] Thomas L.C., *Constrained Markov decision processes as multi-objective problems*, Inst. Math. Appl. Conf. Ser., (1983), 77-94.

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

[42] Wakuta K.; Togawa K., *Solution procedures for multi-objective Markov decision processes*, Optimization, **43** (1998), 29-46.

[43] Waltz F.M., *An engineering approach: hierarchical optimization criteria*, IEEE Trans. Autom. Control, **12** (1967), 179-180.

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

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

[46] Zadeh L.A., *Optimality and non-scalar-valued performance criteria*, IEEE Trans. Autom. Control, **8** (1963), 59-60.