Math Meth Oper Res (2002) 55:359–369

Mathematical Methods
of Operations Research

A multicriteria competitive Markov decision process

A. M. Rodríguez-Chía[a], J. Puerto[b] & F. R. Fernández[b]

[a]Facultad de Ciencias del Mar. Universidad de Cádiz. Pol. Río San Pedro, 11510 Puerto Real (Cádiz), Spain (e-mail: antonio.rodriguezchia@uca.es)
[b]Facultad de Matemáticas. Univ. de Sevilla. C/Tarfia s/n. 41012 Sevilla, Spain
(e-mail: puerto@us.es)

Manuscript received: October 2001/Final version received: January 2002

Abstract. In this paper, we deal with a multicriteria competitive Markov decision process. In the decision process there are two decision makers with a competitive behaviour, so they are usually called players. Their rewards are coupled because depend on the actions chosen by both players in each state of the process. We propose as solution of this game the set of Pareto-optimal security strategies for any of the players in the original problem. We show that this solution set can be obtained as the efficient solution set of a multicriteria linear programming problem.

Key words: Game Theory, multicriteria games, vector linear programming

1 Introduction

Recently, much attention has been paid to game problems in which the payoff is a multiple noncomparable criteria vector [7, 8, 15]. One of the reasons is that this approach represents better some real world applications of Game Theory [1, 14]. In fact, each competitive situation that can be modeled as a scalar zero-sum game has its counterpart as a multicriteria zero-sum game when more than one scenario has to be compared simultaneously. In these situations, once the same strategy has to be used in different scenarios, conflicting interests appear between different decision markers as well as within each individual. For instance, the production policies of two firms which are competing for a market can be seen as a scalar game. However, when they compete simultaneously in several markets the multicriteria approach has to be adopted.

Since the payoff is represented by a vector, there not exists a total order among the payoffs. Hence, the classical concept of solution of scalar games can not be used for this problem. For this reason, new solution concepts have been proposed in recent years [2, 7, 9, 15] and have been compared with existing ones. Particularly, the concept of Pareto-optimal security strategy (POSS) becomes very important in order to solve matrix multicriteria games [3, 7].

On the other hand, the games considered in this paper are stochastic games that are a generalization of the Markov decision processes to the case of two decision makers [6]. Hence, the fortunes of each player are coupled because the probability transition and the rewards depends of the actions chosen by both players. Although multicriteria versions of the classical Markov decision process has been considered in the literature (see [10, 11]) the multicriteria version of the stochastic game has not been considered before. Therefore, in this paper we obtain the set of POSS for this kind of games, and we give an easy method to compute these solutions.

The paper is organized in three sections. In the second section we give the notation used in the paper. In Section 3, we present the multicriteria stochastic game that we are going to study and show that their POSS solution set coincides with the set of efficient solutions of a multicriteria linear programming problem. Finally, this result is illustrated with an example.

2 Definitions and notations

We shall consider a process that is observed at discrete time points $t = 1, 2, 3, \ldots$ that will sometimes be called stages. At each time t, the state of the process will be denoted by S_t. We assume that S_t is a random variable that can take on values from the finite set $S = \{1, \ldots, N\}$ which from now on will be called the state space. The sentence "the process is in state s at time t" will be synonymous with the event $\{S_t = s\}$.

We also assume that the process is controlled by two controllers or decision makers who will be referred to as player 1 and player 2, respectively. Thus, if the process is in state $s \in S = \{1, \ldots, N\}$ at time t, player 1 and 2 independently choose actions $a^1 \in A^1(s)$ and $a^2 \in A^2(s)$ and receive rewards vectors $(r_1^1(s, a^1, a^2), \ldots, r_k^1(s, a^1, a^2))$ and $(r_1^2(s, a^1, a^2), \ldots, r_k^2(s, a^1, a^2))$, respectively. Notice that the main difference with respect to the standard competitive Markov decision processes is that in our model each player receives k $(k > 1)$ different rewards. These rewards may be non-comparable quantities as money, number of employees, etc. Furthermore, we assume that the stationary transition probabilities depend on the actions of one player (this is usually called the single controller case, see [4, 5]). If we consider that such player is the player 1, the transition probabilities are given by

$$P(s'|s, a^1) = P(S_{t+1} = s' \mid S_t = s, A_t^1 = a^1)$$

for all $t = 0, 1, 2, \ldots$. S_t is the state at time t, and A_t^1 denotes the action chosen by player 1 at time t.

Note that the fact that the rewards and the transition probabilities depend on the actions of both players, as well as on the current state, implies that the "fate" of the two players is coupled in this process, even though their choices of actions are independent of one another.

The property that the decisions of each controller in state s are invariant with respect to the time of visit to s sometimes is called the stationarity of the strategy.

Let F_S be the set of stationary strategies of player 1 and G_S the ones of player 2. Note that, if $f = (f(1), \dots, f(N)) \in F_S$, then each $f(s)$ is a $m^1(s)$-dimensional probability vector, where $m^1(s) = |A^1(s)|$, the cardinality of $A^1(s)$. Denote $f(s) := (f(s,1), \dots, f(s, m^1(s)))$ where

$$f(s,a^1) = \text{Probability that player 1 chooses action } a^1 \in A^1(s) \text{ in state } s \in S \text{ whenever is visited,}$$

verifying that $\sum_{a^1=1}^{m^1(s)} f(s,a^1) = 1$. In the same way, $g = (g(1), \dots, g(N)) \in G_S$, then each $g(s)$ is a $m^2(s)$-dimensional probability vector, where $m^2(s) = |A^2(s)|$, the cardinality of $A^2(s)$. Denote $g(s) := (g(s,1), \dots, g(s, m^2(s)))^T$ (notice that the superscript T either over a vector or a matrix represents the corresponding transpose) where

$$g(s,a^2) = \text{Probability that player 2 chooses action } a^2 \in A^2(s) \text{ in state } s \in S \text{ whenever is visited,}$$

verifying that $\sum_{a^2=1}^{m^2(s)} g(s,a^2) = 1$.

A strategy $f(g)$ will be called pure or deterministic if $f(s,a^1) \in \{0,1\}$ for all $a^1 \in A^1(s)$, $s \in S$ $(g(s,a^2) \in \{0,1\}$, for all $a^2 \in A^2(s)$, $s \in S)$. That is, for each $s \in S$ a pure control selects one particular action a_s^j with $j = 1,2$ with probability 1 in state s whenever this state is visited.

It can be easily seen that a strategy f defines a probability transition matrix

$$P(s'|s,f) = \sum_{a^1=1}^{m^1(s)} f(s,a^1) p(s'|s,a^1) := f(s) P(s'|s).$$

Whenever a superscript 1 (or 2) is associated to a symbol in this section, it is there to denote a quantity associated with player 1 (or 2). Now, we denote by $(R_{1,t}^1, \dots, R_{k,t}^1)$ $((R_{1,t}^2, \dots, R_{k,t}^2))$ the reward vector at time t to player 1 (player 2), and $(r_1^1(s,f,g), \dots, r_k^1(s,f,g))$ $((r_1^2(s,f,g), \dots, r_k^2(s,f,g)))$ denoting the immediate expected reward vector to player 1 (player 2) in the state s, corresponding to a strategy pair $(f,g) \in F_S \times G_S$. The immediate expected reward in the state s for player 1 and 2, by choosing actions $a^1 \in A^1(s)$ and $a^2 \in A^2(s)$ is given by the vectors $(r_1^1(s,a^1,a^2), \dots, r_k^1(s,a^1,a^2))$ and $(r_1^2(s,a^1,a^2), \dots, r_k^2(s,a^1,a^2))$, respectively. In this paper, we consider a stochastic zero-sum game, that is,

$$r_l^1(s,a^1,a^2) + r_l^2(s,a^1,a^2) = 0 \quad \forall l = 1, \dots, k$$

for all $s \in S$, $a^1 \in A^1(s)$, $a^2 \in A^2(s)$. Thus, we may drop the superscript 1 and 2 in the reward functions by defining

$$r_l(s,a^1,a^2) := r_l^1(s,a^1,a^2) = -r_l^2(s,a^1,a^2) \quad \forall l = 1, \dots, k.$$

Therefore, for the strategies f, g in the state s we obtain the following expected rewards:

$$r_l(s,f,a^2) = \sum_{a^1=1}^{m^1(s)} f(s,a^1) r_l(s,a^1,a^2) := f(s) R_l(s,a^2)$$

$$r_l(s,a^1,g) = \sum_{a^2=1}^{m^2(s)} r_l(s,a^1,a^2)g(s,a^2) := R_l(s,a^1)g(s) \tag{1}$$

$$r_l(s,f,g) = \sum_{a^1=1}^{m^1(s)} \sum_{a^2=1}^{m^2(s)} f(s,a^1)r_l(s,a^1,a^2)g(s,a^2) := f(s)R_l(s)g(s) \tag{2}$$

for $l = 1,\dots,k$ and for all $s \in S$. We denote by:

$$r_l(f,g) = (r_l(1,f,g),\dots,r_l(N,f,g))^T, \tag{3}$$

the vector of the l-th reward in the states of the process.

3 Multicriteria β-discounted Markov decision model

Let $\{R_t\}_{t=0}^{\infty}$ denote the sequence of random reward vector for the period $[t,t+1)$, where $R_t = (R_{1,t},\dots,R_{k,t})$. It should be clear that once an initial state s as well as strategies f and g are specified, then so is the probability distribution of R_t for every $t = 0,1,2,\dots$. Thus, the expectation of R_t is also well defined and will be denoted by

$$\begin{aligned}\mathrm{E}_{sfg}(R_t) &:= (\mathrm{E}_{sfg}(R_{1,t}),\dots,\mathrm{E}_{sfg}(R_{k,t}))\\ &:= (\mathrm{E}_{fg}(R_{1,t}\,|\,S_0 = s),\dots,\mathrm{E}_{fg}(R_{k,t}\,|\,S_0 = s)).\end{aligned}$$

The expected reward in the criterion l, at stage t resulting from (f,g) and an initial state s, $\mathrm{E}_{sfg}(R_t)$, is

$$E_{sfg}(R_{l,t}) = \sum_{s'=1}^{N} p_t(s'|s,f)r_l(s',f,g) := P_t(s,f)r_l(f,g) \quad \forall l = 1,\dots,k$$

where $p_t(s'|s,f)$ is the t-step transition probability from s to s' in the Markov chain defined by f and $r_l(f,g)$ was defined in (3).

Definition 3.1. *The overall discounted value of a strategy pair $(f,g) \in F_S \times G_S$ from the initial state s and for each $l = 1,\dots,k$ will be given by*

$$\mathrm{v}_{\beta,l}(s,f,g) := \sum_{t=0}^{\infty} \beta^t \mathrm{E}_{sfg}(R_{l,t}) \quad \forall l = 1,\dots,k, \forall s \in S$$

where $\beta \in [0,1)$ is called the discount factor. We denote

$$\mathrm{v}_{\beta,l}(f,g) = (\mathrm{v}_{\beta,l}(1,f,g),\dots,\mathrm{v}_{\beta,l}(N,f,g))^T.$$

Definition 3.2. *The multicriteria competitive discounted Markov decision process for a strategy pair $(f,g) \in F_S \times G_S$ and the initial state s, is the model that uses as criterion the vector;*

$$(\mathrm{v}_{\beta,1}(f,g),\dots,\mathrm{v}_{\beta,k}(f,g)),$$

and it is denoted by Γ_β.

It is well-known from Markov chain theory that the t-th power of $P(f)$ contains all such t-step transition probabilities, that is,

$$P^t(s,f) = (p_t(1|s,f),\dots,p_t(N|s,f))$$

$$P^t(f) = (p_t(s'|s,f))_{s=1,s'=1}^{N,N}.$$

Using the notation above, we obtain that

$$\mathrm{v}_{\beta,l}(s,f,g) = \sum_{t=0}^{\infty} \beta^t P^t(s,f) r_l(f,g) \quad \forall l = 1,\dots,k, \forall s \in S. \tag{4}$$

This expression can be rewritten as

$$\mathrm{v}_{\beta,l}(f,g) = \sum_{t=0}^{\infty} \beta^t P^t(f) r_l(f,g) \quad \forall l = 1,\dots,k, \tag{5}$$

where $P^0(f) := I_N$ the $N \times N$ identity matrix. It is well known that $(I - \beta P(f))$ is an invertible matrix and that

$$(I - \beta P(f))^{-1} = I + \beta P(f) + \beta^2 P^2(f) + \beta^3 P^3(f) + \cdots$$

Substituting the expression above into (5) we obtain the following compact matrix expression for the discounted value vector of f and g

$$\mathrm{v}_{\beta,l}(f,g) = (I - \beta P(f))^{-1} r_l(f,g) \quad l = 1,\dots,k,$$

or equivalently:

$$\mathrm{v}_{\beta,l}(f,g) = r_l(f,g) + \beta P(f) \mathrm{v}_{\beta,l}(f,g) \quad l = 1,\dots,k.$$

Recall that $\mathrm{v}_{\beta,l}(f,g) = (\mathrm{v}_{\beta,l}(1,f,g),\dots,\mathrm{v}_{\beta,l}(N,f,g))^T$ and that its components are then given by

$$\mathrm{v}_{\beta,l}(s,f,g) = r_l(s,f,g) + \beta P(s,f) \mathrm{v}_{\beta,l}(f,g) \quad l = 1,\dots,k \ s = 1,\dots,N.$$

Within the space of strategies $F_S \times G_S$, for players 1 and 2 we need to decide on a pair (f,g) of strategies that constitutes a "solution" to the game. It is clear that the ideal solution would be (f^*,g^*) such that

$$\mathrm{v}_{\beta,l}(f,g^*) \leq \mathrm{v}_{\beta,l}(f^*,g^*) \leq \mathrm{v}_{\beta,l}(f^*,g) \quad \forall l = 1,\dots,k \quad \text{and}$$
$$\forall (f,g) \in F_S \times G_S.$$

However, the ideal pair (f^*,g^*) may not exist in the most cases due to the vectorial character of $\mathrm{v}_{\beta,l}$ (see [2] for conditions in multicriteria matrix games). Therefore, we need to propose an alternative solution concept that can be applied in any case. In order to do that we use the concept of security levels. Every strategy $g \in G_S$ ($f \in F_S$) defines security levels $\overline{\mathrm{v}}_{\beta,l}(s,g)$ for all $s \in S$ ($\underline{\mathrm{v}}_{\beta,l}(s,f)$) as the payoffs with respect to every criterion $\mathrm{v}_{\beta,l}(s,f,g)$ $l = 1,\dots,k$ when player 2 (respectively player 1) bets to minimize (respectively, maximize) the criterion. Hence,

$$\overline{\mathrm{v}}_{\beta,l}(s,g) = \max_{f \in F_S} \mathrm{v}_{\beta,l}(s,f,g) \quad l = 1,\dots,k, \forall s \in S,$$

$$\underline{\mathrm{v}}_{\beta,l}(s,f) = \min_{g \in G_S} \mathrm{v}_{\beta,l}(s,f,g) \quad l = 1,\dots,k, \forall s \in S.$$

The security level vectors are k-tuples denoted by

$$\overline{\mathrm{v}}_{\beta}(s,g) = (\overline{\mathrm{v}}_{\beta,1}(s,g),\dots,\overline{\mathrm{v}}_{\beta,k}(s,g)) \quad \forall s \in S.$$

$$\underline{\mathrm{v}}_{\beta}(s,f) = (\underline{\mathrm{v}}_{\beta,1}(s,f),\dots,\underline{\mathrm{v}}_{\beta,k}(s,f)) \quad \forall s \in S.$$

Notice that the security level $\overline{\mathrm{v}}_{\beta,l}(s,g)(\underline{\mathrm{v}}_{\beta,l}(s,f))$ represents the maximum loss (minimum reward) that player 2 (player 1) can get when he chooses the strategy $g(f)$. Thus, a possible solution for player 2 (player 1) would be to find the strategy $g^*(f^*)$ such that $\overline{\mathrm{v}}_{\beta,l}(s,g^*) = \min_{g \in G_S} \overline{\mathrm{v}}_{\beta,l}(s,g)$ $(\underline{\mathrm{v}}_{\beta,l}(s,f^*) = \max_{f \in F_S} \underline{\mathrm{v}}_{\beta,l}(s,f))$. However, since we are considering a vectorial optimization problem these g^* or f^* must be understood as Pareto-optimal solutions.

Let us denote by

$$\overline{\mathrm{v}}_{\beta,l}(g) = (\overline{\mathrm{v}}_{\beta,l}(1,g),\dots,\overline{\mathrm{v}}_{\beta,l}(N,g))^T \quad \forall l = 1,\dots,k.$$

$$\underline{\mathrm{v}}_{\beta,l}(f) = (\underline{\mathrm{v}}_{\beta,l}(1,f),\dots,\underline{\mathrm{v}}_{\beta,l}(N,f))^T \quad \forall l = 1,\dots,k,$$

and

$$\overline{\mathrm{v}}_{\beta}(g) = (\overline{\mathrm{v}}_{\beta,1}(g),\dots,\overline{\mathrm{v}}_{\beta,k}(g))$$

$$\underline{\mathrm{v}}_{\beta}(f) = (\underline{\mathrm{v}}_{\beta,1}(f),\dots,\underline{\mathrm{v}}_{\beta,k}(f)).$$

We note in passing that, for a given strategy $g \in G_S$ for player 2, the security levels $\overline{\mathrm{v}}_{\beta,l}(g)$ for $l = 1,\dots,k$, might be obtained for player 1 by different strategies. Notice that in the scalar case ($k = 1$) the POSS solution coincide with the classical min max solution. In addition, for a general value of $k > 1$ when there exists the ideal solution, (f^*, g^*), to the game we have that

$$\mathrm{v}_{\beta}(s,f^*,g^*) = \overline{\mathrm{v}}_{\beta}(s,g^*) = \underline{\mathrm{v}}_{\beta}(s,f^*) \quad \forall s \in S.$$

Using our notation, we now adapt the definition of POSS given in [7] to this different class of games.

Definition 3.3. *A strategy $g^* \in G_S$ is a Pareto-optimal security strategy (POSS) for player 2 iff there is no $g \in G_S$ such that $\overline{\mathrm{v}}_{\beta}(g^*) \geq \overline{\mathrm{v}}_{\beta}(g), \overline{\mathrm{v}}_{\beta}(g^*) \neq \overline{\mathrm{v}}_{\beta}(g)$. Similarly, one can define POSS for player 1.*

POSS always exist provided that the criterion vectors $(v_{\beta,l}(f,g))$ are continuous and the decision space is compact (see Corollary 3.2.1. in [12]).

Once we have defined the POSS solution concept, we characterize the whole set of Pareto-optimal security strategies. To do that, we propose the following multicriteria linear programming problem that we will use in the next theorem:

$$\min \quad \sum_{s=1}^{N} v_1(s), \ldots, \sum_{s=1}^{N} v_k(s)$$

$s.t.$

$$\begin{aligned}
& v_l(s) \geq R_l(s, a^1)g(s) + \beta P(s, a^1)v_l \quad \forall s \in S, \forall a^1 \in A^1(s), l = 1, \ldots, k \\
& \sum_{a^2 \in A^2(s)} g(s, a^2) = 1 \quad \forall s \in S \\
& g(s, a^2) \geq 0 \quad \forall a^2 \in A^2(s), \forall s \in S \\
& v_l = (v_l(1), \ldots, v_l(N))^T, \quad l = 1, \ldots, k.
\end{aligned} \tag{6}$$

Theorem 3.1. *The Pareto-solution set of Problem (6) coincides with the Pareto-optimal security strategies set of Game Γ_β.*

Proof. Let us consider the following problem

$$\min \quad v_1, \ldots, v_k$$

$s.t.$

$$\begin{aligned}
& v_l(s) \geq r_l(s, f, g) + \beta P(s, f)v_l \quad \forall s \in S, \forall f \in F_S, l = 1, \ldots, k \\
& \sum_{a^2 \in A^2(s)} g(s, a^2) = 1 \quad \forall s \in S \\
& g(s, a^2) \geq 0 \quad \forall a^2 \in A^2(s), \forall s \in S \\
& v_l = (v_l(1), \ldots, v_l(N))^T, \quad l = 1, \ldots, k
\end{aligned}$$

Since the strategies chosen by both players in each state are independent, the problem above can be equivalently formulated as follows;

$$\min \quad \sum_{s=1}^{N} v_1(s), \ldots, \sum_{s=1}^{N} v_k(s)$$

$s.t.$

$$\begin{aligned}
& v_l(s) \geq r_l(s, f, g) + \beta P(s, f)v_l \quad \forall s \in S, \forall f \in F_S, l = 1, \ldots, k \\
& \sum_{a^2 \in A^2(s)} g(s, a^2) = 1 \quad \forall s \in S \\
& g(s, a^2) \geq 0 \quad \forall a^2 \in A^2(s), \forall s \in S \\
& v_l = (v_l(1), \ldots, v_l(N))^T, \quad l = 1, \ldots, k
\end{aligned} \tag{7}$$

Any feasible solution $(v_l(s))_{\substack{1\le l\le k\\ s\in S}}$ satisfies for any $g\in G_S$,

$$v_l(s)\ge r_l(s,f,g)+\beta P(s,f)v_l \quad \forall s\in S, \forall f\in F_S$$

The above expression can be equivalently written after rearranging the components of $(v_l(s))_{\substack{1\le l\le k\\ s\in S}}$ as

$$v_l\ge r_l(f,g)+\beta P(f)v_l \quad \forall f\in F_S.$$

It implies that

$$\begin{aligned} v_l &\ge r_l(f,g)+\beta P(f)v_l\\ &\ge r_l(f,g)+\beta P(f)r_l(f,g)+\beta^2P^2(f)v_l\\ &\vdots \qquad\qquad \vdots\\ &\ge r_l(f,g)+\beta P(f)r_l(f,g)+\beta^2P^2(f)r_l(f,g)+\beta^3P^3(f)r_l(f,g)+\cdots\\ &=[I-\beta P(f)]^{-1}r_l(f,g)=\mathrm{v}_{\beta,l}(f,g) \quad \forall f\in F_S. \end{aligned}$$

Thus,

$$v_l\ge \overline{\mathrm{v}}_{\beta,l}(g) \quad \forall l=1,\dots,k. \tag{8}$$

Hence, notice that the solutions of Problem (7) give us strategies with a worse objective value than the security strategies for any $g\in G_S$.

Therefore, given $g\in G_S$, we have that $\overline{\mathrm{v}}_{\beta,l}(g)\ge \mathrm{v}_{\beta,l}(f,g)$ for all $f\in F_S$. Besides, since $\mathrm{v}_{\beta,l}(f,g)=(I-\beta P(f))^{-1}r_l(f,g)$ then $\overline{\mathrm{v}}_{\beta,l}(g)\ge (I-\beta P(f))^{-1}\cdot r_l(f,g)$ $\forall f\in F_S$, that is, $\overline{\mathrm{v}}_{\beta,l}(g)\ge r_l(f,g)+\beta P(f)\overline{\mathrm{v}}_{\beta,l}(g)$ $\forall f\in F_S$. Thus, we obtain that for a given $g\in G_S$ the vector $(\overline{\mathrm{v}}_{\beta,1}(g),\dots,\overline{\mathrm{v}}_{\beta,k}(g),g)$ is a feasible solution in Problem (7).

Moreover, let $(\mathrm{v}_1^*,\dots,\mathrm{v}_k^*,g^*)$ be a Pareto-optimal solution of Problem (7). By (8) since $g^*\in G_S$ we must have that $\mathrm{v}_l^*\ge\overline{\mathrm{v}}_{\beta,l}(g^*)$ $l=1,\dots,k$. Therefore, since $(\overline{\mathrm{v}}_{\beta,1}(g^*),\dots,\overline{\mathrm{v}}_{\beta,k}(g^*),g^*)$ is feasible we must have that $\mathrm{v}_l^*=\overline{\mathrm{v}}_{\beta,l}(g^*)$ $l=1,\dots,k$. Otherwise, $(\mathrm{v}_1^*,\dots,\mathrm{v}_k^*,g^*)$ would not be Pareto-optimal. This argument proves that the Pareto-optimal security levels and the Pareto-optimal security strategies for the multicriteria β-discounted Markov decision problem are given by the Pareto-optimal solutions of Problem (7).

Finally, we prove that Problem (7) is equivalent to Problem (6). Using (2), the first constraint of Problem (7) can be written as:

$$v_l(s)\ge r_l(s,f,g)+\beta P(s,f)v_l=f(s)R_l(s)g(s)+\beta f(s)P(s)v_l$$
$$\forall s\in S, \forall f\in F_S,$$

where $P(s)=(p(s'|s,a^1))_{a^1=1,s'=1}^{m^1(s),N}$. Since, the second part of this inequality is linear in $f(s)$ and $f(s)$ is a $m^1(s)$ dimensional probability vector, that is,

$f(s) \geq 0$ and $\sum_{a^1=1}^{m^1(s)} f(s,a^1) = 1$, we have that this constraint is valid to any $a^1 \in A^1(s)$. Thus, it can be formulated equivalently as:

$$v_l(s) \geq r_l(s,a^1,g) + \beta P(s,a^1)v_l \quad \forall s \in S, \forall a^1 \in A^1(s).$$

Hence, Problem (7) can be rewritten equivalently as;

$$\begin{aligned}
\min \quad & \sum_{s=1}^{N} v_1(s), \ldots, \sum_{s=1}^{N} v_k(s) \\
s.t. \quad & \\
& v_l(s) \geq r_l(s,a^1,g) + \beta P(s,a^1)v_l \quad \forall s \in S, \forall a^1 \in A^1(s), l = 1,\ldots,k \\
& \sum_{a^2 \in A^2(s)} g(s,a^2) = 1 \quad \forall s \in S \\
& g(s,a^2) \geq 0 \quad \forall a^2 \in A^2(s), \forall s \in S \\
& v_l = (v_l(1),\ldots,v_l(N))^T, \quad l = 1,\ldots,k. \qquad (9)
\end{aligned}$$

Notice, that using the equality (1), this problem is equivalent to Problem (6). □

It is worth noting the relationship existing between POSS solutions for this kind of multicriteria and minmax strategies in the scalar case. In fact, both can be obtained solving adequate linear programs: in the scalar case the optimal solutions are obtained using simplex algorithm and in the multiple criteria case using multicriteria simplex algorithm (Adbase [13]).

Example 3.1. Let $S = \{1,2\}$, $A^1(s) = A^2(s) = \{1,2\}$ for $s \in S, \beta = 0.7$ and the reward and transition data be

$$(R_1(1), R_2(1)) = \begin{pmatrix} (10,6) & (-6,4) \\ (-4,0) & (8,3) \end{pmatrix}$$

$$(R_1(2), R_2(2)) = \begin{pmatrix} (-2,0) & (5,3) \\ (4,2) & (-10,-10) \end{pmatrix}$$

For $s = 1$,

$$P(1) = (p(s'|1,a^1))_{a^1=1,s'=1}^{2,2} = \begin{pmatrix} 0.5 & 0.5 \\ 0.8 & 0.2 \end{pmatrix}$$

For $s = 2$,

$$P(2) = (p(s'|2,a^1))_{a^1=1,s'=1}^{2,2} = \begin{pmatrix} 0.3 & 0.7 \\ 0.9 & 0.1 \end{pmatrix}$$

For these data the formulation of Problem (6) is as follows:

$$\min \quad v_1(1)+v_1(2), v_2(1)+v_2(2)$$

s.t

$$v_1(1) \geq 10g(1,1) - 6g(1,2) + 0.7(0.5v_1(1) + 0.5v_1(2))$$

$$v_1(1) \geq -4g(1,1) + 8g(1,2) + 0.7(0.8v_1(1) + 0.2v_1(2))$$

$$v_2(1) \geq 6g(1,1) + 4g(1,2) + 0.7(0.5v_2(1) + 0.5v_2(2))$$

$$v_2(1) \geq 0g(1,1) + 3g(1,2) + 0.7(0.8v_2(1) + 0.2v_2(2))$$

$$v_1(2) \geq -2g(2,1) + 5g(2,2) + 0.7(0.3v_1(1) + 0.7v_1(2))$$

$$v_1(2) \geq 4g(2,1) - 10g(2,2) + 0.7(0.9v_1(1) + 0.1v_1(2))$$

$$v_2(2) \geq 0g(2,1) + 3g(2,2) + 0.7(0.3v_2(1) + 0.7v_2(2))$$

$$v_2(2) \geq 2g(2,1) - 10g(2,2) + 0.7(0.9v_2(1) + 0.1v_2(2))$$

$$g(1,1) + g(1,2) = 1$$

$$g(2,1) + g(2,2) = 1$$

$$g(1,1), g(1,2), g(2,1), g(2,2) \geq 0$$

The extreme Pareto-optimal solutions $((v_1(1), v_1(2), v_2(1), v_2(2), g(1,1), g(1,2), g(2,1), g(2,2)))$ of the problem above are:

$$\{(6.86, 7.11, 12.31, 8.58, 0.5, 0.5, 0.4, 0.6), (44.62, 83.09, 9.99, 7.13, 0, 1, 0.48, 0.52), (10.46, 8.32, 11.76, 8.24, 0.38, 0.62, 0.42, 0.58)\}.$$

Notice that the last four components of these vectors give the Pareto-optimal security strategies for player 2.

4 Concluding remarks

In this paper, we have presented an extension of a multicriteria Markov decision process where there exist two decision makers with opposite objectives.

In order to solve this game, we note that its objective function is not linear what implies additional difficulties to deal with. However, we show that the POSS solution set of this game coincides with the Pareto-solution set of a multicriteria linear programming problem. Therefore, we reduce the resolution of this game to solve a multicriteria linear programming problem, which can be done by well-known algorithms (see [13]).

Acknowledgements. The research of the authors is partially supported by Spanish research grants PB97-0707, BMF2001-2378 and BMF2001-4028.

References

[1] Bergstresser K, Yu PL (1977) Domination structures and multicriteria problem in n-person games. Theory and Decision 8(1):5–48

[2] Corley HW (1985) Games with vector payoffs. Journal of Mathematical Analysis and Applications 47(4):491–498

[3] Fernández FR, Puerto J (1996) Vector linear programming in zero-sum multicriteria matrix games. Journal of Optimization Theory and Applications 89(1):115–127

[4] Filar J, Raghavan TES (1984) A matrix game solution to a single-controller stochastic game. Mathematics of Operations Research 9:356–362

[5] Filar J (1986) Quadratic programming and the single-controller stochastic game. Journal of Mathematics Analysis and Applications 113:136–147

[6] Filar J, Vrieze K (1997) Competitive Markov decision processes. Springer Verlag

[7] Ghose D, Prassad R (1989) Solution concepts in two-person multicriteria games. Journal of Optimization Theory and Applications 63(2):167–189

[8] Ghose D (1991) A neccesary and sufficient condition for Pareto-optimal security strategies in multicriteria matrix games. Journal of Optimization Theory and Applications 68(3):463–480

[9] Nieuwenhuis JW (1983) Some minimax theorems in vector-valued functions. Journal of Optimization Theory and Applications 40(3):463–475

[10] Novák J (1989) Linear programmaing in vector criterion Markov and Semi-Markov decision processes. Optimization 20(5):651–670

[11] Novák J (1991) VMDP-program for solving optimality problems in vector criterion Markov and Semi-Markov decision processes. Optimization 22(2):239–248

[12] Sawaragi Y, Nakayama H, Tanino T (1985) Theory of multiobjective optimization. Academic Press

[13] Steuer RE (1986) Multiple criterion optimization: theory, computation and applications. Wiley, New York

[14] Szidarovsky F, Gershon ME, Duckstein L (1986) Techniques for multiobjective decision making in systems management. Elsevier, Amsterdam, Holland

[15] Wang SY (1993) Existence of a pareto equilibrium. Journal of Optimization Theory and Applications 79(2):373–384