Southwest Fisheries Center Administrative Report No. 8H, 1977

THE LINEAR VECTOR MAXIMIZATION PROBLEM AND MULTIOBJECTIVE MARKOV DECISION PROCESSES

Roy Mendelssohn
Southwest Fisheries Center
National Marine Fisheries Service, NOAA
Honolulu, HI 96812

DRAFT FOR COMMENT

April 1977

Markov decision processes have been studied extensively (see Howard 1960; Manne 1960; Denardo 1968, 1977, for example) due to the ability of generally available programming procedures, such as linear programming, to solve a wide class of problems.[1/] These models consider finite state models, with states 1, 2, ..., N; observed at equally spaced epochs. Associated with each state i is a non-empty decision set D_i. Whenever a state i is observed, some decision $k \in D_i$ must be chosen. Reward R_i^k is received immediately, and a transition is made with probability P_{ij}^k. It is assumed that $\Sigma_{j=1}^N P_{ij}^k = 1$ for all i, k, so that the matrix P^k is stochastic. Assume also that P^k has only one ergodic chain (Denardo 1977 shows how to relax this assumption). Let V^n be the value of following some stationary policy. Then it is well known that:

$$V^n = ng + w + o(1) \qquad \text{as} \qquad n \to \infty$$

It is assumed that it is desired to optimize g, the gain, and then out of all gain-optimal policies, choose the one that has the largest bias w. Let Δ be the set of all stationary policies. Then:

$$g^* = \max \{g^\delta | \delta \varepsilon \Delta\}$$

$$w_i^* = \max \{w_\ell^\delta | \delta \varepsilon \Delta,\ g^\delta = g^*\} \qquad i = 1, \ldots, N$$

It is proven in Manne (1960) or Denardo (1977) that a gain-optimal stationary policy must satisfy the following linear program:

[1/] The notation and discussion on Markov decision problems follows Denardo (1977).

Minimize g

subject to: Program (1)

$$g + z_i - \Sigma_j P^k_{ij} z_j \geq R^k_i \qquad \text{all } i, k$$

$$g, z_i \text{ unrestricted}$$

or equivalently the dual program:

$$\text{Maximize } \Sigma_i \Sigma_k x^k_i R^k_i$$

subject to: Program (2)

$$\Sigma_i \Sigma_k x^k_i = 1$$

$$\Sigma_k x^k_i - \Sigma_j \Sigma_k x^k_j P^k_{ij} = 0 \qquad i = 2, \ldots, N$$

$$x^k_i \geq 0 \qquad \text{all } i, k$$

Suppose instead of one reward each period, we are concerned about a vector of rewards $R^k_i = (R^k_{i,\,1}, \ldots, R^k_{i,\,j})$, while retaining the rest of the model as described. We consider again only stationary policies. Sobel (1977) proves that an ordinal sequential game possesses an undominated solution amongst stationary policies if it has one at all. Restricting ourselves to stationary policies is thus reasonable, and in many applications for infinite horizon problems stationary policies are the real policies of interest. Arguing as before, it is true for each objective ℓ, $1 \leq \ell \leq j$, that the total expcected value of each stationary policy satisfies

$$V^n_\ell = ng_\ell + w_\ell + o(1)$$

Therefore, for each objective, its gain-optimal policy satisfies programs (1) or (2). Denote the constraints in each of these programs, for each objective ℓ, as C^1_ℓ, C^2_ℓ.

Let $J^1(x), \ldots, J^j(x)$ be functions mapping a subset of $\mathbb{R}^j$ into $\mathbb{R}$ for $x \in X$, and let $J(x) = (J^1(x), \ldots, J^j(x))$. Let Ω be the set of feasible values of $J(x)$. Then a particular value of $J(x)$, say J^*, is undominated if there is no feasible $J(x)$ such that

$$J(x) \geq J^* \quad \underline{2/}$$

Let Λ be the set of $J(x)$'s that are undominated, and let C be the corresponding subset of X. The operator vmax maps Ω into Λ, that is vmax $J(x)$ maps the set of feasible returns into the set of undominated returns. At times I use the slightly abused notation that vmax $J(x)$ maps X into C, that is it maps the set of feasible solutions into the set of undominated solutions.

If each J^i is a linear function and X is a convex polyhedron, then the problem is often called the linear vector maximization problem (see Evans and Steuer 1973; Yu 1974; Zeleny 1974). Each stationary policy δ produces a vector gain

$$g^\delta = (g_1^\delta, \ldots, g_j^\delta)$$

We desire the set of policies δ that satisfies:

$$\text{vmax } g^\delta$$

2/When comparing two j-vectors: $x \geqq j$ implies $x^i \geq y^i$ for all ij $x \geq y$ implies $x \geqq y$ and $x \neq y$; $x > y$ implies $x^i > y^i$ for all i.

However, this is clearly equivalent to:

$$\text{vmin } g$$

$$\text{subject to} \quad C_\ell^1 \qquad \ell = 1, \ldots, j \qquad (H_a)$$

or equivalently:

$$\text{vmax } (\Sigma_i \Sigma_k x_i^k R_{i,\,1}^k, \ldots, \Sigma_i \Sigma_k x_i^k R_{i,\,j}^k)$$

$$\text{subject to} \quad C_\ell^2 \qquad \ell = 1, \ldots, j \qquad (H_b)$$

The two programming problems H_a or H_b reduce the multiobjective Markov decision process to a linear vector maximization problem. For calculating actual efficient sets, this is a significant result, since what progress has been made for developing algorithms for multiobjective problems, has been made in the area of linear problems (see Evans and Steuer 1973; Yu 1974; Zeleny 1974; Benson 1976; Lin 1976 for example). As these algorithms become more efficient, it should be possible to calculate efficient sets for quite general multiobjective Markov decision processes.

II. Discretizing Multiobjective Dynamic Programs

A more common situation, particularly in fields of my interest (fisheries) is that there are j objective functions, which are continuous functions of a state that is contained in a compact subset of Euclidean space, and a continuous random transition function. Assume we can restrict ourselves in the infinite model to stationary policies (again, conditions that insure that the Pareto optimal policy is stationary can be found in Sobel 1977). Also assume that in the infinite model we are willing to discretize the state space, so that the return functions become piecewise linear. It is well known for the single objective case (see for example Sobel 1971) that this discretization leaves the original problem equivalent to a Markov decision model, and the corresponding linear program is obtained as in the last section, with $R^k_{i,\ell}$ being the value of the return function J_ℓ at state i. Thus it is at least theoretically possible to calculate the set of Pareto optimal stationary policies for a wide class of problems. Given other special structure in the problems (such as being "separable" as in Denardo 1968) it may be possible to reduce the resulting linear vector maximization problems to the point where they are of solvable dimensions.

I hope to explore these possibilities in a later paper, using applications that arise from managing fisheries under the Fishery Conservation and Management Act of 1976.

LITERATURE CITED

BENSON, H. P.

1976. A bicriterion mathematical programming approach for nutrition planning in developing nations. Ph.D. Dissertation, Northwestern Univ., Evanston, Ill., 209 p.

DENARDO, E. V.

1968. Separable Markovian decision problems. Manage. Sci. 14: 451-462.

1977. A Markov decision problem. In Dynamic programming: theory and application, chapter 5. Prentice Hall

EVANS, J. P., and R. G. STEUER.

1973. A revised simplex method for linear multiple objective programs. Math. Programming 5:54-72.

HOWARD, R. A.

1960. Dynamic programming and Markov processes. John Wiley & Sons, Inc., New York

LIN, J. G.

1976a. Multiple-objective programming: Lagrange multiplers and method of proper equality constraints. Proc. 1976 Joint Automat. Contr. Conf. July 1976, p. 517-523.

MANNE, A.

1960. Linear programming and sequential decisions. Manage. Sci. 6:259-276.

SOBEL, M. J.

1971. Production smoothing with stochastic demand II: Infinite horizon case. Manage. Sci. 17:724-735.

1977. Undominated solutions of sequential games. Unpublished manuscript. Yale Univ., New Haven, Conn.

YU, P. L.

1974. Cone convexity, cone extreme points and nondominated solutions in decision problems with multiobjectives. J. Optim. Theory and Its Appl. 14:319-377.

ZELENY, M.

1974. Linear multiobjective programming. Springer-Verlag, New York.