The Annals of Mathematical Statistics
1969, Vol. 40, No. 5, 1635–1660

DISCRETE DYNAMIC PROGRAMMING WITH SENSITIVE DISCOUNT OPTIMALITY CRITERIA[1]

BY ARTHUR F. VEINOTT, JR.

Stanford University

1. Introduction. This paper is concerned with stationary finite state and action Markovian decision processes where future rewards are discounted and the transition matrices are substochastic. The discrete time parameter case is treated in Sections 2–4 with analogous results being obtained for the continuous time parameter case in Section 5.

In Section 2 we generalize some results of Shapley [32], Bellman [4], Howard [20], Blackwell [5], Eaton and Zadeh [15], Derman [8], and Denardo [7] which are concerned with the use of the methods of successive approximation and policy improvement in finding a policy π that maximizes the expected infinite horizon discounted reward $V_\rho(\pi)$ where ρ is the rate of interest. Our contribution is to the case where $-1 < \rho \leqq 0$, e.g., corresponding to inflation. One main new result in this section, Corollary 4, is that the series defining $V_\rho(\pi)$ converges absolutely for every (some) π for all one period rewards if and only if the same assertion is true when π ranges over the stationary policies. This result is proved by dynamic programming methods.

Up to now we have supposed the interest rate to be fixed. A stronger concept of optimality would be to find a policy that maximizes $V_\rho(\cdot)$ for a sufficiently small interval of interest rates. Blackwell [5] has shown that there is a stationary policy maximizing $V_\rho(\cdot)$ for all $\rho > 0$ close enough to zero. Recently Miller and the author [30] discovered a policy improvement algorithm for finding such a policy. The algorithm exploits the fact that for stationary policies the Laurent expansion about the origin of the return $V_\rho(\cdot)$ has a simple form. Analogous results hold in the transient case where one seeks to maximize $V_\rho(\cdot)$ for all $\rho < 0$ close enough to zero. Section 3 consists of an expository development of the Laurent expansion mentioned above together with some new results on properties of its coefficients.

In Section 4 we introduce the following new optimality criteria. For each $n = -1, 0, 1, \cdots$, a policy π^* is called $n^\pm$ discount optimal if (for n^-, we consider only the transient case)

$$\liminf_{\rho \to 0\pm} |\rho|^{-n}[V_\rho(\pi^*) - V_\rho(\pi)] \geqq 0, \quad \text{for all} \quad \pi.$$

The sensitivity of this criterion increases with n. And when n is as large as the number of states, the criterion is shown to be equivalent to Blackwell's criterion

Received 3 September 1968; revised February 1969.

[1] This research was supported by the National Science Foundation under Grant GK-1420, the Office of Naval Research under Contract Nonr-225(77) (NR-347-010), and the IBM Research Center.