The Annals of Probability
1981, Vol. 9, No. 2, 293–301

DECISION PROCESSES WITH TOTAL-COST CRITERIA[1]

BY STEPHEN DEMKO AND THEODORE P. HILL

Georgia Institute of Technology

By a decision process is meant a pair (X, Γ), where X is an arbitrary set (the state space), and Γ associates to each point x in X an arbitrary nonempty collection of discrete probability measures (actions) on X. In a decision process with nonnegative costs depending on the current state, the action taken, and the following state, there is always available a Markov strategy which uniformly (nearly) minimizes the expected total cost. If the costs are strictly positive and depend only on the current state, there is even a stationary strategy with the same property.

In a decision process with a fixed goal g in X, there is always a stationary strategy which uniformly (nearly) minimizes the expected time to the goal, and, if X is countable, such a stationary strategy exists which also (nearly) maximizes the probability of reaching the goal.

1. Introduction. Suppose to each element x of a set X is associated a nonempty collection $\Gamma(x)$ of discrete probability measures (transition probabilities) on X. During the course of the process, when one is at state x he chooses (as a function only of the past) an element γ in $\Gamma(x)$, and the next state is determined according to the distribution of γ. Various objective functions associated with such processes have been studied extensively, among them: maximizing average reward (e.g., [3, 7, 13]); maximizing total discounted reward [1, 3, 7, 13]; minimizing finite-horizon costs [3, 8, 13]; maximizing stop rule expectations [4, 5, 15, 16]; minimizing total cost [2, 7, 13, 14]; and maximizing the probability of reaching (or minimizing the time to) a goal [4, 6, 15, 16]. This paper is concerned with the last two objectives; decision processes with total-cost criteria, and goal problems.

In a total-cost decision process a nonnegative cost is associated with each "move," and the objective is to find strategies which nearly minimize, over all possible strategies, the expected total cost.

Decision processes have been studied extensively under various assumptions: that the decision set is finite [3, 8, 13]; that optimal strategies exist [3, 7]; that the spaces involved (state, strategy, decision) are suitably well structured (e.g., Borel, compact, measurable, convex, etc.) [14, 16]; or that precisely one ergodic class exists [3, 7, 13].

In general, however, the total-cost decision process with discrete transition probabilities satisfies none of the above assumptions, and does not admit optimal, or even ϵ-optimal, stationary strategies.

In Section 3, however, it is shown (Theorem 3.2) that in every such process with nonnegative costs, there is available a Markov strategy which uniformly (nearly) minimizes the expected total cost. If the costs are strictly positive and depend only on the current state (Theorem 3.1), there is even a stationary strategy with the same property. In Section 4, this theorem is used to establish several results concerning goal problems.

In a goal problem one has a fixed goal state $g \in X$, and the objective is to find strategies which "best" enable one to reach this goal.

Dubins and Savage proved ([4], Theorem 3.9.2) that in every finite state goal problem there is available a stationary family of strategies which will uniformly (nearly) maximize

Received September 18, 1979.

[1] Partially supported by AFOSR Grant F49620-79-C-0123.

AMS 1970 *subject classification.* Primary 60G99, 62C05.

Key words and phrases. Gambling theory, dynamic programming, decision theory, stationary strategy, Markov strategy, total-cost criteria.

the probability, under all possible strategies, of reaching the goal. Ornstein gave an alternate proof of this result ([10], L1) and extended it ([10], Theorem B) to all problems where X is countable. Sudderth [15] further extended the conclusion to a much larger class of problems, including many with uncountable state space and finitely additive probability measures. Hill [6] showed that if the state space is finite, there is available a stationary family which not only (nearly) maximizes the probability of reaching the goal, but also (nearly), minimizes the expected time to the goal.

The first main result of Section 4, (Theorem 4.1), states that in every goal problem with discrete transition probabilities there is always available a stationary strategy which uniformly (nearly) minimizes the expected time to the goal. The second, (Theorem 4.2), states that if the state space is countable, there is available a stationary strategy which uniformly both (nearly) maximizes the probability of reaching the goal, and (nearly) minimizes the expected time to the goal.

2. Decision processes.

DEFINITION. A *decision process* is a pair (X, Γ), where X is a set and Γ associates to each point x in X a nonempty collection $\Gamma(x)$ of discrete probability measures on X. (As usual, γ is a discrete probability measure on X if there is a countable sequence $x_1, x_2, \ldots$ of elements in X with $\sum_{n=1}^{\infty} \gamma\{x_n\} = 1$.)

In the terminology of Dubins and Savage [4], a decision process is simply a gambling house with discrete gambles. In dynamic programming terminology, X represents the state space, and $\Gamma(x)$ the actions available at x. No assumption is made either on the size or structure of X, nor on the size or structure of $\Gamma(x)$, other than the fact that each set $\Gamma(x)$ consists only of discrete probability measures, which automatically includes all problems with countable X and countably additive probabilities.

(Allowing $\Gamma(x)$ to contain nondiscrete measures would in general seem to necessitate either additional structural or measurability assumptions (e.g., [15, 16]), or else relaxation of the countable additivity of the strategic measures involved (e.g., [4, 12, 15]).)

Much of the notation will follow that of Dubins and Savage [4]. For a set X, $\mathscr{P}(X)$ will denote the discrete probability measures on X. The Dirac delta-measure at x will be denoted by $\delta(x)$. A strategy is a function from X^*, the free monoid generated by X, to $\mathscr{P}(X)$, that is, from the finite sequences in X (including the empty sequence "$\varnothing$"), to the discrete probability measures on X.

The same symbol, σ, will be used to denote both a strategy and the (countably additive) probability measure generated by σ on the product sigma-algebra on X^N (X endowed with the discrete sigma algebra.)

A *strategy* σ *in* Γ *at* x is a strategy such that $\sigma(\varnothing) \in \Gamma(x)$, and $\sigma(x_1, \ldots, x_n) \in \Gamma(x_n)$ for all $x_1, x_2, \ldots, x_n \in X$ and all $n \in N$. A strategy σ is *Markov* [5] if $\sigma(x_1, \ldots, x_n) = \sigma(x_1', \ldots, x_n')$ whenever $x_n = x_n'$, and is *stationary* if $\sigma(x_1, \ldots, x_n) = \sigma(x_1', \ldots, x_m')$ whenever $x_n = x_m'$. The *conditional strategy given* $x_1, \ldots, x_n$, $\sigma[x_1, \ldots, x_n]$, is defined by $\sigma[x_1, \ldots, x_n](x) = \sigma(x_1, x_2, \ldots, x_n, x)$.

(All the results in this paper would also hold if the definition of strategy were changed so that σ was a function $\sigma(x_1, \gamma_1, x_2, \gamma_2, \ldots, x_n)$ of past actions as well as states.)

3. Decision processes with total-cost criteria. The purpose of this section is to define total-cost decision processes; to prove (Theorem 3.1) that in every such process with strictly positive costs depending only on the current state, there always is available a stationary strategy which uniformly (nearly) minimizes, under all possible strategies, the expected total cost; and to prove (Theorem 3.2) that in every such process with nonnegative costs depending on the current state, the action taken, and the following state, there always is available a Markov strategy which uniformly (nearly) minimizes the expected total cost.

DEFINITION. A *total-cost decision process* is a triple (X, Γ, c), where (X, Γ) is a decision process, and $c: X \to [0, \infty)$.

DEFINITION. Let $c: X \to [0, \infty)$. Then $\tilde{c}: X^N \to [0, \infty]$ is the function $\tilde{c}(x_1, x_2, \ldots) = \sum_{n=1}^{\infty} c(x_n)$.

LEMMA 3.1. *$\tilde{c}$ is measurable with respect to the product sigma algebra on X^N (X is endowed with the discrete sigma algebra), and is thus integrable (in the wide sense) with respect to each strategy.*

PROOF. Routine.

DEFINITION. For a strategy σ available at x, $K(\sigma)$, *the expected total cost using* σ, is given by $K(\sigma) = c(x) + \int \tilde{c}\, d\sigma$. The *minimal expected total cost starting at* x, $K(x)$, is given by $K(x) = \inf\{K(\sigma) : \sigma \text{ is in } \Gamma \text{ at } x\}$.

DEFINITION. For a subset A of X, let $N(A): X^N \to N \cup \{\infty\}$ be the function $N(A)(x_1, x_2, \ldots) = |\{n : x_n \in A\}|$.

LEMMA 3.2. *Let $a, b > 0$, and let A be any subset of $\{x : (1/a) \le c(x)$ and $K(x) < b\}$. If σ is any strategy at x satisfying $K(\sigma) < K(x) + \epsilon$, then $E_\sigma(N(A))$ $(= \int N(A)\, d\sigma) < a(b + 3\epsilon)$.*

PROOF.
Case 1. $x \in A$. Then

$$K(x) + \epsilon > K(\sigma) = c(x) + \int \tilde{c}\, d\sigma \ge \int \tilde{c}\, d\sigma \ge (1/a) E_\sigma(N(A)), \tag{1}$$

where the first inequality in (1) follows by hypothesis; the equality by the definition of $K(\sigma)$; the second inequality since $c(x) > 0$; and the last since $c(x) \ge (1/a)$ for all x in A. Thus

$$E_\sigma(N(A)) < a(K(x) + \epsilon) < a(b + 3\epsilon), \tag{2}$$

where the first inequality in (2) follows from (1); and the second since $x \in A$.

Case 2. $x \notin A$. Let σ be any strategy at x satisfying $E_\sigma(N(A)) \ge a(b + 3\epsilon)$. It will be shown that σ can not be ϵ-optimal, that is, for any such strategy σ at x, $K(\sigma) > K(x) + \epsilon$.

Let $t: X^N \to N \cup \{\infty\}$ be the hitting time of A (i.e., $t(x_1, x_2, \ldots) = \min\{j : x_j \in A\}$ if such a j exists, and $= \infty$ otherwise).

Let $\tilde{c}_t$ be the cost incurred from time t on, that is, $\tilde{c}_t(x_1, x_2, \ldots) = \sum \{c(x_n) : n \ge t(x_1, x_2, \ldots)\}$. Clearly $\tilde{c}_t$ is measurable relative to the product sigma algebra, and thus also integrable (in the wide sense) with respect to each strategy.

For each $x' \in A$, let $\sigma_{x'}$ be a strategy at x' satisfying $K(\sigma_{x'}) \le K(x') + \epsilon$, and define the new strategy σ' at x as follows. $\sigma' = \sigma$ up to time t, and $\sigma[x_1, \ldots, x_t] = \sigma_{x_t}$. It will be shown that the expected total cost using σ' is at least 2ϵ less than that using σ, thereby proving that σ can not be ϵ-optimal.

Clearly

$$K(\sigma) - K(\sigma') = \int \tilde{c}_t\, d\sigma - \int \tilde{c}_t\, d\sigma'. \tag{3}$$

By the definitions of A and σ, it follows that

$$\int \tilde{c}_t \, d\sigma \geq (1/a)E_\sigma(N(A)) \geq (1/a)a(b + 3\epsilon) \geq \sup\{K(x'):x' \in A\} + 3\epsilon. \tag{4}$$

On the other hand,

$$\int \tilde{c}_t \, d\sigma' \leq \sup\{K(\sigma_{x'}):x' \in A\} \leq \sup\{K(x'):x' \in A\} + \epsilon. \tag{5}$$

Combining (3), (4) and (5) yields $K(\sigma) - K(\sigma') \geq 2\epsilon$, proving that σ is not ϵ-optimal. □

The following proposition is a stability or perturbation result for strictly positive total-cost decision processes. It states that in any such process one may change the costs slightly (but not uniformly) thereby obtaining a new problem whose expected total costs are uniformly close to that of the original problem.

PROPOSITION 3.1. *Let (X, Γ, c) be a total-cost decision process with $c > 0$. Given positive ϵ there exists $d:X \to (0, \infty)$ such that if $|c'(x) - c(x)| \leq d(x)$ for all x, then $K(x) - \epsilon \leq K'(x) \leq K(x) + \epsilon$ for all x, (where K' is the expected total cost for (X, Γ, c').)*

PROOF. Without loss of generality, assume that $K(x) < \infty$ for all $x \in X$. For $m, n = 1, 2, 3, \ldots$, let $A_{m,n} = \{x \in X: c(x) \in [(m + 1)^{-1}, m^{-1}) \cup [m, m + 1), \text{ and } n - 1 \leq K(x) < n\}$. The sets $\{A_{m,n}\}$ are disjoint, and $X = \cup_{m,n\geq 1} A_{m,n}$.

Fix $\epsilon > 0$, and define $d:X \to (0, \infty)$ by $d(x) = \epsilon \cdot 2^{-m-n}/(m + 1)(n + 1)$ if $x \in A_{m,n}$. To establish the proposition it suffices to show that for each $x \in X$ there is a strategy σ at x satisfying $\int (\tilde{d}) \, d\sigma < \epsilon$.

Fix $x \in X$, and σ any strategy at x satisfying $K(\sigma) < K(x) + 1/3$. Then

$$\int (\tilde{d}) \, d\sigma = \sum_{m,n\geq 1} E_\sigma(N(A_{m,n})) \, d(A_{m,n}) \leq \sum_{m,n\geq 1} \epsilon(m + 1)(n + 1) \cdot 2^{-m-n}(m + 1)^{-1}(n + 1)^{-1} = \epsilon, \tag{6}$$

where the inequality in (6) follows from Lemma 3.2. □

Without strict positivity, the preceding result may fail even in very simple cases. Consider

EXAMPLE 3.1. $X = \{a, b\}$, $\Gamma(a) = \{\delta(b)\}$, $\Gamma(b) = \{\delta(a)\}$, $c(a) = c(b) = 0$. Then $K(a) = K(b) = 0$, but for any $c' > 0$, $K'(a) = K'(b) = +\infty$

The following lemma is a close analog of [13], Theorem 6.10, or [3], Theorem 1, page 23.

LEMMA 3.3. *Let (X, Γ, c) be a total-cost decision process. Then $K(x)$ satisfies*

$$K(x) = c(x) + \inf\{\textstyle\sum \{\gamma(x')K(x'):x' \in X\}:\gamma \in \Gamma(x)\}. \tag{7}$$

PROOF. If $K(x) = \infty$, then $\sum \{\gamma(x')K(x'):x' \in X\} = \infty$ for all $\gamma \in \Gamma(x)$, and both sides of (7) are infinite. Suppose $K(x) < \infty$.

("$\geq$"). Fix $\epsilon > 0$ and find σ in Γ at x such that $K(\sigma) \leq K(x) + \epsilon$. Then

$$\begin{aligned} K(x) \geq K(\sigma) - \epsilon &= c(x) + \textstyle\sum \{K(\sigma[x'])\sigma(\varnothing)(x'):x' \in X\} - \epsilon \\ &\geq c(x) + \textstyle\sum \{K(x')\sigma(\varnothing)(x'):x' \in X\} - \epsilon \\ &\geq c(x) + \inf\{\textstyle\sum \{K(x')\gamma(x'):x' \in X\}: \gamma \in \Gamma(x)\} - \epsilon \end{aligned} \tag{8}$$

where the first inequality in (8) follows by choice of σ; the equality by definition of $K(\sigma)$ and by conditioning on x_1; the second inequality by definition of K; and the last since $\sigma(\emptyset) \in \Gamma(x)$. Since ϵ was arbitrary, this completes this portion of the proof.

("$\leq$"). Fix $\epsilon > 0$, and for each $x' \in X$ let $\sigma_{x'}$ be a strategy in Γ at x' with $K(\sigma_{x'}) \leq K(x') + \epsilon$. Choose $\tilde{\gamma} \in \Gamma(x)$ so that $\sum \{\tilde{\gamma}(x')K(x'):x' \in X\} \leq \inf\{\sum \{\gamma(x')K(x'):x' \in X\}:\gamma \in \Gamma(x)\} + \epsilon$. Define σ in Γ at x by $\sigma(\emptyset) = \tilde{\gamma}$, and $\sigma[x'] = \sigma_{x'}$ for each $x' \in X$. Then

$$
\begin{aligned}
K(x) \leq K(\sigma) &= c(x) + \textstyle\sum \{K(\sigma[x'])\tilde{\gamma}(x'):x' \in X\} \\
&\leq c(x) + \textstyle\sum \{K(x')\tilde{\gamma}(x'):x' \in X\} + \epsilon \\
&\leq c(x) + \inf\{\textstyle\sum \{\gamma(x')K(x'):x' \in X\}:\gamma \in \Gamma(x)\} + 2\epsilon,
\end{aligned}
\tag{9}
$$

where the first inequality in (9) follows by definition of K; the equality as in (8); the second inequality by definition of σ and $\sigma_{x'}$; and the last inequality by choice of $\tilde{\gamma}$. Since ϵ was arbitrary, this completes the proof. □

LEMMA 3.4. *Suppose $\Gamma'(x) = \{\gamma_x\}$ for all x (i.e., (X, Γ') is simply a Markov chain with stationary transition probabilities), and let K' be the minimal expected total cost for (X, Γ', c). Then $\{K'(x)\}$ is the minimal nonnegative solution of the system of equations*

$$
k(x) = c(x) + \textstyle\sum \{\gamma_x(x')k(x'):x' \in X\}. \tag{10}
$$

(That is, $\{K'(x)\}$ satisfies (10), *and if $\{K''(x)\}$ is any other nonnegative solution of* (10), *then $K'(x) \leq K''(x)$ for all x in X.)*

PROOF. That $\{K'(x)\}$ is a solution of (10) follows immediately from Lemma 3.3.

Let $M = (m_{x,x'})$ be the transition matrix for the Markov chain (X, Γ'), (i.e., $m_{x,x'} = \gamma_x(x')$).

It is easy to see that $\{K'(x)\}$ satisfies, in matrix notation, $K' = \sum_{n=0}^{\infty} M^n c$. By an obvious extension of results in [9], Chapter 1.2 (to include uncountably infinite matrices M which have only countably infinite nonzero entries in each row), $\sum_{n=0}^{\infty} M^n c$ is the minimal nonnegative solution of (10). □

Contrary to Parzen's claim ([11], page 239), even if the times to absorption in the recurrent states are almost everywhere finite, the solution of (10) need not be unique.

EXAMPLE 3.2. Let $X = \{0, 1, 2, \ldots\}$, $\Gamma(0) = \{\delta(0)\}$, $\Gamma(n) = \{\delta(0)/2 + \delta(n + 1)/2\}$, and $c(0) = 0$, $c(n) = 1$ for $n \geq 1$. One solution (the "principal solution") of (10) is $K(0) = 0$, $K(n) \equiv 2$ for $n \geq 1$, but another solution is $\tilde{K}(0) = 0$, $\tilde{K}(n) = 2 + 2^n$.

THEOREM 3.1. *If (X, Γ, c) is a total-cost decision process with $c(x) > 0$ for all x, then for each positive ϵ there is available in Γ a stationary strategy σ such that $K(\sigma[x]) \leq K(x) + \epsilon$ for all $x \in X$.*

PROOF. Fix $\epsilon > 0$, find $d:X \to (0, \infty)$ as in Proposition 3.1, and let K' be the minimal expected total cost for $(X, \Gamma, c + d)$. By Proposition 3.1, $\{K'(x)\}$ satisfies

$$
K'(x) \leq \epsilon + K(x) \qquad \text{for all } x \in X. \tag{11}
$$

By Lemma 3.3, $\{K'(x)\}$ is a solution of

$$
k(x) = c(x) + d(x) + \inf\{\textstyle\sum \{\gamma(x')K(x'): x' \in X\}: \gamma \in \Gamma(x)\}. \tag{12}
$$

For each $x \in X$, choose $\gamma_x \in \Gamma(x)$ so that $\{K'(x)\}$ also satisfies

$$
k(x) = c(x) + d(x) - e(x) + \textstyle\sum \{\gamma_x(x')k(x'): x' \in X\}, \tag{13}
$$

where $0 \leq e(x) < d(x)$ for all $x \in X$ (possible by (12)).

Define Γ' by $\Gamma'(x) = \{\gamma_x\}$, let K'' be the minimal expected total cost for $(X, \Gamma', c + d - e)$, and let K''' be that for (X, Γ', c). Then

$$K(x) \le K'''(x) \le K''(x) \le K'(x) \le K(x) + \epsilon \quad \text{for all } x \in X, \tag{14}$$

where the first inequality in (14) follows since $\Gamma \supset \Gamma'$; the second since $c < c + d - e$; the third by Lemmas 3.3 and 3.4 (since $\{K'(x)\}$ satisfies (13)); and the last inequality by (11).

Define σ by $\sigma(x) = \gamma_x$ for all $x \in X$. Clearly $K(\sigma[x]) = K'''(x)$, and the conclusion follows by (14). □

In general, optimal strategies do not exist.

EXAMPLE 3.3. $X = N$, $\Gamma(n) = \{\delta(m): m \in N\}$, $c(n) = 1/n$. Then $K(1) = 1$, but for any strategy σ at 1, $K(\sigma) > 1$.

Although nonnegativity of c is sufficient for good Markov strategies (Theorem 3.2 below), it is not sufficient for good stationary strategies.

EXAMPLE 3.4. $X = \{a, b\}$, $\Gamma(a) = \{((n-1)/n)\delta(a) + (1/n)\delta(b): n \in N\}$, $\Gamma(b) = \{\delta(a)\}$, $c(a) = 0$, $c(b) = 1$. Then $K(a) = 0$, but for every stationary strategy σ, $K(\sigma[a]) = +\infty$.

If the costs depend on the order of succession of states visited, (e.g., cost $c(x, x')$ is incurred when going from x to x'), then even strict positivity is not sufficient for the existence of good stationary strategies.

EXAMPLE 3.5. $X = N$, $\Gamma(1) = \{\delta(n): n \in N\}$, $\Gamma(n) = \{\delta(1)\}$ for $n \ge 2$, $c(i, j) = 1/j$ if $i = 1$, $= 1/i$ if $j = 1$, and $= 1$ otherwise. Then $K(1) = 0$, but for any stationary strategy σ, $K(\sigma[1]) = +\infty$.

In many dynamic programming formulations of cost problems (e.g., [1, 3, 13]), the costs are allowed to depend upon the present state, the action taken, and the following state. Although good stationary strategies may not exist in general (Example 3.5), if the costs are nonnegative good Markov strategies always exist.

DEFINITIONS. A *generalized total-cost decision process* is a triple (X, Γ, c), where (X, Γ) is a decision process, and $c: X \times \mathcal{P}(X) \times X \to [0, \infty)$. The *(conditional) expected total cost using strategy* σ is $\underline{\hat{K}(\sigma[x_1])} = \sum_{n=1}^{\infty} \sum \{c(x_n, \sigma(x_1, \cdots, x_n), x_{n+1}) \cdot \sigma(x_1)(x_2). \sigma(x_1, x_2)(x_3) \cdots \sigma(x_1, \cdots, x_n), (x_{n+1}): x_j \in X, j = 1, \cdots, n+1\}$, and the *minimal expected total cost starting at* x is $\hat{K}(x) = \inf\{\hat{K}(\sigma): \sigma \text{ is in } \Gamma \text{ at } x\} = \inf\{\hat{K}(\sigma[x]): \sigma \text{ is in } \Gamma\}$.

THEOREM 3.2. *If (X, Γ, c) is a generalized total-cost decision process, then for each positive ϵ there is available a Markov strategy σ such that $\hat{K}(\sigma[x]) \le \hat{K}(x) + \epsilon$ for all x in X.*

PROOF. Fix $\epsilon > 0$ and define the (nongeneralized) total-cost decision process (X', Γ', c') as follows. Let $X' = (X \times N) \cup (X \times \mathcal{P}(X) \times X \times N)$ and define Γ' by $\Gamma'(x, n) = \{\sum_{x' \in X} \gamma(x')\delta(x, \gamma, x', n): \gamma \in \Gamma(x)\}$ and $\Gamma'(x, \gamma, x', n) = \{\delta(x', n+1)\}$.

Define c' by $c'(x, n) = \epsilon \cdot 2^{-2n}$ and $c'(x, \gamma, x', n) = c(x, \gamma, x') + \epsilon \cdot 2^{-2n}$

There is a natural one-to-one correspondence between strategies σ in Γ at x and strategies σ' in Γ' at $(x, 1)$ given by $\sigma(x_1, \cdots, x_n)(x) = \sigma'((x, \sigma(\emptyset), x_1, 1), (x_1, 2), (x_1, \sigma(x_1), x_2, 2), (x_2, 3), \cdots, (x_n, n+1))\, (x_n, \sigma(x_1, \cdots, x_n), x, n+1)$.

The definitions of $\hat{K}$ and c' give that

$$|\hat{K}(\sigma[x]) - K(\sigma'[x, 1])| \le \epsilon \qquad \text{for all } x \in X, \tag{15}$$

(where K is the minimal expected total cost for (X', Γ', c')) and hence that

(16) $$|\hat{K}(x) - K(x, 1)| \le \epsilon \qquad \text{for all } x \in X.$$

By construction, (X', Γ', c') is a (nongeneralized) decision process with strictly positive costs, so by Theorem 3.1 there is available a stationary strategy σ' in Γ' such that $K(\sigma'[x']) \le K(x') + \epsilon$ for all $x' \in X'$.

In particular, it follows that

(17) $$K(\sigma'[x, 1]) \le K(x, 1) + \epsilon \qquad \text{for all } x \in X.$$

Let σ be the strategy in Γ corresponding to σ' (i.e., $\sigma(x_1, \cdots, x_n) = \gamma$ if and only if $\sigma'(x_n, n + 1) = \sum_{x' \in X} \gamma(x')\delta(x_n, \gamma, x', n + 1)$). Clearly σ is Markov, and by (15), (16) and (17), satisfies $\hat{K}(\sigma[x]) \le \hat{K}(x) + 2\epsilon$. □

4. Goal problems. The purpose of this section is to prove (Theorem 4.1) that in any decision process with a goal, there always exist stationary strategies which uniformly (nearly) minimize the expected time to the goal, and (Theorem 4.2) that if the state space is countable, there even exist stationary strategies which simultaneously (nearly) minimize the expected time to the goal and (nearly) maximize the probability of reaching the goal.

DEFINITIONS. As in [6], a *goal problem* is a triple (X, Γ, g) where (X, Γ) is a decision process and $g \in X$ is a "goal" state. The *time to the goal*, T, is the function $T: X^N \to N \cup \{\infty\}$ defined by $T(x_1, x_2, \cdots) = \min\{j: x_j = g\}$ if such a j exists, and $= \infty$ otherwise. The *expected time to the goal using strategy* σ, $W(\sigma)$, is $\int T\, d\sigma$, and the *minimal expected time to the goal from* $x \in X$, $W(x)$, is $W(x) = \inf\{W(\sigma): \sigma \text{ is in } \Gamma \text{ at } x\}$ if $x \ne g$, and $W(g) = 0$.

THEOREM 4.1. *If (X, Γ, g) is a goal problem then for each positive ϵ there is available a stationary strategy σ such that $W(\sigma[x]) \le W(x) + \epsilon$ for all $x \in X$.*

PROOF. Define the total-cost decision process (X', Γ', c) as follows: $X' = X \cup N$ (N disjoint from X); $\Gamma'(x) = \Gamma(x)$ if $x \ne g$, $\Gamma(g) = \{\delta(1)\}$ and $\Gamma(n) = \{\delta(n + 1)\}$ for $n \ge 1$; $c(x) = 1$ if $x \ne g$, $c(g) = 1$, and $c(n) = 1/2^n$ for $n \ge 1$.

Fix $\epsilon > 0$. Since the costs are strictly positive and depend only on the current state, Theorem 3.1 guarantees the existence of a stationary strategy σ' in Γ' satisfying $K(\sigma'[x]) \le K(x) + \epsilon$ for all $x \in X'$, where K is the minimal expected total cost for (X', Γ', c).

Let σ be any stationary strategy in Γ satisfying $\sigma(x) = \sigma'(x)$ if $x \in X$. Since $K(x) = W(x) + 2$, and $K(\sigma) = W(\sigma) + 2$, σ satisfies $W(\sigma[x]) \le W(x) + \epsilon$ for all $x \in X$. □

Theorem 4.1 generalizes [6], Theorem 5.1, which proved the same result for finite X. Clearly g can be replaced by any non-empty subset G of X and the same conclusion will follow.

Precisely the same proof shows that if (X, Γ, g, c) is a goal problem with nongoal costs bounded away from zero (i.e., $c(x) \ge a > 0$ for all $x \ne g$), then for each positive ϵ there is available a stationary strategy which uniformly (nearly) minimizes the expected cost to the goal. If the costs are not bounded away from zero, some strategies which will yield small expected total cost may never reach the goal, and it is not known if there always exist uniformly ϵ-optimal stationary strategies.

It is important to note that in Theorem 4.1 the state space X is completely arbitrary, in contrast to Ornstein's example ([10], Theorem A) of a decision process in which stationary strategies are not uniformly adequate if one's objective is to maximize the *probability* of reaching the goal.

If X is countable, however, there are always available stationary strategies which are nearly optimal in both senses, that is, which simultaneously (nearly) minimize the expected time to the goal and (nearly) maximize the probability of reaching the goal.

DEFINITION. Let $V(\sigma)$ be the probability of reaching the goal using σ, and let $V(x) = \sup\{V(\sigma): \sigma \text{ is in } \Gamma \text{ at } x\}$ for $x \neq g$, and $V(g) = 1$.

THEOREM 4.2. *If (X, Γ, g) is a goal problem with X countable, then for each positive ϵ there is available a stationary strategy σ satisfying both*

$$V(\sigma[x]) \geq V(x) - \epsilon \qquad \text{for all } x \in X \tag{18}$$

and

$$W(\sigma[x]) \leq W(x) + \epsilon \qquad \text{for all } x \in X. \tag{19}$$

PROOF. Fix $\epsilon > 0$, and find a stationary strategy σ' satisfying (18) (possible by [10], Theorem B), and a stationary strategy σ'' satisfying (19) (possible by Theorem 4.1). Let $S = \{x \in X: W(x) < \infty\}$, and define the stationary strategy σ by $\sigma(x) = \sigma'(x)$ if $x \notin S$, and $= \sigma''(x)$ if $x \in S$. It is easy to check that σ satisfies both (18) and (19). □

Theorem 4.2 strengthens [6], Theorem 5.1, which proves (16) and (17) for finite X, and also strengthens Ornstein's result ([10], Theorem B), which proves only (16). For uncountable X, Ornstein's example ([10], Theorem A) shows that it is not possible in general to find a stationary strategy satisfying even (18).

The following proposition, typical of the considerable simplifications that occur when X is finite, is a generalization of the well-known fact that null recurrent states do not exist in finite Markov chains with stationary transition probabilities.

PROPOSITION 4.1. *If (X, Γ, g) is a goal problem with finite X and with $V(x) = 1$ for all x, then $W(x) < \infty$ for all x.*

PROOF. It is even true that there is a stationary strategy σ in Γ with $W(\sigma[x]) < \infty$ for all x. Suppose $|X| = n$. Let $S_1 = \{x \in X: \exists\, \gamma \in \Gamma(x) \text{ with } \gamma(g) > 0\}$, and $S_k = \{x \in X: \exists\, \gamma \in \Gamma(x) \text{ with } \gamma(S_{k-1}) > 0\} \setminus S_{k-1}$. Since $V(x) = 1$ for all $x \in X, \cup_{k=1}^{n} S_k = X$. For each $x \in X$, choose $\gamma_x \in \Gamma(x)$ such that γ_x satisfies $\gamma_x(g) > 0$ if $x \in S_1$, and $\gamma_x(S_{k-1}) > 0$ if $x \in S_k$. Define the stationary strategy σ in Γ by $\sigma(x) = \gamma_x$ for all x. Under σ, X is a finite Markov chain with (stationary transition probabilities and) a single ergodic class containing g. The finiteness of X guarantees that the expected first passage time to g is everywhere finite, that is, $W(\sigma[x]) < \infty$ for all x. □

The following easy example shows that the conclusion of Proposition 4.1 may fail if X is infinite.

EXAMPLE 4.1. $X = \{0, 1, 2, 3, \cdots\}$, $g = 1$, $\Gamma(1) = \{\delta(1)\}$, $\Gamma(j) = \{\delta(j-1)\}$ for $j > 1$, and $\Gamma(0) = \{\sum_{k=1}^{\infty} 2^{-k}\delta(2^k)\}$. Clearly $V(x) = 1$ for all x, but $W(0) = \infty$.

5. Applications to infinite systems of equations.

The purpose of this section is to mention the relationship of the results of Section 3 to the classical study of infinite systems of linear equations of the form

$$w_i = b_i + \textstyle\sum_{j=1}^{\infty} m_{ij} w_j \qquad i = 1, 2, \cdots. \tag{20}$$

The following perturbation result for such systems seems to be new. It states that if the matrix (m_{ij}) is stochastic, then one may change the b_i slightly (but not uniformly), thereby obtaining a new system with solutions *uniformly* close to that of the original.

PROPOSITION 5.1. *If $b_i > 0$ for all i, and the matrix (m_{ij}) is stochastic, then for each positive ϵ there is a function d: $N \to (0, \infty)$ with the following property: if b' satisfies $|b'_i - b_i| < d_i$ for all i, then for every solution $\{w_i^*\}$ of (20) there is a solution $\{w'_i\}$ of*

$$w_i = b'_i + \textstyle\sum_{j=1}^{\infty} m_{ij} w_j \tag{21}$$

$$\textit{satisfying } w_i^* - \epsilon \leq w'_i \leq w_i^* + \epsilon \qquad \textit{for all } i.$$

PROOF. If $\{w_i^*\}$ is the minimal nonnegative (i.e., principal) solution of (20), then the result follows exactly as in the proof of Theorem 3.1 for the special case $|\Gamma(x)| = 1$ for all x. If $\{w_i^*\}$ is any other solution of (20), the result is easily seen to follow since the homogeneous systems associated with (20) and (21) are identical. □

Theorem 3.1 may be used to generalize the notion of the principal solution [9] of a system of linear equations to include systems of the form

$$k(x) = c(x) + \inf\{\sum \{\gamma(x')k(x'): x' \in X\}: \gamma \in \Gamma(x)\}, \tag{22}$$

where (X, Γ) is a decision process and $c(x) > 0$ for all x.

DEFINITION. The *principal solution* of (22) is the minimal nonnegative solution; that is, a nonnegative solution $\{s(x)\}$ of (22) is the principal solution if every other nonnegative solution $\{\hat{s}(x)\}$ is such that $s(x) \leq \hat{s}(x)$ for all $x \in X$.

Clearly the principal solution, if it exists, is unique. The question of existence is answered by the following proposition.

PROPOSITION 5.2. *If (X, Γ, c) is a total-cost decision process with $c(x) > 0$ for all x, then the following are equivalent*:

(i) $K(x) = \inf\{K(\sigma): \sigma$ *is in* Γ *at* $x\}$;

(ii) $\{K(x)\}$ *is the principal solution of* (20);

(iii) $K = \inf\{\sum_{n=0}^{\infty} M^n c$: M *is the transition probability matrix associated with some stationary strategy in* $\Gamma\}$.

PROOF. Equivalence of (i) and (iii) follows from Theorem 3.1. Equivalence of (ii) and (iii) follows as in the proof of Lemma 3.4. □

Acknowledgment. The authors are grateful to Robert Kertz for several useful conversations.

REFERENCES

[1] BLACKWELL, D. (1965). Discounted dynamic programming. *Ann. Math. Statist.* **36** 226–235.

[2] DERMAN, C. (1962). On sequential decisions and Markov chains. *Management Sci.* **9** 16–24.

[3] DERMAN, C. (1970). *Finite State Markovian Decision Processes.* Academic Press, New York.

[4] DUBINS, L. AND SAVAGE, L. (1976). *Inequalities for Stochastic Processes.* Dover, New York.

[5] HILL, T. (1979a). On the existence of good Markov strategies. *Trans. Amer. Math. Soc.* **247** 157–176.

[6] HILL, T. (1979b). On reaching a goal quickly. AFOSR Technical Report, Georgia Inst. Technology.

[7] HORDIJK, A. (1974). *Dynamic Programming and Markov Potential Theory.* Math. Centre Tracts, Amsterdam.

[8] HOWARD, R. (1960). *Dynamic Programming and Markov Process.* Technology Press, Cambridge, Mass.

[9] KANTOROVITCH, L. AND KRYLOV, V. (1964). *Approximate Methods of Higher Analysis.* Noordhoff, Groningen.

[10] ORNSTEIN, D. (1969). On the existence of stationary optimal strategies. *Proc. Amer. Math. Soc.* **20** 563–569.

[11] PARZEN, E. (1965). *Stochastic Processes.* Holden-Day, San Francisco.

[12] PURVES, R. AND SUDDERTH, W. (1976). Some finitely additive probability. *Ann. Probability* **4** 259–276.

[13] ROSS, S. (1970). *Applied Probability Models with Optimization Applications.* Holden-Day, San Francisco.

[14] SCHÄL, M. (1975). On dynamic programming. Compactness of the space of policies. *Stochastic Processes Appl.* **3** 345–364.

[15] SUDDERTH, W. (1969a). On the existence of good stationary strategies. *Trans. Amer. Math. Soc.* **135** 399–414.

[16] SUDDERTH, W. (1969b). On measurable gambling problems. *Ann. Math. Statist.* **42** 260–269.

SCHOOL OF MATHEMATICS
GEORGIA INSTITUTE OF TECHNOLOGY
ATLANTA, GEORGIA 30332