ZASTOSOWANIA MATEMATYKI
APPLICATIONES MATHEMATICAE
XV, 4 (1977)

A. NOWAK (Katowice)

STATIONARY OPTIMAL PROCESS IN DISCOUNTED DYNAMIC PROGRAMMING

In this paper we investigate a discounted dynamic programming problem. Our model is a slight generalization of the models of Blackwell [2] and Strauch [9]. Any policy π, together with an initial distribution p, generates a random process of successive states. This process is called *optimal* if the policy π maximizes the expectation of the total discounted reward. It is proved that under certain assumptions there exists an optimal process which is stationary. The proof is based on representation of a dynamic programming problem in measure spaces and on the Fan-Kakutani fixed-point theorem.

Sutherland [10] studied similar problems in a deterministic model of the economy. He proved the existence of a stationary optimal program.

1. Preliminaries. In this section we give general notation and definitions following those of [2].

A *standard Borel space* (abbreviated to *SB-space*) is a Borel subset of a Polish space, endowed with the induced topology and the Borel σ-field. Let X and Y be non-empty SB-spaces. By XY we mean the Cartesian product of X and Y. We always consider XY with the product topology and with the product σ-field.

A *probability measure on* X is a probability measure defined over the Borel σ-field of X. The set of all probability measures on X is denoted by $P(X)$. A *transition probability from* X *to* Y is a function $q(\cdot|\cdot)$ such that, for each $x \in X$, $q(\cdot|x)$ is a probability measure on Y and, for each Borel subset $B \subset Y$, $q(B|\cdot)$ is a measurable function on X. The set of all transition probabilities from X to Y is denoted by $Q(Y|X)$. For $p \in P(X)$ and $q \in Q(Y|X)$, $m := pq$ denotes the probability measure on XY such that

$$m(AB) = \int_A q(B|x)p(dx)$$

for all Borel subsets $A \subset X$, $B \subset Y$. Conversely, every $m \in P(XY)$ has a factorization $m = pq$, where $p \in P(X)$ is the marginal distribution of the first coordinate variable under m, and $q \in Q(Y|X)$ is a version of the

conditional distribution of the second coordinate variable, given the first. A measure $p \in P(X)$ is *invariant under the transition probability* $q \in Q(X \mid X)$ if $pq(XA) = p(A)$ for all Borel subsets $A \subset X$.

By $C(X)$ we denote the set of all real-valued bounded continuous functions on X. $C(X)$ with the sup norm is a Banach space. Let X be compact. Then $P(X)$ is a subset of $C^*(X)$, dual to $C(X)$. The space $P(X)$ is metrizable and compact in the weak* topology of $C^*(X)$ (see Partharasaty [7], p. 43 and 46). A sequence $\{p_n\}$ of probability measures on X converges to $p \in P(X)$ in this topology iff, for each $u \in C(X)$,

$$\lim_n \int_X u(x)p_n(dx) = \int_X u(x)p(dx).$$

Denote by N the set of positive integers.

A *multifunction* φ from X to Y is a function defined on X, whose values are non-empty subsets of Y. We call φ *closed* if its graph

$$\{(x, y) \in XY\colon\ y \in \varphi(x)\}$$

is closed. φ is *upper* (*lower*) *semicontinuous* if, for any closed (open) subset $B \subset Y$, the set

$$\{x \in X\colon\ \varphi(x) \cap B \neq \emptyset\}$$

is closed (open). φ is *continuous* if it is upper and lower semicontinuous. Every compact-valued and upper semicontinuous multifunction is closed (see Berge [1], Theorem 6, p. 117). If Y is compact and φ is closed, then φ is upper semicontinuous (see [1], p. 118).

Let φ be a multifunction from X to X. An element $x \in X$ is a *fixed point* of φ if $x \in \varphi(x)$. We use the Fan-Kakutani fixed-point theorem (see Fan Ky [3]):

Let X be a convex compact subset of a locally convex topological linear space. Then every convex compact-valued and upper semicontinuous multifunction φ from X to X has a fixed point.

2. Model. A *discounted dynamic programming model* is specified by a system of six objects $(S, A, \varphi, q, r, \beta)$ defined as follows:

(i) S is a non-empty SB-space, the *set of states* of some system.

(ii) A is an SB-space, the *set of actions*.

(iii) φ is a multifunction from S to A, $\varphi(s)$ is the *set of all admissible actions* if the system is in state s. We assume that the graph of φ,

$$G := \{(s, a) \in SA\colon\ a \in \varphi(s)\},$$

is a Borel subset of SA, and φ has a *measurable selection*, i.e. there is a measurable function $g\colon S \to A$ such that $g(s) \in \varphi(s)$ for all $s \in S$.

(iv) q is a transition probability from SA to S, the *law of motion* of the system.

(v) r is a bounded from above, real-valued, measurable function on SAS, the *reward function*.

(vi) β is a *discount factor*, $0 \leqslant \beta < 1$.

If the system is in state s and we take an action $a \in \varphi(s)$, then the system moves to a new state s' according to the probability distribution $q(\cdot \mid s, a)$, and we receive a reward $r(s, a, s')$. The process is then repeated from the new state s'. We discount our future rewards with the factor β, so that a reward of one unit, n stages in the future, is worth β^n now. We intend to maximize the expectation of the total discounted reward over the infinite future.

We write $H_1 := S$ and $H_{n+1} := GH_n$ for $n \in N$. H_n is the set of all histories of the system at time n. A *policy* π is a sequence $\{\pi_1, \pi_2, \ldots\}$, where

$$\pi_n \in Q(A \mid H_n) \quad \text{and} \quad \pi_n\big(\varphi(s_n) \mid h\big) = 1$$
$$\text{for all } h = (s_1, a_1, s_2, \ldots, s_n) \in H_n,\ n \in N.$$

If we use a policy π, we choose the n-th action according to the probability distribution $\pi_n(\cdot \mid h)$, where h is the history of the system up to time n. A policy π is *Markov* if each $\pi_n \in Q(A \mid S)$; in this case the action at time n depends only on the integer n and on the n-th state of the system. A *stationary policy* is a Markov policy such that $\pi_n = \sigma$ for some $\sigma \in Q(A \mid S)$. The stationary policy defined by σ is denoted by $\sigma^{(\infty)}$. Denote by Π the set of all policies. The set of all Markov policies is denoted by Π_{M}.

Any policy $\pi \in \Pi$, together with an initial distribution $p \in P(S)$, defines the probability measure

$$e_{p,\pi} := p\pi_1 q\pi_2 q \ldots$$

on $H := SASA \ldots$ (More precisely, we must first extend each π_n to a transition probability from $SASA \ldots S$ ($2n-1$ factors) to A.) By $\mathrm{E}_{p,\pi}$ we denote the expectation under $e_{p,\pi}$. The *total reward function* is defined on H by

$$R(h) := \sum_{n \in N} \beta^{n-1} r(s_n, a_n, s_{n+1}), \qquad \text{where } h = (s_1, a_1, s_2, a_2, \ldots).$$

Denote by $\hat{s}_n$ and $\hat{a}_n$ the projection from H into the n-th state space and the n-th action space, respectively. The random variables $\hat{s}_n$ and $\hat{a}_n$ describe the state of the system and the action at time n.

Any pair (p, π), where $p \in P(S)$ and $\pi \in \Pi$, defines the probability measure $e_{p,\pi} \in P(H)$ and, therefore, the random process $\{\hat{s}_n\}$. $\mathrm{E}_{p,\pi} R$ is the expected reward corresponding to this process. A process generated by

the pair (p, π^*) is called *optimal* if

$$\mathrm{E}_{p,\pi^*} R = \sup_{\Pi} \mathrm{E}_{p,\pi} R.$$

This process maximizes our expected reward for the initial distribution p. For any $\pi \in \Pi$ and $p \in P(S)$, there exists $\pi' \in \Pi_{\mathrm{M}}$ such that $\mathrm{E}_{p,\pi} R = \mathrm{E}_{p,\pi'} R$ (see Strauch [9], Theorem 4.1). Thus we may restrict our attention to Markov policies.

Let $\pi = \{\pi_n\}$ be a Markov policy, and let T be a Borel subset of S. By the properties of conditional expectations, we have

$$e_{p,\pi}(\hat{s}_{n+1} \in T | \hat{s}_1, \ldots, \hat{s}_n) = \pi_n q(AT | \hat{s}_n) = e_{p,\pi}(\hat{s}_{n+1} \in T | \hat{s}_n).$$

Hence, $\{\hat{s}_n\}$ is a Markov process whose transition probabilities are given by

$$\mu_\pi^n(\cdot | s) := \pi_n q(A \cdot | s), \qquad s \in S, \ n \in N.$$

Consider the process $\{\hat{s}_n\}$ generated by an initial distribution p and a stationary policy $\sigma^{(\infty)}$. This is a Markov process with the transition probability

$$\mu_\sigma(\cdot | s) := \sigma q(A \cdot | s)$$

independent of n. The process $\{\hat{s}_n\}$ is *stationary* if, for any $n \in N$, the probability distribution of the random vector $(s_t, s_{t+1}, \ldots, s_{t+n})$ does not depend on t. The process generated by $(p, \sigma^{(\infty)})$ is stationary if and only if the initial distribution p is invariant under μ_σ.

The *optimal reward function* v^* is defined by

$$v^*(s) := \sup_{\Pi} \mathrm{E}_{p,\pi}(R | \hat{s}_1 = s), \qquad s \in S.$$

$v^*(s)$ is the *optimal expected reward* if the system starts from the state s. In general, v^* is not measurable. Strauch has shown (see [9], Theorems 7.1 and 8.2) that v^* is *universally measurable*, i.e. measurable with respect to the completion of every $p \in P(S)$, and satisfies the *optimality equation*

$$v^*(s) = \sup_{a \in \varphi(s)} \int_S \big(r(s, a, t) + \beta v^*(t)\big) q(dt | s, a), \qquad s \in S.$$

For any $p \in P(S)$,

$$\sup_{\Pi} \mathrm{E}_{p,\pi} R = \int_S v^*(s) p(ds) \tag{1}$$

(see Hinderer [4], Theorem 14.2, p. 100).

Remarks. I. Hinderer [4] calls a policy π $\overline{p}$-*optimal* if the process generated by (p, π) is optimal.

II. If a dynamic programming model satisfies assumptions A1-A3 (stated in Section 4), and the reward function r is upper semicontinuous,

then for any $p \in P(S)$ there exists a (deterministic) stationary policy $\sigma^{(\infty)}$ such that the generated process is optimal (see [6]).

3. Associate deterministic problem. From our dynamic programming problem we can obtain an equivalent deterministic problem by considering the probability distribution on S as the new state of the system.

More precisely, consider the model $(P(S), P(G), \psi, f, w, \beta)$, where ψ is a multifunction from $P(S)$ to $P(G)$ defined by

$$\psi(p) := \{m \in P(G): m(TA \cap G) = p(T) \text{ for all Borel subsets } T \subset S\},$$

f is a function from $P(G)$ to $P(S)$ given by

$$f(m)(T) := mq(GT)$$

(T is a Borel subset of S), w is the real-valued function on $P(G)$ defined by

$$w(m) := \int_G \Big(\int_S r(s, a, t) q(dt \mid s, a) \Big) m\big(d(s, a)\big),$$

$P(S)$ is the set of states, $P(G)$ is the set of actions, and $\psi(p)$ is the set of all admissible actions in state p.

If we take an action $m \in \psi(p)$ in state p, then the system moves to a new state $p' = f(m)$, and we receive a reward $w(m)$. The process is then repeated from the new state p'. Future rewards are discounted with the discount factor β, the same as in the stochastic model. We intend to maximize the total discounted reward over the infinite future.

A *program* starting from $p \in P(S)$ is a sequence $\{m_n\}$ of probability measures on G satisfying $m_n \in \psi(p_n)$, where $p_1 := p$, and $p_n := f(m_{n-1})$, $n > 1$. A program is *stationary* if m_n does not depend on n. The stationary program defined by $m \in P(G)$ is denoted by $m^{(\infty)}$. Let $M(p)$ denote the set of all programs starting from p. The *optimal reward function* in the deterministic model is given by

$$V(p) := \sup_{M(p)} \sum_{n \in N} \beta^{n-1} w(m_n), \qquad p \in P(S).$$

It satisfies the *optimality equation*

$$V(p) = \sup_{m \in \psi(p)} \big[w(m) + \beta V\big(f(m)\big)\big], \qquad p \in P(S).$$

A program $\{m_n^*\} \in M(p)$ is called *optimal* if

$$\sum_{n \in N} \beta^{n-1} w(m_n^*) = V(p).$$

The program $\{m_n\} \in M(p)$ is *optimal* if it satisfies

$$V(p_n) = w(m_n) + \beta V\big(f(m_n)\big), \qquad n \in N, \tag{2}$$

where $p_1 := p$, and $p_n := f(m_{n-1})$, $n > 1$.

We show that there is a one-to-one correspondence between Markov processes $\{\hat{s}_n\}$ and programs $\{m_n\}$. For the process generated by (p, π), where $p \in P(S)$ and $\pi \in \Pi_M$, we define a sequence of probability measures on G:

$$m_1 := p\pi_1, \quad m_{n+1} := f(m_n)\pi_{n+1}, \quad n \in N. \tag{3}$$

It is clear that $\{m_n\} \in M(p)$. Conversely, consider a program $\{m_n\}$ starting from p. Each m_n has a factorization $m_n = p_n\pi_n$, where $p_n \in P(S)$, $p_1 = p$, $\pi_n \in Q(A \mid S)$, and $\pi_n(\varphi(s) \mid s) = 1$ for $s \in S$ (cf. [6], proof of the Lemma). $\pi := \{\pi_n\}$ is a Markov policy and satisfies (3). Note that $p_n := f(m_{n-1})$ is the distribution of the random variable $\hat{s}_n$.

We next show that the process $\{\hat{s}_n\}$ is stationary iff the corresponding program is stationary. Let the process generated by $(p, \sigma^{(\infty)})$ be stationary, and let $m := p\sigma$. Since p is invariant under the transition probability μ_σ,

$$f(m)(T) = p\sigma q(GT) = p\mu_\sigma(ST) = p(T)$$

for all Borel subsets $T \subset S$. Hence, $m^{(\infty)}$ is the program corresponding to $(p, \sigma^{(\infty)})$. Conversely, for any stationary program $m^{(\infty)}$, there exist $p \in P(S)$ and $\sigma \in Q(A \mid S)$ such that $m = p\sigma$ and $\sigma^{(\infty)}$ is a policy. Since $m^{(\infty)}$ is a program, $f(m) = p$. Then p is invariant under μ_σ, and the process generated by $(p, \sigma^{(\infty)})$ is stationary.

Rewards associated with the process $\{\hat{s}_n\}$ generated by (p, π) and the program $\{m_n\}$ defined by (3) are equal to

$$\mathrm{E}_{p,\pi} R = \sum_{n \in N} \beta^{n-1} w(m_n).$$

Therefore, for $p \in P(S)$,

$$\sup_{\Pi} \mathrm{E}_{p,\pi} R = V(p). \tag{4}$$

Hence, the process $\{\hat{s}_n\}$ with an initial distribution p is optimal iff the corresponding program $\{m_n\} \in M(p)$ is optimal.

Remark. Jeanjean [5] and Schäl [8] used a similar representation of a dynamic programming problem in measure spaces.

4. Stationary process. In this section we prove the existence of a stationary process of successive states of the system. We assume the following:

A1. *The set of states S is compact.*

A2. *The multifunction φ is upper semicontinuous and compact-valued.*

A3. *The transition probability q is continuous, i.e. for each $u \in C(S$ the function v: $G \to R$ defined by*

$$v(s, a) := \int_S u(t) q(dt \mid s, a) \tag{5}$$

is continuous.

THEOREM 1. *If a discounted dynamic programming problem satisfies assumptions* A1-A3, *then there exist an initial distribution* p *and a stationary policy* $\sigma^{(\infty)}$ *such that the generated process is stationary.*

Proof. It suffices to show the existence of a stationary program in the associate deterministic model.

Not every $m \in P(G)$ generates a stationary program. Let ψ_1 be the multifunction from $P(G)$ to $P(G)$ defined by

$$\psi_1(m) := \psi\big(f(m)\big).$$

LEMMA 1. *A measure* $m \in P(G)$ *defines a stationary program iff* m *is a fixed point of* ψ_1.

This lemma is an immediate consequence of the definition of a program.

We show that $P(G)$ and ψ_1 satisfy the assumptions of the Fan-Kakutani fixed-point theorem.

$P(G)$ is a convex subset of $C^*(G)$, dual to $C(G)$. Under assumptions A1 and A2 the set

$$\varphi(S) := \bigcup_{s \in S} \varphi(s)$$

is compact (see Berge [1], Theorem 3, p. 116). Hence, G is compact as a closed subset of $S\varphi(S)$. Then $P(G)$ is compact and metrizable in the weak* topology of $C^*(G)$. Throughout the remainder of the proof we consider the spaces $P(G)$ and $P(S)$ with the weak* topology.

Now we prove that the multifunction ψ_1 is upper semicontinuous and compact-valued. Since $P(G)$ is compact, it suffices to show that ψ_1 is closed. By the definition of ψ_1, if f is continuous and ψ is closed, then ψ_1 is closed.

We first show that the function f is continuous. Let $m_n, m \in P(G)$, $\lim_n m_n = m$, and let $p_n := f(m_n)$, $p := f(m)$. For any $u \in C(S)$

$$\int_S u(t)p_n(dt) = \int_G \Big(\int_S u(t)q(dt \mid s, a) \Big) m_n\big(d(s, a)\big) = \int_G v(s, a)m_n\big(d(s, a)\big),$$

where v is defined by (5). Since v is continuous, we have

$$\lim_n \int_S u(t)p_n(dt) = \int_G v(s, a)m\big(d(s, a)\big) = \int_S u(t)p(dt).$$

Hence

$$\lim_n f(m_n) = f(m).$$

Next we prove that the multifunction ψ is closed. Let

$$p_n, p \in P(S), \quad m_n, m \in P(G), \quad m_n \in \psi(p_n) \quad \text{for } n \in N,$$

$$\lim_n p_n = p \quad \text{and} \quad \lim_n m_n = m.$$

We must show that $m \in \psi(p)$. Consider the probability measure p' on S defined by

$$p'(T) := m(TA \cap G),$$

where T is a Borel subset of S. Note that $m \in \psi(p)$ if and only if $p' = p$. For any $u \in C(S)$ and $n \in N$,

$$\int_S u(s) p_n(ds) = \int_G u(s) m_n\big(d(s, a)\big).$$

Hence

$$\int_S u(s) p(ds) = \int_G u(s) m\big(d(s, a)\big) = \int_S u(s) p'(ds).$$

Thus it follows that $p' = p$.

Let $m_1, m_2 \in \psi(p)$ for some $p \in P(S)$, and let $0 \leqslant \lambda \leqslant 1$. We show that $\lambda m_1 + (1-\lambda) m_2 \in \psi(p)$. For any Borel subset $T \subset S$,

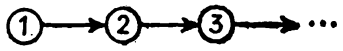
$$\big(\lambda m_1 + (1-\lambda) m_2\big)(TA \cap G) = \lambda m_1(TA \cap G) + (1-\lambda) m_2(TA \cap G) = p(T).$$

Therefore, $\psi(p)$ is a convex subset of $P(G)$.

By the Fan-Kakutani theorem, the multifunction ψ_1 has a fixed point. Hence, there exists a stationary program in the deterministic problem, which completes the proof.

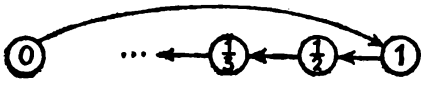
Theorem 1 shows that in every system satisfying our assumptions there is a possibility of long-run stationary behaviour. Simple examples show that we cannot weaken assumptions of this theorem.

Example 1. We have



Let $S = N$, $\varphi(s) = \{1\}$ for all s, $q(T \mid s, 1) = I_T(s+1)$, where I_T is the indicator function of $T \subset S$. The model satisfies all assumptions of Theorem 1 except the compactness of S. It is obvious that there is no stationary process in S.

Example 2. We have



Let

$$S = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}, \quad \varphi(s) = \{1\}, \quad q(T \mid 0, 1) = I_T(1),$$

and

$$q\left(T \middle| \frac{1}{n}, 1\right) = I_T\left(\frac{1}{n+1}\right) \quad \text{for } n \in N.$$

This time all assumptions except the continuity of q are satisfied, and again there is no stationary process.

Example 3. This is the reformulation of Example 2. Let

$$S = A = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}, \quad \varphi(0) = \{1\}, \quad \varphi\left(\frac{1}{n}\right) = \left\{\frac{1}{n+1}\right\} \text{ for } n \in N,$$

$$q(T \mid s, a) = I_T(a) \quad \text{for } T \subset S,\ s \in S,\ a \in A.$$

All assumptions of Theorem 1 except the upper semicontinuity of φ are satisfied, and there is no stationary process in S.

5. Stationary optimal process. We show the existence of an optimal process which is stationary. Consider the following assumptions:

A2′. *The multifunction φ is continuous and compact-valued.*

A4. *The reward function r is continuous.*

We need the following lemma:

LEMMA 2 (cf. Jeanjean [5], Theorem 3.1.2). *Under assumptions* A1, A2′, A3 *and* A4 *the optimal reward function v^* is continuous.*

Proof. Let L be the operator defined on $C(S)$ by

$$Lu(s) := \sup_{a \in \varphi(s)} v(s, a), \quad u \in C(S),\ s \in S,$$

where

$$v(s, a) := \int_S \big(r(s, a, t) + \beta u(t)\big) q(dt \mid s, a).$$

We show that v is continuous.

Let (s_n, a_n), $(s_0, a_0) \in G$, and $\lim_n (s_n, a_n) = (s_0, a_0)$. Now

$$|v(s_n, a_n) - v(s_0, a_0)| \leqslant \int_S |r(s_n, a_n, t) - r(s_0, a_0, t)|\, q(dt \mid s_n, a_n) +$$

$$+ \left| \int_S [r(s_0, a_0, t) + \beta u(t)][q(dt \mid s_n, a_n) - q(dt \mid s_0, a_0)] \right|.$$

By the continuity of q, the second term on the right-hand side of this inequality converges to zero as $n \to \infty$. Since r is continuous on the compact GS, for any $\varepsilon > 0$ there exists $n_0 \in N$ such that

$$|r(s_n, a_n, t) - r(s_0, a_0, t)| < \varepsilon \quad \text{for all } n \geqslant n_0,\ t \in S.$$

Then the first term also converges to zero.

From the continuity of v and assumption A2′ it follows that Lu is continuous (see Berge [1], p. 122). Hence, L is an endomorphism of $C(S)$. It is easily established that L is a contraction. By the Banach fixed-point theorem, there exists $u_0 \in C(S)$ such that $Lu_0 = u_0$. Since v^* is a unique bounded solution of the optimality equation (see Strauch [9], Theorem 8.2), $u_0 = v^*$. This completes the proof.

THEOREM 2. *If a discounted dynamic programming problem satisfies assumptions* A1, A2′, A3 *and* A4, *then there exist an initial distribution p on S and a stationary policy $\sigma^{(\infty)}$ such that the generated process $\{\hat{s}_n\}$ is stationary and optimal.*

Proof. By the correspondence between the stochastic and deterministic models, it suffices to prove the existence of a stationary optimal program. Let ψ_2 be the multifunction from $P(G)$ to $P(G)$ defined by

$$\psi_2(m) := \{m' \in \psi_1(m):\ V(f(m)) = w(m') + \beta V(f(m'))\}.$$

LEMMA 3. *A measure $m \in P(G)$ defines a stationary optimal program iff m is a fixed point of ψ_2.*

This is an immediate consequence of Lemma 1 and (2).

We apply the Fan-Kakutani theorem to $P(G)$ and ψ_2. $P(G)$ is a convex compact subset of $C^*(G)$ endowed with the weak* topology (see the proof of Theorem 1). Now we show that the multifunction ψ_2 is closed and convex-valued.

Since the function $\int_S r(s, a, t) q(dt|s, a)$ is continuous on G (see the proof of Lemma 2), w is continuous on $P(G)$. It follows from (1) and (4) that

$$V(p) = \int_S v^*(s) p(ds), \quad p \in P(S).$$

By virtue of Lemma 2, v^* is continuous on S. Thus V is continuous on $P(S)$. In the proof of Theorem 1 we have shown that f is continuous and ψ_1 is closed. The function $w(\cdot) + \beta V(f(\cdot))$ is continuous, and hence attains its supremum on $\psi_1(m)$. By the deterministic optimality equation,

this supremum is equal to $V(f(m))$. Consequently, $\psi_2(m)$ is non-empty for $m \in P(G)$.

Let

$$m_n, m \in P(G), \quad m'_n \in \psi_2(m_n) \quad \text{for } n \in N,$$

$$\lim_n m_n = m \quad \text{and} \quad \lim_n m'_n = m'.$$

Since ψ_1 is closed, $m' \in \psi_1(m)$. For each $n \in N$,

$$V(f(m_n)) = w(m'_n) + \beta V(f(m'_n)).$$

By the continuity of w, f and V, we have

$$V(f(m)) = w(m') + \beta V(f(m')).$$

Hence, $m' \in \psi_2(m)$, which proves that ψ_2 is closed.

Let $m_1, m_2 \in \psi_2(m)$ for some $m \in P(G)$, $0 \leqslant \lambda \leqslant 1$, and $m' := \lambda m_1 + (1-\lambda)m_2$. Since $\psi_1(m)$ is convex, $m' \in \psi_1(m)$. It is easily established that the conditions

$$V(f(m)) = w(m_i) + \beta V(f(m_i)) \quad (i = 1, 2)$$

imply

$$V(f(m)) = w(m') + \beta V(f(m')).$$

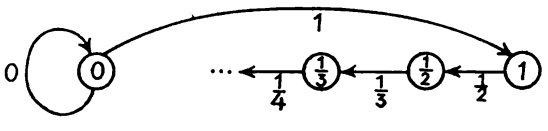
Consequently, $m' \in \psi_2(m)$.

In order to complete the proof, it suffices to apply the Fan-Kakutani fixed-point theorem and Lemma 3.

Theorem 2 can be applied to the stochastic model of economic growth. There are many random factors which affect the production, e.g. weather, prices on external markets, technological progress. In every stationary stochastic economy satisfying our assumptions there exists at least one optimal growth process which is stationary.

The following example shows that assumption A2′ cannot be weakened to A2:

Example 4. We have



Let

$$S = A = \left\{0, 1, \frac{1}{2}, \frac{1}{3}, \ldots\right\}, \quad \varphi(0) = \{0, 1\}, \quad \varphi\left(\frac{1}{n}\right) = \left\{\frac{1}{n+1}\right\}, \ n \in N,$$

$$q(T \mid s, a) = I_T(a) \text{ for } T \subset S, \ s \in S, \ a \in A, \quad r(s, a, t) = t.$$

All assumptions of Theorem 2, except the continuity of φ, are satisfied and there is no stationary optimal process. Note that φ is upper semicontinuous.

Remark. In general, the *best stationary process* (i.e. maximizing one-step expected reward among stationary processes) is not optimal over the infinite future. In Example 4, the sequence $\{0, 0, 0, \ldots\}$ is a unique stationary process, but it is not optimal.

References

[1] C. Berge, *Espaces topologiques*, Paris 1959.

[2] D. Blackwell, *Discounted dynamic programming*, Ann. Math. Statist. 36 (1965), p. 226-235.

[3] Fan Ky, *Fixed point and minimax theorems in locally convex topological linear spaces*, Proc. Nat. Acad. Sci. U.S.A. 38 (1952), p. 121-126.

[4] K. Hinderer, *Foundations of non-stationary dynamic programming with discrete time parameter*, Lectures Notes in Operation Research and Math. Systems 33, Berlin 1970.

[5] P. Jeanjean, *Optimal growth with stochastic technology in a multisector economy*, Collaborative Research on Economic Systems and Organization, Tech. Report No. 16, University of California, Berkeley 1972.

[6] A. Nowak, *On a general dynamic programming problem*, Coll. Math. 37 (1977), p. 131-138.

[7] K. R. Partharasaty, *Probability measures on metric spaces*, New York 1967.

[8] M. Schäl, *On dynamic programming: Compactness of the space of policies*, Stoch. Processes Appl. (to appear).

[9] R. E. Strauch, *Negative dynamic programming*, Ann. Math. Statist. 37 (1966), p. 871-890.

[10] W. Sutherland, *On optimal development in multi-sectoral economy: The discounted case*, Rev. Economic Studies 37 (1970), p. 585-596.

INSTITUTE OF MATHEMATICS
SILESIAN UNIVERSITY
40-007 KATOWICE

Received on 10. 6. 1974;
revised version on 20. 1. 1976

A. NOWAK (Katowice)

STACJONARNY PROCES OPTYMALNY W PROBLEMIE PROGRAMOWANIA DYNAMICZNEGO Z DYSKONTEM

STRESZCZENIE

W pracy rozpatrywany jest markowowski proces decyzyjny z dyskretnym czasem. Dowolna polityka π, wraz z rozkładem początkowym p, definiuje proces stochastyczny w przestrzeni stanów. Proces ten nazywamy *optymalnym*, gdy polityka π maksymalizuje wartość oczekiwaną całkowitej dyskontowanej wypłaty. W głównym twierdzeniu pracy podajemy warunki wystarczające na to, aby istniał stacjonarny proces optymalny. Jest to przeniesienie na przypadek stochastyczny wyniku Sutherlanda [10]. Nasz dowód oparty jest na twierdzeniu Fana-Kakutaniego o punkcie stałym. Pomocniczo wykazana została równoważność rozpatrywanego modelu stochastycznego i modelu deterministycznego zbudowanego w przestrzeniach miar.