Z. Wahrscheinlichkeitstheorie verw. Geb. 21, 279–288 (1972)

On Continuous Dynamic Programming with Discrete Time-Parameter

MANFRED SCHÄL

1. Introduction

A rigorous foundation of stochastic dynamic programming was given by Blackwell [2] and Strauch [11], who treated stationary models. The decision model which is taken as a basis of the present work is a slight generalization of the model of Blackwell and Strauch allowing the discount factor to depend on the state of the system and the selected action. Thus we include models arising from Markov renewal processes or semi-Markoff processes, respectively, as well as from stopping and search problems. This model is a special case of a non-stationary decision model as defined by Hinderer [5], but preserves the stationary structure of the model of Blackwell and Strauch. Thus, on the one hand a series of results obtained by Hinderer [5] and [6], e.g. the universal measurability of the optimal return and the validity of the optimality equation, apply to our model. On the other hand, results of Blackwell and Strauch ([2, 3, 11]) concerning the stationary character, e.g. the optimality of stationary plans, can be generalized to our model by using many of their ideas. In [9] it was investigated to what extent it is justified to confine ourselves to stationary plans. The main purpose of the present paper is to give sufficient conditions for the existence of optimal and ε-optimal plans. We assume the reward, the discount factor, and the transition law to depend continuously on the actions. Then, under certain convergence conditions on the expected total return under admissible plans, there exists a stationary ε-optimal plan and, if moreover the sets of admissible actions are compact, there exists a stationary optimal plan. Similar results were obtained by Maitra ([7, 8]) who on the one hand presumes a weaker form of continuity (more precisely the upper semi-continuity) but on the other hand assumes the reward and the transition law to depend continuously on both the states and the actions. As to the convergence conditions imposed on the expected return, we will essentially treat the negative bounded case (in the terminology of Strauch). However the assumption that the reward is negative will be generalized to a large extend, thus including the so-called discounted case. These conditions were found by Hinderer in [6] for the more general non-stationary model and were adjusted to the model of the present work.

2. The Decision Model

The decision model is a tupel $((S, \mathfrak{S}), (A, \mathfrak{A}), D, p, q, \beta, r)$ of the following meaning:

(i) $(S, \mathfrak{S})$ stands for the state space. $(S, \mathfrak{S})$ is assumed to be a standard Borel space, i.e. S is a non-empty Borel subset of a Polish (complete, separable, metric) space and $\mathfrak{S}$ is the system of all Borel subsets of S.

(ii) The standard Borel space $(A, \mathfrak{A})$ is the space of actions.

(iii) $D \in \mathfrak{S} \otimes \mathfrak{A}$ is assumed to contain the graph of a measurable map of S into A. D_s, the section of D at s, is called the set of admissible actions if the system is in state s.

(iv) p is a probability measure on S, the so-called initial distribution.

(v) The so-called transition law q is a transition probability from $S \times A$ to S. $q(s, a, \cdot)$ is the distribution of the state next visited by the system if the system is in state s and the action a is taken.

(vi) β is a bounded measurable map of $S \times A \times S$ into the set of the non-negative numbers and can be interpreted as a discount factor.

(vii) The reward r is an extended real valued measurable function on $S \times A \times S$ such that

$$\bar{r}(s, a) = \int q(s, a, dt)\, r(s, a, t) \tag{2.1}$$

exists for any $s \in S$, $a \in A$.

It is sufficient that q is defined on D and β and r are defined on $D \times S$; their definition can be extended to $S \times A \times S$ in an arbitrary manner.

We write $H_1 = S$, $H_{n+1} = D \times H_n$, $n \in \mathbb{N}$[1]. As usual, a randomized plan $\pi = (\pi_n)$ is defined as a sequence of transition probabilities π_n from H_n to A such that $\pi_n(s_1, a_1, \dots, s_n, D_{s_n}) = 1$ for any $(s_1, a_1, \dots, s_n) \in H_n$, $n \in \mathbb{N}$. A deterministic plan is a sequence $f = (f_n)$ of measurable maps f_n: $H_n \to A$ such that $f_n((s_1, a_1, \dots, s_n)) \in D_{s_n}$ for any $(s_1, a_1, \dots, s_n) \in H_n$, $n \in \mathbb{N}$. Obviously a deterministic plan can be described by a randomized plan π where the probabilities $\pi_n(h, .)$ are concentrated at the point $f_n(h)$. A plan π or f is called a Markov plan, if π_n or f_n, respectively, does not depend on $(s_1, a_1, \dots, a_{n-1})$. Thus, a randomized Markov plan $\pi = (\pi_n)$ is given by a sequence of transition probabilities π_n from S to A and a deterministic Markov plan $f = (f_n)$ is determined by a sequence $f_n \in D^S$, where

$$D^S = \{g;\ g\colon (S, \mathfrak{S}) \to (A, \mathfrak{A}), g(s) \in D_s \quad \text{for } s \in S\}.$$

A deterministic Markov plan is called a stationary plan if $f_n \equiv f$ for some $f \in D^S$ and all n. For such a plan (f_n) we write $f^{(\infty)}$. The initial distribution, the transition law and a plan σ define a probability measure P_σ on

$$(H, \mathfrak{H}) = (S \times A \times S \times A \times \cdots, \mathfrak{S} \otimes \mathfrak{A} \otimes \mathfrak{S} \otimes \mathfrak{A} \otimes \cdots)$$

and therefore a random process $(\hat{s}_1, \hat{a}_1, \hat{s}_2, \dots)$ (cf. [5] p. 80). $\hat{s}_n$ and $\hat{a}_n$ stand for the projection from H into the n-th state space and the n-th action space, respectively, i.e. the random variables $\hat{s}_n$ and $\hat{a}_n$ describe the state of the system and the action at time n. If the history of the system up to the $(n+1)$-th stage is $(s_1, a_1, \dots, s_{n+1})$, then one will receive during the n-th time period

$$r_n(s_1, a_1, \dots, s_{n+1}) = \beta(s_1, a_1, s_2) \dots \beta(s_{n-1}, a_{n-1}, s_n)\, r(s_n, a_n, s_{n+1}) \tag{2.2}$$

[1] Let $\mathbb{N}$ denote the set of the positive integers.

(especially $r_1 \equiv r$). As the reward depends on the whole history at time n, the decision model cannot be regarded as a Markovian decision model in the sense of Blackwell and Strauch.

3. Selection Theorem

The proof of the existence of an optimal or ε-optimal plan rests on a selection theorem which is proved in [10] and will be quoted below. The selection theorem will play the same role as the selection theorem of Dubins and Savage in the paper [7] of Maitra. A preliminary result for the proof of the selection theorem in [10] is the following lemma which will also be used in Section 5.

Throughout this paper, we shall make the following

Assumption A. D_s is compact for $s \in S$ or D_s is closed for $s \in S$ and A is locally compact. There is a denumerable set $A' \subset A$ such that $A' \cap D_s$ is dense in D_s for any $s \in S$. (A' is independent of s.)

Note that if A is assumed to be locally compact we still assume that A is separable. Then A has a countable base. The condition that $A' \cap D_s$ is dense in D_s is likely to be satisfied in any practical problem. In the important case where $D_s = A$ for $s \in S$, Assumption A is satisfied if A is locally compact. Then the separability of A ensures the existence of such a set A'.

Lemma 3.1. *If u is a real measurable functlon on D such that $u(s, \cdot)$ is continuous for any $s \in S$, then $\sup_{a \in D_s} u(s, a)$ is a measurable (possibly extended real valued) function on S.*

The *proof* of Lemma 3.1 rests on the fact that

$$\sup_{a \in D_s} u(s, a) = \sup_{a \in A' \cap D_s} u(s, a).$$

Theorem 3.2. *Let u satisfy the condition of Lemma* 3.1.

a) *If D_s is compact for $s \in S$, then there exists some $f \in D^S$ such that*

$$u(s, f(s)) = \max_{a \in D_s} u(s, a).$$

b) *If the function $u(s, \cdot)$ attains its supremum on D_s for $s \in S$, then there exists some $f \in S^S$ such that*

$$u(s, f(s)) = \max_{a \in D_s} u(s, a).$$

c) *If ε is a strictly positive measurable function on S, then there exists some $f \in D^S$ such that*

$$u(s, f(s)) \geqq \sup_{a \in D_s} u(s, a) - \varepsilon(s).$$

4. Convergence and Continuity Assumptions

In this section we first introduce a condition which ensures that the decision model with infinite horizon can be approximated by a model with finite horizon in the following sense: the positive part of the tail of the total return can be neglected such that the method of proof described by Strauch [11] as "the method of improving the tail" is applicable.

We write $\mathrm{a}^{\pm}=\max(0, \pm a)$ for any real a,

$$R_{n,\pm}=\sum_{\nu=n}^{\infty} r_\nu^{\pm} \quad\text{and}\quad W_{n,\pm}=\sup_{\pi} E_\pi(R_{n,\pm}|\hat{s}_1).$$

Throughout the paper (exept for Theorem 6.3 below), we shall make the

Assumption B. $W_{1,+}(s)<\infty$, $W_{n,+}(s)\to 0$ $(n\to\infty)$ for $s\in S$.

Assumption B holds if the reward is negative of if the reward is bounded from above and the following condition K is satisfied.

$$K\colon\ \sup_{\pi,s} E_\pi\big(\beta(\hat{s}_1,\hat{a}_1,\hat{s}_2)\dots\beta(\hat{s}_{k-1},\hat{a}_{k-1},\hat{s}_k)|\hat{s}_1=s\big)<1 \qquad \text{for some } k\in\mathbb{N}.$$

As has been shown by Hinderer in [6], the total return $R=\sum_{\nu=1}^{\infty} r_\nu$ exists $[\mathfrak{H}, P_\pi]$ for any plan π. Moreover,

$$I_\pi=E_\pi(R|\hat{s}_1) \quad\text{and}\quad I_\pi^n=E_\pi\left(\sum_{\nu=1}^{n} r_\nu|\hat{s}_1\right)$$

are defined. $I_\pi(s)$ and $I_\pi^n(s)$ are the expectations of the total return or of the return up to the n-th stage if we start in s. By Lemma 4.2 in [9] we have (if we choose p as the probability concentrated at any point $s\in S$):

Lemma 4.1. $I_\pi^n\to I_\pi$ $(n\to\infty)$.

If P and Q are two propabilities on $(S,\mathfrak{S})$, we write $\|P-Q\|$ for the total variation of the set function $P-Q$. Then $\|P-Q\|=2\sup_{B\in\mathfrak{S}}|P(B)-Q(B)|$. Throughout this paper we make the

Assumption C. 1) $\bar{r}(s,\cdot)$ is a continuous function on D_s for $s\in S$.

2) $\|q(s,a',\cdot)-q(s,a,\cdot)\|\to 0$ as $a'\to a$ for $s\in S$.

3) $\beta(s,\cdot,t)$ is a continuous function on D_s for $s,t\in S$.

In many practical problems $r(s,a,t)$ does not depend on t, hence $r\equiv\bar{r}$. If not so, a simple sufficient condition for Assumption C.1 is the following: If r is bounded and $r(s,\cdot,t)$ is a continuous function on D_s for any $s,t\in S$, then $\bar{r}(s,\cdot)$ is continuous. The proof is similar to the proof of Lemma 5.2 below. A well-known theorem of Scheffé (cp. [1]) yields a simple sufficient condition for Assumption C.2: If the probability measures $q(s,a,\cdot)$, $a\in D_s$, are dominated by a σ-finite measure μ (μ may depend on s), i.e. if $q(s,a,dt)=p(s,a,t)\,\mu(dt)$ for some non-negative function p, and if $p(s,\cdot,t)$ is continuous on D_s for $t\in S$, then $\int|p(s,a',t)-p(s,a,t)|\,\mu(dt)\to 0$ as $a'\to a$ and hence $\|q(s,a',\cdot)-q(s,a,\cdot)\|\to 0$.

It is clear that Assumption C is satisfied if A is countable. Then the sets D_s are compact if they are finite. Thus the results of the present paper can be regarded as a generalization of the corresponding results where A is assumed to be countable or "essentially countable by some plan π" in the terminology of Blackwell [2].

5. The Operators L and U

In this section we introduce some operators which are well-known in the theory of dynamic programming. We make use of the following notation. Let supp(Q) denote the support of the probability measure Q on S. Then we write for an arbitrary function u on S

$$\|u\|_s=\sup\{|u(t)|, t\in \bigcup_{a\in D_s} \operatorname{supp}(q(s,a,\cdot))\}$$

and define the set M of measurable function u on S by $M=\{u, \|u\|_s<\infty$ for $s\in S\}$. Obviously every bounded measurable function is contained in M. In many practical problems, it may happen that for any $s\in S$ there exists a compact set K^s such that $q(s,a,K^s)=1$, i.e. $\operatorname{supp}(q(s,a,\cdot))\subset K^s$ for $a\in D_s$. Then for example every continuous function is contained in M. As usual we write $\|u\|=\sup_{x\in X}|u(x)|$ for any extended real valued function u defined on any set X.

For any $u\in M$ we define

$$Lu(s,a)=\bar{r}(s,a)+\int q(s,a,dt)\,\beta(s,a,t)\,u(t) \qquad (s,a)\in D$$
$$L_f u(s)=Lu(s,f(s)) \qquad f\in D^S$$
$$U u(s)=\sup_{a\in D_s} Lu(s,a).$$

We may interpret $Lu(s,a)$ as the expected return if we are in state s, take action a, and receive a terminal return of $u(t)$ at the resulting state t. If in the above definition r is replaced by 0, then we write $\tilde{L}$ and $\tilde{U}$, respectively.

The following lemma interrelates the operators L, U and $\tilde{L}$, $\tilde{U}$. It is easily proved by induction on n.

Lemma 5.1. *Let $u, v\in M$.*

(a) $L^n_f(u+v)=L^n_f u+\tilde{L}^n_f v$ *if* $L^\nu_f u, \tilde{L}^\nu_f v\in M \quad 1\leqq\nu\leqq n-1$,

(b) $U^n(u+v)\leqq U^n u+\tilde{U}^n v$ *if* $U^\nu u, \tilde{U}^\nu v\in M \quad 1\leqq\nu\leqq n-1$.

Further, we can prove

Lemma 5.2. *If $u\in M$, then $Lu(s,\cdot)$ is continuous on D_s for any $s\in S$ and $Uu(\cdot)$ is measurable.*

Proof. In view of Lemma 3.1, it suffices to prove the continuity of $\tilde{L}u(s,\cdot)$. Now,

$$\begin{aligned}
&|\tilde{L}u(s,a)-\tilde{L}u(s,a')|\\
&\quad\leqq\left|\int q(s,a,dt)[\beta(s,a,t)-\beta(s,a',t)]\,u(t)\right|\\
&\qquad+\left|\int q(s,a,dt)\,\beta(s,a',t)\,u(t)-\int q(s,a',dt)\,\beta(s,a',t)\,u(t)\right|\\
&\quad\leqq\cdots+\|q(s,a,\cdot)-q(s,a',\cdot)\|\,\|\beta\|\,\|u\|_s.
\end{aligned}$$

From the dominated convergence theorem it is clear that the first term converges to zero as $a'\to a$. Also, by Assumption C.2, the second term converges to zero.

6. The Functions v^* and u_∞

As a consequence of Theorem 3.5 in Hinderer [6] one obtains

$$\sup_\pi I^n_\pi=U^n\,0=:u_n. \tag{6.1}$$

We are interested in the two functions

$$u_\infty=\varliminf u_n, \qquad v^*=\sup_\pi I_\pi$$

and their relationships. u_n is the optimal return if we terminate at the n-th stage with no terminal return and v^* is the optimal return from infinite stage play. It is known (cf. Theorems 3.2, 3.3 in Hinderer [6] or Theorem 7.1 in [9]) that u_n, $n \in \mathbb{N}$, and v^* are universally measurable and that v^* satisfies the optimality equation

$$v^* = U v^*. \tag{6.2}$$

Another optimality equation is $W_{n+1,+} = \tilde{U} W_{n,+}$ which implies

$$W_{n+1,+} = \tilde{U}^n W_{1,+}. \tag{6.3}$$

(6.3) derives from Eq. (2.10) in Hinderer [6] and is proved by modifying the original decision model such that $r_\nu \equiv 0$ $1 \leqq \nu \leqq n, n+1$ and $r_\nu = r_\nu^+$ $\nu > n, n+1$.

Theorem 6.1. $\lim u_n = u_\infty$ *exists and* $u_\infty \geqq v^*$. *If* r *is bounded or more generally if* $\|u_n\|_s < \infty$, $s \in S$, $n \in \mathbb{N}$, *then* u_∞ *is measurable. If* $\sup_n \|u_n\|_s < \infty$ *for* $s \in S$, *then* $U u_\infty \leqq u_\infty$.

Proof. In order to show $\overline{\lim}\, u_n = \underline{\lim}\, u_n$, we make use of

$$\sup_{n \geqq m} u_n \leqq u_m + W_{m+1,+}. \tag{6.4}$$

By use of Assumption B, we conclude that

$$\begin{aligned} \overline{\lim}\, u_m &= \lim_m \sup_{n \geqq m} u_n \\ &\leqq \underline{\lim}\, u_m + \lim W_{m+1,+} = \underline{\lim}\, u_m \end{aligned}$$

which establishes the existence of $\lim u_n$. Since $I_\pi^n \leqq u_n$, we know by Lemma 4.1 that $\lim_n I_\pi^n = I_\pi \leqq \lim_n u_n$ for any π, hence $v^* \leqq u_\infty$. If $\|u_n\|_s < \infty$, $s \in S$, $n \in \mathbb{N}$, then it is easily established that u_n is measurable on making use of Lemma 5.2. Hence u_∞ is measurable. If moreover $\sup_n \|u_n\|_s < \infty$, then $u_\infty \in M$. By means of the dominated convergence theorem, we finally obtain

$$\begin{aligned} \sup_a L u_\infty(s, a) &= \sup_a \left(\lim_n L u_n(s, a)\right) \\ &\leqq \underline{\lim}_n \sup_a L u_n(s, a) \\ &= \lim_n u_{n+1}(s), \end{aligned}$$

hence $U u_\infty \leqq u_\infty$.

The assumption $\sup_n \|u_n\|_s < \infty$, $s \in S$, is used in the present paper several times. Therefore, we give a sufficient condition in the following

Lemma 6.2. *If* $\|v^*\|_s < \infty$ *and* $\|W_{1,+}\|_s < \infty$, $s \in S$, *then* $\sup_n \|u_n\|_s < \infty$, $s \in S$.

This lemma is a consequence of the following inequalities, which derive from Theorem 6.1 and (6.4):

$$W_{1,+} \geqq u_n \geqq u_\infty - W_{n+1,+} \geqq v^* - W_{1,+}. \tag{6.5}$$

The conditions of Lemma 6.2 are satisfied if r is bounded and condition K holds. The awkward case that Assumption B holds and $\|W_{1,+}\|_s=\infty$ for some $s\in S$ is not likely to arise in any applications of the theory. If we know that $\|W_{1,+}\|_s<\infty$, e.g. if $r\leqq 0$, then it is clear that $\|v^*\|_s<\infty$ if a plan σ can be found such that $\|I_\sigma\|_s<\infty$.

It is known that the functions u_∞ and v^* are identical in the "positive case". This fact is generalized in the following theorem. It is to be noticed that the proof of Theorem 6.3 will be carried through without use of Assumption B.

Theorem 6.3. *Suppose that $W_{1,-}(s)<\infty$ for $s\in S$.*

(a) *If $\sup\limits_{a\in D_s}\bar{r}(s,a)\geqq 0$ for $s\in S$, then $u_n\uparrow u_\infty=v^*$ and v^* is measurable.*

(b) *If $W_{n,-}(s)\to 0$ as $n\to\infty$ for $s\in S$, then $u_\infty=v^*$. If moreover $\|u_n^-\|_s<\infty$, $n\in\mathbb{N}$, $s\in S$, then v^* is measurable.*

Proof. First we note that if for any measurable function u the negative part u^- is contained in M, then $U u$ is measurable by Lemma 5.2; for we can write $U u=\sup\limits_n U\min(n,u)$.

(a) By assumption we have $UO=u_1\geqq 0$ so that $U^n u_1\geqq U^n O$, i.e. $u_{n+1}\geqq u_n$. Since $u_n\geqq 0$, u_n is measurable for any n and hence u_∞ is measurable. It can now be shown as in Blackwell [3] that u_∞ is the smallest non-negative fixed point of U. To identify v^* with u_∞, it suffices to show $v^*\geqq 0$. By Theorem 7.1b in [9] v^* is a fixed point of U. Choosing p as the probability concentrated at s, we obtain from Lemma 4.2 in [9] that $I_\pi^n\to I_\pi$. Since $I_\pi^n\leqq u_n$, we have $I_\pi\leqq u_\infty$ for any π and hence $v^*\leqq u_\infty$, as in Theorem 6.1. But u_∞ is the smallest non-negative fixed point so that $v^*=u_\infty$. In order to show that $v^*\geqq 0$, choose $\varepsilon>0$ and pick ε_n, $n\in\mathbb{N}$, such that $\sum\limits_{\nu=1}^{\infty}\|\beta\|^\nu\varepsilon_\nu\leqq\varepsilon$. From Theorem 3.2 there exists some $f_n\in D^S$ such that $r(s,f_n(s))\geqq \sup\limits_a r(s,a)-\varepsilon_n\geqq-\varepsilon_n$. Proceeding inductively we obtain

$$I_\pi^n=L_{f_1}\ldots L_{f_n}0\geqq-\sum_{i=1}^{n}\|\beta\|^{i-1}\varepsilon_i\geqq-\varepsilon \quad \text{where } \pi=(f_1,f_2,f_3,\ldots).$$

Now $I_\pi^n\to I_\pi$, so that $I_\pi\geqq-\varepsilon$. Thus $v^*\geqq-\varepsilon$ for any $\varepsilon>0$, which completes the proof of (a).

(b) The first assertion of (b) is Theorem 7.1c in [9]. Now the fact that $u_n^-\in M$ implies the measurability of u_{n+1}, $n\in\mathbb{N}$. Thus u_∞ is measurable.

It is to be noticed that Assumption B as well as the assumption of Theorem 6.3b are satisfied if r is bounded and condition K is satisfied, especially if we have the "discounted case". Using the terminology of Dubins and Savage [4] and Strauch [11] we say that U conserves $v\in M$ if $Uv\geqq v$. Such functions are lower bounds for v^* as is shown in

Theorem 6.4. *Suppose that U conserves $v\in M$ and $v\leqq c\,W_{1,+}$ for some $c\geqq 0$.*

(a) *If D_s is compact for $s\in S$, then there exists a stationary plan $f^{(\infty)}$ such that $I_{f^{(\infty)}}\geqq v$.*

(b) *For any $\varepsilon>0$, there exists a deterministic Markov plan π such that $I_\pi\geqq v-\varepsilon$.*

Proof. (a) Referring to Theorem 3.2 a, we know that there exists some $f \in D^S$ such that

$$L_f v(s) = L v(s, f(s)) = \max_{a \in D_s} L v(s, a) = U v(s) \geqq v(s).$$

Proceeding inductively, we obtain by Lemma 5.1 and Eq. (5.3)

$$\begin{aligned} v &\leqq L_f^n v \\ &= L_f^n 0 + \tilde{L}_f^n v \\ &\leqq L_f^n 0 + c \tilde{L}_f^n W_{1,+} \\ &\leqq L_f^n 0 + c \tilde{U}^n W_{1,+} \\ &= L_f^n 0 + c W_{n,+}. \end{aligned}$$

Now the fact that $L_f^n 0 = I_{f^{(\infty)}}^n \to I_{f^{(\infty)}}$ and $W_{n,+} \to 0$ implies $v \leqq I_{f^{(\infty)}}$.

(b) Pick ε_n, $n \in \mathbb{N}$, such that $\sum_{\nu=1}^{\infty} \|\beta\|^\nu \varepsilon_\nu \leqq \varepsilon$. By Theorem 3.2c there exists some $f_n \in D^S$ such that $L_{f_n} v \geqq U v - \varepsilon_n \geqq v - \varepsilon_n$. It is easily established inductively that

$$L_{f_1} \dots L_{f_n} v \geqq v - \sum_{\nu=1}^{n} \|\beta\|^{\nu-1} \varepsilon_\nu \geqq v - \varepsilon.$$

Paralleling the proof of a), we finally obtain $I_\pi \geqq v - \varepsilon$ where $\pi = (f_1, f_2, f_3, \dots)$.

Corollary 6.5. *If* $\sup_n \|u_n\|_s < \infty$, $s \in S$, *and* U *conserves* u_∞, *then* $U u_\infty = u_\infty$ *and* $u_\infty = v^*$.

Proof. By Theorem 6.1 we know that $u_\infty \in M$, $u_\infty \geqq v^*$, and $U u_\infty \leqq u_\infty$. On identifying $v = u_\infty$ and $c = 1$ in Theorem 6.4, however, we have $v^* \geqq u_\infty$.

7. Optimal Plans

In this section, we shall give sufficient conditions for the existence of an optimal plan. A plan π is called optimal if $I_\pi = v^*$. π is said to be ε-optimal for some $\varepsilon > 0$ if $I_\pi \geqq v^* - \varepsilon$.

Theorem 7.1. *Suppose that* v^* *is (Borel-)measurable and* $\|v^*\|_s < \infty$ *for* $s \in S$. *If* $L v^*(s, \cdot)$ *attains its supremum on* D_s *for* $s \in S$, *then there exists a stationary optimal plan* $f^{(\infty)}$.

Proof. By Lemma 5.2, we know that $L v^*(s, \cdot)$ is continuous since $v^* \in M$. An appeal to Theorem 3.2 b proves the existence of an $f \in D^S$ such that $L_f v^* = U v^* = v^*$. Proceeding inductively we obtain

$$\begin{aligned} v^* &= L_f^n v^* = L_f^n 0 + \tilde{L}_f^n v^* \\ &\leqq I_{f^{(\infty)}}^n + \tilde{L}_f^n W_{1,+} \\ &\leqq I_{f^{(\infty)}}^n + \tilde{U}^n W_{1,+} \\ &= I_{f^{(\infty)}}^n + W_{n,+}. \end{aligned}$$

Now, the obvious passage to the limit proves $v^* \leqq I_{f^{(\infty)}}$.

Note that Theorem 6.3 yields sufficient conditions for the measurability of v^*. If we require that D_s is compact, Theorem 7.1 quarantees the existence of an optimal plan. We shall give here another proof which can be carried through without the assumption of the measurability of v^*. It is clear, however, that the existence of an optimal plan (or more generally the existence of ε-optimal plans for $\varepsilon > 0$) imply the measurability of v^*.

Theorem 7.2. *If D_s is compact and $\sup_n \|u_n\|_s < \infty$ for $s \in S$, then there exists a stationary optimal plan $f^{(\infty)}$ and $u_\infty = v^*$.*

Proof. In view of Corollary 6.5 and Theorem 6.4, it suffices to prove that U conserves u_∞. Fix some $s \in S$. By Lemma 5.2, the function $L u_{n-1}(s, \cdot)$ is continuous for $n \geqq 2$ and hence attains its supremum on D_s in a_n, say. Now there exists a subsequence (a_{n_ν}) of (a_n) such that $a_{n_\nu} \to a \in D_s$ (say). Set $\bar{m} = \sup_n \|u_n\|_s$ and choose $\varepsilon > 0$. Then there exists some ν_0 such that for $\nu \geqq \nu_0$ $|\bar{r}(s, a) - \bar{r}(s, a_{n_\nu})| < \varepsilon/2$ and

$$\|q(s, a, \cdot) - q(s, a_{n_\nu}, \cdot)\| \, \|\beta\| + \int q(s, a, dt)\, |\beta(s, a, t) - \beta(s, a_{n_\nu}, t)| < \varepsilon/2\bar{m}.$$

These inequalities imply

$$|L u_{n_\nu - 1}(s, a) - L u_{n_\nu - 1}(s, a_{n_\nu})| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2\bar{m}} \, \|u_{n_\nu - 1}\| \leqq \varepsilon.$$

Thus,

$$\begin{aligned} L u_{n_\nu - 1}(s, a) &\geqq L u_{n_\nu - 1}(s, a_{n_\nu}) - \varepsilon \\ &= U u_{n_\nu - 1} - \varepsilon = u_{n_\nu} - \varepsilon. \end{aligned}$$

From the dominated convergence theorem we know that $L u_{n_\nu - 1}(s, a) \to L u_\infty(s, a)$. Hence $U u_\infty(s) \geqq L u_\infty(s, a) \geqq u_\infty(s) - \varepsilon$ and, since ε is an arbitrary positive number, $U u_\infty(s) \geqq u_\infty(s)$.

8. ε-Optimal Plans

Let c be any positive function on S. Then $C_f(s) = \sum_{\nu=0}^{\infty} \tilde{L}_f^\nu c(s)$ (where $\tilde{L}_f^0 c = c$) can be interpreted as the total return under plan $f^{(\infty)}$ if we start in $\hat{s}_1 = s$ and receive a return of $c(\hat{s}_n)$ at the n-th stage independently of the action $\hat{a}_n$.

Lemma 8.1. *If U conserves $v \in M$ and $v \leqq \gamma W_{1,+}$ for some $\gamma \geqq 0$, then, for any positive function c on S, there exists a stationary plan $f^{(\infty)}$ such that $I_{f^{(\infty)}} \geqq v - C_f$.*

Proof. From Theorem 3.2c we conclude that there exists some $f \in D^S$ such that $L_f v \geqq U v - c$, and hence $L_f v \geqq v - c$. By use of Lemma 5.1, it is easily established inductively that $L_f^n v \geqq v - \sum_{\nu=0}^{n-1} \tilde{L}_f^\nu c$. We can now show exactly as in the proof of Theorem 6.4 that $I_{f^{(\infty)}} \geqq v - C_f$.

Theorem 8.2. *If r is bounded and condition K is satisfied then there exists a stationary ε-optimal plan $f^{(\infty)}$.*

Proof. It is easily established that the assumption imply that

$$\bar{m} = \sup \left\| \sum_{\nu=0}^{\infty} \tilde{L}_f^{\nu} 1 \right\| < \infty \quad \text{and} \quad \|v^*\| < \infty.$$

From Theorem 6.3b we conclude that v^* is measurable. On identifying $v = v^*$ and $c \equiv \varepsilon/\bar{m}$ for any $\varepsilon > 0$, we know by Lemma 8.1 that there exists some stationary plan $f^{(\infty)}$ such that

$$I_{f^{(\infty)}} \geqq v^* - \frac{\varepsilon}{\bar{m}} \sum_{\nu=0}^{\infty} \tilde{L}_f^{\nu} 1 \geqq v^* - \varepsilon.$$

References

1. Billingsley, P.: Convergences of probability measures. New York: John Wiley & Sons, Inc. 1968.
2. Blackwell, D.: Discounted dynamic programming. Ann. math. Statistics **36**, 226–235 (1965).
3. – Positive dynamic programming. Proc. 5th Berkeley Sympos. math. Statist. Probab. **I**, 415–418 (1965).
4. Dubins, L.E., Savage, L.J.: How to gamble if you must. New York: McGraw-Hill 1965.
5. Hinderer, K.: Foundations of non-stationary dynamic programming with discrete time-parameter. Lecture Notes in Operations Research and Mathematical Systems, vol. 33. Berlin-Heidelberg-New York: Springer 1970.
6. – Instationäre dynamische Optimierung bei schwachen Voraussetzungen über die Gewinnfunktionen. Abh. math. Sem. Univ. Hamburg **36**, 208–223 (1971).
7. Maitra, A.: Discounted dynamic programming on compact metric spaces. Sankhyá **30**, Ser. A, 211–216 (1968).
8. – A note on positive dynamic programming. Ann. math. Statistics **40**, 316–319 (1969).
9. Schäl, M.: Ein verallgemeinertes stationäres Entscheidungsmodell der dynamischen Optimierung. Vol. X, 145–162: Methods of operations research, ed. R. Henn. Meisenheim: Anton Hain 1971.
10. – Ein Auswahlsatz für Optimierungsprobleme. (To be published.)
11. Strauch, R.E.: Negative dynamic programming. Ann. math. Statistics **37**, 871–890 (1966).

Dr. Manfred Schäl
Institut für Mathematische Stochastik
der Universität
D-2000 Hamburg 13, Rothenbaumchaussee 45
Germany

(Received February 22, 1971)