A necessary and sufficient condition of optimality for Markovian control problems

By D. VERMES in Szeged

To Professor Béla Sz.-Nagy on his 60th birthday

1. Introduction

The vital importance of the Markov property of processes in control theory was first pointed out by R. BELLMAN in [1]. As a consequence of this perception he was able to give a necessary and sufficient condition for the optimality of control for discrete-time Markov decision processes. For continuous-time processes there are no general results known. In the deterministic case the Bellman equation could only be proved to be a sufficient condition of optimality (cf. [1], [2]). On the consideration of stochastic problems there arise additional difficulties from the absence of a sufficiently general definition of the controlled Markov process. FLEMING [6], [7], MANDL [12], WONHAM [15] have given conditions for the optimality of the control of diffusion processes, governed by stochastic differential equations, GRIGELIONIS and SHIRYAEV [8] have investigated the properties of the optimal expense function of processes for which the value of the control parameter has been allowed to change only in fixed "switching times", while KUSHNER [11] has regarded families of Markov processes with the strategy space as an index set and has given a sufficient condition of optimality for them.

The aim of the present paper is to prove a necessary and sufficient optimality condition for the control of general discrete or continuous-time Markov processes, by employing the functional analytic theory of Markov processes, which was developed by HILLE, YOSIDA, FELLER, DYNKIN, and others in the last 25 years.

First we give a sufficiently general formulation of the optimal stochastic control problem. Controlled Markov processes are defined similarly to [11], [13] as families of processes. Unless otherwise stated, our considerations hold both for discrete- and continuous-time processes, i.e., both the set N of the non-negative integers and the set R^+ of the non-negative reals are allowed to be the time axis T. In our investigations we shall necessarily consider time dependent controls and so time-inhomogeneous (non-autonomous) processes. The theory of one parameter semigroups of operators — which will be extensively used throughout the paper — is only adequate to describe time-homogeneous processes. But in [13] it was proved, that to every inhomogeneous process on the state space E', there exists an equivalent homogeneous process on the product state space $E = T \times E'$. This way we do not restrict the generality by assuming that our processes are homogeneous and their state space is of the form $T \times E'$. Studying Markov processes we shall refer to the monographs of DYNKIN [4], [5]. For control strategies we allow measurable mappings of the state space into the control region, that is, we consider problems with Markov feed-back control policies. While in the deterministic case the effectivity of open loop and feed-back strategies are the same, for stochastic problems one is obliged to consider closed loop controls (cf. [3]). In the present paper we call a control strategy optimal, if it minimizes the expected loss of the integral form for an arbitrary initial state.

The main result of this paper, proved in the third chapter is a necessary and sufficient condition for the existence of optimal control strategies. The theorem is formulated in the form of a generalized boundary value problem and also presents a characterisation of the optimal strategy. It can be regarded as a comman generalization of the results of papers [1], [6], [7], [8], [11], [12], [13] and [15]. Two corollaries of the main theorem are given in the fourth chapter. They specialize the main theorem for problems having some local control dependence, and this allows a simplification of the optimality conditions. The recursive equations of the dynamic programming and the optimality condition of FLEMING for controlled diffusion processes are also derived as examples for the application of the optimality theorem.

For the computation of an optimal control strategy, or for the verification of its existence, by the optimality theorem given below one has to solve a generalized boundary value problem. Considerations about the existence of its solution can be found in the papers of KRYLOV [9], [10]. The results of the present paper, especially Corollary 2, can be used with success to prove the optimality of a strategy given in advance, e.g., by means of a necessary condition of optimality, and they serve to an extent as substitutions for the existence theorems of optimal control.

The author would like to thank Prof. P. H. MÜLLER, Prof. H. LANGER, Dr. sc. V. NOLLAU from the Technische Universität Dresden and Prof. K. TANDORI from the University Szeged for their valuable suggestions during the course of this work.

2. Optimal Markovian control problems

First we make some further assumptions about the structure of the time axis, the state space and the control region, which will enable us to apply the results of [4], [5] to our problems. The time axis T will be considered as a measurable topological space with its usual topology $\mathscr{C}_T$ and the σ-field of its Borel sets $\mathscr{T}$. The state space E is defined as the topological product of the time axis and some topological measurable space $(E', \mathscr{C}', \mathscr{E}')$, together with the product σ-field $\mathscr{E} = \mathscr{T} \otimes \mathscr{E}'$. Further on we assume that all open sets and also all one-point-sets of E are measurable. We shall call the measurable space $(D, \mathscr{D})$ with measurable one-point-sets the control region, while $\mathbf{B}(E, \mathscr{E})$ denotes the Banach space of all bounded measurable real valued functions on $(E, \mathscr{E})$, with the usual supremum norm.

Suppose we are given an open subset $G \subset E$, a class $\mathscr{U}$ of measurable mappings from $(G, \mathscr{C} \cap G, \mathscr{E} \cap G)$ into $(D, \mathscr{D})$, and a family of (homogeneous) right-continuous strong Markov processes $\{\Xi^U;\, U \in \mathscr{U}\}$ with $\Xi^U = (\xi^U, \zeta^U, \mathscr{M}_t^U, \mathbf{P}_x^U)$ on the state space $(E, \mathscr{C}, \mathscr{E})$, stopped at the first exit from G. In this paper we call $\{\Xi^U;\, U \in \mathscr{U}\}$ a controlled Markov process with the target set $E \setminus G$ and the class $\mathscr{U}$ of admissible control strategies (or policies) if the following conditions are satisfied:

a) If $U_1 \in \mathscr{U}$, $U_2 \in \mathscr{U}$ and if, I is an arbitrary (finite or infinite) subinterval of T, and

$$U(s, x') := \begin{cases} U_1(s, x') & \text{for} \quad s \in I,\ x \in E \\ U_2(s, x') & \text{for} \quad s \notin I,\ x \in E \end{cases}$$

then $U \in \mathscr{U}$.

b) For the transition function of Ξ^U we have

$$P^U(t, (s, x'),\ (T \setminus \{s+t\}) \times E') = 0 \tag{2.1}$$

for any $U \in \mathscr{U}$, $x = (s, x') \in E$, $t \in T$ ($\{v\}$ is the set containing the only point v).

c) The first exit time τ^U of the process Ξ^U from G is a Markov time for Ξ^U.

d) If I denotes an arbitrary interval in T then $U_1(x) = U_2(x)$ implies

$$\mathbf{P}_x^{U_1}(\xi_t^{U_1} \in \Gamma) = \mathbf{P}_x^{U_2}(\xi_t^{U_2} \in \Gamma) \tag{2.2}$$

for all $x \in I \times E'$, $\Gamma \subset I \times E'$, $\Gamma \in \mathscr{E}$, $t \in T$.

e) The domains $\mathbf{D}(A^U)$ of the weak infinitesimal Operators A^U of the processes Ξ^U stopped at τ^U coincide for all $U \in \mathscr{U}$. (We denote this common domain by $\mathbf{D}$.)

Let there be given a controlled Markov process $\{\Xi^U;\, U \in \mathscr{U}\}$ such that, for all $x \in E$, $U \in \mathscr{U}$ we have $\mathbf{E}_x^U < K$ for some $K \succ 0$ ($\mathbf{E}_x^U$ denotes the expectation w.r. to the measure $\mathbf{P}_x^U$), and non-negative functions $p \in \mathbf{B}(E \setminus G, \mathscr{E} \cap (E \setminus G))$, $q \in \mathbf{B}(G \times D, (\mathscr{E} \cap G) \otimes \mathscr{D})$ (we shall also write $q^U(x) := q(x, U(x))$). For all $x \in E$ we define the performance functionals on $\mathscr{U}$ by

$$J_x(U) := \mathbf{E}_x^U \left\{ p(\xi_{\tau^U}^U) + \int_0^{\tau^U} q(\xi_t^U, U(\xi_t^U))\, dt \right\} \tag{2.3}$$

Our aim is to find a strategy $U^* \in \mathscr{U}$, such that

$$J_x(U^*) \leqq J_x(U)$$

for all $x \in E$, $U \in \mathscr{U}$. A strategy with this property is said to be optimal.

Remarks. For the definitions and propositions cited here cf. DYNKIN [5]. We call $\Xi=(\xi, \zeta, \mathcal{M}_t, \mathbf{P}_x)$ a homogeneous Markov process on the state space $(E, \mathscr{C}, \mathscr{E})$, if:

a) ζ is a function defined on a sample space Ω with values in $T\cup\{+\infty\}$.
b) ξ is a partial map from $T\times\Omega$ into $(E, \mathscr{C}, \mathscr{E})$, and $\xi(t, \omega)=\xi_t(\omega)$ is defined for all $\omega\in\Omega$, $t\in T\cap(0, \zeta(\omega))$. For a fixed $\omega_0\in\Omega$ the function $\xi_t(\omega_0)$ is called a trajectory of process Ξ.
c) $\mathcal{M}_t$ is a σ-field on $\Omega_t:=\{\omega:\zeta(\omega)>t\}$ $(t\in T)$.
d) $\mathbf{P}_x$ is for all $x\in E$ a functional defined on a σ-field $\mathcal{M}$ of subsets of Ω with $\mathcal{M}\supset\bigcup_{t\in T}\mathcal{M}_t$.

And for these elements the following conditions are satisfied:

(1) If $s\leqq t$ and $A\in\mathcal{M}_s$ then $A\cap\{\omega:\zeta(\omega)>t\}\in\mathcal{M}_t$.
(2) $\{\xi_t\in\Gamma\}:=\{\omega:\xi_t(\omega)\in\Gamma\}\in\mathcal{M}_t$ $(t\in T, \Gamma\in\mathscr{E})$.
(3) $\mathbf{P}_x$ is a probability measure on $\mathcal{M}$ for all $x\in E$.
(4) For all $\Gamma\in\mathscr{E}$, $t\in T$,

$$P(t, x, \Gamma):=\mathbf{P}_x(\xi_t\in\Gamma)$$

is an $\mathscr{E}$-measurable function of x. P is called the transition function of the process Ξ.
(5) $P(0, x, E\setminus\{x\})=0$ for all $x\in E$.
(6) For all $s, t\in T$, $\Gamma\in\mathscr{E}$ we have

$$\mathbf{P}_x(\xi_{t+s}\in\Gamma|\mathcal{M}_t) = P(s, \xi_t, \Gamma)$$

$\mathbf{P}_x$ a.e. on Ω_t.
(7) For all $t\in T$, $\omega\in\Omega_t$ there exists an $\omega_t\in\Omega$ such that

$$\zeta(\omega_t) = \zeta(\omega)-t \quad\text{and}\quad \xi_s(\omega_t) = \xi_{s+t}(\omega) \quad\text{for}\quad 0\leqq s<\zeta(\omega)-t.$$

A Markov process is said to be right-continuous if all of its trajectories are right-continuous. Every right-continuous Markov process is measurable, that is,

$$\{(s, \omega): s\leqq t,\ \omega\in\Omega_t,\ \xi_s(\omega)\in\Gamma\}\in(\mathscr{T}\cap[0, t])\otimes\mathcal{M}_t \tag{2.4}$$

for all $t\in T$, $\Gamma\in\mathscr{E}$, where $\mathscr{T}\cap[0, t]$ denotes the restriction of $\mathscr{T}$ to interval $[0, t]$.

A mapping τ of Ω into $T\cup\{+\infty\}$ is called a Markov time for $\Xi=(\xi, \zeta, \mathcal{M}_t, \mathbf{P}_x)$ if

$$\tau(\omega)\leqq\zeta(\omega)$$

and for all $t\in T$

$$\{\omega:\tau(\omega)>t\}\in\mathcal{M}_t\cap\mathscr{N}$$

Where $\mathcal{N}$ denotes the σ-field generated by all sets of the form $\{\omega:\xi_t(\omega)\in\Gamma\}$ for $\Gamma\in\mathcal{E}$, $t\in T$. We set $A\in\mathcal{M}_\tau$, if $A\subset\Omega_\tau:=\{\omega:\tau(\omega)<+\infty\}$ and for any $t\in T$

$$A\cap\{\omega:\tau(\omega)\leqq t\}\in\mathcal{M}_t$$

By this definition $\mathcal{M}_\tau$ is a σ-field on Ω_τ. ([15]. 3.16.)

A measurable Markov process is said to be strongly Markovian if for an arbitrary Markov time τ and for all $t\in T$, $x\in E$, $\Gamma\in\mathcal{E}$

(2.5) $$\mathbf{P}_x(\xi_{\tau+t}\in\Gamma|\mathcal{M}_\tau) = P(t,\xi_\tau,\Gamma)$$

holds $\mathbf{P}_x$ a.e. on Ω_τ ([5]. 3.18).

The function $\tau=\tau(\omega)$ defined by

$$\tau(\omega):=\sup\{t:\{\xi_s(\omega):s\leqq t\}\subset G\}$$

is called the first exit time from the set G ([5], 4.1). Conditions for the first exit time to be Markov are given by Dynkin in [5], Chapter 4.

Let $\Xi=(\xi,\zeta,\mathcal{M}_t,\mathbf{P}_x)$ be a right-continuous strong Markov process and τ the Markov time of the first exit from the set G. If we set

$$\tilde{\zeta}(\omega):=\begin{cases}\zeta(\omega) & \text{for}\quad \tau(\omega)=\zeta(\omega)\\ +\infty & \text{for}\quad \tau(\omega)<\zeta(\omega)\end{cases}$$

$$\tilde{\xi}_t(\omega):=\xi_{\min[t,\tau(\omega)]}(\omega)\qquad (0\leqq t<\tilde{\zeta}(\omega))$$

$$\tilde{\mathcal{M}}_t:=\{A\in\mathcal{M}:A\subset\{\tilde{\zeta}>t\}\quad\text{and}\quad A\cap\{\tau>t\}\in\mathcal{M}_t\}$$

then the process $\tilde{\Xi}=(\tilde{\xi},\tilde{\zeta},\tilde{\mathcal{M}}_t,\mathbf{P}_x)$ is also a right-continuous strong Markov process, and is said to arise of Ξ by stopping it at τ ([5], 10.4).

A sequence of functions $f_n\in\mathbf{B}(E,\mathcal{E})$ is said to tend weakly to $f\in\mathbf{B}(E,\mathcal{E})$ if for every signed measure φ of bounded variation, defined on the σ-field $\mathcal{E}$

$$\int f_n d\varphi\to\int f d\varphi$$

holds as $n\to\infty$. In $\mathbf{B}(E,\mathcal{E})$ the weak convergence of f_n to f is equivalent to

(i) $f_n(x)\to f(x)$ for all $x\in E$ as $n\to\infty$, and
(ii) there exists a $K>0$ such that $\|f_n\|\leqq K$ for all $n\in N$.

With every Markov process $\Xi=(\xi,\zeta,\mathcal{M}_t,\mathbf{P}_x)$ we can associate a semigroup of linear operators $\{T_t\}_{t\in T}$ on the Banach space $\mathbf{B}(E,\mathcal{E})$ defined by

(2.6) $$T_t f(x):=\mathbf{E}_x f(\xi_t)$$

where $\mathbf{E}_x$ means the expectation w.r. to the measure $\mathbf{P}_x$. A function $f\in\mathbf{B}(E,\mathcal{E})$ is said to be weakly continuous (w.r. to the semigroup T_t) if

$$w\text{-}\lim_{t\downarrow 0} T_t f = f$$

If T is the set of all non-negative reals, the weak infinitesimal operator of $\{T_t\}_{t\in T}$ is defined by

$$Af:=w\text{-}\lim_{t\downarrow 0}\frac{T_t f-f}{t}$$

for all $f\in\mathbf{B}(E,\mathscr{E})$ such that the right-hand side tends weakly to a weakly continuous function. If T equals the set of all non-negative integers, we define A for all $f\in\mathbf{B}(E,\mathscr{E})$ by

$$Af:=T_1 f-f$$

and call it infinitesimal operator of T_t as well.

For every controlled Markov process $\{(\xi^U,\zeta^U,\mathscr{M}_t^U,\mathbf{P}_x^U);\ U\in\mathscr{U}\}$, we can construct by tranformation of the sample space a controlled Markov process $\{\hat{\xi},\hat{\zeta},\hat{\mathscr{M}}_t,\mathbf{P}_x^U);\ U\in\mathscr{U}\}$ such that for every $U\in\mathscr{U}$ the processes $\hat{\Xi}^U$ and Ξ^U are equivalent (Dynkin [4]). (E.g. we set $\hat{\Xi}^U$ for the canonical process of Ξ^U, cf. [4] Lemma 2. 3). By this we may omit index U of $\xi,\zeta,\mathscr{M}_t$ without loss of generality. The U-independence of the exit time τ is a consequence of its definition and the fact that ξ does not depend on U.

3. A necessary and sufficient condition of optimality

Theorem. *Let G, $\{\Xi^U;\ U\in\mathscr{U}\}$, q and p given as in the second section and let q^U be weakly continuous w.r. to T_t^U $(U\in\mathscr{U})$. Then there exists an optimal strategy $U^*\in\mathscr{U}$ if and only if the boundary value problem*

(3. 1) $$\min_{U\in\mathscr{U}}(A^U f+q^U)(x)=0 \quad \textit{for} \quad x\in G,$$

(3. 2) $$f(x)=p(x) \quad \textit{for} \quad x\in E\setminus G$$

possesses a solution $f\in\mathbf{B}(E,\mathscr{E})$. In this case the minimum occurs in (3. 1) *for U^*.*

Remarks. (3. 1) means in detail that for any element x of G

(3. 3) $$A^{U^*}f(x)+q^{U^*}(x)=0,$$

(3. 4) $$A^U f(x)+q^U(x)\geqq 0 \quad (U\in\mathscr{U}).$$

A^U denotes the weak infinitesimal operator of the processes arising from Ξ^U by stopping at the first exit from G.

If q^U is continuous then right-continuity of Ξ^U implies weak continuity of q^U. ([5]. Lemma 2. 2.)

Proof. Sufficiency: If $f^* \in \mathbf{D}(A^U)$, and τ is a Markov time for process Ξ^U, then by [5], Theorem 5. 1 we have

$$(3.5) \qquad f^*(x) = \mathbf{E}_x^U f^*(\xi_\tau) - \mathbf{E}_x^U \int_0^\tau A^U f^*(\xi_t)\,dt$$

Let f^* be the required solution of the boundary value problem (3. 1)—(3. 2). Since $f^* \in \mathbf{D} = \mathbf{D}(A^U)$ $(U \in \mathcal{U})$, and the first exit time τ from G is a Markov time for all Ξ^U $(U \in \mathcal{U})$, (3. 5) holds for every $U \in \mathcal{U}$. Let us observe that the right continuity of the processes Ξ^U and the fact that G is open imply $\xi_\tau \in E \setminus G$. By virtue of this and of equations (3. 5), (3. 2), (3. 3) we have

$$(3.6) \qquad f^*(x) = \mathbf{E}_x^{U^*}\Big\{f^*(\xi_\tau) - \int_0^\tau A^{U^*} f^*(\xi_t)\,dt\Big\} = \mathbf{E}_x^{U^*}\Big\{p(\xi_\tau) + \int_0^\tau q^{U^*}(\xi_t)\,dt\Big\}.$$

Analogously but with the aid of (3. 4) instead of (3. 3) we get for all $U \in \mathcal{U}$

$$(3.7) \qquad f^*(x) = \mathbf{E}_x^U f^*(\xi_\tau) - \int_0^\tau A^U f^*(\xi_t)\,dt \leqq \mathbf{E}_x^U\Big\{p(\xi_\tau) + \int_0^\tau q^U(\xi_t)\,dt\Big\}.$$

On account of (2. 3) the relations (3. 6) and (3. 7) imply for arbitrary $U \in \mathcal{U}$, $x \in E$:

$$J_x(U^*) = \mathbf{E}_x^{U^*}\Big\{p(\xi_\tau) + \int_0^\tau q^{U^*}(\xi_t)\,dt\Big\} \leqq \mathbf{E}_x^U\Big\{p(\xi_\tau) + \int_0^\tau q^U(\xi_t)\,dt\Big\} = J_x(U).$$

But the last equation implies that U^* is optimal. Hence the sufficiency of our assumption is proved.

Necessity: Let U^* denote the optimal control strategy, and let us introduce f^* by $f^*(x) := J_x(U^*)$. By the boundedness of p, q and $\mathbf{E}_x^U \tau$ we find that $f^* \in \mathbf{B}(E, \mathcal{E})$.

First we shall prove equation (3. 3) and that $f^* \in \mathbf{D}$. Eq. (2. 3), the definition of the stopped processes and that of f^* show that

$$(3.8) \qquad T_h^{U^*} f^*(x) = \mathbf{E}_x^{U^*} \chi_{\{\tau > h\}} f^*(\xi_h) + \mathbf{E}_x^{U^*} \chi_{\{\tau \leqq h\}} f^*(\xi_\tau) =$$

$$= \mathbf{E}_x^{U^*} \chi_{\{\tau > h\}} \mathbf{E}_{\xi_h}^{U^*} p(\xi_\tau) + \mathbf{E}_x^{U^*} \chi_{\{\tau > h\}} \mathbf{E}_{\xi_h}^{U^*} \int_0^\tau q^{U^*}(\xi_t)\,dt + \mathbf{E}_x^{U^*} \chi_{\{\tau \leqq h\}} p(\xi_\tau)$$

(χ_A denotes the characteristic function of the set A). For the first term of the right-hand side of (3. 8) we obtain by applying [5] Theorem (3. 1) that

$$\mathbf{E}_x^{U^*} \chi_{\{\tau > h\}} \mathbf{E}_{\xi_h}^{U^*} p(\xi_\tau) \doteq \mathbf{E}_x^{U^*} \chi_{\{\tau > h\}} \mathbf{E}_x^{U^*}\big(\theta_h p(\xi_\tau) | \mathcal{M}_h\big)$$

where the operator θ_t is defined by

$$\theta_t \eta(\omega) := \eta(\omega_t)$$

(η is any random variable; cf. (7) in the definition of Markov processes.) Taking into account the $\mathcal{M}_h$ measurability of $\chi_{\{\tau>h\}}$, the basic properties of conditional expectations, and the definition of θ_h, we find:

$$\text{(3. 9)} \qquad \mathbf{E}_x^{U^*}\chi_{\{\tau>h\}}\mathbf{E}_x^{U^*}\left(\theta_h p(\xi_\tau)|\mathcal{M}_h\right) = \mathbf{E}_x^{U^*}\mathbf{E}_x^{U^*}\left(\chi_{\{\tau>h\}}\theta_h p(\xi_\tau)|\mathcal{M}_h\right) =$$

$$= \mathbf{E}_x^{U^*}\chi_{\{\tau>h\}}p\left(\xi_{\tau(\omega_h)}(\omega_h)\right) = \mathbf{E}_x^{U^*}\chi_{\{\tau>h\}}p\left(\xi_{h+\tau(\omega)-h}(\omega)\right) = \mathbf{E}_x^{U^*}\chi_{\{\tau>h\}}p(\xi_\tau).$$

For the second term of (3. 8) we find with $q_0:=\chi_G\cdot q^{U*}$ that

$$\mathbf{E}_x^{U^*}\chi_{\{\tau>h\}}\mathbf{E}_{\xi_h}^{U^*}\int_0^\tau q^{U^*}(\xi_t)\,dt = \int_0^\infty \mathbf{E}_x^{U^*}\chi_{\{\tau>h\}}\mathbf{E}_{\xi_h}^{U^*}q_0(\xi_t)\,dt =$$

$$= \int_0^\infty \mathbf{E}_x^{U^*}\mathbf{E}_{\xi_h}^{U^*}q_0(\xi_t)\,dt - \int_0^\infty \mathbf{E}_x^{U^*}\chi_{\{\tau\le h\}}q_0(\xi_\tau)\,dt.$$

The change of the order of integration was allowed by the measurability of the processes (cf. (2. 4)) and by Fubini's theorem. Let us observe that $q_0(\xi_\tau)=0$, then by means of definition of T_t^{U*} we obtain:

$$\text{(3. 10)} \quad \mathbf{E}_x^{U^*}\chi_{\{\tau>h\}}\mathbf{E}_{\xi_h}^{U^*}\int_0^\infty q^{U^*}(\xi_t)\,dt = \int_0^\infty \mathbf{E}_x^{U^*}\mathbf{E}_{\xi_h}^{U^*}q_0(\xi_t)\,dt =$$

$$= \int_0^\infty T_{t+h}^{U^*}q_0(x)\,dt = \int_h^\infty T_t^{U^*}q_0(x)\,dt = \mathbf{E}_x^{U^*}\int_h^\tau q_0^*(\xi_t)\,dt =$$

$$= \mathbf{E}_x^{U^*}\int_0^\tau q^{U^*}(\xi_t)\,dt - \int_0^h T_t^{U^*}q_0(x)\,dt.$$

Substituting (3. 9), (3. 10) into (3. 8) and taking into account of the definition of f^* we get

$$T_h^{U^*}f^*(x) = \mathbf{E}_x^{U^*}\left\{p(\xi_\tau)+\int_0^\tau q^{U^*}(\xi_t)\,dt\right\} - \int_0^h T_t^{U^*}q_0(x)\,dt = f^*(x) - \int_0^h T_t^{U^*}q_0(x)\,dt.$$

On account of the definition of A^{U^*} we obtain hence that

$$A^{U^*}f^* = w\text{-}\lim_{h\downarrow 0}\frac{T_h^{U^*}f^*-f^*}{h} = w\text{-}\lim\frac{1}{h}\int_0^h T_t^{U^*}q_0\,dt = -q_0$$

Since q_0 is weakly continuous we get that $f^*\in\mathbf{D}$ so (3. 3) is proved.

Now we prove inequality (3. 4) indirectly. Let us assume, there exist a strategy $U_0\in\mathcal{U}$ and state $x_0\in E$ such that

$$A^{U_0}f^*(x_0)+q^{U_0}(x_0)<0.$$

From the weak continuity (w.r. to $T_t^{U_0}$) of q^{U_0} and $A^{U_0}f^*$ it follows the existence of a $t_0>0$ such that for all $0\leqq t<t_0$ we have

$$T_t^{U_0}(A^{U_0}f^*+q^{U_0})(x_0)<0.$$

But with the notation $\tau_0(\omega):=\min[t_0,\tau(\omega)]$ we obtain the inequality

$$\mathbf{E}_{x_0}^{U_0}\int_0^{\tau_0}(A^{U_0}f^*+q^{U_0})(\xi_t)\,dt<0. \tag{3.11}$$

Let us denote by U_1 the control strategy

$$U_1(s,x'):=\begin{cases}U^*(s,x') & \text{for}\quad s\geqq s_0+t_0,\quad x'\in E',\\ U_0(s,x') & \text{for}\quad s<s_0+t_0,\quad x'\in E',\end{cases}$$

where s_0 denotes the time-component of x_0, more precisely $x_0=(s_0,x_0')$ with some $x_0'\in E'$. Since $J(U^*)=f^*\in\mathbf{D}$ and τ_0 is a Markov time, we may apply [5] Theorem 5. 1 (cf. also Eq. (3. 5)) to Ξ^{U_1} and we get

$$J_{x_0}(U^*)=f^*(x_0)=\mathbf{E}_{x_0}^{U_1}\Big\{f^*(\xi_{\tau_0})+\int_0^{\tau_0}A^{U_1}f^*(\xi_t)\,dt\Big\} \tag{3.12}$$

In virtue of the definition of $J(U_1)$ we obtain:

$$J_{x_0}(U_1)=\mathbf{E}_{x_0}^{U_1}\Big\{p(\xi_\tau)+\int_0^{\tau}q^{U_1}(\xi_t)\,dt\Big\}= \tag{3.13}$$

$$=\mathbf{E}_{x_0}^{U_1}\Big\{p(\xi_\tau)+\int_{\tau_0}^{\tau}q^{U_1}(\xi_t)\,dt\Big\}+\mathbf{E}_{x_0}^{U_1}\Big\{\int_0^{\tau_0}q^{U_1}(\xi_t)\,dt\Big\}.$$

For the first term of (3. 13) it holds the decomposition

$$\mathbf{E}_{x_0}^{U_1}\Big\{p(\xi_\tau)+\int_{\tau_0}^{\tau}q^{U_1}(\xi_t)\,dt\Big\}=$$

$$=\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau_0=\tau\}}p(\xi_\tau)+\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau>\tau_0\}}p(\xi_\tau)+\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau>\tau_0\}}\int_{t_0}^{\tau}q(\xi_t)\,dt$$

The second and the third term of the last equation can be transformed analogously to formulae (3. 9), (3. 10) and we obtain

$$\mathbf{E}_{x_0}^{U_1}\Big\{p(\xi_\tau)+\int_{\tau_0}^{\tau}q^{U_1}(\xi_t)\,dt\Big\}=$$

$$=\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau_0=\tau\}}p(\xi_\tau)+\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau>\tau_0\}}\mathbf{E}_{\xi_{\tau_0}}^{U_1}p(\xi_\tau)+\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau>\tau_0\}}\mathbf{E}_{\xi_{\tau_0}}^{U_1}\int_0^{\tau}q^{U_1}(\xi_t)\,dt.$$

By the properties (2. 1), (2. 2) and the definition of U_1 this equals

$$\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau_0=\tau\}}\,p(\xi_\tau)+\mathbf{E}_{x_0}^{U_1}\chi_{\{\tau_0<\tau\}}\,\mathbf{E}_{\xi_{\tau_0}}^{U_1}\Big\{p(\xi_\tau)+\int_0^\tau q^{U_1}(\xi_t)\,dt\Big\}=$$

$$=\mathbf{E}_{x_0}^{U_1}\Big[\chi_{\{\tau_0=\tau\}}\,p(\xi_\tau)+\chi_{\{\tau_0<\tau\}}\,\mathbf{E}_{\xi_{\tau_0}}^{U^*}\Big\{p(\xi_\tau)+\int_0^\tau q^{U^*}(\xi_t)dt\Big\}\Big]=\mathbf{E}_{x_0}^{U_1}f^*(\xi_{\tau_0}).$$

In this way for (3. 13) we get that

$$J_{x_0}(U_1)=\mathbf{E}_{x_0}^{U_1}\Big\{f^*(\xi_{\tau_0})+\int_0^\tau q^{U_1}(\xi_t)\,dt\Big\}. \tag{3. 13'}$$

By subtraction (3. 12) from (3. 13'), and by the definition of U_1 we obtain

$$J_{x_0}(U_1)-J_{x_0}(U^*)=\mathbf{E}_{x_0}^{U_1}\{f^*(\xi_{\tau_0})-f^*(\xi_{\tau_0})\}+\mathbf{E}_{x_0}^{U_1}\int_0^{\tau_0}(A^{U_1}f^*+q^{U_1})(\xi_t)\,dt=$$

$$=\mathbf{E}_{x_0}^{U_0}\int_0^{\tau_0}(A^{U_0}f^*+q^{U_0})(\xi_t)\,dt.$$

But by (3. 11) this means

$$J_{x_0}(U_1)<J_{x_0}(U^*).$$

in contradiction to the supposed optimality of U^*.

The necessity of (3. 2) follows from the fact that $\mathbf{P}_x^U(\tau=0)=1$ for all $U\in\mathscr{U}$, $x\in E\setminus G$. By this our theorem is proved.

Remarks. As we see from the proof, the solution f^* of (3. 1)—(3. 2) has the meaning of expected loss in the state x, if we apply the optimal strategy:

$$f^*(x)=J_x(U^*).$$

From the proof one can see, that we may formulate our theorem in any other topology. E.g. in the strong topology we find: Let q^U be strong continuous (w.r. to T_t^U). Then there exists an optimal control strategy $U^*\in\mathscr{U}$ if and only if the boundary value problem (3. 1)—(3. 2) possesses a solution $f^*\in\mathbf{B}(E,\mathscr{E})$. In this case A^U means the strong infinitesimal operator of P^U.

4. Applications of the main theorem

In this part of our paper let us assume that $\mathscr{U}$ contains all constant strategies $U(x)\equiv d$ $(d\in D)$.

We denote by $\mathbf{D}_L$ the set of all functions $f\in\mathbf{D}$ such that for any x the equality $U_1(x)=U_2(x)$ implies

$$A^{U_1}f(x)=A^{U_2}f(x) \tag{4. 1}$$

i.e. $\mathbf{D}_L:=\{f\in\mathbf{D}: U_1(x)=U_2(x)\Rightarrow A^{U_1}f(x)=A^{U_2}f(x)\}$.

Corollary 1. *Given a controlled Markov process* $\{\Xi^U; U\in\mathcal{U}\}$ *with* $\mathbf{D}=\mathbf{D}_L$, *if* q^U *is weakly continuous w.r. to* T_t^U *then for the existence of an optimal control strategy* $U^*\in\mathcal{U}$ *is necessary and sufficient that there exist an* $f^*\in\mathbf{B}(E,\mathcal{E})$ *and a* $U'\in\mathcal{U}$, *which satisfy the relations*

$$A^{U'(x)}f^*(x)+q(x,U'(x))=0 \quad \text{for} \quad x\in G, \tag{4.2}$$

$$A^d f^*(x)+q(x,d)\geqq 0 \quad \text{for} \quad d\in D,\ x\in G, \tag{4.3}$$

$$f^*(x)=p(x) \quad \text{for} \quad x\in E\setminus G \tag{4.4}$$

(A^d *denotes the infinitesimal operator of* Ξ^U *for which* $U\equiv d$).

Proof. Necessity: Let $U^*\in\mathcal{U}$ be optimal. Then (4. 2) follows with $U':=U^*$ from (3. 3), if we observe, that by (4. 1)

$$A^U f(x) = A^{U(x)}f(x) \tag{4.5}$$

holds for all $U\in\mathcal{U}, f\in\mathbf{D}=\mathbf{D}_L$.

Let us suppose the existence of an $x_0\in E$, $d_0\in D$ such that

$$A^{d_0}f^*(x)+q^{d_0}(x_0)<A^{U'}f^*(x_0)+q^{U'}(x_0)=0$$

where $f^*(x)=J_x(U')$. But since $U\equiv d_0$ is an admissible strategy, according to the theorem the last inequality contradicts the optimality of U'.

Sufficiency: For an arbitrary $x\in E$ there exists a $U'(x)\in D$ such that

$$A^d f^*(x)+q(x,d)\geqq A^{U'(x)}f^*(x)+q(x,U'(x))=0$$

Put $U^*(x):=U'(x)\in\mathcal{U}$. By (4. 5) the equation (3. 3) holds true for U^*. Let $U\in\mathcal{U}$ arbitrary, then (3. 4) follows from (4. 3), and Theorem implies the statement of Corollary 1.

Corollary 2. *Let* $U^*\in\mathcal{U}$, $f^*:=J(U^*)\in\mathbf{D}_L$ *and let* $T_t^U q^U$ *be weakly continuous. Then* U^* *is optimal if and only if for* U^* *and* f^* *relations* (4. 2)—(4. 4) *hold true (with* $U^*\equiv U'$*).*

Proof. Analogous to the proof of Corollary 1.

Remarks: We can restate Corollary 1 in the following form: There exists an optimal control strategy if and only if there exists an $f^*\in\mathbf{B}(E,\mathcal{E})$ such that for all $x\in G$ we have

$$\min_{d\in D}[A^d f^*(x)+q(x,d)]=(A^{U^*(x)}f^*)(x)+q(x,U^*(x))=0 \tag{4.6}$$

and for all $x\in E\setminus G$ we have $f(x)=p(x)$, furthermore the strategy U^* for which the minimum occurs in (4. 6) belongs to $\mathcal{U}$.

Worth of noting is that this way Corollary 1 states the optimal policy U^* to be independent on the class of admissible control strategies. This implies that exactly those classes have optimal policies which contain the strategy U^* obtained by minimization of equation (4. 6) (assuming that such a minimization exists).

From practical point of view the main advantage of Corollary 1 over Theorem is that the minimization in it has to be carried out only over the control region, the cardinality of which is generally smaller than that of the class of the admissible strategies.

Corollary 2 is useful if we want to decide about a given strategy U^* (which has been determined earlier e.g. by the application of a necessary condition of optimality) whether it is optimal or not. In this case we have to prove the truth of (4. 1) only for $f^*=J(U^*)$, and have the advantage of minimizing over the control region.

Finally, for the illustration of the theorems above, we give two examples which show how to derive the results of Bellman and Fleming from those of this paper.

Discrete-time dynamic programming:

Let the time be discrete. $T=N$, let the sets D, E' be finite and denote by $\mathcal{U}$ the set of all functions from $T\times E'$ in D. Furthermore for all $U\in\mathcal{U}$ let Ξ^U be a Markov chain on the state space $E=T\times E'$ with

$$P^d(i,(k,x'),\{i\}\times E)=\chi_{\{k+i\}}(j)$$

for all $k,i,j\in T$, $x'\in E'$, $d\in D$ (cf. (1. 1)). For G we choose a bounded subset of $T\times E'$. From Corollary 1 it follows that a control U^* is optimal iff there is an f^* such that

$$(5.5)\quad f^*(k,x')=\min_{U(k,x')\in D}\left[E^{U(k,x')}_{(k,x')}f^*(k+1,\xi'_{k+1})+q\big(k,x',U(k,x')\big)\right.\quad\text{for}\quad (k,x')\in G,$$

$$(5.6)\qquad f^*(k,x')=p(k,x')\qquad\text{for}\quad (k,x')\in E\setminus G$$

In (5. 5)—(5. 6) we recognize the well-known recursive equations of the stochastic dynamic programming [1].

Controlled diffusions (cf. FLEMING [6], [7], MANDL [12]).

Let $T=R^+$, $E'\subset R^{n-1}$, $D\subset R^m$ and let $\mathcal{U}$ the class of all measurable bounded functions from E into D. Let the functions b', σ', p, q be continuously differentiable and bounded together with their first order partial derivatives, and let $a':=\frac{1}{2}\sigma'\sigma'^{\mathrm{T}}$ (where σ'^{T} is the matrix transposed of σ') such that the eigenvalues of $a'(x)$ are bounded from below by some $c>0$ for all $x'\in E'$. We introduce the notations

$$b=\begin{pmatrix}1\\ b'\end{pmatrix},\quad \sigma=\begin{pmatrix}0&0\\0&\sigma'\end{pmatrix},\quad a=\tfrac{1}{2}\sigma\sigma^T,\quad w_t=\begin{pmatrix}0\\ w'_t\end{pmatrix};$$

where w'_t is a Brownian motion process on E'. Then the Ito integral equation

$$\xi_t = x + \int_0^t b(\xi_t, U(\xi_t))\,dt + \int_0^t \sigma(\xi_t)\,dw_t$$

determines a Markov process for all $x \in E$, $U \in \mathcal{U}$ with the infinitesimal generator:

$$A^U f(x) = \sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial^2 f(x)}{\partial x_i \partial x_j} + \langle a(x, U(x));\, f_x(x)\rangle$$

where f_x denotes the gradient of f and $\langle ; \rangle$ the inner product in R^n. From this we see that $\mathbf{D} = \mathbf{D}_L$ holds, and by Corollary 1 it follows:

Necessary and sufficient for the optimality of U^* is the existence of an $f^* \in \mathbf{B}(E, \mathcal{E})$ satisfying

$$\sum_{i,j=1}^{n} a_{ij}(x) \frac{\partial f^*}{\partial x_i \partial x_j} + \min_{d \in D} [\langle a(x,d);\, f_x^*(x) + q(x,d)] \quad \text{for} \quad x \in G,$$

$$f(x) = p(x) \quad \text{for} \quad x \in E \setminus G.$$

References

[1] R. Bellman, *Dynamic Programming* (Princeton, 1957).

[2] В. Г. Болтянский, Достаточные условия оптимальности и обоснование метода динамического программирования, *Изв. АН. СССР., сер. матем.*, **28** (1964), 481—514.

[3] S. E. Dreyfus, Introduction to stochastic optimization and control, *Stochastic Optimization and Control*, Proc. Adv. Seminar, University of Wiscosin, 1967 (New York, 1968).

[4] E. B. Dynkin, *Die Grundlagen der Theorie der Markoffschen Processe* (Berlin, 1961).

[5] E. B. Dynkin, *Markov Processes* (Berlin, 1965).

[6] W. H. Fleming, Duality and a priori estimates in Markovian optimization problems, *J. Math. Anal. Appl.*, **16** (1961), 254—279. Erratum *ibid.*, **19** (1966), 204.

[7] W. H. Fleming, Optimal continuous-parameter stochastic control, *SIAM Review*, **11** (1969), 470—509.

[8] Б. И. Григелионис—А. Н. Ширяев, Об управяемых марковских процессах и задаче Стефана, *Пробл. перед. Инф.*, **4** (1968), 60—72.

[9] Н. В. Крылов, Об одном классе нелинейных уравнений в пространстве измеримых функций, *Матем. сб.*, **80 (122)** (1969), 253—265.

[10] Н. В. Крылов, Управление марковскими процессами и пространства *W*, *Изв. АН. СССР, сер. матем.*, **35** (1971), 224—255.

[11] H. J. Kushner, *Stochastic Stability and Control* (New York, 1967).

[12] P. Mandl, *Analytical Treatment of One-dimensional Markov Processes* (Praha, Berlin, 1968).

[13] V. Nollau and D. Vermes, Zur optimalen Steuerung inhomogener Markoffscher Prozesse *Math. Operationsforsch. Statist.* (to appear).

[14] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko, *The Mathematical Theory of Optimal Processes* (New York, 1962).

[15] W. M. Wonham, Random differential equations in control theory, *Probabilistic Methods ni Applied Mathematics*, Vol. II (New York, 1970).

(Received September 28, 1972)