On optimal control of semi-Markov processes

By D. VERMES in Szeged

1. Introduction

The system we consider changes its states stochastically at random points of the continuous time. Each of its paths s_t is a step function of the time $t \in [0, \infty)$ with values in M dimensional Euclidean space E. The probability $\Pi_x(\Gamma, \vartheta)$ of the event that the state $x \in E$ will be followed by an element of the set $\Gamma \subset E$ and that the system stays in x no longer than ϑ, depends on the state x but not on the earlier ones. A process s_t of this kind is said to be semi-Markovian (c.f. [2]). It is known that a semi-Markov process is Markovian iff the sojourn time in any state x is independent of the following state and is exponentially distributed, i.e. $\Pi_x(\Gamma, \vartheta) = Q_x(\Gamma) \exp\{-\lambda(x)\vartheta\}$ with some $\lambda > 0$, where $Q_x(\Gamma) = \Pi_x(\Gamma, \infty)$.

In the first part of the present paper we show that the investigation of general semi-Markov processes can be attributed to the study of special Markov processes. More precisely, we prove that the vector process, built up from s_t and the time difference y_t between t and the last jump moment preceding t, is Markovian. Further we give an explicit expression for the infinitesimal generator of the extended process in the case when the sojourn time is independent of the following state.

The second part of the paper deals with the optimal control of semi-Markov processes. Suppose the probabilities $\Pi_x(\Gamma, \vartheta)$ depend besides x, Γ and ϑ on a decision d too, i.e. they are of the form $\Pi_x^d(\Gamma, \vartheta)$. The value of d can be freely chosen from a set D, the so called *decision space,* at any time moment. This way we can influence the dynamics of the process. In the sequel we consider the case when the choice of the actual value of d is based upon the current state and the actual value of y. In other words d is chosen according to a function $u(x, y)$, called a strategy.

Generally we will influence the system in order to obtain an, in some sense ideal, process dynamics. Suppose there is given a subset G of the state space, and that the process reaches its goal when it enters the complementary set of G, the *target set.* Another interpretation of G is that the system gets damaged when its state leaves G, the set of *admissible states.* Whatever the intuitive meaning of G is, it is uninteresting for the controller, what happens after the state of the process has left it; hence we observe the process only until it leaves G. The time of the first exit from the set G will be denoted by τ. Assume the expense, which constitutes the basis of the judgement of the quality of the different strategies, consists of two parts. The first one depends on the terminal state, where the process leaves G for the first time; we designate it by $p(x, y)$. The second expense component is the integral of the so-called "differential" costs $q(x, y, d)$ over the time interval $[0, \tau)$. The "differential expense" $q(x, y, d)$ arises when the process has already been staying in the state x for a time y, and the decision d is made. Clearly, the value of the expense depends both on the initial state and on the chance. A strategy u is said to be optimal, if it minimizes the expected cost for every initial state.

In the third section of the present paper we prove a necessary and sufficient optimality theorem for semi-Markov processes. The optimality condition is formulated in form of a boundary value problem relative to the results in [3]. In the fourth section we specialize our theorem to Markov jump processes, and we obtain a more simple optimality condition, than that of derived from the main theorem of [3].

The results of the paper, formulated for finite dimensional state and decision space can be generalized to an arbitrary measurable state space E and a topological measurable decision space D, without any additional difficulties.

2. Markov equivalents of semi-Markov processes

We denote by $\mathbf{R}^+$ the set of all non-negativ reals, $\mathbf{R}^+$ will serve as the time axis of our processes. Let $(\Omega, \mathcal{S}, P)$ be a probability space, and denote $(E, \mathcal{E})$ a subspace of the M-dimensional Euclidean space $\mathbf{R}^M$ with the σ-field $\mathcal{E}$ of its Borel sets. Assume v_n ($n \in \mathbf{N}$, where $\mathbf{N}$ stands for the set of all non-negative integers) are independent, identically distributed $\mathbf{R}^+$ valued random variables, while the E valued variables ξ_n ($n \in \mathbf{N}$) constitute a (homogeneous) Markov chain, i.e.

$$\Pi_x(B, t) := P(\xi_{n+1} \in B,\ v_n \leqq t | \xi_n = x) =$$

$$= P(\xi_{n+1} \in B,\ v_n \leqq t | \xi_n = x,\ \xi_{n-1} = x_{n-1}, \dots, \xi_0 = x_0,\ v_{n-1} = t_{n-1}, \dots, v_0 = t_0)$$

for arbitrary $n \in \mathbf{N}$; $x_0, \dots, x_{n-1}, x \in E$; $t_0, \dots, t_{n-1}, t \in \mathbf{R}^+$ and $B \in \mathcal{E}$, and Π_x is independent of n. Further on we assume that Π_x is a probability measure on the space $E \times \mathbf{R}^+$ ($\times$ denotes the Cartesian product), and that the image of Ω with respect to the variables ξ_n is measurable i.e. $\xi_n(\Omega) \in \mathcal{E}$ for arbitrary $n \in \mathbf{N}$. ξ_n can be interpreted as the state of the system between the n-th and $n+1$-th jumps, while v_n is the sojourn time in the n-th state, i.e. the time difference between the $n+1$-th and n-th jump moments. If we define

$$\xi_n(\omega) := \sum_{i=0}^{n-1} v_i(\omega) \quad \text{and} \quad N(t, \omega) := \sup \{n : \eta_n(\omega) \leqq t\}$$

then η_n means the time of the n-th jump, and $N(t)$ the number of jumps until the time t. We introduce the notation $s_t(\omega) := \xi_{N(t,\omega)}(\omega)$ and call the continuous time stochastic process $S = \{(s_t, \Pi_x) : t \in \mathbf{R}^+, x \in E\}$ the semi-Markov process corresponding to the measures Π_x.

Observe that as a consequence of its definition, $s_t(\omega)$ is a right-continuous function of the time for any $\omega \in \Omega$. We suppose for the whole paper that $N(t, \omega)$ is finite for any $\omega \in \Omega$ and $t \in \mathbf{R}^+$. Conditions guaranteeing this property are given in [2].

We denote by T the common range of the variables v_i ($i \in \mathbf{N}$) and assume T to be right-open. We introduce the notations $x_t(\sigma, \omega) := \xi_{N(t+\sigma,\omega)}(\omega)$ and $y_t(\sigma, \omega) := t + \sigma - \eta_{N(t+\sigma,\omega)}(\omega)$ where $t \in \mathbf{R}^+$, $\sigma \in T$, $\omega \in \Omega$. Here we have $x_t(0, \omega) = s_t(\omega)$ and $y_t(0, \omega)$ means the time difference between the moment t of observation and the moment of the last jump before t. $x_t(\sigma, \omega)$ and $y_t(\sigma, \omega)$ arise by shifting of the t-functions $x_t(0, \omega)$ and $y_t(0, \omega)$ to the left by σ time units. They can be interpreted as the x- and y-trajectories, respectively, when we know that at the beginning of the observation the time σ had already been passed since the preceding jump, or more roughly speaking the last jump before $t=0$ was at "$t = -\sigma$".

Suppose the distribution of the variable v_0 is supported on T whatever the initial condition $\xi_0 = x$ is. In other words, $\bar{T}$ is the smallest closed set such that $P(v_0 \in \bar{T} | \xi_0 = x) = 1$ holds true for any $x \in E$. We define the measures $P_{x,y}$ ($x \in E$, $y \in T$) on the product space $T \times \Omega$ by $P_{x,y}(A) = P(A_y | \xi_0 = x,\ v_0 > y)$ where A_y denotes the section of the set $A \subset T \times \Omega$, i.e. $A_y = \{\omega \in \Omega : (y, \omega) \in A\}$. Then $P_{x,y}$ means the probability of the event A_y under the condition that we know, at the beginning of the observation the process had already stayed in the state x for a time y.

Let we denote by $\mathscr{C}^+$ the topology on the space $E \times T$ which is the product of the weakest topology on E and of the right-side topology on T. In other words, the sets $\{(x, y) \in E \times T : x = x_0, y_0 \leqq y < y + \varepsilon\}$ ($\varepsilon > 0$) constitute a basis of neighbourhoods of the point $(x_0, y_0) \in E \times T$. The σ-field of Borel sets of T will be denoted by $\mathscr{T}$. $\mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ or simply $\mathscr{B}$ abbreviates the Banach space of all bounded real-valued functions, defined on $E \times T$ which are measurable with respect to the product σ-field $\mathscr{E} \times \mathscr{T}$. The norm $\|f\|$ of a function $f \in \mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ is defined by $\|f\| := \sup_{x \in E, y \in T} |f(x, y)|$. The space $\mathscr{B}(E, \mathscr{E})$ is defined analogously.

Theorem 1. *The stochastic process $Z = \{(x_t, y_t), P_{x,y} : x \in E, y \in T, t \in \mathbf{R}^+\}$ is a homogeneous strong Markov process.*

Proof. If we make use of the Markov property of the chain ξ_n ($n \in \mathbf{N}$) and the independence of the variables ν_n ($n \in \mathbf{N}$) a simple computation proves the relation

$$P(x_t \in B,\ y_t \leqq \vartheta | x_{s_n} = x^{(n)},\ y_{s_n} = y^{(n)};\ \dots;\ x_{s_1} = x^{(1)},\ y_{s_1} = y^{(1)}) = \tag{1}$$
$$= P(x_t \in B,\ y_t \leqq \vartheta | x_{s_n} = x^{(n)},\ y_{s_n} = y^{(n)})$$

for any $n \in \mathbf{N}$ $0 \leqq s_1 < s_2 <, \dots, < s_n < t$; $y^{(1)}, \dots, y^{(n)} \in T$; $x^{(1)}, \dots, x^{(n)} \in E$; $B \in \mathscr{E}$. (1) implies that Z is a homogeneous Markov process.

By the definition of x_t and y_t the function $(x_t(\sigma, \omega), y_t(\sigma, \omega))$ with values in the topological space $(E \times T, \mathscr{C}^+)$ is a right-continuous function of the time for any fixed $(\sigma, \omega) \in T \times \Omega$. We show that for any $\mathscr{C}^+$-continuous function $f \in \mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ and $t \geqq 0$ the function $E_{x,y} f(x_t, y_t)$ is also $\mathscr{C}^+$-continuous, i.e. the process is Fellerian in the topology $\mathscr{C}^+$. ($E_{x,y}$ denotes the expectation with respect to the measure $P_{x,y}$.) If $f \in \mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ is a $\mathscr{C}^+$-continuous function then $h(\omega, y) := f(\xi_{N(t+y,\omega)}(\omega), t+y-\eta_{N(t+y,\omega)}(\omega))$ can be uniformly approximated by finite sums of characteristic functions, which are right-continuous w.r.t. the second variable for every fixed $\omega \in \Omega$. (E.g. by $g_n(\omega, y) = \frac{k}{n}$ if $h(\omega, y) \in \left[\frac{k}{n} - \frac{n}{2n}, \frac{k}{n} + \frac{1}{2n}\right)$, where $k = 0, \pm 1, \dots, \pm n \|f\|$.) If $\chi(\omega, y)$ is a characteristic function, right-continuous in y for any fixed $\omega \in \Omega$, then

$$\int_\Omega \chi(\omega, y) P(d\omega | \nu_0 > y,\ \xi_0 = x) =$$
$$= [P(\nu_0 > y | \xi_0 = x)]^{-1} \int_\Omega \chi(\omega, y) P(d\omega \cap \{\nu_0 > y\} | \xi_0 = x)$$

is a right-continuous function of y, since $P(\nu_0 > y | \xi_0 = x)$ is right-continuous. Thus $\int g_n(\omega, y) P(d\omega | \nu_0 > y, \xi_0 = x)$ is $\mathscr{C}^+$-continuous with respect to (x, y). Since h is the uniform limit of g_n as $n \to \infty$,

$$E_{x,y} f(x_t, y_t) = \int_\Omega h(\omega, y) P(d\omega | \xi_0 = x,\ \nu_0 > y)$$

is $\mathscr{C}^+$-continuous.

We have shown Z to be right-continuous and Fellerian in the topology $\mathscr{C}^+$. This implies by [1], Theorem 3.10, the strong Markov property of Z and Theorem 1 is proved.

It is known (c.f. [1]), that as a consequence the Markov property (1) of the process Z, the operators $\{T_t : t \in \mathbf{R}^+\}$, defined by $(T_t f)(x, y) := E_{x,y} f(x_t, y_t)$ for any $f \in \mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$, constitute a semigroup of linear contractions. We say that the sequence $\{f_n\}$ of elements of the Banach space $\mathscr{B}$ tends weakly to $f \in \mathscr{B}$ ($^w\lim f_n = f$) if the numerical sequence $f_n(x, y)$ converges to $f(x, y)$ at any point $(x, y) \in E \times T$ and $\|f_n\|$ is uniformly bounded (with respect to $n \in \mathbf{N}$). The (weak) infinitesimal generator A of the semigroup T_t is the linear operator defined by the expression

$Af := {}^{w}\lim_{t \downarrow 0} \frac{1}{t}(T_t f - f)$ for all $f \in \mathscr{B}$ for which the weak limit on the right-hand side exists and satisfies ${}^{w}\lim_{t \downarrow 0} T_t Af = Af$.

Since $P_{x,y}(\nu_0 \leqq t) := P(\nu_0 \leqq t + y | \xi_0 = x, \nu_0 > y)$ is in $\mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ we may define the function a by $a(x, y) := {}^{w}\lim_{t \downarrow 0} \frac{1}{t} P_{x,y}(\nu_0 \leqq t)$ if the weak limit exists. The relation $P_{x,y}(\nu_0 \leqq 0)$ implies $a(x, y) = \frac{d^+}{dt} P_{x,y}(\nu_0 \leqq t)|_{t=0}$. (Here and in the sequel $\frac{d^+}{dt} f$ or f_t^+ denote the right-hand side derivative of the function f with respect to the variable t.) Denote $Q_{x,y}(B)$ $(B \in \mathscr{E})$ the probability of the event that the process jumps into a point $x' \in B$ after it has left x i.e. $Q_{x,y}(B) := \Pi_x(B, \infty | \nu_0 > y)$.

Theorem 2. *If the function a is $\mathscr{C}^+$-continuous and the sojourn time is independent of the following state then the weak infinitesimal generator A of the process is given by the expression*

$$(2) \qquad (Af)(x, y) = f_y^+(x, y) - a(x, y) f(x, y) + a(x, y) \int_E f(x', 0) Q_{x,y}(dx')$$

for any $\mathscr{C}^+$-continuous function $f \in \mathscr{B}$ with $\mathscr{C}^+$-continuous Af and uniformly locally Lipschitzian with respect to y, i.e. such that

$$\sup_{\substack{x \in E, y \in T \\ y+t \in T}} |f(x, y+t) - f(x, y)| < K \cdot t$$

holds true for some $K > 0$, $t_0 > 0$ and for every t, $0 \leqq t < t_0$.

Proof. The following decomposition holds true for every $f \in \mathscr{B}$

$$\frac{T_t f(x, y) - f(x, y)}{t} = P_{x,y}(\nu_0 > t) \frac{f(x, y+t) - f(x, y)}{t} - \frac{P_{x,y}(\nu_0 \leqq t)}{t} f(x, y) +$$

$$(3) \qquad + \frac{P(N(t+y) = 1 | \xi_0 = x, \nu_0 > y)}{t} \int_\Omega f(\xi_1, t + y - \nu_0) P(d\omega | \xi_0 =$$

$$= x, \nu_0 > y, N(t+y) = 1) +$$

$$+ \frac{P(N(t+y) \geqq 2 | \xi_0 = x, \nu_0 > y)}{t} \cdot E(f(x_t, y_t) | \xi_0 = x, \nu_0 > y, N(t+y) \geqq 2).$$

The definition of $N(t)$ and the independence of ν_0 and ν_1 imply

$$P(N(t+y) \geqq 2 | \xi_0 = x, \nu_0 > y) = P(\nu_0 + \nu_1 \leqq t + y | \xi_0 = x, \nu_0 > y) =$$

$$(4) \qquad = \int_{[0,t) \times E} P(\nu_0 \leqq t + h | \xi_1 = x') P(\nu_0 - y \in dh, \xi_1 \in dx' | \xi_0 = x, \nu_0 > y) \leqq$$

$$\leqq \sup_{x' \in E} P(\nu_1 \leqq t | \xi_1 = x') P(\nu_0 \leqq t + y | \xi_0 = x_0, \nu_0 = y) \leqq [\sup_{x \in E, y \in T} P_{x,y}(\nu \leqq t)]^2.$$

Since $E(f(x_t, y_t)|\xi_0=x, \nu_0>y, N(t+y)\geqq 2)$ is bounded, and $\lim_{t\downarrow 0} P_{x,y}(\nu\leqq t)=0$ holds for any $x\in E$, $y\in T$, the last term of the decomposition (3) tends to zero if $t\downarrow 0$. The definition of $N(t)$ and inequality (4) imply

$$P_{x,y}(\nu_0 \leqq t)-[\sup_{x\in E, y\in T} P_{x,y}(\nu \leqq t)]^2 \leqq P(N(t) = 1|\xi_0 = x, \nu_0 > y) \leqq P_{x,y}(\nu_0 \leqq t),$$

and hence ${}^w\lim_{t\downarrow 0} \frac{1}{t} P(N(t)-1|\xi_0=x, \nu_0>y)=a(x, y)$ holds true. From the condition that the sojourn time is independent of the following state we obtain

$$P(\xi_1\in B, \nu_0 \leqq \vartheta+y|\xi_0 = x, \nu_0 > y, N(t) = 1) =$$

$$= P(\nu_0 \leqq \vartheta+y|\xi_0 = x, y < \nu_0 \leqq y+t < \nu_1)\cdot Q_x(B),$$

and consequently,

$$\int_\Omega f(\xi_1, t+y-\nu_0)P(d\omega|\xi_0 = x, \nu_0 > y, N(t) = 1) =$$

$$= \int_0^t \int_E f(x', t+y-h)Q_x(dx')P(\nu_0-y\in dh|\xi_0 = x, y < \nu_0 \leqq y+t < \nu_1).$$

If $t\downarrow 0$ in the last expression then

$$\lim_{t\downarrow 0} \int_\Omega f(\xi_1, t+y-\nu_0)P(d\omega|\xi_0 = x, \nu_0 > y, N(t) = 1) = \int_E f(x', 0)Q_x(dx')$$

holds true, since $P(\nu_0\leqq t_1+y|\xi_0=x, y<\nu_0\leqq t+y<\nu_1)=1$ and f is $\mathscr{C}^+$-continuous. The last limit relation holds also in the weak sense, since the integral is bounded by $\|f\|$ independently of t.

The differentiability and the uniform Lipschitz property of f implies the weak convergence of $\frac{1}{t}[f(x, y+t)-f(x, y)] \to f_y^+(x, y)$ for $t\downarrow 0$. Thus if $t\downarrow 0$ then each component of the decomposition (3) tends weakly to the corresponding component of (2). Next we show that ${}^w\lim_{t\downarrow 0} T_tAf=Af$. A decomposition analogous to (3) holds true for $(T_tAf)(x, y)-Af(x, y)$. Because of the absence of the divisor t, the last three terms of this decomposition tend weakly to zero if $t\downarrow 0$, while the $\mathscr{C}^+$-continuity of Af implies the first term to converge weakly to zero.

To finish the proof we have to show that every $f\in\mathscr{B}$ from the domain of A with $\mathscr{C}^+$-continuous image Af is differentiable and uniformly Lipschitz continuous. The last two conditions were used only in the proof of the convergence of the first term of (3). This way the left side and also the right side up to its first term converge weakly for every from the domain of A. Since ${}^w\lim_{t\downarrow 0} P_{x,y}(\nu_0>t)=1$ holds, and

$\frac{1}{t}[f(x, y+t)-f(x, y)]$ converges weakly too if $t\downarrow 0$, what is equivalent to the differentiability and Lipschitz property of f, and the proof is finished.

We call two Markov processes equivalent (c.f. [1]), if they are defined on the same state space, and their transition functions coincide. Right-continuous processes are determined by their weak infinitesimal generators uniquely up to equivalence. Consequently if the assumptions of Theorem 2 are fulfilled, the function a and the measures Q_x $(x\in E)$ determine the process Z and this way also the semi-Markov process S up to equivalence.

3. Optimal control of semi-Markov processes

Suppose we are given a family of semi-Markov processes $\{S^d: d\in D\}$ satisfying the conditions of Theorems 1 and 2, and determined by the function $a(x, y, d)$ and the measures $Q^d_{x,y}$. The decision space D is a measurable subset of $\mathbf{R}^L$. $\mathscr{D}$ and $\mathscr{C}_D$ denote respectively the induced σ-field and the induced topology on D. The decision (or control) parameter d can be freely chosen by the controller at every moment. But we only suppose that the decisions are made on the basis of the observation of the state x_t and of the time y_t. In other words d is chosen according a function $u: E\times T\to D$. If u is measurable and $(\mathscr{C}^+, \mathscr{C}_D)$-continuous, then it determines by $Q^{u(x,y)}_{x,y}$ and $a(x, y, u(x, y))$ a new Markov process $Z^u=\{(x_t, y_t), P^u_{x,y}: x\in E, y\in T, t\in\mathbf{R}^+\}$ and hence a new semi-Markov process S^u too. (According to a remark at the end of Section 2 of [3] the trajectories x_t, y_t need not be indexed by u.) We call a measurable, $(\mathscr{C}^+, \mathscr{C}_D)$-continuous function $u: E\times T\to D$ a (feed-back) control strategy, the set U of all strategies the strategy space, while S^u and Z^u are called the processes governed by the strategy u.

Observe, that if A^u and A^d denote the weak generators of the processes Z^u and Z^d, respectively, then the relation

$$A^u f(x, y) = A^{u(x,y)} f(x, y) \tag{5}$$

holds true for every f from the domain of A^u and for every $x\in E$, $y\in T$. The relation (5) implies that if two strategies u_1 and u_2 are equal on a set $B\subset E\times T$, then the processes Z^{u_1} and Z^{u_2} coincide on B.

Let there be given set $G'\subset E\times T$, open in the $\mathscr{C}^+$ topology, and such that for all processes Z^u $(u\in U)$ the time $\tau(\sigma, \omega):=\inf\{t: (x_t(\sigma, \omega), y_t(\sigma, \omega))\notin G'\}$ of the first exit from G' is a Markov time (c.f. [1]). The complementary set of G' is the target set.

Suppose the funtions $p: (E\times T)\setminus G'\to[0, \infty)$ and $q: G'\times D\to[0, \infty)$ are bounded, measurable, $\mathscr{C}^+$ and $\mathscr{C}^+\times\mathscr{C}_D$ continuous, respectively. We are looking for a strategy

$u^* \in U$ minimizing the functional

$$J_{x,y}(u) := E^u_{x,y}\{p(x_\tau, y_\tau) + \int_0^\tau q(x_t, y_t, u(x_t, y_t))\,dt\}$$

for any initial state $x \in E$, $y \in T$, such that $J(u^*)$ is bounded.

We introduce the notation $Bg(x, y) := \inf_{d \in D} [A^d g(x, y) + q(x, y, d)]$, where

$$A^d f(x, y) = f_y^+(x, y) - a(x, y, d) f(x, y) + a(x, y, d) \int_E f(x', 0)\, Q_x^d(dx').$$

The following theorem gives a complete characterization of the optimal strategy.

Theorem 3. *A strategy $u^* \in U$ is optimal iff the boundary value problem*

(6) $$A^{u^*} f(x, y) + q(x, y, u(x, y)) = Bf(x, y) = 0 \quad \text{if} \quad (x, y) \in G',$$

(7) $$f(x, y) = p(x, y) \quad \text{if} \quad (x, y) \notin G'$$

possesses a bounded, measurable solution f^*.

The proof of Theorem 3 is based on the following lemma, proved in [4].

Lemma. *If $\varkappa$ is a Markov time with $\varkappa \leqq \tau$, then for any $u \in U$ the relation*

(8) $$J_{x,y}(u) = E^u_{x,y}\{J_{x_\varkappa, y_\varkappa}(u) + \int_0^\varkappa q(x_t, y_t, u(x_t, y_t))\,dt\}$$

holds true.

Proof (Theorem 3). Let $u^* \in U$ be an optimal strategy, and let $f^*(x, y) := J_{x,y}(u^*)$ The Feller property of Z^u implies $\mathscr{C}^+$-continuity of f^*. (6) is equivalent relations

(9) $$A^{u^*} f^*(x, y) + q(x, y, u^*(x, y)) = 0 \quad \text{for all} \quad (x, y) \in G',$$

(10) $$A^d f^*(x, y) + q(x, y, d) \geqq 0 \quad \text{for all} \quad (x, y) \in G', \; d \in D.$$

To prove (9) let us apply the lemma to the strategy u^* and to the time $\varkappa = h \in \mathbf{R}^+$. With the abbreviation $q^u(x, y) := q(x, y, u(x, y))$ we obtain:

$$T_h^{u^*} f^*(x, y) - f^*(x, y) = E^{u^*}_{x,y} J_{x_h, y_h}(u^*) - J_{x,y}(u^*) = -E^{u^*}_{x,y} \int_0^h q^{u^*}(x_t, y_t)\,dt,$$

and hence

$$A^{u^*} f^* = {}^{w}\lim_{h \downarrow 0} \frac{1}{h}(T_h f^* - f^*) = {}^{w}\lim_{t \downarrow 0} \frac{1}{h} \int_0^h T_t^{u^*} q^{u^*}\,dt = -q^{u^*}.$$

Since q^{u^*} is $\mathscr{C}^+$-continuous, the last relation shows that f^* is in the domain of A^{u^*} and proves (9). Equation (7) holds, since $P_{x,y}(\tau = 0) = 1$ holds for any pair $(x, y) \notin G'$.

Next we prove relation (10). Suppose there is a decision d_0 and a point-pair $(x_0, y_0)\in G'$, for which (10) is false. Then for any $u\in U$ with $u(x_0, y_0)=d_0$ the relation

$$A^u f^*(x_0, y_0)+q(x_0, y_0, u(x_0, y_0)) < 0$$

holds true. We define for arbitrary $t\in T$ the strategy $u_t\in U$:

$$u_t(x, y):=\begin{cases} d_0 & \text{if} \quad x=x_0 \quad \text{and} \quad y_0 \leqq y < y_0+t, \\ u^*(x, y) & \text{elsewhere.}\end{cases}$$

Since for any $u\in U$ the relations ${}^w\lim\limits_{ht\downarrow 0} T_h^u A^u f^* = A^u f^*$ and ${}^w\lim\limits_{t\downarrow 0} T_t^u q^u = q^u$ hold, there exists for any $t>0$ a $t_0>0$ such that for all $0\leqq h<t_0$

$$T_h^{u_t}[A^{u_t} f^*(x_0, y_0)+q^{u_t}(x_0, y_0)] < 0$$

holds true. We introduce the notations $u_0:=u_{\min[t_1, t_0]}$ for a fixed $t_1\in T$, and $\varkappa(\sigma, \omega):=$ $=\min[\tau, \tau_0, \min(t_1, t_0)]$, where τ_0 denotes the time of the first exit from the point x_0. Since the strategies u_{t_1} and u_0 coincide on the set $\{x_0\}\times[y_0, y_0+\min(t_1, t_0))$ so do the processes $Z^{u_{t_1}}$ and Z^{u_0}, and the last inequality can be rewritten in the form

$$\text{(11)} \qquad E^{u_0}_{x_0, y_0} \int_0^{\varkappa} [A^{u_0} f^*(x_0, y_t)+q^{u_0}(x_0, y_t)]\, dt < 0.$$

If we apply Dynkin's formula (corollary of Theorem 5.1 of [1]) to the function $f^*=J(u^*)$, to the process Z^{u_0} and to the Markov time $\varkappa$, we obtain

$$J_{x_0, y_0}(u^*) = E^{u_0}_{x_0, y_0}\Big\{J_{x_\varkappa, y_\varkappa}(u^*) - \int_0^{\varkappa} A^{u_0} f^*(x_t, y_t)\, dt\Big\}.$$

The application of the lemma to u_0 and $\varkappa$ implies

$$J_{x_0, y_0}(u_0) = E^{u_0}_{x_0, y_0}\Big\{J_{x_\varkappa, y_\varkappa}(u_0) + \int_0^{\varkappa} q^{u_0}(x_0, y_t)\, dt\Big\}.$$

We denote by μ the Markov time of the first entrance in the set $H_0=\big[\{x_0\}\times$ $\times[y_0, y_0+\min(t_1, t_0))\big]\cap G'$. (Clearly, $\varkappa$ is the time of the first exit from H_0.) Let we define $\Omega_0:=\{(\sigma, \omega)\in T\times\Omega: \mu\geqq\tau\}$. Intuitively Ω_0 means the set of the elementary events, for which (x_t, y_t) leaves G' before crossing the set H_0. Applying the lemma to μ and the strategies u^* and u_0 respectively, with the aid of the decomposition $\Omega':=T\times\Omega=\Omega_0\cup(\Omega'\setminus\Omega_0)$, we obtain for any $(x, y)\in G'$

$$J_{x, y}(u^*) = E^{u^*}_{x, y}\chi_{\Omega_0}\Big\{p(x_\tau, y_\tau)+\int_0^{\tau} q^{u^*}(x_t, y_t)\, dt\Big\}+$$
$$+E^{u^*}_{x, y}\chi_{\Omega'\setminus\Omega_0}\int_0^{\mu} q^{u^*}(x_t, y_t)\, dt + J_{x_0, y_0}(u^*)E^{u^*}_{x, y}\chi_{\Omega'\setminus\Omega_0}$$

and

$$J_{x,y}(u_0) = E^{u_0}_{x,y}\chi_{\Omega_0}\{p(x_\tau, y_\tau) + \int_0^\tau q^{u_0}(x_t, y_t)\,dt\} +$$
$$+ E^{u_0}_{x,y}\chi_{\Omega'\setminus\Omega_0}\int_0^\mu q(x_t, y_t)\,dt + J_{x_0 y_0}(u_0)\,E^{u_0}_{x,y}\chi_{\Omega'\setminus\Omega_0},$$

where χ_A denotes the indicator function of the set A. Since the strategies u_0 and u^*, and hence also the processes Z^{u_0} and Z^{u^*} coincide outside of H_0, the relation

$$E^{u^*}_{x,y}\chi_{\Omega_0}\{p(x_\tau, y_\tau) + \int_0^\tau q^{u^*}(x_t, y_t)\,dt\} = E^{u_0}_{x,y}\chi_{\Omega_0}\{p(x_\tau, y_\tau) + \int_0^\tau q^{u_0}(x_t, y_t)\,dt\}$$

holds true for any $(x, y) \notin H_0$. Further on, since if $(\sigma, \omega) \in \Omega' \setminus \Omega_0$ the trajectory $(x_t(\sigma, \omega), y_t(\sigma, \omega))$ does not cross the set H_0 before μ,

$$E^{u^*}_{x,y}\chi_{\Omega'\setminus\Omega_0}\int_0^\mu q^{u^*}(x_t, y_t)\,dt = E^{u_0}_{x,y}\chi_{\Omega'\setminus\Omega_0}\int_0^\mu q^{u_0}(x_t, y_t)\,dt$$

holds true for any $(x, y) \notin H_0$. Again by the same argument we obtain the relation

$$E^{u^*}_{x,y}\chi_{\Omega'\setminus\Omega_0} = P^{u^*}_{x,y}(\mu < \tau) = P^{u_0}_{x,y}(\mu < \tau) = E^{u_0}_{x,y}\chi_{\Omega'\setminus\Omega_0}$$

for any $(x, y) \notin H_0$. The coupling of the last five relations implies

$$J_{x_0, y_0}(u_0) - J_{x_0, y_0}(u^*) = E^{u_0}_{x_0, y_0}\{J_{x_\varkappa, y_\varkappa}(u_0) - J_{x_\varkappa, y_\varkappa}(u^*)\} +$$
$$+ E^{u_0}_{x_0, y_0}\int_0^\varkappa [A^{u_0} f^*(x_0, y_t) + q^{u_0}(x_0, y_t)]\,dt =$$
$$= [J_{x_0, y_0}(u_0) - J_{x_0, y_0}(u^*)]\,E^{u_0}_{x_0, y_0}E^{u_0}_{x_\varkappa, y_\varkappa}\chi_{\Omega'\setminus\Omega_0} +$$
$$+ E^{u_0}_{x_0, y_0}\int_0^\varkappa [A^{u_0} f^*(x_0, y_t) + q^{u_0}(x_0, y_t)]\,dt.$$

With the abbreviation $\alpha := E^{u_0}_{x_0, y_0}E^{u_0}_{x_\varkappa, y_\varkappa}\chi_{\Omega'\setminus\Omega_0} = E^{u_0}_{x_0, y_0}P^{u_0}_{x_\varkappa, y_\varkappa}(\mu < \tau)$ the last equality can be written in the form

$$(1-\alpha)[J_{x_0, y_0}(u_0) - J_{x_0, y_0}(u^*)] = E^{u_0}_{x_0, y_0}\int_0^\varkappa [A^{u_0} f^*(x_0, y_t) + q^{u_0}(x_0, y_t)]\,dt < 0,$$

where the inequality follows from (11). Since $0 \leqq \alpha \leqq 1$ the last relation contradicts the assumed optimality of u^*, and hence relation (5) is proved.

To finish the proof we have to show, that the solvability of the boundary value problem (6), (7) is sufficient for the strategy u^* to be optimal. For this we refer to [2], where the proof is given for more general processes.

4. Optimal control of Markov jump processes

As it is known, a semi-Markov process is Markovian, iff the sojourn time in any state is exponentially distributed, independently of the following state, i.e. $P(v_n \leqq \vartheta | \xi_n = x) = \exp\{-\lambda(x)\vartheta\}$ with some $\lambda(x) > 0$. As a consequence of the exponential distribution of v, the relation $P(v_n \leqq \vartheta + y | \xi_n = x, v_n > y) = \exp\{-\lambda(x) \cdot \vartheta\}$ holds true, and therefore, $a(x, y) = \lambda(x)$ holds independently of y. It arises the question, when is it possible to control a family of Markov processes optimally by strategies based only upon the observation of the current state but not upon the sojourn time. The same question was extensively studied in [4]. In the sequel we show that the main result of [4] can be obtained as an easy consequence of Theorem 3. More precisely we show that if the expense components do not depend on the time the process has already spent in its current state, then conditions relative to those of Theorem 3 are necessary and sufficient for optimality.

Suppose we are given a family $\{X^d : d \in D\}$ of Markov jump processes determined by the reverse expected sojourn times $\lambda(x, d)$ with $\lambda(x, d) \leqq K$, and by the jump probabilities Q_x^d. Denote U_M the class of all measurable strategies $u: E \to D$. Suppose $G \subset E$ is a set, such that the first exit time τ from G is Markovian, and the functions $p: E \setminus G \to [0, \infty)$, $q: G \times D \to [0, \infty)$ are bounded and measurable. A strategy $u^* \in U_M$ is said to be optimal, if it minimizes the functionals

$$J_x(u) = E_x^u\{p(x_\tau) + \int_0^\tau q(x_t, u(x_t))\,dt\}$$

under all strategies $u \in U_M$ for any $x \in E$. We introduce the notation $Bf(x) := \inf_{d \in D} [A^d f(x) + q(x, d)]$ for all functions $f \in \mathscr{B}(E, \mathscr{E})$, where

$$A^d f(x) = -\lambda(x, d) f(x) + \lambda(x, d) \int_E f(x') Q_x^d(dx'). \tag{12}$$

Then we can state the following

Theorem 4. *A strategy u^* is optimal in U_M iff the boundary value problem*

$$A^{u^*} f(x) + q(x, u^*(x)) = Bf(x) = 0 \quad (x \in G) \tag{13}$$

$$f(x) = p(x) \quad (x \in E \setminus G) \tag{14}$$

possesses a bounded solution f^*.

Proof. If λ is bounded, then the infinitesimal generator of the process X^d is defined for all functions $f \in \mathscr{B}(E, \mathscr{E})$ and is given by (12). Consequently, if we extend the process X_t to Z_t onto the space $E \times T$, then the generator A_{ext}^d of the latter will be given by $A_{\text{ext}}^d\, g(x, y) = g_y^+(x, y) - \lambda(x, y) g(x, y) + \lambda(x, y) \int_E g(x', 0) Q_x^d(dx')$ for all functions $g \in \mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ uniformly Lipschitzian with respect to y. $\mathscr{B}(E, \mathscr{E})$ can be embedded in $\mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ as the subspace of all functions, constant with respect to y.

If a function $f^* \in \mathscr{B}(E, \mathscr{E})$ suffices (12)—(13), then $g^* \in \mathscr{B}(E \times T, \mathscr{E} \times \mathscr{T})$ defined by $g^*(x, y) := f^*(x)$ for any $y \in T$ is a solution of (6)—(7), since $g_y \equiv 0$. Hence, the optimality of u^* follows from the statement of Theorem 3.

To prove that if u^* is optimal in U_M then $f^*(x) := J_x(u^*)$ is a solution of (13)—(14) we can repeat the proof of the corresponding part of Theorem 3. We have only to show that u_0 can be chosen from U_M too. Set $u_0(x) = u^*(x)$ if $x \neq x_0$ and $u_0(x_0) = d_0$, and the rest of the proof can be carried out analogously to that of Theorem 3.

We remark that Theorem 4 applies to Markov jump processes, a similiar result can be derived from the results of [3]. But an essential difference is that in Theorem 4 the operators A^u are simple integral operators, while with the methods of [3] one can derive optimality conditions with unbounded operators only, even in the case of a jump process.

References

[1] E. B. Dynkin, *Markov processes* (Berlin, Heidelberg, New York, 1965).

[2] R. Pyke, Markov renewal processes, *Ann. Math. Stat.*, **32** (1961), 1231—1259.

[3] D. Vermes, A necessary and sufficient condition of optimality for Markovian control problems, *Acta Sci. Math.*, **34** (1973), 401—413.

[4] D. Vermes, Zur optimalen Steuerung sprungartiger Markov Prozesse, *Beiträge zur Analysis* (Berlin), *to appear.*

(Received December 29, 1973)