数理解析研究所講究録
第 99 巻 1970 年 2-28

Optimal Control of Markov Processes with Average Cost Criterion

Masao Nakamura*)
Keio University

1. Case of Markov Processes

1. Formulation of the problem

Let $(\Omega, \mathcal{F}, P)$ be a probability space with points ω and consider the following stochastic differential equation

(1) $$dx(t)=a(x(t),u(t))dt+b(x(t),u(t))dB(t)$$
$$x(0)=\xi$$

where:

1) $B(t,\omega)=(B_1(t,\omega),B_2(t,\omega),\cdots,B_n(t,\omega))^T$[1)] is a vector Wiener process with its elements mutually independant,

2) $\xi(\omega)$ is a vector random varialble being independent of $B(t,\omega)-B(0,\omega)$ for any $t\in[0,\infty)$,

3) $x(t,\omega)=(x_1(t,\omega),x_2(t,\omega),\cdots,x_n(t,\omega))^T$ takes a value in R^n[2)],

4) $a(x,u)=(a_1(x,u),a_2(x,u),\cdots,a_n(x,u))^T$ and $b(x,u)=\{b_{ij}(x,u);\ 1\leq i,j\leq n\}$ are continuous in u in R^m and satisfy the Lipschitz condition with respect to x in R^n uniformly in u (to be precise this condition need to hold only for u in U and x in X where U and X are specified below),

5) $u(t,\omega)=(u_1(t,\omega),u_2(t,\omega),\cdots,u_m(t,\omega))^T$ is an element of C defined below.

For a given set $U\subseteq R^m$, consider all the stochastic processes of $u(t,\omega)$ such that $u(t,\omega)\in U$[3)] for all ω in $\Delta\subseteq\Omega$.

Assumption 1. $P(\Delta)>0$.

---

*). Graduate Student at the Dpt. of Administration Engrg.
1). T denotes a traspose.
2). n-dimensional Euclidean space.
3). U may not be a stochastic process. For if $U=U(t,\omega)$, the function $u(\cdot)$ of C" defined later will depend on (t,ω), and the following theory will not apply.

Let $(\Delta, \Delta\mathcal{F}, P_\Delta)$ [4] be a probability space with points ω reduced to Δ, where P_Δ is defined by $P_\Delta(\cdot)=P(\cdot)/P(\Delta)$.

Definition. Admissible classes of control; C, C', C".

1). C is the class of stochastic processes $\{u(t,\omega)\}$ [†] such that

a). $u(\cdot,\omega) \in U$ for all $\omega \in \Delta$,

b). $\{\omega; u(t,\omega) \leq c\} \in \Delta\mathcal{F}\{x(s,\omega), s \leq t\}$ [5] for any $c \in U$ and any $t \in [0,\infty)$.

2). C' is the class of stochastic processes $\{u(t,\omega)\}$ in C such that

c). $\{\omega; u(t,\omega) \leq c\} \in \Delta\mathcal{F}\{x(t,\omega)\}$ for any $c \in U$ and any $t \in [0,\infty)$.

3). C" is the class [6] of stochastic processes $\{u(t,\omega)\}$ in C' such that

d). $u(t,\omega)$ is a B-measurable function of $x(t,\omega)$, i.e. there exists a Baire function $u(x)$ satisfying $u(t,\omega)=u(x(t,\omega))$ for a.a. ω in Δ [7] and any $t \in [0,\infty)$.

Thus by definition $C \supseteq C' \supseteq C''$.

Assumption 2. $C'' \neq \phi$ [8].

The existence and uniqueness of the solution to (1) for every u in C is proved in the similar way as is found in Theorem 66.1 of [6] *), by which the solution $x(t,\omega)$ of (1) exists a.a. ω uniquely for any $t \in [0,\infty)$ and u in C. It is also shown by this theorem that

---

4). $\Delta\mathcal{F}$ is the Borel field of sets of $\mathcal{F}$ contained in Δ.

5). $\Delta\mathcal{F}\{x(s,\omega), s \leq t\}$ is the Borel field generated by $\{x(s,\omega), 0 \leq s \leq t\}$.

6). The element of C" is considered as a feedback control law. As an element of C-C' we may consider, for example, $u(t,\omega)=F(x(t-n\varphi,\omega),\cdots,x(t-\varphi,\omega),x(t,\omega))$ where $\varphi(>0)$ is a (random or nonrandom) constant, $n>0$, and F is a function which describes the controller.

7). In the following we write a.a. ω.

8). ϕ denotes a null set.

*). Numbers in brackets refer to the references cited at the end of this paper.

†). We assume that $u(t,\omega)$ is Lebesque integrable with respect to t.

$x(t,\omega)$ is measurable in ω with respect to P_Δ measure and in t with respect to Lebesgue measure, and that it will be a Markov process with an infinitecimal operator $A_u=\sum_{i=1}^{n} a_i(x,u)\frac{\partial}{\partial x_i}+\frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{n} b_{ik}(x,u)b_{jk}(x,u)\frac{\partial^2}{\partial x_i \partial x_j}$. For further results about (1), refer to [2] and the Appendix of [7].

Let X be the given state space in R^n ($X \subseteq R^n$).

Assumption 3. For a.a.ω and any $t\in[0,\infty)$, $x(t,\omega)\in X$.

In the sequel we consider only $(\Delta, \Delta\mathcal{F}, P_\Delta)$ instead of $(\Omega, \mathcal{F}, P)$, and the distribution function of transition probability corresponding to u in C is denoted by $P_u(t,x;s,E)$ for $0\le t\le s<\infty$, $x\in X$, and $E\in\mathcal{B}$, where $\mathcal{B}$ is the Borel field over X.

Our purpose is to find an element of C which minimizes the cost criterion which depends on $u\in C$

$$Q_u(x(0))=\overline{\lim_{T\to\infty}}\frac{1}{T}\int_0^T E^{x(0)}k(x(t,\omega),u(t,\omega))dt^{9)}$$

where $k(x,u)$ is assumed to be continuous in x and u, and $x(x,u)>0$ for any $x\in X$ and $u\in U$.

In the fields of operations research and system science there are many problems where the state of the controlled system is approximately described by (1), and its control is to be decided by the average cost criterion for a long time period. Our result shows that under certain conditions a stationary control, i.e. an element of C" is optimal over the wider class C, as might be expected intuitively.

2. Sufficient Condition for Optimality

At first the following assumptions are made.

Assumption 4. $x(t,\omega)$ is completely observable for ant $t\in[0,\infty)$ a.a.ω.

---

9). $E^{x(0)}$ denotes a conditional expectation (over Δ) given $x(0)$.

Assumption 5. $k(x,u)<\infty$ for any x in X and u in U.

Here let us denote a subset of U, $\{u(t,\omega)\in U;\omega\in\Delta\}$ for each t, by C(t), and a subset of X, $\{x(t,\omega)\in X;\omega\in\Delta\}$ for each t, by $\Xi(t)$. Note that $\Xi(t)\in\mathcal{B}$.

Assumption t. C(t) is compact.

Assumption 7. There exists a $T<\infty$ such that for any $t>T$, at least one u satisfying the following (2) does not depend on t explicitely a.a. ω, i.e. it has a form : $u(t,\omega)=u(x(t,\omega))$ for $t>T$.

Note that by Assumption 7 $x(t,\omega)$ will (a.a. ω) enter an ergodic subset of X. for a certain u.

Under the assumptions that we have made the following theorem holds. The derivation of the theorem is an analog of that presented for a discrets time system with a denumerable state space by Derman [4].

Theorem 1. If there exists a pair $\{g(\text{constant}),\ v(x);x\in X\}$ bounded and v(x) being B-measurable in x such that for an arbitrary small $\Delta t>0$, and for sufficiently large t

$$(2)\qquad g\,\Delta t+v(x(t,\omega))=\inf_{u\in C(t)}\{k(x(t,\omega),u)\,\Delta t+ \\ +\int_X P_u(t,x(t,\omega);t+\Delta t,dy)v(y)\}\quad \text{a.a.}$$

then $u^*(t,\omega)\in C''$ such that for any u in C

$$\varrho_u(x(0))-\varrho_{u^*}(x(0))\geq 0 \qquad \text{a.a. } \omega.$$

exists uniquely and $g=\varrho_{u^*}(x(0))$ a.a. ω.

Remark. By As umption 5 and by the assumptions on v(x) and k(x,u), the right side[10] of (2) is (a.a. ω) well defined for any $t\in[0,\infty)$.

Proof. Let $\{u_n^*\}$, n=1,2,$\cdots$ be such that $\{k(x(t,\omega),\ u_n^*)\,\Delta t+ +\int_X P_{u_n^*}(t,x(t,\omega);\ t+\Delta t,\ dy)v(y)\}$ converges to the right side of (2). (Note that u^* which satisfies (2) belongs to C' for sufficiently large t.) Then for each n, by Assumption 7, we may consider that it has a form of $u_n^*(t,\omega)=u_n^*(x(t,\omega))$.

---

10). Note that the infimum exists uniquely.

Now let us define the distance in C" by $\sup_{x\in X}|u_1(x)-u_2(x)|$, where u_1 and u_2 are arbitrary elements of C".

To show that $u_n^*(\cdot)$ is B-measurable, we consider a set E_n defined by $E_n=\{x;u_n^*(x)\leq c\}$ for any c in U.Since to any $u_n^*(x(t,\omega))\in U$ corresponds a $x(t,\omega)\in X$ for any t a.a.ω , $E_n\in\mathcal{B}$ by the definition of $\mathcal{B}$, and $\{u_n^*(\cdot),\ n=1,2,\cdots\}$ is B-measurable by Theorem 20.2D of [6].

Since C(t) is compact, $\prod_{t\in[0,\infty)} C(t)$ is also compact by Tychonoff's theorem and thers exists a convergent subsequence $\{u_{n_\nu}^*,\ \nu=1,2,\cdots\}$ such that $\lim_{\nu\to\infty} u_{n_\nu}^*=u^*$.Since each $u_{n_\nu}^*$ is a Baire function, so is u^*.Therefore $u^*\in C"$.

Since $\{(k(x(t,\omega),u_{n_\nu}^*)\Delta t + \int_X P_{u_{n_\nu}^*}(t,x(t,x(t,\omega);t+\Delta t,dy)v(y)\}$
$\nu=1,2,\cdots$ is a subsequence of a convergent sequence
$\{k(x(t,\omega),u_n^*)\Delta t + \int_X P_{u_n^*}(t,x(t,\omega);t+\Delta t,dy)v(y)\}$ $\quad n=1,2,\cdots$
its limit coincides with that of the latter, i.e. the infimum of (2) is attained by u^*.

For any element of $u\in C"$, $x(t,\omega)$ is a stationary Markov process,[11) so that the transition probability distribution is described as $P(x(0);t,x(t,\omega)\in E)$ for $E\in\mathcal{B}$.Thus (2) becomes for u^*

$$g\Delta t + v(x)=k(x,u^*(x))\Delta t + \int_X P_{u^*}(x:\Delta t,dy)v(y) .$$

Multiplying both sides of the above by $\int_X P_{u^*}(x(0);t,dx)$

$$g\Delta t + \int_X P_{u^*}(x(0);t,dx)v(x)= \int_X P_{u^*}(x(0);t,dx)k(x,u^*(x))\Delta t + \\ + \int_X P_{u^*}(x(0);t+\Delta t,dy)v(y) .$$

By taking the time average in the above, we obtain

$$g\Delta t + \overline{\lim_{T\to\infty}}\frac{1}{T}\int_0^T\int_X P_{u^*}(x(0);t,dx)v(x)=\varrho_{u^*}(x(0))\Delta t + \\ + \overline{\lim_{T\to\infty}}\frac{1}{T}\int_0^T\int_X P_{u^*}(x(0);t+\Delta t,dy)v(y)$$

and

$$g\Delta t=\varrho_{u^*}(x(0))\Delta t \qquad \text{or} \qquad g=\varrho_{u^*}(x(0)) .$$

To prove the optimality of u^* over C, we define the optimal total

---

11). Theorem 66.1 and 66.2 of [6].

cost[12] up to time $t+\Delta t$ for each ω, by $g_{t+\Delta t}(x(t+\Delta t,\omega))$: i.e. for sufficiently small $\Delta t>0$,

$$g_{t+\Delta t}(x(t+\Delta t,\omega))=\inf_{u\in C(t)}\{k(x(t,\omega),u)\Delta t+\int_{\Xi(t)}P_u(t,x(t,\omega);\ t+\Delta t,dy)g_t(y)\}$$

(3)

$$g_0(x(0))=0.$$

As is shown in Appendix, there exists (a.a. ω) a unique solution of (3) on $[0,\infty)\times\Xi(t)$ for any $t\in[0,\infty)$. That $g_{t+\Delta t}(x(t+\Delta t,\omega))$ denotes an optimal total (expected) cost up to time $t+\Delta t$ for a.a ω. comes from the principle of optimality.

Next, we prove the existence of $M(\Delta t)<\infty$ such that for some finite $g'\geq 0$ and any $t\in[0,\infty)$:

(4) $\quad (t+\Delta t)g'-M\leq g_{t+\Delta t}(x(t+\Delta t,\omega))\leq M+(t+\Delta t)g'$ a.a. ω.

To prove the second inequality, assume that it does not hold, i.e. for at least one $u\in C, t\in[0,\infty)$ and ω with positive provability $g_{t+\Delta t}(x(t+\Delta t,\omega))>M+(t+\Delta t)g'$ holds for all finite M independent of t. Let t^* be such that this holds and let M be such that $g_{t^*+\Delta t}(x(t^*+\Delta t,\omega)<M$. Then $(t^*+\Delta t)g'<0$, contradicting that $g'\geq 0$. The first inequality of (4) follows in a similar way.

Thus from (4) it follows that for any $t\in[0,\infty)$, $|g_{t+\Delta t}(x(t+\Delta t,\omega))/ /(t+\Delta t)-g'|\leq M/(t+\Delta t)$ a.a. ω, and letting $t\to\infty$, $\lim_{t\to\infty}\{g_{t+\Delta t}(x(t+\Delta t,\omega))/ /(t+\Delta t)\}=g'$ a.a. ω. This shows that g' is an optimal average cost. Next, we will show that $g=g'$. At first, note that by (2) and (3)

---

12). Let $g_{t+\Delta t}(x(t+\Delta t,\omega))$ be an optimal total (expected) cost up to time $t+\Delta t$ incurred by the process which passes $x(t,\omega)$ at time t, where $t\in[0,\infty)$. Then

$$g_{t+\Delta t}(x(t+\Delta t,\omega))=\inf_{u\in C(t)}\{E^{x(t,\omega)}\int_0^{t+\Delta t}k(x(\tau,\omega)d\tau\} =$$
$$=\inf_{u\in C(t)}\{E^{x(t,\omega)}\int_t^{t+\Delta t}k(x(\tau,\omega),u)d\tau+\int_{\Xi(t)}p_u(t,x(t,\omega);t+\Delta t,dy)g_t(y)\}$$

$g_{t+\Delta t}(x(t+\Delta t,\omega))=g\Delta t+v(x(t,\omega))+\ell(\text{const.})$ holds for sufficiently large t.

Assume that $g'>g$. Then as $g=Q_{u*}(x(0))$ a.a. ω, contradicting the definition of $g_{t+\Delta t}(x(t+\Delta t,\omega))$. Now assume that $g'<g$. Then by the definition of g', $g_{t+\Delta t}(x(t+\Delta t,\omega))=g_t(x(t,\omega))+g'\Delta t+o(\Delta t)$ hold for sufficiently small Δt. Thus $g_{t+\Delta t}(x(t+\Delta t,\omega))=g_t(x(t,\omega))+$ $+g'\Delta t+o(\Delta t)>g'\Delta t+v(x(t,\omega))+\ell$. Put $v(x)\triangleq v(x)+o(\Delta t)$ anew. A contradiction follows from (2) and (3). Thus $g=g'$.

Let $u(t,\omega)$ be an arbitrary element of C and its corresponding total (expected) cost up to time t be $h_t(x(t,\omega))$, then for a.a. ω,

$$\underline{\lim}_{t\to\infty}(h_t(x(t,\omega))/t)\geq\underline{\lim}_{t\to\infty}(g_{t+\Delta t}(x(t+\Delta t,\omega))/(t+\Delta t))=g$$

proving that for any u in C

$$Q_u(x(0))-Q_{u*}(x(0))\geq 0 \qquad \text{a.a. } \omega.$$

Q.E.D.

Example. Let $n=m=1$, $a(x,u)=u$, $b(x,u)=1$, $k(x,u)=u+e^{-x}$, $X=[0,\infty)$, $C(t)=$ $=[1-1/(t+\varepsilon),1]\cap[0,\infty)$ where $\varepsilon>0$ is sufficiently close to 1, and $B(t,\omega)$ is a Wiener process with a reflecting barrier of 0. It is assumed that $P(x(0)=0)=P(B(0)=0)=1$. Consider a control u_1 such that $u_1(t,\omega)=0$ $(0\leq t\leq 1-\varepsilon)$, $u_1(t,\omega)=1-1/(t+\varepsilon)$ $(1-\varepsilon\leq t<\infty)$. Note that $u_1\in C-C''$. For $0\leq t\leq 1-\varepsilon$, $k(x_{u_1}(t,\omega),u_1(t,\omega))>0$ since $x_{u_1}(t,\omega)=$ $=B(t,\omega)$ is (a.a. ω) small. For $t\geq 1-\varepsilon$, $x_{u_1}(t,\omega)=t-\log(t+\varepsilon)-$ $-(1-\varepsilon)+B(t,\omega)$, whence $x_{u_1}(t,\omega)\to\infty$, (a.a. ω) as $t\to\infty$. Thus $Q_{u_1}(x(0))=1$.

Now let $\{g,v(x);x\in X\}$ be $\{1,0\}$ and consider (2), i.e.

(A) $$\Delta t=\inf_{u\in C(t)}\{(u+e^{-x(t,\omega)})\Delta t\}.$$

Let u_2 be such that $u_2(x)=1$ for all x in X. Then $x_{u_2}(t,\omega)=t+B(t,\omega)$. By substituting $x_{u_2}(t,\omega)$ into the brace of (A), we obtain for

---

(continued)

$$=\inf_{u\in C(t)}\left[E^{x(t,\omega)}\{k(x(t,\omega),u)\Delta t+o(\Delta t)\}+\int_{\Xi(t)}P_u(t,x(t,\omega);t+\Delta t,dy)g_t(y)\right]=\inf_{u\in C(t)}\{k(x(t,\omega),u)\Delta t+o(\Delta t)+\int_{\Xi(t)}P_u(t,x(t,\omega);t+\Delta t,dy)g_t(y)\}.$$

Thus we obtain (3) by omitting $o(\Delta t)$. (See Appendix.)

$t \approx \infty$, $t=\inf\{\Delta t\}$, meaning that for sufficiently large t, u_2 attains the infimum of the right side of (A). It follows that u_2 is optimal by Theorem 1, and that $\varrho_{u_2}(x(0)) = g = 1$.

Thus it has been shown in this example that under the assumption of Theorem 1, though there may be optimal controls in C–C", there is an optimal control in C" which attains the same average cost as is incurred by those in C–C".

Since Theorem 1 contains an unspecified quantity Δt, it does not provide us with a concrete method to obtain u^* in C. For this reason the following theorem is useful.

Theorem 2. If the assumptions of Theorem 1, with $v(x)$ and $Av(x)$ being twice differentiable and continuous in $x \in X$ respectively, hold, then (2) is equivalent to the following (5):

(5) $$g = \inf_{u \in C''} \{ k(x,u(x)) + A_u v(x) \}$$

and u^* which attains the infimum of the right side of (5) is (a.a. ω) optimal over C. Note that A_u is an infinitecimal operator of our (stationary) Markov process corresponding to u in C" and is given by: $A_u = \sum_{i=1}^{n} a_i(x,u)\frac{\partial}{\partial x_i} + \frac{1}{2}\sum_{i,j=1}^{n}\sum_{k=1}^{n} b_{ik}(x,u)b_{jk}(x,u)\frac{\partial^2}{\partial x_i \partial x_j}$.

Proof. To prove the theorem it suffices to show the equivalence of (2) and (5). Since the right side of (2) is attained infimum by an element of C", (2) is written as

$$g\Delta t = \inf_{u \in C''} \{k(x(t,\omega),u(x(t,\omega)))\Delta t + \int_X P_u(t,x(t,\omega);\ t+\Delta t,dy)v(y) - \int_X p_u(t,x(t,\omega);t,dy)v(y)\}$$

or for x in X

$$g\Delta t = \inf_{u \in C''} [k(x,u(x))\Delta t + \{ \int_X p_u(t,x;t+\Delta t,dy) - \int_X P_u(t,x;t,dy)\} v(y)] .$$

Since $\Delta t > 0$, dividing both sides by Δt, and letting $\Delta t \longrightarrow 0$, we obtain (5).

To prove the converse, we use Dynkin's formula ([2]) that

$$E^x \int_0^t Av(x(s,\omega))ds = E^x v(x(t,\omega)) - v(x).$$

Then for sufficiently small $\Delta t > 0$,[13)]

$$E^{x(t,\omega)} Av(x(t,\omega))\Delta t = E^{x(t,\omega)} v(x(t+\Delta t,\omega)) - v(x(t,\omega)) + o(\Delta t)$$

or

$$A_u v(x)\Delta t = \int_X P_u(t,x;t+\Delta t,dy)v(y) - v(x) + o(\Delta t).$$

Substituting the above into (5) multiplied by Δt, it follows that

$$g\Delta t = \inf_{u\in C''}\{k(x,u(x))\Delta t + (\int_X P_u(t,x;t+\Delta t,dy)v(y) - v(x))$$
$$+o(\Delta t)\}$$

or $$g\Delta t + v(x) = \inf_{u\in C''}\{k(x,u(x))\Delta t + \int_X P_u(t,x;t+\Delta t,dy)v(y) + o(\Delta t)\}.$$

Thus putting $v(x) \hat{=} v(x) + o(\Delta t)$ anew proves the assertion.

Q.E.D.

Remark. Theorem 4.1 of Wonham [8] has a similar form as Theorem 2 above, but for narrower classes of u.

Remark. Suppose that $\{g, v(x)\}$ satisfying (5) exists, where $n \geq 2$. Then (5) becomes a partial differential equation of $v(x)$, i.e.

$$g = k(x,u^*(x)) + A_{u^*}v(x)$$

for x in X. The boundary condition of $v(x)$ is not specified.

Example. Let $m,n=1$, $a(x,u)=1-u$, $b(x,u)=1$, $k(x,u)=e^{-x}+u$, $X=[0,\infty)$. $U=[0,1]$ and $B(t,\omega)$ is assumed to be a Wiener process with 0 being a reflecting barrier. Note that the infinitecimal operator is defined on $(0,\infty)$. Then (5) becomes

$$(6)\quad g = \inf_{u\in C''}\left[e^{-x} + u(x) + \left\{(1-u(x))\frac{d}{dx} + \frac{1}{2}\frac{d^2}{dx^2}\right\}v(x)\right].$$

Let $P(x(0)=1)=1$. Then for $u(x)=0$ $(x\in X)$, $x_u(t,\omega)=1+t+B(t,\omega)$ a.a. ω, where we have assumed that $P(B(0,\omega)=0)=1$. Thus $Q_u(1)=$
$= \lim_{T\to\infty}\frac{1}{T}\int_0^T E^{x(0)}(e^{-(1+t+B(t,\omega))}+0)dt \leq \lim_{T\to\infty}\frac{1}{T}\int_0^T e^{-(1+t)}dt = \lim_{T\to\infty}\frac{1}{T}$
$e^{-1}\int_0^T e^{-t}dt = 0$, where $E^{x(0)}e^{-B(t,\omega)} \leq 1$ is used. Since $Q_{u'}(1)$ is nonnegative for any u' in C by assumptions on $k(x,u)$, it follows that $u(x)=$ $=0$ $(x\in X)$ is optimal over C. For this $u(x)$, (6) becomes

---

13). $A_u v(x)$ is assumed to be continuous in $x\in X$. (for all u in C").

$$0=g=0+\frac{dv}{dx}+\frac{1}{2}\frac{d^2v}{dx^2}$$

or $v(x)=c_1e^{-2x}+c_2$ where we take c_1 and c_2 as bounded oonstants. Then $v(x)$ is bounded over X, showing that for this system (5) is necessary too for optimality. Sufficiency for the optimality of $u(x)=0$ $(x \in X)$ is seen by the existence of a pair $(g,v(x))=$ $=(0,c_1e^{-2x}+c_2)$ satisfying (6). The right side of (6) in which this $(g,v(x))$ is substituted is minimized by $u(x)=0$ $(x \in X)$ if c_1 is taken to be nonnegative.

3. Policy improvement Procedure

Hereafter it is assumed that the assumptions of Theorem 1 are satisfied. The methods of proving the following assertions are the continuous version of the work of Derman [4] for denumerable Markov chains.

Assumption 8. For any u in C", its corresponding Markov process $x(t,\omega)$ is positive recurrent, i.e. for any x in X and for any measurable function $f(x)$, there exists $P_u(y)$ such that $\lim_{t\to\infty}\int_X \frac{1}{t}\int_0^t P_u(x;\tau,dy)f(y)d\tau=\int_X P_u(dy)f(y)$.

Assumption 9. For some u_i in C" exists $\{g(\text{constant}),\ v(x);x \in X\}$ bounded with $v(x)$ twice differentiable such that $g=k(x,u_i(x))+A_{u_i}v(x)$ and $A_{u_i}v(x) \geq 0$.

Let χ be $\chi=[x \in X;k(x,u_i(x)+A_{u_i}v(x) \neq \inf_{u\in C''}\{k(x,u)+A_uv(x)\}]$. We may assume that $\chi \neq \phi$. Unless so, u_i is optimal over C by Theorem 2.

Let u_i' be such that

$$u_i'(x)=u_i(x) \qquad x \in X-\chi$$
$$u_i'(x)=u^o(x) \qquad x \in \chi$$

where u^o is such that

(7) $\quad k(x,u^o(x))+A_{u^o}v<k(x,u_i(x))+A_{u_i}v.$

Here let us assume that:

Assumption 10. For $P_u(\cdot)$ in Assumption 8, $\inf_{u\in C''}\int_A P\,(dx)>0$ for any nonzero subset A in X.

This assumption shows that any nonzero subset of X is positive recurrent.

Lemma 1.If Assumption 1 - 10 hold, $Q_{u_i'} < Q_{u_i}$.

Proof.Let $\varepsilon(x)$ be the right side of (7) minus the left side of (7),i.e. $\varepsilon(x)=(k(x,u_i(x))-A_{u_i}v)-(k(x,u^o(x))-A_{u^o}v)$.Then $\varepsilon(x)>0$ if $x \in \mathcal{X}$, and $\varepsilon(x)=0$ if $x \in X-\mathcal{X}$.

Since

$$\int_X P_{u_i'}(x(0);t,dx)\ \varepsilon(x)=\int_X P_{u_i'}(x(0);t,dx)\left[g-\left\{k(x,u_i'(x))+A_{u_i'}v\right\}\right] =$$
$$= g-\int_X P_{u_i'}(x(0);t,dy)A_{u_i'}v - \int_X P_{u_i'}(x(0);t,dx)k(x,u_i'(x)),$$

averaging the above by T,

$$(1/T)\int_0^T\int_X P_{u_i'}(x(0);t,dx)\ \varepsilon(x)=g-(1/T)\int_0^T\int_X P_{u_i'}(x(0);t,dx)A_{u_i'}v\ dt-$$
$$-(1/T)\int_0^T\int_X P_{u_i'}(x(0);t,dx)k(x,u_i'(x))dt,$$

and letting $T\to\infty$, it follows by Assumptions 8 - 10 and $\mathcal{X}\neq\phi$ that

$$\int_X P_{u_i'}(dx)\ \varepsilon(x) + \int_X P_{u_i'}(dx)A_{u_i'}v=g-Q_{u_i'}(x(0))>0.$$

It is noted that the interchange of integrals above comes from Fubini's theorem.

Q.E.D.

Lemma 2.Under the assumptions of Lemma 1,if u_i is optimal over C", then u_i is optimal over C.

Proof.If u_i is optimal over C", then $\mathcal{X}=\phi$, and (5) holds.Then Theorem 2 applies.

Q.E.D.

In the following it will be assumed that:

Assumption 11.For any u in C" there exists $\{g^u,\ v^u(x)\}$ which satisfies Assumption 9.

Assumption 12.U is compact.

Theorem 3.If Assumptions 1 - 12 hold, there exists a u* in C" which is optimal over C.

Proof.By Assumption 11, for any u in C", $g^u=k(x,u(x)) + A_u v^u(x)$, $x \in X$.Let $g^*=\inf_{u\in C''}\{g^u\}$. Let $\{u_n(x)\}\in C''$ be such that $\lim_{t\to\infty} g^{u_n}=g^*$

uniformly in $x \in X$.Since $\prod_{x \in X} U_x$ is compact by Tychonoff's theorem, there exists a subsequence $\{u_{n_\nu}\}$, $\nu=1,2,\cdots$ such that $\lim_{\nu \to \infty} u_{n_\nu}=u^*$ uniformly in x.Since C" consists of Baire functions, u* belongs to C".Let v* be such that $\lim_{\nu \to \infty} v^{u_{n_\nu}}=v^*(x)$[14].Then by Assumptions 9 and 11, $v^*(x)$ is bounded.Since A_u is continuous in u[15], it follows that $A_{u_{n_\nu}} \to A_{u^*}$ as $\nu \to \infty$ .Since $g^{u_{n_\nu}}$ is a subsequence of g^{u_n} , it follows that $g^{u_{n_\nu}} \to g^*$ as $\nu \to \infty$.

Consider $g^{u_{n_\nu}}=k(x,u_{n_\nu}(x)) + A_{u_{n_\nu}} v^{u_{n_\nu}}(x)$, and letting $\nu \to \infty$,

$g^* = \lim_{\nu \to \infty} g^{u_{n_\nu}} = k(x,u^*(x)) + A_{u^*}v^*(x)$,it follows that $\{g^{u^*}, v^{u^*}(x)\} = \{g^*, v^*(x)\}$.

For this u* in C", (5) holds.Otherwise, by Lemma 1, policy improvement is possible, resulting a smaller average cost than g*, which contradicts that $g^* = \inf_{u \in C''} g^u$.Thus (5) holds, proving the existence of u* in C" being optimal over C.

Q.E.D.

Let $\varepsilon^u(x)$ be such that

$$\varepsilon^u(x)=(k(x,u(x))+A_u v^u(x))-(k(x,u'(x))+A_{u'}v^u(x))$$
$$=g^u - (k(x,u'(x)) + A_{u'}v^u(x))$$

where $u(x) \in C''$, and u'(x) is an improved policy by the iteration.

Lemma 3.Assume that Assumptions 1 - 12 hold.Let $u_1 \in C''$ be arbitrary and $\{u_n\}$ be a sequence of the policy improvement procedure.Then $\lim_{n \to \infty} \varepsilon^{u_n}(x) = 0$, and $\lim_{n \to \infty} A_{u_{n+1}} v^{u_n}(x) = 0$ for all $x \in X$.

Proof.As in the proof of Lemma 1, using Assumption 8,

$$(*) \quad g^{u_n} - g^{u_{n+1}} = \int_X P_{u_{n+1}}(dx)\varepsilon^{u_n}(x) + \int_X P_{u_{n+1}}(dx) A_{u_{n+1}} v^{u_n} > 0.$$

Since $\{g^{u_n}\}$ is a decreasing sequence and $g^{u_n} > 0$, $\lim_{n \to \infty} g^{u_n}$ exists. Thus $g^{u_n} - g^{u_{n+1}} \to 0$ as $n \to \infty$, and the right side of (*) goes to 0 as $n \to \infty$.By Assumption 10,$\varepsilon^{u_n}(x) \to 0$ and $A_{u_{n+1}} v^{u_n}(x) \to 0$ as $n \to \infty$.

Q.E.D.

---

14). It is seen that this convergence does not imply the ordinary convergence of $\{v^{u_{n_\nu}}, \nu=1,2,\cdots\}$.

15). Note that a(x,u) and b(x,u) are continuous in $u (\in U)$.

Theorem 4. If Assumptions 1 - 12 hold, for any $u_1 \in C''$ a sequence of the policy improvement procedure $\{u_n\}$ converges to u^* in C'' which is optimal over C.

Proof. Let $\{u_n\}$ be a policy improvement sequence. Then using the compactness arguments and the property of Baire functions as we did in the proof of Theorem 3, a convergent subsequence $\{u_{n_\nu}\}$ such that $\lim_{\nu\to\infty} u_{n_\nu} = u^*$ uniformly in $x \in X$ and $u^* \in C''$ can be chosen. Let g^* and v^* be such that[16] $\lim_{\nu\to\infty} g^{u_{n_\nu}} = g^*$ and $\lim_{\nu\to\infty} v^{u_{n_\nu}}(x) = v^*(x)$. Note that $\lim_{\nu\to\infty} \varepsilon^{u_{n_\nu}}(x)=0$, and $\lim_{\nu\to\infty} A_{u_{n_{\nu+1}}} v^{u_{n_\nu}}(x)=0$.

For any u_{n_ν}, by Assumption 11, $g^{u_{n_\nu}} = k(x,u_{n_\nu}(x)) + A_{u_{n_{\nu+1}}} v^{u_{n_\nu}}(x)$. Letting $\nu\to\infty$, $g^* = k(x,u^*) + A_{u^*}v^*(x)$. Thus $g^* = g^{u^*}$ and $v^* = v^{u^*}$.

Since

$$g^* \geq \overline{\lim_{\nu\to\infty}} \inf_{u\in C''} \{ k(x,u(x)) + A_u V^{u_{n_\nu}}(x)\}$$

and

$$g^{u_{n_\nu}} \leq \inf_{u\in C''}\{k(x,u(x)) + A_u v^{u_{n_\nu}}(x)\} + \varepsilon^{u_{n_\nu}}(x) \quad + A_{u_{n_{\nu+1}}} v^{u_{n_\nu}}(x)$$

hold, using Lemma 3 and letting $\nu\to\infty$ in the above, we obtain

$$g^* = \lim_{\nu\to\infty} \inf_{u\in C''} \{k(x,u(x)) + A_u v^{u_{n_\nu}}(x)\}.$$

However

$$\lim_{\nu\to\infty} \inf_{u\in C''} \{k(x,u(x)) + A_u v^{u_{n_\nu}}(x)\} \leq \lim_{\nu\to\infty}\{k(x,u(x)) + A_u v^{u_{n_\nu}}(x)\}.$$

Thus

$$\begin{aligned} g^* &= \lim_{\nu\to\infty} \inf_{u\in C''} \{ k(x,u(x)) + A_u v^{u_{n_\nu}}(x)\} \\ &\leq \inf_{u\in C''} \lim_{\nu\to\infty}\{k(x,u(x)) + A_u v^{u_{n_\nu}}(x)\} \\ &= \inf_{u\in C''} \{k(x,u(x)) + A_u v^*(x)\}. \end{aligned}$$

If we put $u=u^*$, $g^* = k(x,u^*(x)) + A_{u^*}v^*(x)$ holds, showing that (5) holds. This proves the theorem.

Q.E.D.

Remark. Let $b(x,u) = 0$ for all $x \in X$ and $u \in U$. Then our system becomes deterministic, and Assumption 9 and Theorem 2 take a form

---

16). See the footnote 14.

of

$$g = \inf_{u \in C''}\{k(x,u(x)) + \sum_{i=1}^{n} a_i\ (x,u(x))\frac{\partial v(x)}{\partial x_i}\}$$

$$\sum_{i=1}^{n} a_i\ (x,u(x))\ \frac{\partial V(x)}{\partial x_i} \geq 0$$

which is the Hamilton-Jacobi equation.The second condition shows that the value of $v(x(t))$ is nondecreasing along the trajectory of $x(t)$ as $t \to \infty$.

Remark. The relationbetween $v(x(t,\omega))$ and $g_t(x(t,\omega))$ can be interpreted as follows.Suppose that $g_{t+\Delta t}(y)$ can be expanded into a Taylor series about any $t \in [0,\infty)$ uniformly in $y \in X$:

$$g_{t+\Delta t}(y) = g_t(y) + \frac{\partial g_t(y)}{\partial t}\ \Delta t + o(\Delta t).$$

Then, considering that an optimal control exists in C", from (3)

$$\frac{\partial g_t(x(t+\Delta t,\omega))}{\partial t}\ \Delta t + o(\Delta t) = \inf_{u \in C''}\{k(x(t,\omega),u(x(t,\omega))\Delta t +$$

$$+ \int_X P_u(t,x(t,\omega);t+\Delta t\ ,dy)g_t\ (y) - g_t\ (y)\}.$$

Dividing both sides by Δt, and letting $\Delta t \to 0$, with the use of a.a.ω continuity of $x(t,\omega)$, we obtain

$$\frac{\partial g_t(x(t,\omega))}{\partial t} = \inf_{u \in C''}\{k(x(t,\omega),u(x(t,\omega)) + A_u g_t(x(t,\omega))$$

where it is assumed that $g_t(x)$ is twice differentiable in $x \in X$.

Thus under appropriate conditions an optimal total (expected) cost satisfies a partial differential equaion along $x=x(t,\omega)$:

$$\frac{\partial g_t(x)}{\partial t} = \inf_{u \in C''}\{k(x,u(x)) + A_u g_t(x)\}.$$

The above result shows that $g_t(x(t,\omega))$ and $v(x(t,\omega))$ coincide if we take into account that for an optimal control

$$g = \left.\frac{\partial g_t(x)}{\partial t}\right|_{x=x(t,\omega)} \qquad \text{a.a.}\ \omega \qquad \text{for sufficiently large } t.$$

II. Case of Finite Markov Chains

1. Formulation of the Problem

In this and following sections the optimal control of the following linear stochastic difference equation

(8) $x_{n+1} = ax_n + u_n + B_n$

is considered where for each integer $n \geq 0$:

a; integer gain of our system

x_n; state of our system at n

u_n; control at n

B_n; random noise at n

and where u_n and B_n are specified below.

It is assumed that u_n and x_n may take values in the specified integer sets U and X in R^1 respectively,[17] where X is assumed to be a finite set.

Let $\{B_n(\omega), n \geq 0\}$ be a mutually independent random variables over the probability space $(\Omega, \mathcal{F}, P)$ with points ω such that

$P(B_n = 1) = p \quad (0 < p < 1)$

and $P(B_n = -1) = 1-p$.

Note. To extend the following theory to the case of $P(B_n = b_k) = p_k$, where $0 \leq p_k \leq 1$, $\sum_{k=1}^{m} p_k = 1$, is not difficult.

To define the admissible control, let us consider all the stochastic processes of $\{u_n(\omega)\}$ such that $u.(\omega) \in U$[18] for all $\omega \in \Delta$ in Ω.

Assumption 13. $P(\Delta) > 0$.

Let the probability space reduced to Δ be $(\Delta, \Delta\mathcal{F}, P_\Delta)$ with points

---

17). We may extend our theory to a vector control process such that $x_{n+1} = Ax_n + Bu_n + B_n$ where: $x. \in R^n$, $u. \in R^m$, B_n is a n-dimensional stochastic process, A and B are nxn and nxm matrices rspectively.

18). See the footnote 3.

ω, where P_Δ is defined by $P_\Delta(\cdot) = P(\cdot)/P(\Delta)$.

Definition. Admissible classes of control; C, C', C".

1). C is the class of stochastic processes $\{u_n(\omega)\}$ such that

a). $u.(\omega) \in U$ for all $\omega \in \Delta$,

b). $\{\omega; u_t(\omega) = c\} \in \Delta\mathcal{F}\{x_s(\omega), s \leq t\}$ for any $c \in U$ and any $t \in [0,1,2,\cdots)$.

2). C' is the class of stochastic processes $\{u_n(\omega)\}$ in C' such that

c). $\{\omega; u_t(\omega) = c\} \in \Delta\mathcal{F}\{x_t(\omega)\}$ for any $c \in U$ and any $t \in [0,1,2,\cdots)$.

3). C" is the class of stochastic processes $\{u_n(\omega)\}$ in C' such that

d). $u_t(\omega)$ is a B-measurable function of $x_t(\omega)$, i.e. there exists a Baire function $u(x)$ satisfying $u_t(\omega) = u(x_t(\omega))$ for a.a. ω in[19] Δ and any $t \in [0,1,2,\cdots)$.

Thus by definition $C \supseteq C' \supseteq C''$.

Assumption 14. $C'' \neq \Phi$.

Assumption 15. $x_0(\omega)$ is measurable with respect to P_Δ measure.

By Assumption 13 and by the definition of C the solution $\{x_n(\omega), n \geq 0\}$ to (8) is measurable with respect to P_Δ measure.

Now suppose that to each elemetn of x in X and c in U is attached a cost $g(x,c)$ which is positive and finite, i.e. if the state and the control are $x_n(\omega)$ and $u_n(\omega)$ at time n respectively, we need to pay $g(x_n(\omega), u_n(\omega))$ then.

Though our interest is a total amout of the cost incurred by each $\{x_n(\omega)\}$ and $\{u_n(\omega)\}$, $\sum_{t=0}^{\infty} g(x_t(\omega), u_t(\omega))$ diverges a.a. ω by the definition of the cost, therefore instead, we consider the average cost $\varrho_u(x_0) = \overline{\lim_{T\to\infty}} (1/T)\, E^{x_0} \sum_{t=0}^{T} g(x_t(\omega), u_t(\omega))$ which depends on the choice of $\{u_n(\omega)\}$ in C.

Thus our problem is to find an optimal control $\{u_n^*(\omega)\}$ which

---

19). See the footnote 7.

minimizes the average cost $Q_u(x_0)$ a.a.ω ,where $x_0(\omega)$ is assumed to be independent from $\{B_n(\omega), n \geq 0\}$

2. Algorithm of the Solution

Since $\{x_n(\omega)\}$ is a measurable stochastic process for every u in C as was pointed out, a conditional probability of $x_{n+1}(\omega) = i$ given $x_0, x_1, \cdots, x_n$ and $u_0, u_1, \cdots, u_n$ is defined as $P(x_{n+1}=i/x_s, u_s; 0 \leq s \leq n)$. For u in C" it is seen that $\{x_n(\omega)\}$ will be a finite Markov chain with stationary transition probabilities: $p_{ij} = P(x_{n+1} = j/x_n = i); i, j \in X$.

Before seeking an optimal control of our problem the following version of Derman's Theorem 1 ([3]) is stated without proof.

Theorem 5. If Assumptions 13 - 15 hold, there axists u* in C" which is optimal (a.a.ω) over C.

By this theorem we have only to consider the elements of C" to obtain an optimal control. The following theorem gives us an algorithm to obtain an optimal control.

Theorem 6. Under the assumptions of Theorem 5, an optimal control u* is obtained in finite steps.

Proof. We first compute the transition probability of a Markov chain for u in C". For m and N such that $m \in X, m+N \in X$;

$$
\begin{aligned}
(9) \qquad P_{m,m+N} &= P(x_{n+1}=m+N/x\ =m) \\
&= P(ax_n+u_n + B_n = m + N/x_n=m) \\
&= P(am + u(m) + B_n = m + N) \\
&= P(B_n = (1-a)m - u(m) + N) \\
&= \begin{cases} p & \text{if } (1-a)m-u(m)+N= 1, \\ 1-p & \text{if } (1-a)m-u(m)+N=-1, \\ 0 & \text{if } (1-a)m-u(m)+N \neq \pm 1. \end{cases}
\end{aligned}
$$

(9) is written in terms of u(m) as:

$$(10)\qquad p_{m,m+N} = \begin{cases} p & \text{if } u(m)=(1-a)m+N-1, \\ 1-p & \text{if } u(m)=(1-a)m+N+1, \\ 0 & \text{if } u(m)= \text{otherwise}, \end{cases}$$

where by Assumptions 13 - 14 there exists at least one u in C" for which $(1-a)m + N \pm 1 \in U$ always holds.

By (10) and that X is finite, the number of elements of C" is finite whether U is a finite set or not. Let us denote by $P_u = \{ p^u_{ij} \}$ the transition probability matrix corresponding to each u in C. Let us consider a partition of X[20] as $X = X^P_1 + \cdots + X^P_n + X^N_1 + \cdots + X^N_m$ where X^P_i denotes a positive class and X^N_i a null class respectively. Then compute $\pi^u_j (j \in X)$ and g^u by

$$\pi^u_j \geq 0, \qquad \sum_{j \in X} \pi^u_j = 1, \qquad \pi^u_j = \sum_{i \in X} \pi^u_i \, p^u_{ij}$$

$$(11)\qquad g^u = \begin{cases} \sum\limits_{j \in \bigcup_{k=1} X^P_k} {}_s p^u_{ij} \; g^u_j & \text{if } x_0 = i \in \bigcup\limits_{k=1}^{m} X^N_k \\ g^u_i & \text{if } x_0 = i \in X^P_k (1 \leq k \leq n) \end{cases}$$

where

$$(12)\qquad g^u_i = \sum_{j \in X^P_k} \pi^u_j g(j, u(j)) \quad i \in X^P_k$$

$${}_s p^u_{ij} = \lim_{n \to \infty} p^{u,(n)}_{ij}$$

and where $p^{u,(n)}_{ij}$ is a n-step transition probability. It is easily seen by the ergodic theorem for Markov chains that $g^u = Q_u(x_0)$. Thus by evaluating g^u for each u in C", we can find u* minimizing the average cost in finite steps.

A.E.D.

Remark. ${}_s p^u_{ij} = \pi_j \eta_i$ where $\eta_i - \sum_{v \in \bigcup_{k=1}^{m} X^N_k} p_{iv} \eta_v = \sum_{v \in \bigcup_{k=1} X^P_k} P_{iv}$.

To compute g^u_i $(i \in X)$ above we may use the policy improvement procedure as is stated in:

Theorem 7. g^u_i $(i \in X)$ in Theorem 6 satisfy

---

20) A partition of X depends on u.

(13) $$g_i^u = \sum_{j\in X} p_{ij}^u \quad g_j^u$$
$$v_i^u + g_i^u = (i,u(i)) + \sum_{j\in X} p_{ij}^u \, v_j^u$$

which are solved for v_i^u and g_i^u by setting one of v_i^u 's ($i \in X_k^P$) for each k ($1 \le k \le n$) to zero.

Proof.This theorem follows directly from the arguments found in Chapter 6 of Howard [5] .

Remark.We now have two methods to obtain u*.We can choose one of them that provides us with a less task of calculations.

Let us assume that $P(x_0=i) = p_i$ ($0 \le p_i \le 1$, $\sum_{i\in X} p_i = 1$) is known a priori.Theorem 5 will have the following corollary:

Theorem 8.Let $P(x_0=i)= p_i$, as stated above, be known a priori, and assume that other assumptions of Theorem 6 hold.Then u* in C" which is optimal (with respect to a priori distribution of x_0) over C is obtained in finite steps.

Proof.It suffices to note that g^u is given by

(14) $$g^u = \sum_{i\in\bigcup_{k=1}^{n} X_k^P} p_i \cdot g_i^u + \sum_{i\in\bigcup_{k=1}^{n} X_k^u} p_i \sum_{j\in X} {}_s p_{ij}^u \quad \cdot \quad g_j^u \quad .$$

Q.E.D.

3. Nemerical Example

Suppose that X = (0,1,2,3), U =($-\infty$,··· ,-1,0,1,···,∞), $P(x_0=0)=1$ a=1, and $p=\frac{1}{2}$.The cost is assumed to be:g(0)=2,g(1)=6,g(2)=4 and g(3)=8.Thus in this example the cost is independent of control.As will be seen in the following, Assumptions 13 - 15 are satisfied and an optimal control exists in C".

Using (10) we obtain the following sixteen elements of C".This follows from Assumptions 13 - 14 which imply that P_u for each u in C" is a stochastic matrix,i.e.each u(x) ($x\in X$) may take only values of:u(0)=1,2;u(1)=0,1;u(2)=-1,0;u(3)=-2,-1.Elements of C" are shown in Table 1.P_u for each u in C" are given in Appendix.

Table 1.

x \ C''	u^1	u^2	u^3	u^4	u^5	u^6	u^7	u^8	u^9	u^{10}	u^{11}	u^{12}	u^{13}	u^{14}	u^{15}	u^{16}
0	1	2	1	2	1	1	2	2	1	1	2	2	1	1	2	2
1	0	0	1	1	0	1	0	1	0	0	0	0	1	1	1	1
2	-1	-1	-1	-1	0	0	0	0	-1	0	-1	0	-1	0	-1	0
3	-2	-2	-2	-2	-2	-2	-2	-2	-1	-1	-1	-1	-1	-1	-1	-1

Next we calculate $\{\pi_j^u, j \in X\}$. They are given in Table 2.

Table 2.

π_j^u \ C''	u^1	u^2	u^3	u^4	u^5	u^6	u^7	u^8	u^9	u^{10}	u^{11}	u^{12}	u^{13}	u^{14}	u^{15}	u^{16}
π_0^u	1/2	1/3	1/2	1/4	1/3	1/4	1/4	1/6	1/2	1/4	2/7	1/6	1/2	0	0	0
π_1^u	0	1/6	0	1/4	1/6	1/4	1/4	1/3	0	1/4	3/14	1/3	0	1/2	1/2	1/2
π_2^u	1/2	1/3	1/2	1/4	1/3	1/4	1/4	1/6	1/2	1/4	5/14	1/6	1/2	0	0	0
π_3^u	0	1/6	0	1/4	1/6	1/4	1/4	1/3	0	1/4	1/7	1/3	0	1/2	1/2	1/2

Then compute g_i^u for each u in C" and for each i in positive subclasses of X. For example, for u^1, X is decomposed as $X = X_1^P + X_1^N + X_2^N$ where $X_1^P = \{0,2\}$, $X_1^N = \{1\}$, and $X_2^N = \{3\}$. Thus the initial state 0 is in a positive subclass and $g^{u^1} = 3$. by (12). For u^{16}, X is decomposed as $X = X_1^P + X_1^N + X_2^N$ where $X_1^P = \{1,3\}$, $X_1^N = \{0\}$, $X_2^N = \{2\}$. Thus in this case x_0 is in a null subclass, so we need ${}_5p_{01}^{u^{16}} = \frac{1}{2}$, and ${}_5p_{03}^{u^{16}} = \frac{1}{2}$. Then since $g_1^{u^{16}} = g_3^{u^{16}} = 7$, we obtain $g^{u^{16}} = \frac{1}{2}(7+7) = 7$. In this way we can compute all g^u for u in C". They are given in Table 3.

Table 3.

$u \in C''$	u^1	u^2	u^3	u^4	u^5	u^6	u^7	u^8	u^9	u^{10}	u^{11}	u^{12}	u^{13}	u^{14}	u^{15}	u^{16}
$\sum_{j \in X} \pi_j^u q(j)$	3	4 1/3	3	5	4 1/3	5	5	5 2/3	3	5	4 3/7	5 2/3	3	7	7	7

From Table 3.it is seen that an optimal control is given by either u^1, u^3, u^9, or u^{13}.If one of these control laws is used, the expected average cost incurred by the process will be 3, being not larger than any other control law in C.

To obtain g_i^u ($u \in C''$), we may use the policy improvement procedure shown in Theorem 7 instead of calculating $\pi_j^u, j \in X$.For u^5 and u^{16} this method is illustrated below:

For u^5,(13) becomes;

$$g^{u5} = g_j^{u5} = g_i^{u5} \quad (i,j \in X), \qquad \begin{aligned} v_0^{u5} + g^{u5} &= 2 + \tfrac{1}{2}(v_0^{u5} + v_2^{u5}) \\ v_1^{u5} + g^{u5} &= 6 + \tfrac{1}{2}(v_0^{u5} + v_2^{u5}) \\ v_2^{u5} + g^{u5} &= 4 + \tfrac{1}{2}(v_1^{u5} + v_3^{u5}) \\ v_3^{u5} + g^{u5} &= 8 + \tfrac{1}{2}(v_0^{u5} + v_2^{u5}). \end{aligned}$$

Putting $v_3^{u5} = 0$, we obtain $g_i^{u5} = 4\frac{1}{3}$ for all i in X.

For u^{16} ,(13) becomes;

$$\begin{aligned} g_0^{u16} &= \tfrac{1}{2}(g \quad + g \quad) \\ g_1^{u16} &= \tfrac{1}{2}(\quad '' \quad) \\ g_2^{u16} &= \tfrac{1}{2}(\quad '' \quad) \\ g_3^{u16} &= \tfrac{1}{2}(\quad '' \quad) \end{aligned} \qquad \begin{aligned} v_0^{u16} + g_0^{u16} &= 2 + \tfrac{1}{2}(v_1^{u16} + v_3^{u16}) \\ v_1^{u16} + g_1^{u16} &= 6 + \tfrac{1}{2}(\quad '' \quad) \\ v_2^{u16} + g_2^{u16} &= 4 + \tfrac{1}{2}(\quad '' \quad) \\ v_3^{u16} + g_3^{u16} &= 8 + \tfrac{1}{2}(\quad '' \quad) \end{aligned}$$

Putting $v_3^{u16} = 0$, we obtain $g_1^{u16} = g_3^{u16} = 7$.

In this way g_i^u can be obtained by either method for all u in C".

Note that g_i^u in the same positive class coincide for all i, and that π_i^u for a null class are zero.

As is found in this example, the optimal control is not always unique.

A simulation study for some of u's in C". is given in Figure 1.A sequence $\{B_n(\omega)\}$ used for this case is given in Table 4.

Table 4.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$B_n(\omega)$	1	1	-1	1	-1	-1	1	1	1	-1	1	-1	1	1	-1

n	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
$B_n(\omega)$	1	-1	-1	-1	1	1	-1	-1	1	-1	1	-1	1	-1	-1

4. Appendix

4-1. Existance and Uniqueness of a Solution to (3)

The existence and uniqueness of a solution ot equation (3)

$$(*)\quad g(0,x(0)) = 0$$

$$g(t+\Delta t,x(t+\Delta t))=\inf_{u\in C(t)}\left\{ k(x(t),u)\Delta t +\int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\, g(t,y)\right\}$$

for sufficiently small $\Delta t>0$ and $t\in[0,\infty)$ is shown as follows. Note that ω is fixed and $g_{t+\Delta t}(x(t+\Delta t,\omega))$ in (3) is denoted by $g(t+\Delta t,x(t+\Delta t))$ here, and that $g(t+\Delta t,x(t+\Delta t))$ is defined on $[0,\infty)\times\Xi(t)$.

Let us define $g_n(t+\Delta t,x(t+\Delta t))$ for integers $n\geq 0$ by:

$$g_0(t,x) = 0 \quad (t\in[0,\infty),x\in X)$$

$$g_{n+1}(t+\Delta t,x(t+\Delta t)) = \inf_{u\in C(t)}\left\{k(x(t),u)\Delta t +\int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\, g_n(t,y)\right\}$$

Then for $n\geq 1$,

$$\begin{aligned}
&|g_{n+1}(t+\Delta t,x(t+\Delta t))-g_n(t+\Delta t,x(t+\Delta t))| \\
&= \Big|\inf_{u\in C(t)}\{ k(x(t),u)\Delta t +\int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)g_n(t,y)\} - \\
&\qquad -\inf_{u\in C(t)}\{k(x(t),u)\Delta t +\int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)g_{n-1}(t,y)\}\Big| \\
&\leq \sup_{u\in C(t)}\int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\,|g_n(t,y) - g_{n-1}(t,y)| \\
&\leq \sup_{x(t)\in\Xi(t)} |g_n(t,x(t)) - g_{n-1}(t,x(t))|
\end{aligned}$$

where an inequality $\inf_u (a_u + b_u) - \inf_u (a_u + c_u) \leq \sup_u (b_u - c_u)$ has been used, and

$$|g_1(t+\Delta t,x(t+\Delta t))-g_0(t+\Delta t,x(t+\Delta t))|$$

$$= \Big|\inf_{u\in C(t)}\{k(x(t),u)\Delta t +\int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)g_0(t,y)\} - \inf_{u\in C(t)} k(x(t),u)\Delta t\Big|$$

$$\leq \sup_{u\in C(t)} \int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\ |g_0(t,y)|$$

$$\leq \sup_{x(t)\in \Xi(t)} |g_0(t,x(t))|.$$

Define $v_n(c)$ for any c by

$$v_n(c) = \sup_{\substack{t+\Delta t\leq c \\ x(t+\Delta t)\in \Xi(t+\Delta t)}} |g_{n+1}(t+\Delta t,x(t+\Delta t)) - g_n(t+\Delta t,x(t+\Delta t))|, \quad (c>0)$$

$$v_n(c)=0 \quad \text{if } c<0.$$

Then

$$\begin{aligned} v_n(c) &\leq \sup_{t+\Delta t\leq c,\, x(t)\in\Xi(t)} |g_n(t,x(t)) - g_{n-1}(t,x(t))| \\ &= \sup_{t\leq c-\Delta t,\, x(t)\in\Xi(t)} |g_n(t,x(t)) - g_{n-1}(t,x(t))| \\ &= v_{n-1}(c-\Delta t), \end{aligned}$$

hence $v_{n+1}(c)\leq v_n(c-\Delta t)\leq v_{n-1}(c-2\Delta t)\leq \cdots \leq v_0(c-n\Delta t)$. Since $v.(\cdot)\geq 0$ and $\Delta t>0$, it follows that $\lim_{n\to\infty} v_n(c)=0$, meaning that $g_{n+1}(t+\Delta t,x(t+\Delta t))$ converges to some $g(t+\Delta t,x(t+\Delta t))$ uniformly in t for $t+\Delta t\leq c$.

That $g(t+\Delta t,x(t+\Delta t))$ is a solution of (*) is shown as follows:

$$|g(t+\Delta t,x(t+\Delta t)) - \inf_{u\in C(t)} \{k(x(t),u)\Delta t + \int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\, g(t,y)\}|$$

$$\begin{aligned} = |g(t+\Delta t,x(t+\Delta t)) &- g_{m+1}(t+\Delta t,x(t+\Delta t)) + \\ &+ \inf_{u\in C(t)} \{k(x(t),u)\Delta t + \int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\, g_m(t,y)\} - \\ &- \inf_{u\in C(t)} \{k(x(t),u)\Delta t + \int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\, g\ (t,y)\}| \end{aligned}$$

$$\leq |g - g_{m+1}| + |\inf\{\ \} - \inf\{\}|$$

$$\leq |g(t+\Delta t,x(t+\Delta t)) - g_{m+1}(t+\Delta t,x(t+\Delta t))| - \\ - \sup_{x(t)\in\Xi(t)} |g_m(t,x(t)) - g(t.x(t))|$$

$$\longrightarrow 0 \quad \text{as } m\to\infty.$$

Now by definition of $\{g_n(t+\Delta t,x(t+\Delta t)),\ n\geq 0\}$, $g_n(0,x(0))=0$, it follows that $g(0,x(0))=0$. Since

$$\begin{aligned} g(\Delta t,x(\Delta t)) &= \inf_{u\in C(t)} \{k(x(0),u)\Delta t + \int_{\Xi(t)} P_u(0,x;t,dy)\, g(0,y)\} \\ &= \inf_{u\in C(t)} \{k(x(0),u)\Delta t\}, \end{aligned}$$

letting $\Delta t\to 0$, it follows that $g(0,x(0))=0$, proving the continuity of $g(t+\Delta t,x(t+\Delta t))$ at $t+\Delta t=0$.

To prove the uniqueness let $g(t+\Delta t, x(t+\Delta t))$ and $G(t+\Delta t, x(t+\Delta t))$ be two solutions of (*) continuous at $t+\Delta t=0$.

Let $v(c)$ be

$$v(c)=\sup_{t+\Delta t\leq c,\, x(t+\Delta t)\in\Xi(t+\Delta t)} |g(t+\Delta t, x(t+\Delta t))-G(t+\Delta t, x(t+\Delta t))| \qquad (c>0)$$
$$v(c)=0,\ (c\leq 0).$$

Then in a similar way as above, we obtain that

$$\begin{aligned} v(c) &\leq \sup_{\substack{t+\Delta t\leq c\\ x(t+\Delta t)\in\Xi(t+\Delta t)}} \sup_{u\in C(t)} \int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\,|(g(t,y)-G(t,y))| \\ &\leq \sup_{t\leq c-\Delta t,\, x(t)\in\Xi(t)} |g(t,x(t))-G(t,x(t))| \\ &= v(c-\Delta t). \end{aligned}$$

Thus it follows that $v(c)\leq v(c-\Delta t)\leq \cdots \leq v(c-n\Delta t)$. Since $g(t+\Delta t, x(t+\Delta t))$ and $G(t+\Delta t, x(t+\Delta t))$ are confinuous at $t+\Delta t=0$, $\lim_{n\to\infty} v(c-n\Delta t)=0$, and $v(c)=0$ identically, proving the assertion.

Now let us consider the equation (see the foot nate 12)

$$(**)\quad \hat{g}(0,x(0))=0$$
$$\hat{g}(t+\Delta t, x(t+\Delta t)) = \inf_{u\in C(t)} \Big\{ k(x(t),u)\Delta t + \int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\,\hat{g}(t,y) + o(\Delta t)\Big\}.$$

It is clear from the above argument that there exists (a.a.ω) a unique solution $\hat{g}(t+\Delta t, x(t+\Delta t))$ of (**) uniformly in $t\in[0,\infty)$. It follows that for any $t\in[0,\infty)$ and a.a.ω,

$$\begin{aligned} &|\hat{g}(t+\Delta t, x(t+\Delta t)) - g(t+\Delta t, x(t+\Delta t))| \\ &\leq \sup_{u\in C(t)} \int_{\Xi(t)} P_u(t,x;t+\Delta t,dy)\,|\hat{g}(t,x(t))-g(t,x(t))| + |o(\Delta t)| \\ &\leq \sup_{x(t)\in\Xi(t)} |\hat{g}(t,x(t))-g(t,x(t))| + |o(\Delta t)| \\ &\leq \quad \cdots\cdots\cdots \\ &\leq |o(\Delta t)|. \end{aligned}$$

Thus $\displaystyle\lim_{t\to\infty} \frac{|\hat{g}(t+\Delta t, x(t+\Delta t))-g(t+\Delta t, x(t+\Delta t))|}{t+\Delta t} = 0$ (a.a.ω)

or $\lim\limits_{t \to \infty} \dfrac{\hat{g}(t + \Delta t, x(t + \Delta t))}{t + \Delta t} = g$ a.a. ω. It has been shown that we need to consider only (*) instead of (**).

4-2. Measurability of $g(t + \Delta t, x(t + \Delta t, \omega))$ with respect to ω

There may occur a case where $g(t + \Delta t, x(t + \Delta t, \omega))$ is not measurable in ω. In such cases a proposition containing a term a.a. ω may be false. We can circumvent such cases by introducing $\operatorname{ess\,inf}\limits_{u \in C(t)}$ (essential infimum) instead of $\inf\limits_{u \in C(t)}$ in the right side of (*). Then we can replace C(t) by its countable subset D(t), i.e. $\operatorname{ess\,inf}\limits_{u \in C(t)} = \operatorname{ess\,inf}\limits_{u \in D(t)}$, and $g(t + \Delta t, x(t + \Delta t, \omega))$ will be measurable in ω. No change is necessary in other parts of 4-1 and in the proof of Theorem 1.

4-3. P_u for u in C"

$$u^1 \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix} \quad u^2 \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix} \quad u^3 \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix} \quad u^4 \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix}$$

$$u^5 \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix} \quad u^6 \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix} \quad u^7 \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix} \quad u^8 \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \end{bmatrix}$$

$$u^9 \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \end{bmatrix} \quad u^{10} \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \end{bmatrix} \quad u^{11} \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & \frac12 & 0 \end{bmatrix} \quad u^{12} \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \end{bmatrix}$$

$$u^{13} \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \end{bmatrix} \quad u^{14} \begin{bmatrix} \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \end{bmatrix} \quad u^{15} \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ \frac12 & 0 & \frac12 & 0 \\ 0 & \frac12 & 0 & \frac12 \end{bmatrix} \quad u^{16} \begin{bmatrix} 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \\ 0 & \frac12 & 0 & \frac12 \end{bmatrix}$$

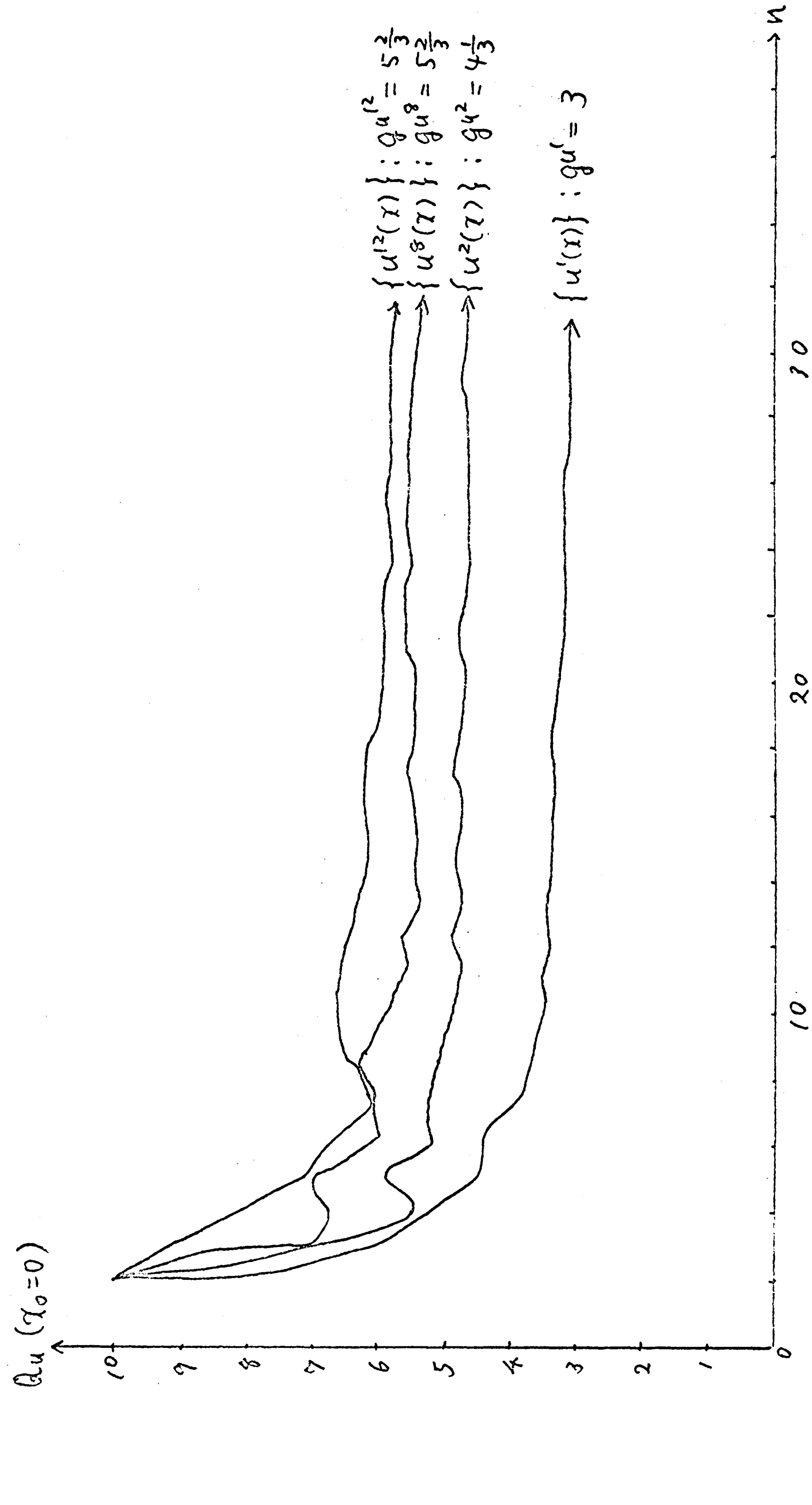
Q_u ($x_0=0$)
10
9
8
7
6
5
4
3
2
1
0
10
20
30
n
$\{u^{12}(x)\}$: $gu^{12} = 5\frac{2}{3}$
$\{u^{8}(x)\}$: $gu^{8} = 5\frac{2}{3}$
$\{u^{2}(x)\}$: $gu^{2} = 4\frac{1}{3}$
$\{u^{1}(x)\}$: $gu^{1} = 3$

Figure 1

References

[1] Chung,K.L., Markov Chains with Stationary Transition Probabilities, Springer (1960).

[2] Дынкин, Е.Б., Марковские Процессы, Москва (1963).

[3] Derman, C., On Sequential Rules and Markov Chains, Management Sci. 9(1963).

[4] Derman, C., Denumerable State Markovian Decision Processes-Average Cost Criterion, Ann.Math.Statist. Vol. 37(1966).

[5] Howard, R.,Dynamic Programming and Markov Processes, Wiley(1960).

[6] Ito,K., Probability Theory (in Japanese), Iwanami(1963).

[7] Kushner,H.J., On the Stochastic Maximum Principle:Fixed Time of Centrol, J.Math.Anal.Appl., 11(1965).

[8] Wonham,W.H., Optimal Control of a Linear System with State-Dependent Noise, SIAM J.Centrol, Vol.5, No.3(1967).

[9] Мандл, П., Об Управлении Необрывающимися Диффузионными Процессами, Теория. Вероят. и её Примен., Том 9 (1964).

[10] Рыков, В.В., Управляемые Марковские Процессы с Конечными Пространствами Состояний и Управлений, Теория. Вероят. и её Примен., Том 11 (1966).

Acknowledgement. I would like to appreciate Professor H.Sakamoto and Dr.H.Yanai of Keio University for their helpful suggestions. I also thank Professor C.Furukawa and Dr.H.Minami of Kyushu University who kindly pointed out the footnotes 3), †) and the reference [10]