数理解析研究所講究録
48巻 1968年 75-86

Non-stochastic forward L.P. and D.P. method in stochastic control.

Hideo Yamagata
(Department of Mathematics
College of Engineering
University of Osaka Prefecture)

§1. Introduction; R. Bellman has created the dynamic programming useful to calculi in many branches. This dynamic programming is also the most powerful constructive method of the numerical treatment in control theory, because it is needed to choose the suitable solution from the results given as the necessary condition for optimality by Pontryagin's maximal principle[2] p.61 etc.. But even dynamic programming is not necessarily effective to the calculation of controls using the digital computer on line with regard to some nonlinear system containing the stochastic terms (see [4] p.1846 in English translation). Namely, since dynamic programming is the backward calculation, the future observations must be used in the calculation of controls for this system. Since these future observations remain in the optimal control for the non-linear stochastic system, this situation happens, and it becomes one of the essential difficulties for the control of the non-linear stochastic system.

In order to calculate a sort of optimal controls for this complicated non-linear system containing the stochastic terms we must give the detailed classification of the mean (i.e. the expectation) used for the construction of the cost function and a sort of extended treatment of dynamic programming. The above difficulty and the requirement for the classification of mean appearing in non-linear system also appears in the linear

stochastic system with bounded control energy in the more simple form. Hence, at the first step we investigate the simple linear stochastic system with bounded control energy which also shows the essential difficulties appearing in [4] p. 1846 in English translation. One of the above linear models with respect to the cost function $\{x(T)\}^2$ (not $M_{\pi,x(0),\xi}\{x(T)\}^2$) shall be shown in §2. Here $x(T)$ is the terminal value of the state. By using this model we will show here the numerical treatment for a sort of optimal control which is attainable by using the future observations without error for the dynamical system omitted the stochastic term, comparing with D.P. problem in [5] p.719 [6] p.228. We call it non-stochastic method. The method for the linear system developed in §2 is the non-stochastic forward L.P. method. It is based on the thought like one appearing in the proof of maximal principle [2] p.100. The non-stochastic forward D.P. method for the non-linear system containing the stochastic terms is shown in § 3.

§2. Division of the control energy and the determination of the sign of controls; We will show here the simple concrete linear system with bounded control energy and with the cost-function $\{x(T)\}^2$, where $x(T)$ is the terminal state. The calculation of control for this system by using the method like one appearing in [3] p.746- p.748 [4] p.1846 (in English translation) is not possible, since the cost function in $\{x(T)\}^2$. We will develop here the non-stochastic forward L.P. method effective to this linear system.

Model 1. Suppose that a, b and c are constants. Our plant is described by the equation $\dot{x} = ax + bu + h(t)$... (1) defined on the interval $[0,T]$.

Here x is one dimensional phase coordinate vector,

u is one dimensional control vector, and h(t) is a one dimensional random perturbation vector with mean zero satisfying $M[h(t)\cdot h^*(\tau)] \equiv W(t)\cdot\delta(t-\tau) \equiv l_1^2\delta(t-\tau)$. The feedback vector for our plant is $y = Cx + \xi(t)$... (2) defined on the interval $[0,T]$, where $\xi(t)$ is a one dimensional random measurement error vector with mean zero satisfying $M[\xi(t)\cdot\xi^*(\tau)] \equiv l_2^2\delta(t-\tau)$. We neglect this $\xi(t)$ in many cases here. Let's take a time sequence $\{t_k\ ;\ k=N-1,\ldots,0\}$ satisfying $0=t_N < t_{N-1} < \cdots\cdots < t_1 < t_0 < T = t_{-1}$ in $[0,T]$. Suppose that $y(t)$ can be observed only at times $t=t_k$; $k=N,\ldots,0$, and that the constant controls u_k in $t \in [t_k, t_{k-1})$; $k=N,\ldots,0$ can be constructed from these y_k; $k=N,\ldots,0$, where $y_k \equiv y(t_k)$ for $k=N,\ldots,0$. Our purpose is to obtain the control vector $\underline{u} = \{u_k;\ k=N,\ldots,0\}$ which optimize (maximize) $M_k L$, $M_{k,x(0)} L$, $M_{x(0)} L$ or L itself constructed from $L = \{x(T)\}^2$ .... (3) under the condition $I = \int_0^T u^2 dt \leq I_0$ ... (4), where $M_k L$ etc. are the mean (expectation) value of L with respect to h etc..

At the first step, we define the following transformation of (1). The elementary solution of (1) (i.e. the solution of $\dot{\Phi}(T,t) = -a\,\Phi(T,t)$ with the initial condition $\Phi(\tau,\tau)=1$) is $\Phi = \exp\{-a(t-T)\}$. By using this Φ, let's construct the function $\rho(t) = x(t)\cdot\exp\{-a(t-T)\}$ satisfying the equation

$$\dot{\rho}(t) = \dot{x}(t)\cdot\exp\{-a(t-T)\} - a\,x(t)\cdot\exp\{-a(t-T)\}$$
$$= b\cdot\exp\{-a(t-T)\}\cdot u + h(t)\cdot\exp\{-a(t-T)\} \quad .....(5).$$

Namely (1) is transformed to (5).

Next let's decompose $\rho(t)$ into the following sum $\rho_1(t)+\rho_2(t)$; $\rho_1(t) \equiv \int_0^t b\exp\{-a(t-T)\}\cdot u(t)dt + K$ ($K \equiv \rho(0)$; constant) and $\rho_2(t)$ is the function satisfying $\dot{\rho}_2(t) = h(t)\exp\{-a(t-T)\}$.

The notations used in [3] p.749 become r_k $(\equiv \int_{t_k}^{t_{k-1}} W(t)^2 \exp\{-2a(t-T)\}dt)$ $\equiv \int_{t_k}^{t_{k-1}} l_1^2 \exp\{-2a(t-T)\}dt = -(l_1^2/(2a))\cdot[\exp\{-2a(t_{k-1}-T)\} - \exp\{-2a(t_k-T)\}]$, $\lambda_k^{-1} \equiv M[\xi_k \xi_k^*] = l_2^2$, $Q(t) \equiv b\exp\{-a(t-T)\}$ and $G(t) \equiv C\exp\{-a(t-T)\}$ in Model 1, and the result obtained in [3] p.746-748 is the following: the controls $u_N^0, \ldots, u_0^0$ satisfying $\sum_{k=N}^{0} u_k^{0\,2}(t_{k-1}-t_k) \leqslant I_0$ which optimize $M_{k,x(0),\xi}\{x(T)\}^2$ are determined in terms of $G_k \equiv G(t_k)$ and $Q_k \equiv Q(t_k)$ by using all observations $y_N, y_{N-1}, \cdots, y_0$. The following three types (determinations) of controls can be considered.

(A) u_k is determined apriori.

(B) $y_N, \cdots, y_k$ (or y_{t_k}) determines u_k.

(C) All $y_N, \cdots, y_0$ (or $y_{t_k}, \cdots, y_0$) are needed for the determination of u_k.

We can easily see in the following that the result in [3] p.746-748 for $M_{k,x(0),\xi}\{x(T)\}^2$ belongs to the case (A).

One of our purposes is to show the difference among the control problems which optimize (maximize) $M_k\{\rho(T)\}^2$, $M_{x(0)}\{\rho(T)\}^2$, $M_{k,x(0)}\{\rho(T)\}^2$ and $\{\rho(T)\}^2$ (i.e. $M_k\{x(T)\}^2$, $M_{x(0)}\{x(T)\}^2$, $M_{k,x(0)}\{x(T)\}^2$ and $\{x(T)\}^2$) under the condition (4).

The control problem which optimize $M_k\{\rho(T)\}^2$ (i.e. $M_k\{x(T)\}^2$) under the condition (4) is equal to the control problem which optimize $\{\rho_1(T)\}^2 = (\sum_{k=N}^{0} u_k \int_{t_k}^{t_{k-1}} b\exp\{-a(t-T)\}dt + K)^2$ under the condition $\sum_{k=N}^{0} u_k^2(t_{k-1}-t_k) \leqslant I_0$ $(K \equiv \rho(0))$, since $M_k\{\rho(T)\}^2 = \{\rho_1(T)\}^2 + \sum_{k=N}^{0} r_k$ holds for the real stochastic function h(t). Next suppose that the distribution of $x(0)$ is normal with mean zero. If $M_{k,x(0)}\{x(T)\}^2$ is treated instead of $M_k\{x(T)\}^2$, $(\sum_{k=N}^{0} u_k \int_{t_k}^{t_{k-1}} b\exp\{-a(t-T)\}dt)^2$ must be used instead of $(\sum_{k=N}^{0} u_k \int_{t_k}^{t_{k-1}} b\exp\{-a(t-T)\}dt + K)^2$, since $M_{k,x(0)}(\sum_{k=N}^{0} u_k \int_{t_k}^{t_{k-1}} b\exp\{-a(t-T)\}dt + K)^2 = (\sum_{k=N}^{0} u_k \int_{t_k}^{t_{k-1}} b\exp\{-a(t-T)\}dt)^2 + M_{x(0)}(K^2)$

holds. Hence we can determine apriori the controls $\tilde{u}_k$ which optimize $M_{k,x(0)}\{x(T)\}^2$ and $M_k\{x(T)\}^2$. Furthermore, we see from the above argument that M_ξ is independent of our problem relating to M_k. $\rho_1(t) \equiv \int_0^t b \exp\{-a(t-T)\}\cdot u(t)dt$ must be also used for the optimal problem concerning to the cost function $M_{x(0)}\{x(T)\}^2$ instead of $\int_0^t b \exp\{-a(t-T)\}\cdot u(t)dt+K$, since $M_{x(0)}\{K\}\neq 0$. Difficulties of the control's determination appearing in the case (C) appears in the optimal problem with regard to Model 1 concerning to L itself. But the use of the observations $\{y_k; k=N,\ldots,0\}$ and the numerical treatment of controls using the digital computer on line seems to be sufficiently powerful, and it seems to be enough to optimize L itself instead of M_k L or $M_{k,x(0)}$ L in a sense. The above arguments bring to the following conclusion.

Theorem 1. (a) The control problem which optimize $M_{k,x(0)}\{x(T)\}^2$ is the special case of the control problem which optimize $M_k\{x(T)\}^2$, and they belong to the case (A).

(b) The control problem which optimize $M_{x(0)}\{x(T)\}^2$ is the special case of the control problem which optimize $\{x(T)\}^2$, and they belong to the case (C).

(c) M_ξ is independent of the problem appearing in (a).

Next, let's construct the amended controls $\tilde{u} \equiv \{\tilde{u}_k; k=N,\cdots,0\}$ which optimize $L \equiv \{x(T)\}^2$ itself by using the already determined controls $u \equiv \{u_k; k=N,\cdots,0\}$. Suppose that the error of the observation can be neglected (i.e. $\xi(t)\equiv 0$) and that $y_k \equiv y(t_k)$ $(k=N,\ldots,0)$ are given. Then $x_k \equiv x(t_k)$ becomes y_k/c, $\rho_k \equiv \rho(t_k)$ becomes $x_k \exp\{-a(t_k-T)\} = y_k \exp\{-a(t_k-T)\}/c$, and

$\rho(t_k)-\rho(t_{k-1})$ becomes $x(t_k)\cdot\exp\{-a(t_k-T)\}-x(t_{k-1})\cdot\exp\{-a(t_{k-1}-T)\}$ $=[y(t_k)\cdot\exp\{-a(t_k-T)\}-y(t_{k-1})\cdot\exp\{-a(t_{k-1}-T)\}]/C$.

Since $\rho(t_k)-\rho(t_{k-1})=-u_k\int_{t_k}^{t_{k-1}}b\exp\{-a(t-T)\}dt-\int_{t_k}^{t_{k-1}}h(t)\exp\{-a(t-T)\}dt$ ..... (6), then $\int_{t_k}^{t_{k-1}}h(t)\exp\{-a(t-T)\}dt=-[y(t_k)\cdot\exp\{-a(t_k-T)\}$ $-y(t_{k-1})\cdot\exp\{-a(t_{k-1}-T)\}]/C-u_k\int_{t_k}^{t_{k-1}}b\exp\{-a(t-T)\}dt$ holds from (6) etc.. $\tilde{u}_N,\ldots,\tilde{u}_0$ becomes the controls which optimize $\left(\sum_{k=N}^{0}\tilde{u}_k\int_{t_k}^{t_{k-1}}b\exp\{-a(t-T)\}dt+K-\sum_{k=N}^{0}\left[\{y(t_k)\cdot\exp\{-a(t_k-T)\}\right.\right.$ $\left.\left.-y(t_{k-1})\cdot\exp\{-a(t_{k-1}-T)\}\}/C+u_k\int_{t_k}^{t_{k-1}}b\exp\{-a(t-T)\}dt\right]\right)^2$ ..... (7) under the condition $\sum_{k=N}^{0}\tilde{u}_k^2(t_{k-1}-t_k)\leqslant I_0$, where K is the initial value $\rho(0)\equiv\{y_N\cdot\exp(aT)\}/C$.

By the transforms $v_k\equiv\sqrt{(t_{k-1}-t_k)/I_0}\cdot\tilde{u}_k$, $\int_{t_k}^{t_{k-1}}b\exp\{-a(t-T)\}dt$ $\equiv\sqrt{(t_{k-1}-t_k)/I_0}\cdot A_k$ and $\tilde{C}\equiv K-\sum_{k=N}^{0}\left[\{y(t_k)\cdot\exp\{-a(t_k-T)\}\right.$ $\left.-y(t_{k-1})\cdot\exp\{-a(t_{k-1}-T)\}\}/C+u_k\int_{t_k}^{t_{k-1}}b\exp\{-a(t-T)\}dt\right]$, the last equation (7) $\left(\sum_{k=N}^{0}A_kv_k+\tilde{C}\right)^2$ ... (8). (8) must be optimized under the condition $\sum_{k=N}^{0}v_k^2\leqslant 1$ .... (9). $\tilde{\underline{v}}\equiv\{\tilde{v}_k;k=N,\ldots,0\}$ which maximize (8) under (9) must satisfy $\tilde{v}_0:\tilde{v}_1:\cdots:\tilde{v}_N\equiv A_0:A_1:\cdots:A_N$. The signs of $v_0, v_1,\ldots,v_N$ which attain $\underset{u=\{u_k\}}{\mathrm{Max}}\{x(T)\}^2$ can be determined by the sign of $\tilde{C}$ which is determined by all y_k and u_k for k=N,...,0. Namely we obtain the following

Theorem 2. Suppose that the vector $\{v_k^-\}\equiv\{-A_k/\sqrt{A_0^2+\cdots+A_N^2}\}$ is denoted by $\underline{v}^-$ and the vector $\{v_k^+\}\equiv\{A_k/\sqrt{A_0^2+\cdots+A_N^2}\}$ is denoted by $\underline{v}^+$. (a) If $\tilde{C}>0$, $\tilde{\underline{v}}=\underline{v}^+$ (i.e. $\tilde{u}_k=\sqrt{I_0/(t_{k-1}-t_k)}\,v_k^+$) maximizes $\{\rho(T)\}^2$ and $\{x(T)\}^2$.

(b) If $\tilde{C}<0$, $\tilde{\underline{v}}=\underline{v}^-$ (i.e. $\tilde{u}_k=\sqrt{I_0/(t_{k-1}-t_k)}\,v_k^-$) maximizes $\{\rho(T)\}^2$ and $\{x(T)\}^2$.

The controls which attain the $\underset{u}{\mathrm{Max}}\,M_{k,x(0)}\{x(T)\}^2$ (or $\underset{u}{\mathrm{Max}}\,M_k\{x(T)\}^2$) can be obtained by the method used in Theorem

2 apriori (Case (A)). It is not the amendment of $\{u_k; k=N,\ldots,0\}$ like Theorem 2. Considering the above Theorem 2 the true optimal control is defined as follows.

Definition 1. After the all observations $y(t_k)$ $(k=N,\ldots,0)$ and all controls $u_k(k=N,\ldots,0)$ have known, we see that if the controls were $\{\tilde{u}_k ; k=N,\ldots,0\}$, the cost function becomes optimum. These controls $\tilde{u} \equiv \{\tilde{u}_k ; k=N,\cdots,0\}$ are called the true optimal control.

$\tilde{u}$ is the amendment of $\{u_k ; k=N,\ldots,0\}$ in a sense in Model 1. It is not necessarily possible but desirable that $\tilde{u}_k$ is obtained from $y_m(m=N,\ldots,k)$ and from the formula of the system in $0 \leqslant t \leqslant T$.

Let's construct a non-stochastic equation omitting the stochastic terms from the stochastic plant equation. Suppose that $y_m(m=N,\ldots,k)$ are the observations at $t=t_m$, and that $\tilde{y}_m(m=k-1,\ldots,0)$ are the estimates of observations at $t=t_m$ for the solution of its non-stochastic equation.

Definition 2. When the control $\bar{u}_k$ is determined by using the true observations $y_m(m=N,\ldots,k)$ and the estimates of observations $\bar{y}_m(m=k-1,\ldots,0)$, we say that $\bar{u}_k$ is determined by the non-stochastic method. If linear programming is the method which determines $\bar{u}_k$, it is called non-stochastic forward L.P. method. If dynamic programming is the method which determines $\bar{u}_k$, it is called non-stochastic forward D.P. method.

If the controls which maximize $L \equiv \{x(T)\}^2$ itself by the meaning of Def. 2 are required, non-stochastic forward I.P. method effective for this purpose becomes the following one

for our Model 1. Namely, the optimal control becomes
$\{\bar{u}_k; k=N,\dots,0\}$ (by $\bar{u}_k \equiv u_k$) satisfying $\max_{u_m;\, k\geq m\geq 0} \left(\sum_{m=k}^{0} u_m \int_{t_m}^{t_{m-1}} b\exp\{-a(t-\tau)\}dt + \rho(t_k)\right)^2$
$\equiv \max_{u_m;\, k\geq m\geq 0} \left(\sum_{m=k}^{0} u_m \int_{t_m}^{t_{m-1}} b\exp\{-a(t-\tau)\}dt + y(t_k)\cdot\exp\{-a(t_k-\tau)\}/c\right)^2$
under the condition $\sum_{m=k}^{0} u_m^2(t_{m-1}-t_m) \leq I_0 - \sum_{m=N}^{k+1} \bar{u}_m^2(t_{m-1}-t_m)$.

Theorem 3. (a) The optimal controls $\{\bar{u}_k; k=N,\dots,0\}$ (in Model 1) calculated by the non-stochastic forward L.P. method satisfy the condition $\sum_{m=k}^{0} \bar{u}_m^2(t_{m-1}-t_m) \leq \left(\sum_{m=k}^{0} A_m^2 / \sum_{m=N}^{0} A_m^2\right)\cdot I_0$ (i.e. $\bar{u}_k^2 \leq A_k^2 / \sum_{m=N}^{0} A_m^2 \cdot I_0/(t_{k-1}-t_k)$).

(b) The optimal (maximal) control $\bar{u}_k$ becomes the following; if $y_k/c > 0$, $u_k \equiv \sqrt{I_0/(t_{k-1}-t_k)} \times -A_k/\sqrt{\sum_{m=N}^{0} A_m^2}$ holds, and if $y_k/c < 0$, $u_k \equiv \sqrt{I_0/(t_{k-1}-t_k)} \times A_k/\sqrt{\sum_{m=N}^{0} A_m^2}$ holds.

$\{\bar{u}_k;\ k=N,\dots,0\}$ is not necessarily the controls which maximize L by the meaning of Def. 1, but an attainable controls which maximize L in a sense. The method in [3] p.746-p.748 under the conditions $u_k^2 \leq A_k^2 / \sum_{m=N}^{0} A_m^2 \cdot I_0/(t_{k-1}-t_k)$ does not attain $\max_u L$ but $\max_u M_{k,x(0),\xi} L$. It is suitable to obtain the results considering M_ξ, but it cannot determine the controls which attain $\max_u L$ in a sense. It seems to be due to that the method in [3] p.746-p.748 does not use the expectation of the future observation.

At last, we will show the relation between D.P. problem in [5] p.719 [6] p.228 and the non-stochastic method in order to show the theoretical standpoint of our non-stochastic method. Here we use the same notations as one used in [6] p.226. Namely suppose that S describes the state's space, A describes the action's space, and Π_n describes the policy which determine $a \in A$ from $S_1 a_1 S_2 a_2 \cdots S_n$. The purpose of D.P. problem is to

optimize the sum of reward functions $\sum_n \gamma(S_{n-1}, a_n, S_n)$. These S, A, Π_n and $\sum_n \gamma(S_{n-1}, a_n, S_n)$ correspond to the following one in our non-stochastic method. Namely $s(\in S)$ corresponds to the solution of $\dot{x} = ax + bu$ with the initial condition $x(t_k) \equiv x_k$, $a(\in A)$ corresponds to u_k under the condition $u_k^2 \leq A_k^2 / \sum_{m=N}^{0} A_m^2 \cdot I_0/(t_{k-1} - t_k)$, and Π_n corresponds to the determination of u_k from $x(t_k)$ obtained through $y = Cx + \xi(t)$ (in our calculation $\xi(t) \equiv 0$). But $\sum_n \gamma(S_{n-1}, a_n, S_n)$ takes the more complicated form. The use of the complicated S and Π_n (but more exact one for our system) gives the more detail results (like max L in a sense) than the usual D.P. method (used in [3] p.746-p.748). Our non-stochastic method has more wide application's domain. Namely it is also effective to the non-linear system containing the stochastic term. It seems to be effective for this system under the partial observation. Even the consideration of $\xi(t)$ seems to be possible.

§3. The calculation of controls for the non-linear system containing stochastic terms;

Here, by using non-stochastic forward D.P. method, let's calculate the controls $\{u_k; k=N,\ldots,0\}$ of the following system which maximize $\{x(T)\}^2$ by the meaning of Def.2.

Model 2. $\begin{cases} \dot{x} = ax(1+kx) + bu + h(T) & \text{for } |u| \leq 1 \\ y = Cx + \xi(t) \end{cases}$

The Bellman's equation for the $\underset{|u|\leq 1}{\mathrm{Max}} \{x(T;p)\}^2$ is obtained by the following procedures.

Namely, since $f(p,T) = f(p + \{ap(1+kp)+bu\}\cdot\Delta t,\ T-\Delta t) + O(\Delta t^2)$ hold, $\partial f(p,T)/\partial T = \underset{|u|\leq 1}{\mathrm{Max}} [\partial f(p,T)/\partial p \{ap(1+kp)+bu\}] = \partial f(p,T)/\partial p \cdot ap(1+kp) + b|\partial f(p,T)/\partial p|$

is obtained from

$$\max_{|u|\le 1}\left[f(p,T)+\partial f(p,T)/\partial p\cdot\{ap(1+kp)+bu\}\Delta t-\partial f(p,T)/\partial T\cdot\Delta t+O(\Delta t^2)\right]\equiv f(p,T),$$

where $f(p,T)\equiv\max_{|u|\le 1}\{x(T;p)\}^2$. Since $f(p,0)=p^2$ holds for $T=0$, p^2 becomes the initial condition for this equation. The equation $\partial f(p,T)/\partial T-\{ap(1+kp)\pm b\}\partial f(p,T)/\partial p=0$ is solved by using the solution of $dT/1=dp/-\{ap(1+kp)\pm b\}=df/0$. Namely f becomes constant on the curves $T=\int\frac{dp}{-ak\{p^2+p/k\pm b/(ak)\}}$. The graph of

$$\frac{dp}{dt}=-ak\{(p+1/(2k))^2-1/(4k^2)\pm b/(ak)\}$$

$$=-ak(p+1/(2k))^2+a/(4k)\mp b$$

is shown in the following Fig. 1. The graph in Fig.1 drawn by the connected line is one useful to $f(p,T)\equiv\max_{|u|\le 1}\{x(T;p)\}^2$.

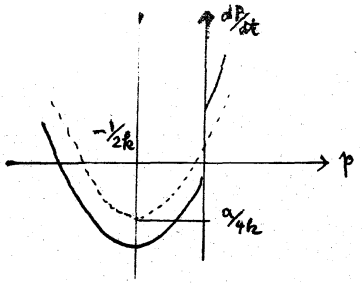

Fig. 1.

We can easily see from Fig.1 that $u=-\text{sign } p\cdot\text{sign } b$ holds. Since the sign of p is obtained by the sign of $x=y/c$ for the observation without error (i.e. $\xi(t)\equiv 0$), the control u which optimize $\{x(T;p)\}^2$ itself by the meaning of non-stochastic forward D.P. method can be determined easily by the above relations.

The method obtained in [4] p.1846 which optimize $M_{k,x(0),\xi}L$ can be also applied by using the expected controls u_k and the expected observations y_k obtained by the relations $y_{m-1}=c\{\tilde{x}_m+[dp/dt]_{p=\tilde{x}_m}(t_{m-1}-t_m)\}\equiv c\tilde{x}_{m-1}$ and $u_m=-\text{sign }\tilde{x}_{m-1}\cdot\text{sign } b$ for $m=k,\dots,0$, where $\tilde{x}_k\equiv y_{k-1}/c$.

At last, we will show the following problems derived from our results.

1. How much differ $\max_u\{x(T)\}^2$ obtained by non-stochastic forward D.P. (or L.P.) method from $\max_u\{x(T)\}^2$ by the meaning of

true maximum?

A sort of experiment seems to be needed for non-linear stochastic system to solve this problem.

2. What is the calculation for the system with the higher dimensional phase coordinate vector etc.?

Perhaps it will be similar as one for our one dimensional Model, to some extend.

3. Can we define the (stochastic) forward D.P. (or L.P.) method which more exact results than non-stochastic forward D.P. (or L.P.) method?

Perhaps it will be possible. But, sometimes, it needs the complicated calculations. If new method is the non-stochastic method, they will not be unified one.

We will try to solve these problems in the next chance.

References

[1] R.Bellman; Dynamic Programming. Princeton Univ. Press (1957).

[2] L.S.Pontryagin etc.; The Mathematical Theory of Optimal Processes. John Wiley & Sons, Inc., New York (1963).

[3] I.A.Boguslavskii; Statistically optimal control of the final state. Automation and Remote Control. Vol.27. No.5 pp. 15-27 (pp.745-757 in English translation)(1966).

[4] ———————; On the equations of stochastic control. ibid. Vol.27 No. 11 pp.9-18 (pp. 1845-1853 in English translation) (1966).

[5] D.Blackwell; Discrete dynamic programming Ann. Math. Statist Vol. 33 pp. 719-726 (1962).

[6] ———————; Discounted dynamic programming. Ann. Math. Statist. Vol. 36 pp. 226-235 (1965).

Acknowledgement; The contents of this paper are the developed results of a part of my lecture (named "On the equation in stochastic control") given at the meeting on linear and non-linear control theory held at Research Institute for Mathematical Science Kyoto University on 26, March 1968. The part of the above lecture does not overlap to one given by Dr. Y. Sunahara at the same meeting. Furthermore, the most part of this paper was lectured, on 25, June 1968, at the meeting in Osaka for bidding Prof. Dr. H.J. Kushner welcome.