U. S. AIR FORCE

PROJECT RAND

RESEARCH MEMORANDUM

DYNAMIC PROGRAMMING AND STOCHASTIC CONTROL PROCESSES

By

Richard Bellman

RM-1904

ASTIA Document Number AD 150651

May 10, 1957

Assigned to ______________________

This is a working paper. It may be expanded, modified, or withdrawn at any time. The views, conclusions, and recommendations expressed herein do not necessarily reflect the official views or policies of the United States Air Force.

The RAND Corporation
1700 MAIN ST. • SANTA MONICA • CALIFORNIA

12

SUMMARY

A fundamental problem in control processes is that of determining optimal feedback control in order to neutralize the effect of "noise"—i.e., of random disturbance. One version of this problem is studied here in discrete form.

Consider a system S specified at any time t by a finite-dimensional vector $x(t)$ satisfying a vector differential equation $dx/dt = f(x,r(t),v(t))$, $x(0) = c$, where c is the initial state, $r(t)$ is a random forcing term possessing a known distribution, and $v(t)$ is a forcing term either chosen, via a feedback process, so as to minimize the expected value of a functional $J(x) = \int_0^T h(x - y,t)dG(t)$, where $y(t)$ is a known function, or chosen so as to minimize the functional defined by the probability that $\max_{0 \le t \le T} h(x - y,t)$ exceed a specified bound.

It is shown how the functional-equation technique of dynamic programming may be used to obtain a new computational and analytic approach to problems of this genre. The memory-capacity constraint of present-day digital computers restricts the successful application of these techniques to first- and second-order systems at the moment, with limited application to higher-order systems.

CONTENTS

DYNAMIC PROGRAMMING AND STOCHASTIC CONTROL PROCESSES

1. INTRODUCTION

In a previous paper [3], we have discussed the application of the theory of dynamic programming [1], [2] to the study of some classes of control processes of deterministic type. In this paper, we indicate the application of these techniques to the computational solution of some stochastic control processes.

A fundamental problem is that of determining optimal feedback control in order to neutralize the effect of "noise"—i.e., of random disturbances. One version of this ubiquitous problem is the following. A physical system S at any time t is specified by a finite-dimensional vector x(t), determined as a function of time by means of the differential equation

$$(1) \qquad \frac{dx}{dt} = g(x) + r(t), \quad x(0) = c,$$

where in many significant cases the function g(x) is not linear. The function r(t) is a random function whose distribution is taken to be known. Problems in which the distribution of r(t) is only partially known and must be determined as time goes on belong to a more difficult domain that will not be entered here. To counteract the influence of r(t), we introduce "feedback control" in the form of a forcing function v(t), so that the equation has the form

$$(2) \qquad \frac{dx}{dt} = g(x) + r(t) + v(t), \quad x(0) = c.$$

In using the term "feedback control," we mean that at each instant of time $v(t)$ can be chosen to depend on the state of the system at that time. An alternative procedure consists of choosing $v(t)$ at the beginning of the process as a deterministic function of time.

Let $y(t)$ be the solution of the unperturbed and uncontrolled equation

$$(3) \qquad \frac{dy}{dt} = g(y), \quad y(0) = c,$$

and let us measure the difference between x and y by means of a functional of the form

$$(4) \qquad J(v) = \int_0^T h(x - y)dG(t).$$

Here $h(z)$ is a scalar function of z. Two cases of particular interest are those where

$$(5) \qquad \text{(a)} \quad h(x - y) = (x - y, x - y), \quad dG(t) = dt,$$

(b) $G(t)$ is a step-function with a single jump at T.

In the former case, we are dealing with the mean-square deviation; whereas in the latter case, we have what is often called "terminal control."

Let us now take the expected value of $J(v)$ over a prescribed class of random functions $r(t)$. The problem is to determine $v(t)$ so as to minimize this expected deviation, taking advantage of the fact that $v(t)$ can be chosen as a function of the actual state of the system at any time.

We shall show that the functional-equation technique of dynamic programming furnishes a feasible computational solution of this problem for second-order systems, without regard to the analytic character of either the differential equation or the functional $J(v)$. More refined computational techniques offer an approach to higher-dimensional systems, a subject we shall not discuss here.

For the case of linear systems and quadratic deviation, those questions can be resolved by means of classical variational techniques; cf. Ref. [6]. Let us mention also the recent work of Booton [7, 8].

In order to illustrate the range of applicability of the functional-equation technique of dynamic programming, we shall consider the problem of minimizing

$$(6) \qquad J_1(v) = \max_{0 \le t \le T} ||x - y||,$$

where $||z||$ is the norm of the vector z defined in a suitable fashion. Here the metric possessing the proper invariant property is the probability that we have $J_1(v) \geq d$.

A treatment of the deterministic version of this problem may be found in Ref. [4].

Finally, we shall briefly discuss the case of correlated random functions, a much more realistic situation in general.

2. A DISCRETE VERSION

Let us consider the following particular problem, which will illustrate the general method. We have the Van der Pol equation

(1) $\dfrac{d^2x}{dt^2} + \lambda(x^2-1)\dfrac{dx}{dt} + x = r(t) + v(t), \quad x(0) = c_1, \quad x^1(0) = c_2,$

with a random forcing term $r(t)$ and the feedback term $v(t) \equiv v(x,dx/dt,t)$. It is desired to determine the function $v(t)$, subject to the constraint

(2) $$|v(t)| \leq 1,$$

so as to minimize the expected value of the functional

(3) $$\int_0^T x^2 dt + |x(T)|,$$

taken over a suitable class of functions $r(t)$.

In order to prepare the problem for eventual computational solution, and simultaneously to avoid any discussion of the concept of random function, let us consider a discrete version of this problem.

In place of the second-order equation in (1), we consider the system

(4) $$\frac{dx}{dt} = y, \quad x(0) = c_1,$$
$$\frac{dy}{dt} = -\lambda(x^2-1)y - x + r(t) + v(t), \; y(0) = c_2.$$

This first-order system is converted to a system of difference equations in the following way. Divide the interval $[0,T]$ into N equal parts of length Δ, so that $N\Delta = T$, and set

(5) $$x(k\Delta) = x_k, \quad y(k\Delta) = y_k, \quad r(k\Delta) = r_k, \quad v(k\Delta) = v_k.$$

In terms of this notation, the equations in (4) are replaced by

(6) $$x_{k+1} = x_k + y_k\Delta, \quad x_o = c_1,$$

$$y_{k+1} = y_k + \left[r_k + v_k - \lambda(x_k^2 - 1)y_k - x_k\right]\Delta, \quad y_o = c_2,$$

for $k = 0,1,2,\ldots,N-1$.

Similarly, in place of the functional in (3), we use the sum

$$J_N = \Delta \sum_{k=0}^{N-1} x_k^2 + |x_N|. \tag{7}$$

To begin with, assume that the r_k are independent random variables drawn from known distributions, which for simplicity we shall assume to be the same.

Processes in which these distributions must be determined on the basis of observation and experimentation as the process continues constitute an important, but considerably more difficult, class that we shall not consider here. The interested reader may consult Refs. [5] and [9].

Our problem here is to determine the sequence $\{v_k\} = \{v_k\{x_k, y_k\}\}$ that minimizes the expected value of J_N.

3. DYNAMIC-PROGRAMMING APPROACH

It is clear that the minimum of the expected value of J_N depends only on the initial state and the duration of the process—i.e., on c_1, c_2, and N. Let us then, for $-\infty < c_1$, $c_2 < \infty$, $N = 1,2,\ldots$, define

$$f_N(c_1,c_2) = \min_v \ \exp_r \ J_N. \tag{1}$$

For $N = 1$, we have

$$f_1(c_1,c_2) = \Delta c_1^2 + |c_1 + c_2\Delta|. \tag{2}$$

The process proceeds as follows. Knowing the distribution of r_k, but not the value of r_k, we must choose v_k, for $k = 0,1,2,\ldots$.

To obtain a recurrence relation connecting the members of the sequence $\{f_k(c_1,c_2)\}$, we employ the principle of optimality, [1]. Thus

$$(3) \qquad f_k(c_1,c_2) = \min_{|v_o|\leq 1} \left[\Delta c_1^2 + \int_{-\infty}^{\infty} P(c_1,c_2;r_o)dG(r) \right],$$

for $k = 2,3,\ldots,N$, where

$$P(c_1,c_2;r_o) = f_{k-1}(c_1 + c_2\Delta, c_2 + \left[r_o + v_o - c_1 - (c_1^2 - 1)c_2\right]\Delta).$$

The solution of the original control problem is thus reduced to the computation of the sequence of 2-dimensional functions $\{f_k(c_1,c_2)\}$, starting with the known function $f_1(c_1,c_2)$, and continuing by means of (3).

The limiting form of this equation is a nonlinear partial differential equation that is used in the study of the structure of the optimal control policy in various cases.

The time required for the computation depends on the accuracy required, which determines the grid used in the (c_1,c_2) plane.

4. MINIMUM OF MAXIMUM DEVIATION

Consider now the problem of minimizing the functional

$$(1) \qquad J = \text{prob} \left\{ \max_{0\leq t\leq T} |x| \geq a \right\},$$

a discrete version of which is

$$(2) \qquad J_N = \text{prob} \left\{ \max \left[|x_o|, |x_1|, \ldots, |x_{N-1}| \right] \geq a \right\}.$$

For $-\infty < c_1, c_2 < \infty$, $N = 1, 2, \ldots,$ define

$$(3) \qquad f_N(c_1, c_2) = \min_v \quad J_N.$$

Since

$$(4) \qquad \max\left[|x_o|, |x_1|, \ldots, |x_{N-1}|\right] = \max\left[|x_o|, \max\left[|x_1|, \ldots, |x_{N-1}|\right]\right],$$

the principle of optimality yields

$$\begin{aligned}(5) \qquad f_k(c_1, c_2) &= \text{prob}\left\{\max\left[|c_1|, \max\left[|x_1|, \ldots, |x_{N-1}|\right]\right] \geq a\right\} \\ &= 1, \quad |c_1| \geq a, \\ &= \min_{\{v_1\}} \text{prob}\left\{\max\left[|x_1|, \ldots, |x_{N-1}|\right] \geq a\right\}, |c_1| < a, \\ &= \min_{v_o} \int_{-\infty}^{\infty} Q(c_1, c_2; r_o)\, dG(r_o),\end{aligned}$$

for $k = 2, 3, \ldots, N$, with

$$Q(c_1, c_2; r_o) = f_{k-1}(c_1 + c_2\Delta, c_2 + [r_o + v_o - c_1 - (c_1^2 - 1)c_2]\Delta),$$

and

$$\begin{aligned}(6) \qquad f_1(c_1, c_2) &= 1, \quad |c_1| \geq a, \\ &= 0, \quad |c_1| < a.\end{aligned}$$

5. TIME-DEPENDENT PROCESSES

Processes which are time-dependent may be treated by the simple expedient of counting time backward; cf. Refs. [3, 4].

6. CORRELATION

Let us now consider the case where the r_i are not independent. The simplest case is that where the distribution of r_{n+1} depends on the value of r_n. In this case, it is clear that an essential part of the information pattern at each stage is the value of r at the preceding stage. Turning to the problem discussed in Sec. 3, let

(1) $\quad f_N(c_1,c_2;r)$ = minimum expected value of J_N, as defined by (2.7), given the initial state (c_1,c_2) and the information that the value of the random variable at the preceding stage was r.

Further, let

(2) $\quad dG(r_{n+1},r_n)$ = distribution function of r_{n+1}, given the value of r_n.

Then the analogue of (3.3) is

(3)
$$f_k(c_1,c_2;r) = \underset{|v_o|\leq 1}{\mathrm{Min}} \left[\Delta c_1^2 + \int_{-\infty}^{\infty} R(c_1,c_2;r_o,r)\,dG(r_o,r)\right],$$

with

$$R(c_1,c_2;r_o,r) = f_{k-1}(c_1+c_2\Delta, c_2+[r_o+v_o-(c_1^2-1)c_2]\Delta).$$

REFERENCES

1. Bellman, R., "Theory of Dynamic Programming," Bull. Amer. Math. Soc., Vol. 60, 1954, pp. 503-515.

2. Bellman, R., Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.

3. Bellman, R., "Dynamic Programming and Control Processes," Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, 1956.

4. Bellman, R., "On the Minimum of Maximum Deviation," Quart. Appl. Math., Vol. XIV, 1957, pp. 419-423.

5. Bellman, R., "A Problem in the Sequential Design of Experiments," Sankhya, Vol. 16, 1956, pp. 221-229.

6. Bellman, R., I. Glicksberg, and O. Gross, "On Some Variational Problems Occurring in the Theory of Dynamic Programming," Rendiconti del Circolo Matematico di Palermo, Vol. 2, 1954, pp. 1-35.

7. Booton, R. C., Jr., "Optimum Design of Final-value Control Systems," Symposium on Nonlinear Circuit Analysis, Polytechnic Institute of Brooklyn, 1956.

8. Booton, R. C., Jr., "Final-value Systems with Gaussian Inputs," Transactions on Information Theory, Vol. IT-2, No. 3, 1956, pp. 173-176.

9. Robbins, H., "Some Aspects of the Sequential Design of Experiments," Bull. Amer. Math. Soc., Vol. 58, 1952, pp. 527-536.