Serdica
Bulgariacae mathematicae publicationes

Сердика
Българско математическо списание

STOCHASTIC OPTIMAL CONTROL FOR A CLASS OF DISCRETE SYSTEMS

STRATIS KOUNIAS

Properties of the optimal control and the optimal cost are studied for a class of discrete stochastic systems.

It is proved, that for independent random inputs, the optimal control and the optimal cost are functions of the last available information on the state variable only and that to each degree of dependence on the random inputs corresponds an additional variable on the optimal cost.

The total cost is taken to be the expended sum of the costs at each stage and the state variable is a function of the state and control variables of the previous stage plus a random error.

Finaly, for convex cost functions it is proved that the optimal cost is convex and the optimal control is continuous function of the state variable.

1. Introduction. In this paper we study the problem of optimal control for a wide class of discrete stochastic systems.

The study of discrete systems ([1; 2; 3]) is very important because the computation of the optimal control and the optimal cost of any control problem, using digital computers, requires the discretization of the problem. Furthermore if deterministic control is a first approximation to actual problems, stochastic control is more realistic.

Throughout this paper we assume that the different coefficients or functions are known and therefore we are not concerned with their estimation. At the beginning we study properties of the discrete stochastic control with additive cost function, and at the end we restrict our attention to convex cost functions. The technique we apply is equivalent to Bellman's principle of optimality.

We assume that the behaviour of the system is governed by the following relation:

$$x_{k-1}=f_k(x_k, u_k)+e_k, \quad k=n, \dots, 1, \tag{1}$$

where:

x_k is a p-dimensional vector representing the value of the state variable at stage k.

u_k is a q-dimensional vector representing the value of the control variable at stage k, furthermore we assume that $u_k \in U_k \subset E_q$ ($E_q = q$-dimensional Euclidian space).

e_k is a p-dimensional random variable representing random disturbances at stage k.

$f_k(x_k, u_k) \equiv z_k$ are known p-dimensional functions.

For computational convenience we define the last stage as the stage zero so that, when we are at stage k there are k stages remaining until the end of the process. The control variable u_k will be taken as a (measurable) function of the presently known values of the state and control variables with $u_k \in U_k$.

The overall cost is defined as the sum of the individual costs, i. e.

$$J(x_{n-1}, \ldots, x_0;\ u_n, \ldots, u_1) = \sum_{k=n}^{1} r_k(x_{k-1}, u_k), \tag{2}$$

where n is the total number of stages and r_k are known (measurable) functions. This includes the case of final stage optimization where $r_k(x_{k-1}, u_k) = 0$ for $k = n, \ldots, 2$.

In this stochastic system the state and control variables are, from relation (1), random variables so that a reasonable criterion will be the minimization of the expected cost, i. e.

$$V_n(x_n) = \min_{[u_n]} E\left(\sum_{k=n}^{1} r_k(x_{k-1}, u_k) \mid x_n\right), \tag{3}$$

where the conditional expectation given x_n is taken with respect to the joint distribution of the random inputs $e_n, \ldots, e_1$, and $[u_k]$ denotes the set of admissible control sequences from stage k until the end of the process, i. e. $[u_k] = \{u_k, \ldots, u_1\}$ with $u_r \in U_r$ for all $r = k, \ldots, 1$.

Taking conditional expectation we have tacitly assumed that at the beginning of the process we know x_n, and then the optimal expected cost V_n is a function of x_n and the total number of stages, i. e. $V_n = V_n(x_n)$.

If at the beginning of the process the value of x_n is not known but is a random variable then we have to take unconditional expectation and the optimal cost will be only a function of the total number of stages n.

The essential difference between deterministic and stochastic control is that in deterministic control, given x_n, we are able to know in advance the exact values of x_k and u_k, $k = n, n-1, \ldots, 1$, whereas in stochastic control we do not know the values of x_n, u_k until we reach stage k and then, apply the optimal control based on the known information up to stage k. In other words, the optimal trajectory in the first case is a function known in advance, whereas in the second case it is a stochastic sequence.

In what follows we examine first the case where at each stage we know the values of the state variable and the control variable up to that stage but the random inputs are independent and then, we prove that the optimal cost and the optimal control are functions of only one variable and that to each degree of dependence of the random inputs corresponds an additional variable. We examine also the case where at stage k we know the values of the state variable up to stage $k+r$, $r \geqq 1$ and we prove that a delay of information of r stages does not increase the dimensionality of the optimal control and the optimal cost. Also some other useful results are proved concerning convex cost functions in which case the optimal control is proved to be convex and the optimal control continuous functions.

2. Independent and dependent inputs. In the case where at each stage, say k, before applying control u_k we know $x_n, \ldots, x_k$ and the controls $u_n, \ldots, u_{k+1}$ we applied up to stage k, then

$$u_k = u_k(x_n, \ldots, x_k;\ u_n, \ldots, u_{k+1}) \quad \text{for all } k = n, \ldots, 1 \tag{4}$$

or equivalently

$$u_k = u_k(x_n, \ldots, x_k) \quad \text{for all } k = n, \ldots, 1. \tag{5}$$

Under these conditions we can write:

$$V_m(x_n, \ldots, x_m) = \min_{[u_m]} E(\sum_{k=m}^{1} r_k(x_{k-1}, u_k) \mid x_n, \ldots, x_m)$$

$$= \min_{[u_m]} E(r_m(x_{m-1}, u_m) + \sum_{k=m-1}^{1} r_k(x_{k-1}, u_k) \mid x_n, \ldots, x_m)$$

$$= \min_{[u_m]} E(Er_m(x_{m-1}, u_m) + E\sum_{k=m-1}^{1} r_k(x_{k-1}, u_k) \mid x_n, \ldots, x_{m-1}) \mid x_n, \ldots, x_m)$$

$$= \min_{[u_m]} E(r_m(x_{m-1}, u_m) + \min_{[u_{m-1}]} E(\sum_{k=m-1}^{1} r_k(x_{k-1}, u_k) \mid x_n, \ldots, x_{m-1}) \mid x_n, \ldots, x_m)$$

$$= \min_{[u_m]} E(r_m(x_{m-1}, u_m) + V_{m-1}(x_n, \ldots, x_{m-1}) \mid x_n, \ldots, x_m),$$

i. e.

$$(6) \qquad V_m(x_n, \ldots, x_m) = \min_{[u_m]} E(r_m(x_{m-1}, u_m) + V_{m-1}(x_n, \ldots, x_{m-1}) \mid x_n, \ldots, x_m)$$

$= u_m^* x_m$ for all $m = n, \ldots, 2$ and

$$(7) \qquad V_1(x_n, \ldots, x_1) = \min_{u_1} E(r_1(x_0, u_1) \mid x_n, \ldots, x_1).$$

From now on u_m^* denotes the optimal control at stage m and $z_m^* = f_m(x_m, u_m^*)$.

2.1. Theorem. *If the random disturbances are independent and the system is determined by* (1) *and* (2), *then* $V_m(x_n, \ldots, x_m) = V_m(x_m), u_m^*(x_n, \ldots, x_m) = u_m^*(x_m)$ *for all* $m = n, \ldots, 1$.

Proof. From (1), (7) and the assumption of independence we have:

$$(8) \qquad V_1(x_n, \ldots, x_1) = \min_{u_1} E(r_1(x_0, u_1) \mid x_n, \ldots, x_1)$$

$$= \min_{u_1} E(r_1(z_1 + e_1, u_1) \mid x_n, e_n, \ldots, e_2) = \min_{u_1} E(r_1(z_1 + e_1, u_1)).$$

So

$$(9) \qquad V_1(x_n, \ldots, x_1) = \min_{u_1} \widehat{r}_1(z_1, u_1).$$

Therefore $u_1^* = u_1^*(x_1)$ and from (9) obtain $V_1(x_n, \ldots, x_1) = \widehat{r}_1(z_1^*, u_1^*) = V_1(x_1)$. Now proceed by induction. Assume $V_{m-1}(x_n, \ldots, x_{m-1}) = V_{m-1}(x_{m-1})$ and $u_{m-1}^*(x_n, \ldots, x_{m-1}) = u_{m-1}^*(x_{m-1})$. Then from (1), (6) and the previous assumptions obtain

$$V_m(x_n, \ldots, x_m) = \min_{[u_m]} E(r_m(x_{m-1}, u_m) + V_{m-1}(x_n, \ldots, x_m) \mid x_n, \ldots, x_m)$$

$$= \min_{[u_m]} E(r_m(z_m + e_m, u_m) + V_{m-1}(x_{m-1}) \mid x_n, e_n, \ldots, e_{m+1})$$

$$= \min_{[u_m]} (\widehat{r}_m(z_m, u_m) + EV_{m-1}(x_{m-1})) = \min_{[u_m]} (\widehat{r}_m(z_m, u_m) + \widehat{V}_{m-1}(z_m)),$$

where $z_m = f_m(x_m, u_m)$, hence $u_m^* = u_m^*(x_m)$ and consequently $V_m(x_n, \ldots, x_m) = V_m(x_m)$, Q. E. D.

So for independent disturbances and full information at each stage, the optimal control and the optimal cost are functions only of the last known value of the state variable. This result was expected but here we have a rigorous proof. The above result is also very desirable because the time and space needed for tabulation of $u^*(x_m)$ and $V_m(x_m)$ are smaller than if they were functions of more than one variable. Below we prove that for each degree of dependence of the random inputs we add one variable.

2.2. Theorem. *If $e_n, e_{n-1}, \ldots, e_1$, form a Markov chain, and the system is determined by* (1) *and* (2), *then u_k^* and V_k can be expressed as functions of two p-dimensional variables.*

Proof. (i) From the assumptions above and (1), (8) we obtain $V_1(x_n, \ldots, x_1) = \min_{u_1} \widehat{r}_1(z_1, u_1, e_2)$, where $e_2 = x_1 - z_2$. Therefore $u_1^* = u_1^*(x_1, e_2) = u_1^*(x_1, z_2^*)$ and $V_1(x_n, \ldots, x_1) = \widehat{r}_1(z_1, u_1^*, e_2) = V_1(x_1, e_2) = V_1(x_1, z_2^*)$. Now proceed by induction as before. Assume $u_{k-1}^* = u_{k-1}^*(x_{k-1}, e_k)$, $V_{k-1}(x_n, \ldots, x_{k-1}) = V_{k-1}(x_{k-1}, e_k)$. Then from (1), (7) and the assumptions of the theorem obtain

$$V_k(x_n, \ldots, x_k) = \min_{[u_k]} E(r_k(z_k + e_k, u_k) + V_{k-1}(z_k + e_k, e_k) \mid x_n, e_n, \ldots, e_{k+1})$$

$$= \min_{[u_k]} (\widehat{r}_k(z_k, u_k, e_{k+1}) + \widehat{V}_{k-1}(e_{k+1}, z_k)).$$

Therefore

$$u_k^* = u_k^*(x_k, e_{k+1}) = u_k^*(x_k, z_{k+1}^*) \tag{10}$$

and

$$V_k(x_n, \ldots, x_k) = V_k(x_k, e_{k+1}) = V_k(x_k, z_{k+1}^*) \tag{11}$$

for $k = n-1, \ldots, 1$. For $k = n$ we have $u_n^* = u_n^*(x_n)$, $V_n = V_n(x_n)$. Now from (10), (11), obtain, $u_{n-1}^* = u_{n-1}^*(x_{n-1}, z_n^*)$, $V_{n-1} = V_{n-1}(x_{n-1}, z_n^*)$, where $z_k^* = f_k(x_k, u_k^*)$ is a p-dimensional variable, then from (10), (11) we have

$$u_k^* = u_k^*(x_k, z_{k+1}^*),\ V_k = V_k(x_k, z_{k+1}^*) \text{ for } k = n-1, \ldots, 1,$$

$$u_n^* = u_n^*(x_n),\ V_n = V_n(x_n) \text{ for } k = n. \tag{12}$$

Hence u_k, v_k can be expressed as functions of two p-dimensional variables.

Q. E. D.

For computational purposes the forms (12) are preferred because we must try to keep the number of variables, through which u_k^*, V_k, are expressed, to a minimum; then we should tabulate z_k^*, $u_k^* = u_k^*(x_k, z_{k+1}^*)$ and $V_k = V_k(x_k, z_{k+1}^*)$ for all $k = n-1, \ldots, 1$.

2.1. Corollary. *If the random disturbances form a Markov sequence of order r, i. e. their joint distribution*

$$F(e_k \mid x_n, e_n, \ldots, e_{k+1}) = F(e_k \mid e_{k+r}, \ldots, e_{k+1})$$

for some $r \geq 1$, then

$$u_k^* = u_k^*(x_k, e_{k+1}, \ldots, e_{k+r}),\ V_k = V_k(x_k, e_{k+1}, \ldots, e_{k+r}) \text{ for } k = n-r, \ldots, 1,$$

$$u_k^* = u_k^*(x_k, e_{k+1}, \ldots, e_n),\ V_k = V_k(x_k, e_{k+1}, \ldots, e_n) \text{ for } k = n, \ldots, n-r+1,$$

where $e_k = x_{k-1} - f_k(x_k, u_k^*) = x_{k-1} - z_k^*$, $k = n, \ldots, 1$.

Proof. The proof is similar to the previous one. Q. E. D.

If we want to save computational effort we must be prepared to increase the total cost. One way of doing this is to find the suboptimal policy in which the control u_k is only a (measurable) function of x_k. The increase in the optimal cost will depend then on the form of the cost function and the joint distribution of the random inputs.

3. Observations with a time lag. In this section we consider the case where there is a time lag in information about the state variable, i. e. at stage k we know $x_n, \ldots, x_{k+r}$ for some integer $r \geq 1$ for all $k = n-r, \ldots, 1$ and we know only x_n from the beginning of the process until stage $n-r$. The control sequence already applied up to stage k is considered known at stage k. Under these conditions (4), (5), (6), (7) become

$$u_k = u_k(x_n),\ k = n, \ldots, n-r, \tag{13}$$

$$u_k = u_k(x_n, \ldots, x_{k+1}),\ k = n-r, \ldots, 1, \tag{14}$$

$$V_k(x_n) = \min_{[u_k]} E(r_k(x_{k-1}, u_k) + V_{k-1}(x_n) \mid x_n),\ k = n, \ldots, n-r+1, \tag{15}$$

$$V_k(x_n, \ldots, x_{k+r}) = \min_{[u_k]} E(r_k(x_{k-1}, u_k) + V_{k-1}(x_n, \ldots, x_{k+r-1}) \mid x_n, \ldots, x_{k+r}), \tag{16}$$

$$k = n-r, \ldots, 2,$$

$$V_1(x_n, \ldots, x_{1+r}) = \min_{u_1} E(r_1(x_0, u_1) \mid x_n, \ldots, x_{1+r}). \tag{17}$$

3.1. Theorem. *If the random disturbances $e_n, \ldots, e_1$ are independent, the system is determined by* (1), (2) *and we have r-steps lag of ivformation about the state variable, then the optimal control u_k^* and the optimal cost V_k, can be expressed as functions of only one p-dimensional variable.*

Proof. Relation (1) can be written as

$$x_{k-1} = g_k(x_{k+r};\ u_{k+r}, \ldots, u_k;\ e_{k+r}, \ldots, e_k) \text{ for } k = n-r, \ldots, 1 \tag{18}$$

$$x_{k-1} = g_k(x_n;\ u_n, \ldots, u_k;\ e_n, \ldots, e_k) \text{ for } k = n, \ldots, n-r+1, \tag{19}$$

where the functions g_k are easily derived from the relations

$$x_{n-2} = f_{n-1}(x_{n-1}, u_{n-1}) + e_{n-1} = f_{n-1}(f_n(x_n, u_n) + e_n, u_{n-1}) + e_{n-1}$$

$$= g_{n-1}(x_n;\ u_n, u_{n-1};\ e_n, e_{n-1})$$

$$x_{k-1} = f_k(x_k, u_k) + e_k = \cdots = g_k(x_n;\ u_n, \ldots, u_k;\ e_n, \ldots, e_k) \tag{20}$$

for $k = n-r-1, \ldots, 1$ and

$$x_{k-1} = f_k(x_k, u_k) + e_k = \cdots = g_k(x_{k+r};\ u_k, \ldots, u_{k+r};\ e_k, \ldots, e_{k+r})$$

for $k = n-1, \ldots, n-r$.

Hence (17) becomes

$$V_1(x_n, \ldots, x_{1+r}) = \min_{u_1} E(r_1(x_0, u_1) \mid x_n, \ldots, x_{1+r}),$$

but $x_0 = g_1(x_{1+r};\ u_{1+r}, \ldots, u_1;\ e_{1+r}, \ldots, e_1)$. Hence

$$V_1 = \min_{u_1} \widehat{r_1}(y_1, u_1),$$

where $y_1 = y_1 \ (x_{1+r};\ u_{1+r}, \ldots, \ u_1) = E(r_1(g_1,\ u_1)\ x_n, e_n, \ldots, e_{1+r})$. Therefore $u_1^* = u_1^*(y_1)$, $V_1 = V_1(y_1)$. Assume now that $V_{k-1} = V_{k-1}(y_{k-1})$ and repeating the same procedure as in the previous theorem we prove the desired result.

If the random inputs are dependent as in Corollary 2.1 then u_k^* and V_k will depend on $e_{k+r+m}, \ldots, e_{k+r+1}$ as well as on y_k^*.

In other words, we have to add one p-dimensional parameter for each degree of dependence but the time lag in information on the state variable does not increase the number of parameters through which u_k^* and V_k can be expressed.

4. Convex cost functions. In this section we study the important case where the cost functions are convex. Throughout this section assume that $u_k \in U_k$, $U_k \subset E_q$, $k = n, \ldots, 1$ and that U_k are convex sets for all $k = n, \ldots, 1$, in the q-dimensional Euclidean space E_q.

4.1 Lemma. *If $F(x, u)$ is a (strictly) convex function of the $p+q$ dimensional variable (x, u) and U is a convex set in E_q then*

$$G(x) = \min_{u \in U} F(x, u)$$

s (strictly) convex in x.

Proof. Given any two points x_1 and x_2 in E_p we have for $0 \leq \lambda \leq 1$,

$$G(\lambda x_1 + (1-\lambda)x_2) = \min_{u \in U} F(\lambda x_1 + (1-\lambda)x_2, u)$$

$$= F(\lambda x_1 + (1-\lambda)x_2, u^*(\lambda x_1 + (1-\lambda)x_2)) \leq F(\lambda x_1 + (1-\lambda)x_2, \lambda u^*(x_1) + (1-\lambda)u^*(x_2)),$$

where $u^*(x)$ is a minimizing value of $F(x, u)$, and $(\lambda u^*(x_1) + (1-\lambda)u^*(x_2)) \in U$ because U is convex. Now $F(x, u)$ is convex in (x, u), hence

$$G(\lambda x_1 + (1-\lambda)x_2) \leq \lambda F(x_1, u^*(x_1)) + (1-\lambda)F(x_2, u^*(x_2))$$

$$= \lambda \min_{u \in U} F(x_1, u) + (1-\lambda) \min_{u \in U} F(x_2, u) = \lambda G(x_1) + (1-\lambda)G(x_2).$$

Q. E. D

4.2 Lemma. *If $G(x, u)$ is (strictly) convex in (x, u), e is a random variable, and $f(x, u)$ is linear in x, u, then $\widehat{G}(x, u) = EG(f(x, u) + e, u)$ is (strictly) convex in (x, u).*

Proof.

$$\widehat{G}(\lambda x_1 + (1-\lambda)x_2, \lambda u_1 + (1-\lambda)u_2) = E(G(f(\lambda x_1 + (1-\lambda)x_2, \lambda u_1 + (1-\lambda)u_2)$$

$$+ e, \lambda u_1 + (1-\lambda)u_2)) = E(G(\lambda f(x_1, u_1) + (1-\lambda)f(x_2, u_2) + e, \lambda u_1 + (1-\lambda)u_2))$$

$$= E(G(\lambda(f(x_1, u_1) + e) + (1-\lambda)(f(x_2, u_2) + e), \lambda u_1 + (1-\lambda)u_2))$$

$$\leq E(\lambda G(f(x_1, u_1) + e, u_1) + (1-\lambda)G(f(x_2, u_2) + e, u_2)) = \lambda\widehat{G}(x_1, u_1) + (1-\lambda)\widehat{G}(x_2, u_2)$$

for all $0 \leq \lambda < 1$.

Q. E. D.

4.1 Theorem. *If*

(i) *At stage k we know $x_n, \ldots, x_k$*

(ii) *$e_n, \ldots, e_1$ are independent*

(iii) *The functions* $r_k(x_{k-1}, u_k)$ *are convex functions of* (x_{k-1}, u_k) *and* $f_k(x, u)$ *linear in* x, u *for all* $k=n, \dots, 1$ *then, the optimal cost* $V_k(x)$ *is convex for all* $k=n, \dots, 1$.

Proof.

$$E(r_1(f_1(x, u_1)+e_1, u_1) \mid x_n, e_n, \dots, e_2) = Er_1(f_1(x, u_1)+e_1, u_1)$$
$$= \widehat{r}_1(f_1(x, u_1), u_1) = \widehat{r}_1(x, u_1)$$

and by lemmas 4.1 and 4.2 the functions

$$\widehat{\widehat{r}}_1(x, u_1) \text{ and } V_1(x) = \min_{u_1 \in U_1} \widehat{\widehat{r}}(x, u_1)$$

are convex. Now assume that $V_{k-1}(x)$ is convex and set

$$F_k(x, u_k) = E(r_k(f_k(x, u_k)+e_k, u_k)+V_{k-1}(f_k(x, u_k)+e_k))$$
$$= \widehat{r}_k(f_k(x, u_k), u_k) + \widehat{V}_{k-1}(f_k(x, u_k)).$$

Each term in the right hand side is convex by lemma 4.2, then $F_k(x, u_k)$ is convex in (x, u_k) as the sum of two convex functions and by lemma 4.1

$$V_k(x) = \min_{u_k \in U_k} F_k(x, u_k)$$

is convex in x. So $V_k(x)$ is convex for all $k=n, , , , 1$. Q. E. D.

If at stage k we have r-step lag of information, then similarly we prove that V_k is a convex function of one p-dimensional variable.

4.1 Corollary. *If*

(i) *The assumptions of theorem 4.1 hold*

(ii) $r_1(x, u)$ *is strictly convex in* (x, u).

Then

(i) V_k *is strictly convex*

(ii) *the optimal control* u_k^* *is continuous for all* $k=n, \dots, 1$.

Proof. (i) From lemmas 4.1, 4.2 we conclude that V_1 is strictly convex and consequently V_k is strictly convex for all $k=n, \dots, 1$.

Notice here that $r_k(x, u)$, $k=n, \dots, 2$ need not be strictly convex.

(ii) The function

$$F_k(x, u_k) = E(r_k(f_k(x, u_k)+e_k, u_k)+V_{k-1}(f_k(x, u_k)+e_k))$$

is strictly convex from theorem 4.1, hence to every $x \in E_p$ there corresponds exactly one $u_k^* \in U_k$, such that $V_k(x) = F_k(x, u_k^*(x))$. Now suppose that there exists a sequence $x_i \in E_p$ with $x_i \to x$ for some $x \in E_p$ as $i \to \infty$ such that $u_i^* = u^*(x_i)$ do not $\to u^*(x)$. We can choose then a subsequence of u_i^* which converges. Let $u_{r_i} = u^*(x_{r_i}) \to u \in U_k$, then $(x_{r_i}, u_{r_i}) \to (x, u)$. But the convex functions are continuous, hence

$$V_k(x_{r_i}) = F_k(x_{r_i}, u_{r_i}^*) \to F_k(x, u),\ V_k(x_{r_i}) \to V_k(x) = F_k(x, u_k(x)).$$

Therefore $V_k(x) = F_k(x, u) = F_k(x, u_k^*(x))$. From the uniqueness of $u^*(x)$ we conclude that $u = u^*(x)$. We proved that every convergent subsequence of $u^*(x_i)$

converges to $u^*(x)$, which implies that $u^*(x_i) \to u^*(x)$ for every $x_i \to x$, i. e. it is continuous.

Q. E. D.

The above corollary implies that even if $r_1(x, u)$ is not strictly convex, we can change $r_1(x, u)$ slightly in the parts where it is not strictly convex to obtain a continuous optimal control or that if $r_1(x, u)$ is not strictly convex then although $u_k^*(x)$ might not be single valued, we can always choose $u_k^*(x)$ such that $u_k^*(x)$ is continuous.

REFERENCES

1. K. Astrom. Introduction to stochastic control theory. New York, 1970.
2. M. Aoki. Optimization of stochastic systems. New York, 1967.
3. R. Bellman. Adaptive control processes: A guided tour. Princeton, 1961.

University of Thessaloniki
Thessaloniki Greece

Received 21. 2. 1978