ON WORST CASE DESIGN STRATEGIES*

T. Başar and P. R. Kumar
Decision and Control Laboratory,
Department of Electrical Engineering and Coordinated Science Laboratory,
University of Illinois, Urbana, Illinois

Abstract

For sequential decision processes, we consider the problem of obtaining the min-max strategy which minimizes the worst case performance. This is a game against nature, where the controller attempts to minimize a specified cost criterion, while nature attempts to maximize it. It is apparently a folk theorem that such a min-max strategy can be obtained by means of a dynamic programming like recursion, even though we have not seen any general proof of this, applicable to stochastic systems, which does not rely on the existence of a saddle point. We prove this theorem, and also examine the precise roles of the strategy sets allowed to the minimizer and the maximizer in determining the upper value of the game. Improvements in the results for the case of deterministic systems, and generalizations to continuous time systems are indicated.

1 Introduction

For simplicity, we shall restrict attention to finite state, finite control systems, where unnecessary technical details do not arise.

Consider a controlled Markov chain with a finite state space $\{1, 2, \ldots, I\}$, and transition probabilities

$$[P_{ij}^t(u, v) : 1 \leq i, j \leq I] \quad \text{for} \quad 0 \leq t \leq N-1,$$

*The research of the first author has been supported by the U.S. Air Force under Contract AFOSR 84-0056, while the research of the second author has been supported by the National Science Foundation under Grant ECS-83-04435 and the U.S. A.R.O. under Contract DAAG-29-85-K-0094.

such that for $u \in U$ and $v \in V$,

$$Prob(x_{t+1} = j | x_0, x_1, \ldots, x_{t-1}, x_t = i, u_0, \ldots, u_{t-1}, u_t = u, v_0, \ldots, v_{t-1}, v_t = v) := P_{ij}^t(u, v).$$

Both U and V are assumed finite. We are given a cost criterion

$$\sum_{t=0}^{N-1} c_t(x_t, u_t, v_t) + c_N(x_N),$$

where x_t is the state at time t, and $u_t \in U$, $v_t \in V$ are chosen by the minimizer and the maximizer, respectively, at time t.

For the minimizer a policy is a sequence $g = \{g_0, g_1, \ldots, g_{N-1}\}$. Each g_t is a function such that

$$g_t : (x^t, u^{t-1}) \mapsto p_t \in P(U)$$

where, here and throughout, $P(A)$ is the set of all probability distributions on the set A, $x^t := (x_0, x_1, \ldots, x_t)$ and a similar interpretation holds for u^{t-1}. When a minimizer uses a policy g, each u_t is chosen randomly according to the probability distribution p_t. Let G_R denote the set of all policies for the minimizer. (The subscript R in G_R stands for "randomized".)

For the maximizer a policy is a sequence $h = \{h_0, h_1, \ldots, h_{N-1}\}$, where each h_t is a mapping such that

$$h_t : (x^t, u^t, v^{t-1}) \mapsto q_t \in P(V).$$

Let H_{DR} denote the class of all such policies. (The subscript D in H_{DR} stands for "delay" and is used to denote the fact that the maximizer can, at each time t, base his decision on u^t, and thus can "delay" choosing v_t until the maximizer has already chosen u_t.) When a maximizer uses the policy h, each v_t is chosen according to the probability distribution q_t.

When the minimizer and maximizer have chosen policies $g \in G_R$ and $h \in H_{\mathrm{DR}}$, the resulting expected cost is

$$J_{g,h}(i) \quad := \quad E^{g,h}\left[\sum_{t=0}^{N-1} c_t(x_t, u_t, v_t) + c_N(x_N) | x_0 = i\right],$$

where $E^{g,h}[\cdot|\cdot]$ denotes the conditional expectation under the probability measure induced on $\{x^N, u^{N-1}, v^{N-1}\}$ by g and h.

Our goal is to determine the value of,

$$\overline{J}(i) \quad := \quad \min_{g \in G_R} \max_{h \in H_{\mathrm{DR}}} J_{g,h}(i),$$

and also to determine a $g^* \in G_R$ which attains the minimum above. $\overline{J}(i)$ will be called the *upper value* and g^* a *min-max* policy.

2 Main Results

It is also of interest to know whether the minimizer and maximizer can restrict their attention to smaller sets of policies than G_R and $H_{\rm DR}$, and for this reason we consider the following subsets. Let $G_{\rm MN} \subseteq G_R$ be the set of all policies $g = \{g_0, g_1, \ldots, \}$ where each g_t depends only on x_t and is moreover a degenerate probability distribution on U, i.e., u_t is chosen as

$$u_t \;\; = \;\; g_t(x_t) \in U.$$

Similarly, let $H_{\rm MN} \subseteq H_{\rm DR}$ be the set of policies $h = \{h_0, h_1, \ldots, \}$ such that each h_t is a mapping $h_t : \{1, 2, \ldots, I\} \to V$. (The subscript MN stands for "Markovian nonrandomized".) For the maximizer it is convenient to define two more intermediate sets of policies $H_{\rm DMN}$ and H_R. $H_{\rm DMN}$ is the set of policies $h = \{h_0, h_1, \ldots, \}$ where each h_t is a mapping $h_t : \{1, 2, \ldots, I\} \times U \to V$. When such an h is adopted, v_t is chosen as $v_t = h_t(x_t, u_t)$. Clearly $H_{\rm MN} \subseteq H_{\rm DMN} \subseteq H_{\rm DR}$. Also define $H_R \subseteq H_{\rm DR}$ as the set of all policies $h = \{h_0, h_1, \ldots, h_{N-1}\}$ where each h_t is a function only of x_t, with the interpretation that under such an h, $q_t = h_t(x_t) \in P(V)$ determines the probability distribution according to which v_t is chosen.

Our main result is the following.

Theorem 1. *Recursively define the functions* $\{V_N, V_{N-1}, \ldots, V_0\}$ *by the following:*

$$V_N(i) \;\; := \;\; c_N(i) \text{ for } 1 \le i \le I, \tag{1a}$$

$$V_t(i) \;\; := \;\; \min_{u \in U} \max_{v \in V} \{c_t(i,u,v) + \sum_{j=1}^{I} P_{ij}^t V_{t+1}(j)\} \quad \text{for } 1 \le i \le I. \tag{1b}$$

Then, the following results hold.

$$V_0(i) = \overline{J}(i) \text{ for } 1 \le i \le I. \tag{2a}$$

For each $0 \le t \le N-1$ *and* $1 \le i \le I$, *let* $u =: g_t^*(i) \in U$ *attain the minimum in (1b). Then* $g^* := \{g_0^*, g_1^*, \ldots, g_{N-1}^*\} \in G_{\rm MN}$ *is a min-max policy, i.e.,*

$$J_{g^*,h}(i) \le \overline{J}(i) \text{ for all } h \in H_{\rm DR} \text{ and } 1 \le i \le I.$$

For each $0 \le t \le N-1$, $1 \le i \le I$, $u \in U$, *let* $v =: \widehat{h}_t(i,u)$ *attain the inner maximum in (1b). Then* $\widehat{h} := \{\widehat{h}_0, \widehat{h}_1, \ldots, \widehat{h}_{N-1}\} \in H_{\rm DMN}$ *is an optimal policy for the maximizer, i.e.,*

$$J_{g,\widehat{h}}(i) \ge \overline{J}(i) \quad \text{for all} \quad g \in G_R \quad \text{and} \quad 1 \le i \le I. \tag{2b}$$

For each $g = \{g_0, g_1, \ldots, g_{N-1}\} \in G_{\rm MN}$, *let* $h(g) := \{h_{0,g_0}, h_{1,g_1}, \ldots, h_{N-1,g_{N-1}}\} \in H_{\rm MN}$ *be defined by*

$$h_{t,g_t}(i) \;\; := \;\; \widehat{h}_t[i, g_t(i)]. \tag{2c}$$

Then, $h(g)$ *is an optimal threat for the maximizer against* $g \in G_{\rm MN}$ *for the minimizer, i.e.,*

$$J_{g,h(g)}(i) \geq \overline{J}(i) \text{ for all } g \in G_{\rm MN}, 1 \leq i \leq I.$$
$$\min_{g \in G_{\rm MN}} \max_{h \in H_{\rm MN}} J_{g,h}(i) = \min_{g \in G_R} \max_{h \in H_{\rm DR}} J_{g,h}(i) [= \overline{J}(i)] \text{for} 1 \leq i \leq I. \tag{2d}$$

Proof: Suppose the minimizer uses the policy $g^* \in G_{\rm MN}$ defined in (2b). Then, the system evolves according to the transition probabilities,

$$Prob(x_{t+1} = j | x^t, u^t, v^t) \;\; = \;\; P^t_{x_t,j}[g^*_t(x_t), v_t].$$

Defining

$$r^t_{ij}(v) \;\; := \;\; P^t_{ij}[g^*_t(i), v] \tag{3}$$
$$d_t(i, v) \;\; := \;\; c_t[i, g^*_t(i), v], \tag{4}$$

the problem for the maximizer reduces to that of maximizing

$$E^{\bar{h}} \left[\sum_{t=0}^{N-1} d_t(x_t, v_t) + c_N(x_N) | x_0 = i \right]$$

over all $\bar{h} \in H_{\rm DR}$, for the system with transition probabilities given by (3). For each $\bar{h} = \{\bar{h}_0, \bar{h}_1, \ldots, \bar{h}_{N-1}\} \in H_{\rm DR}$ define a policy $h = \{h_0, h_1, \ldots, h_{N-1}\} \in H_R$ by

$$h_t(x^t, v^{t-1}) \;\; := \;\; \bar{h}_t(x^t, u^t, v^{t-1}) \text{ with } u_t := g^*_t(x_t), \text{ for } t = 0, 1, \ldots, N-1.$$

Clearly the system behaves under (g^*, h) in the same way as it behaves under $(g^*, \bar{h})$, and so,

$$E^{\bar{h}} \left[\sum_{t=0}^{N-1} d_t(x_t, v_t) + c_N(x_N) | x_0 = i \right] = E^{h} \left[\sum_{t=0}^{N-1} d_t(x_t, v_t) + c_N(x_N) | x_0 = i \right]. \tag{5}$$

Hence, as long as the minimizer uses g^*, the maximizer can do no better by considering policies in $H_{\rm DR}$ than he could by restricting attention to policies in H_R. Define

$$W_0(i) \;\; := \;\; \max_{h \in H_R} E^{h} \left[\sum_{t=0}^{N-1} d_t(x_t, v_t) + c_N(x_N) | x_0 = i \right]. \tag{6}$$

From standard results on dynamic programming, see Bertsekas [1], we know that this maximum can be obtained recursively through $\{W_N, W_{N-1}, \ldots, W_0\}$ which are defined by,

$$W_N(i) := c_N(i), \tag{7a}$$

$$W_t(i) := \max_{v \in V} \left\{ d_t(i,v) + \sum_{j=1}^{I} r_{ij}^t(v) W_{t+1}(j) \right\} \tag{7b}$$

Moreover, if for each $i \in \{1, 2, \ldots, I\}$ and $t \in \{0, 1, \ldots, N-1\}$, $v =: h_t(i)$ attains the maximum on the right-hand side in (7b), then $h := \{h_0, h_1, \ldots, h_{N-1}\} \in H_{\mathrm{MN}}$ is a maximizing policy.

Hence we know that

$$max_{h \in H_{\mathrm{DR}}} J_{g_*,h}(i) = \max_{h \in H_R} J_{g^*,h}(i) = max_{h \in H_{\mathrm{MN}}} J_{g^*,h}(i) = W_0(i). \tag{8}$$

Now we claim that $W_t = V_t$ for $t = 0, 1, 2, \ldots, N-1$. Clearly from (7a) and (1a), we see that $W_N = V_N$. Proceeding by induction, suppose that $W_{t+1} = V_{t+1}$. Then, from (7b) we have

$$W_t(i) = \max_{v \in V} \left\{ d_t(i,v) + \sum_{j-1}^{I} r_{ij}^t(v) V_{t+1}(j) \right\} \quad 1 \le i \le I.$$

However, by utilizing (4) we get

$$W_t(i) = \max_{v \in V} \left\{ c_t[i, g_t^*(i), v] + \sum_{j=1}^{I} P_{ij}^t[g_t^*(i), v] V_{t+1}(j) \right\}.$$

Since, as defined in (2b), $u =: g_t^*(i)$ attains the minimum on the right hand side of (1b) for every (t, i), it follows that

$$W_t(i) = \min_{u \in U} \max_{v \in V} \left\{ c_t(i,u,v) + \sum_{j=1}^{I} P_{ij}^t(u,v) V_{t+1}(j) \right\} \quad 1 \le i \le I. \tag{9}$$

Comparing (9) and (1b), we obtain that $V_t = W_t$, thereby completing the induction. Hence, we have shown through (8) that

$$\max_{h \in H_{\mathrm{DR}}} J_{g^*,h}(i) = \max_{h \in H_{\mathrm{MN}}} J_{g^*,h}(i) = V_0(i) \text{ for } 1 \le i \le I. \tag{10}$$

Clearly, this implies that

$$\min_{g\in G_R}\max_{h\in H_{\rm DR}} J_{g,h}(i) \quad \leq \quad V_0(i). \tag{11}$$

Now we will show that we can also reverse the inequality in (11). Suppose that the maximizer uses the policy $\widehat{h} \in H_{\rm DMN}$ defined in (2b). Then the system evolves according to the transition probabilities,

$$Prob(x_{t+1}=j|x^t,u^t,v^t) \quad = \quad P^t_{x_t,j}[u_t,\widehat{h}_t(x_t,u_t)].$$

For convenience redefine

$$r^t_{ij}(u) \quad := \quad P^t_{ij}[u,\widehat{h}_t(i,u)], \tag{12}$$

$$d_t(i,u) \quad := \quad c_t[i,u,\widehat{h}_t(i,u)], \tag{13}$$

and the problem for the minimizer reduces to that of minimizing

$$E^g\left[\sum_{t=0}^{N-1} d_t(x_t,u_t)+c_N(x_N)|x_0=i\right],$$

over all $g\in G_R$, for the system with transition probabilities given by (12).

By standard dynamic programming we know, however, that

$$\min_{g\in G_R} E^g\left[\sum_{t=0}^{N-1} d_t(x_t,u_t)+c_N(x_N)|x_0=i\right] \quad = \quad W_0(i),$$

where this time $\{W_N, W_{N-1},\ldots,W_0\}$ are recursively defined by

$$W_N(i) \quad := \quad c_N(i) \tag{14a}$$

$$W_t(i) \quad := \quad \min_{u\in U}\left\{d_t(i,u)+\sum_{j=1}^{I} r^t_{ij}(u)W_{t+1}(j)\right\}. \tag{14b}$$

Moreover, if for each $t\in\{0,1,\ldots,N-1\}$ and $1\leq i\leq I$, $u=g_t(i)$ attains the minimum on the right-hand side of (14b), then $g:=\{g_0,g_1,\ldots,g_{N-1}\}\in G_{\rm MN}$ is a minimizing policy. Hence

$$\min_{g\in G_R} J_{g,\widehat{h}}(i) \quad = \quad \min_{g\in G_{\rm MN}} J_{g,\widehat{h}}(i) = W_0(i). \tag{15}$$

Now we claim that $W_t(i)=V_t(i)$ for $1\leq i\leq I$. The proof is by induction. Clearly, from (14a) and (1a), we see that $W_N=V_N$. Suppose

that $W_{t+1} = V_{t+1}$. Then substituting V_{t+1} for W_{t+1}, and substituting for $r_{ij}^t(u)$ and $d_t(i,u)$ from (12), (13) in (14b), we get

$$W_t(i) \quad = \quad \min_{u \in U} \left\{ c_t[i,u,\widehat{h}_t(i,u)] + \sum_{j=1}^{I} P_{ij}^t[u,\widehat{h}_t(i,u)] V_{t+1}(j) \right\}. \qquad (16)$$

However, since as defined in (2b), $v = \widehat{h}_t(i,u)$ attains the inner maximum on the right-hand side of (1b) for all (i,u), it follows that

$$W_t(i) \quad = \quad \min_{u \in U} \max_{v \in V} \left\{ c_t(i,u,v) + \sum_{j=1}^{I} P_{ij}^t(u,v) V_{t+1}(j) \right\}. \qquad (17)$$

Comparing (17) and (1b), we see that $V_t = W_t$, completing the induction.

Hence we have proved that

$$\min_{g \in G_R} J_{g,\widehat{h}}(i) = \min_{g \in G_{\rm MN}} J_{g,\widehat{h}}(i) = V_0(i) \quad \text{for} \quad 1 \le i \le I, \qquad (18)$$

showing in particular that,

$$\min_{g \in G_R} \max_{h \in H_{\rm DR}} J_{g,h}(i) \ge V_0(i) \quad \text{for} \quad 1 \le i \le I. \qquad (19)$$

From (11) and (19), (2a) is proved. Moreover (10) proves (2b), while (18) proves (2b).

To show (2c) we first note as before that for each fixed $g \in G_{\rm MN}$, given *any* $\bar{h} \in H_{\rm DR}$, there exists an $h \in H_R$ which does as well against g as $\bar{h}$, i.e., for every $g \in G_{\rm MN}$ and $\bar{h} \in H_{\rm DR}$, there exists $h \in H_R$ with $J_{g,\bar{h}}(i) = J_{g,h}(i)$, $1 \le i \le I$. Hence,

$$\max_{\bar{h} \in H_{\rm DR}} J_{g,\bar{h}}(i) = \max_{h \in H_R} J_{g,h}(i) \qquad 1 \le i \le I \text{ for every } g \in G_{\rm MN}. \qquad (20)$$

Now we can use standard dynamic programming, just as earlier, to show that the maximum on the right-hand side of (20) is achieved by $h(g) \in H_{\rm MN}$ defined as in (2b). Hence

$$\max_{\bar{h} \in H{\rm DR}} J_{g,\bar{h}}(i) = J_{g,h(g)}(i) \qquad 1 \le i \le I \text{ for every } g \in G_{\rm MN}.$$

This clearly proves (2c). The last result (2d) is a simple consequence of (2c) and (2b).

Thus the proof of the theorem is completed.

3 Deterministic Systems

Consider now a system evolving according to

$$x_{t+1} \quad = \quad f_t(x_t, u_t, v_t) \tag{21}$$

where $x_t \in \{1, \ldots, I\}$, $u_t \in U$, $v_t \in V$, and U, V are finite sets. The cost criterion is the same as in the previous section.

This class of systems is easily seen to be a special case of the class of probabilistic systems considered in the previous section if one defines

$$P_{ij}^t[f_t(i,u,v),u,v] \quad := \quad 1.$$

However, for deterministic systems, another class of policies is also of interest for the maximizer. Let $H_0 \subseteq H_{\rm DR}$ be the class of all policies $h = \{h_0, h_1, \ldots, h_{N-1}\}$, where each $h_t \in V$. When such an h is adopted, the maximizer chooses $v_t := h_t$ for each t. (The subscript 0 in H_0 stands for "open-loop".)

For the minimizer also, define $G_N \subseteq G_R$ as the set of policies $g = \{g_0, g_1, \ldots, g_{N-1}\}$ where each

$$g_t : (x^t, u^{t-1}) \mapsto U.$$

When such a g is adopted, the minimizer chooses $u_t = g_t(x^t, u^{t-1})$ at each time t. (The subscript N stands for "nonrandomized".)

Note that

$$G_{\rm MN} \subseteq G_N. \tag{22}$$

For deterministic systems as in (21), in addition to the results of Theorem 1, we also have the following result.

Theorem 2. *If the system evolves deterministically as in (21), then*

$$\min_{g \in G_N} \max_{h \in H_0} J_{g,h}(i) \quad = \quad \overline{J}(i) \text{for} 1 \leq i \leq I.$$

Proof: For each fixed $g \in G_N$, the system evolves according to

$$x_{t+1} \quad = \quad f_t[x_t, g_t(x^t, u^{t-1}), v_t].$$

Hence we have,

$$x_{t+1} \quad = \quad s_t(x^t, v_t),$$

for some functions $\{s_0, \ldots, s_{N-1}\}$. Standard dynamic programming for deterministic problems shows that

$$\max_{h \in H_{\rm DR}} J_{g,h}(i) \quad = \quad \max_{h \in H_0} J_{g,h}(i) \text{ for every } g \in G_N,$$

i.e., an open-loop policy is optimal. Hence

$$max_{h \in H_0} J_{g,h}(i) \geq V_0(i) \text{ for every } g \in G_N \text{ and } 1 \leq i \leq I.$$

Hence

$$\min_{g \in G_N} \max_{h \in H_{\rm DR}} J_{g,h}(i) \quad = \quad \min_{g \in G_N} \max_{h \in H_0} J_{g,h}(i) \geq V_0(i) \text{ for } 1 \leq i \leq I.$$

But noting (22) and (2b), the above inequality becomes an equality.

Remark: The fact that the min-max operations commute in some order, for some information patterns, was first pointed out and utilized by Witsenhausen [2] in the context of deterministic sampled linear systems. Friedman [3] and Bertsekas and Rhodes [4] also used this commutativity property in the context of zero-sum differential games and minimax control of uncertain plants, again for deterministic systems and for some special information patterns.

4 Concluding Remarks

The treatment here can be generalized in a straightforward way to the case of arbitrary state and control spaces X, U, V provided only that the min and max in the recursions above are attained, and that the resulting policies g^*, $\widehat{h}$, $h(g)$ etc., are appropriately measurable.

For *continuous* time systems, the notion of a *randomized* strategy is somewhat more difficult to define since randomization is allowed at *each* time t. So, ignoring randomized strategies and the technical difficulties associated with existence and uniqueness of solutions to the system differential equation when the right hand side is not smooth, the results above can be generalized in a straightforward manner to show that:

$$\begin{aligned} \min_{g \in G_N} \max_{h \in H_{DN}} J_{g,h}(x) &= \min_{g \in G_{\rm MN}} max_{h \in H_{DN}} J_{g,h}(x) \\ &= \min_{g \in G_N} \max_{h \in H_{\rm DMN}} J_{g,h}(x) \\ &= \min_{g \in G_{\rm MN}} \max_{h \in H_{\rm MN}} J_{g,h}(x) \\ &= \overline{J}(x). \end{aligned}$$

Furthermore, $\overline{J}(x)$ is given by the partial differential equation counterpart of the recursions (1a), (1b). Moreover, a $g^* \in G_{\mathrm{MN}}$ defined as in (2b), and a $\widehat{h} \in H_{\mathrm{DMN}}$ defined as in (2b), can be shown to satisfy,

$$\max_{h \in H_{DN}} J_{g^*,h}(x) \quad = \quad \min_{g \in G_N} J_{g,\widehat{h}}(x) = \overline{J}(x).$$

The details are left to the reader.

It should be noted that the results presented in this paper fail to have counterparts when the information patterns of the minimizer and the maximizer are stochastic.

5 References

1. D. Bertsekas, *Dynamic programming and stochastic control*, Academic Press, New York (1976).

2. H. Witsenhausen, "A minimax control problem for sampled linear systems," *IEEE Transactions on Automatic Control*, AC-13, pp. 5-21. (1968).

3. A. Friedman, *Differential games*, Wiley, New York, (1971).

4. D. Bertsekas and I. Rhodes, "Sufficiently informative functions and the minimax feedback control of uncertain dynamic systems," *IEEE Transactions on Automatic Control*, AC-18, pp. 117-124 (1973).