

Since here $Q_d = 0$, it follows from (37) that $\nu(g) = k_g$ and, thus, an estimate for ν will be

$$\sup_{\alpha_{jk}} k_g \leq \inf_{u_2 \in U_2} I(0, 0, u_2), \quad j = 1, \dots, s_k, \quad k = 1, 2, \dots, \quad (51)$$

where the coefficient α_{jk} are subjected to the constraints given by Eqs. (45) and (46).

Note that the positive integers $\{s_k\}$ are not determined yet. A simple choice of these integers will be $s_k = 1$, $k = 1, 2, \dots$. By using this sequence, we see that Eq. (50) yields the following inequality:

$$\frac{\sqrt{2} \{ \int_0^1 Q_0^2(x) dx - a_0 \pi \delta^2 \int_0^T | \sum_{k=1}^{\infty} (-1)^k k \alpha_{0k} (1 - t/T) | dt \}}{[\sum_{k=1}^{\infty} \alpha_{0k}^2 \int_0^T (1/T + a_0 \pi^2 k^2 (1 - t/T)^2) dt]^{1/2}} \leq \inf_{u_2 \in U_2} I(0, 0, u_2), \quad (52)$$

where α_{0k} are given by (45).

A similar scheme can be applied for computing lower bounds on $I(f, u_1, u_2)$ for System B.

4. Conclusions

From the two duality theorems (Theorems 2.1 and 2.2) derived in this paper, estimates for lower bounds on the cost functional for System A or B can be computed. Actually, for any $g \in V$, $\nu(g)$ given by Eq. (37) is a lower bound on $I(f, u_1, u_2)$ for System A; and, for any $g \in W$, $\nu(g)$ given by Eq. (38) is a lower bound on $I(f, u_1, u_2)$ for System B. These follow directly from Ineq. (31).

The scheme for computing lower bounds on the cost functional, suggested in Section 3, may be applicable, but it is not claimed here that this is the best scheme. Actually, one can represent elements of the set V (or W) in forms other than that of (49), and by this other schemes would have to be suggested.

References

1. YAVIN, Y., *Lower Bounds on the Cost Functional for a Class of Distributed Systems*, Paper presented at the IFAC Symposium on the Control of Distributed Parameter Systems, Banff, Canada, 1971.
2. KRABS, W., *Duality in Nonlinear Approximation*, Journal of Approximation Theory, Vol. 2, pp. 136-151, 1969.

Additional Aspects of the Stackelberg Strategy in Nonzero-Sum Games¹

M. SIMAAN² AND J. B. CRUZ, JR.³

Communicated by Y. C. Ho

Abstract. The Stackelberg strategy in nonzero-sum games is a reasonable solution concept for games where, either due to lack of information on the part of one player about the performance function of the other, or due to different speeds in computing the strategies, or due to differences in size or strength, one player dominates the entire game by imposing a solution which is favorable to himself. This paper discusses some properties of this solution concept when the players use controls that are functions of the state variables of the game in addition to time. The difficulties in determining such controls are also pointed out. A simple two-stage finite state discrete game is used to illustrate these properties.

1. Introduction

The Stackelberg solution of a two-player nonzero-sum game (Refs. 1-3) assumes that the roles of the players are different. There is a leader and there is a follower. The follower conforms to the policies of the leader by allowing him to determine his strategy first. The leader foresees this and, in effect, controls the entire system.

Let U_1 and U_2 be the sets of admissible controls for Players 1 and 2,

¹ This work was supported in part by the U.S. Air Force under Grant No. AFOSR-68-1579D, in part by NSF under Grant No. GK-36276, and in part by the Joint Services Electronics Program under Contract No. DAAB-07-72-C-0259 with the Coordinated Science Laboratory, University of Illinois, Urbana, Illinois.

² Research Associate, Coordinated Science Laboratory and Department of Electrical Engineering, University of Illinois, Urbana, Illinois.

³ Professor, Coordinated Science Laboratory and Department of Electrical Engineering, University of Illinois, Urbana, Illinois.

respectively, and let $J_1(u_1, u_2)$ and $J_2(u_1, u_2)$ be their corresponding cost functions. If there exists a mapping $T: U_2 \rightarrow U_1$ such that⁴

$$J_1(Tu_2, u_2) \leq J_1(u_1, u_2) \quad \forall u_1 \in U_1 \quad (1)$$

for every $u_2 \in U_2$, then the set

$$D_1 = \{(u_1, u_2) \in U_1 \times U_2: u_1 = Tu_2 \forall u_2 \in U_2\} \quad (2)$$

is called the rational reaction set for Player 1 when Player 2 is the leader. Furthermore, if there is a $(u_{1s2}, u_{2s2}) \in D_1$ such that

$$J_2(u_{1s2}, u_{2s2}) \leq J_2(u_1, u_2) \quad \forall (u_1, u_2) \in D_1, \quad (3)$$

then (u_{1s2}, u_{2s2}) is called a Stackelberg strategy pair when Player 2 is the leader. When Player 1 is the leader, the rational reaction set for Player 2 and the Stackelberg solution are denoted by D_2 and (u_{1s1}, u_{2s1}) , respectively. It is clear that, if D_1 and D_2 intersect, then their common element (u_{1N}, u_{2N}) is the Nash solution of the game. In this case, it follows (Ref. 2) that

$$J_2(u_{1s2}, u_{2s2}) \leq J_2(u_{1N}, u_{2N})$$

when Player 2 is the leader and that

$$J_1(u_{1s1}, u_{2s1}) \leq J_1(u_{1N}, u_{2N})$$

when Player 1 is the leader.

The properties of the open-loop Stackelberg solution for a class of linear quadratic games were discussed in Ref. 2. In this paper, additional properties of this solution are obtained. It is shown that, unlike the case of closed-loop Nash controls (Refs. 4-5), dynamic programming cannot be used to calculate the closed-loop Stackelberg controls. To differentiate closed-loop Stackelberg controls from controls obtained via dynamic programming, the latter are called Stackelberg feedback strategies. Both closed-loop Stackelberg controls and Stackelberg feedback strategies have attractive properties that are discussed via a simple two-stage finite state game. The difficulties in deriving the necessary conditions for the existence of the closed-loop Stackelberg controls and feedback Stackelberg strategies are pointed out and, finally, necessary conditions for the existence of feedback Stackelberg strategies for a class of discrete multistage games are derived.

⁴ It is clear, from the definition of T , that only problems where, for every $u_2 \in U_2$, there corresponds only one element $Tu_2 \in U_1$ such that (1) is satisfied are considered in this paper.

2. Illustrative Example

There are several forms according to which controls in dynamic games can be selected. There are controls that are functions of the time only and known as open-loop controls, and there are controls that are functions of the time and the state of the game as well, and these are known as closed-loop controls. In order to illustrate how the selection of such controls is done, let us consider the following simple two-stage finite state game.⁵

Example 2.1. Consider the game shown in Fig. 1. At every stage and from every state, each player has a choice between two possible controls, 0 and 1. After decisions have been made, the transition and costs borne by the players are shown in Fig. 1, where the first entries in the encircled quantities are the costs borne by Player 1 and the second entries are those borne by Player 2. The subscripts or superscripts

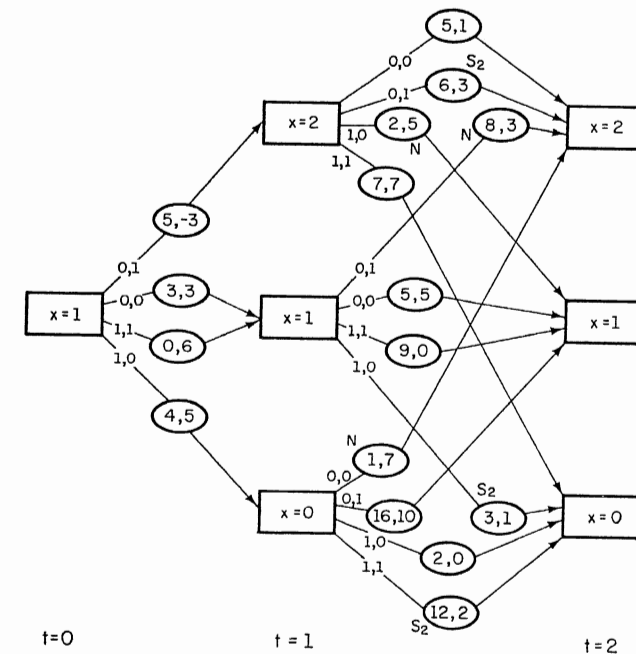


Fig. 1. A two-stage discrete game.

⁵ A similar example was considered in Ref. 5 for the Nash solution.

Table 1

	o_{11}	o_{12}	o_{13}	o_{14}
$u_1(0)$	0	0	1	1
$u_1(1)$	0	1	0	1

o and c will be used to denote open-loop and closed-loop quantities. Consider first the open-loop controls for this game.

(i) *Open-loop controls.* Assume that, before the start of the game, the players have to commit themselves to controls $u_1(t)$ and $u_2(t)$ that are functions of time only; then, each player has four such possible functions to choose from. In other words, the sets of admissible open-loop controls are $U_1^o = \{o_{1j}; j = 1, \dots, 4\}$ and $U_2^o = \{o_{2j}; j = 1, \dots, 4\}$, where o_{ij} are obtained from Table 1 for $i = 1, 2$.

The game with these admissible controls can be represented as a bimatrix game, as shown in Fig. 2. Suppose that Player 2 is the leader; then, for every $o_{2j} \in U_2^o$, Player 1 will choose a control in U_1^o that minimizes J_1 . Thus, the rational reaction set D_{1o} can be easily determined as follows:

$$D_{1o} = \{(o_{13}, o_{21}), (o_{11}, o_{22}), (o_{14}, o_{23}), (o_{13}, o_{24})\},$$

and the Stackelberg strategy is (o_{11}, o_{22}) . The corresponding trajectory is $x(1) = 1$ and $x(2) = 2$, and the costs are $J_{1s2}^o = 11$ and $J_{2s2}^o = 6$. Similarly, the rational reaction set D_{2o} is

$$D_{2o} = \{(o_{11}, o_{23}), (o_{12}, o_{23}), (o_{13}, o_{24}), (o_{14}, o_{21})\},$$

		Player 2			
		o_{21}	o_{22}	o_{23}	o_{24}
Player 1	o_{11}	8, 8	(11, 6)	10, -2	11, 0
	o_{12}	6, 4	12, 3	7, 2	12, 4
	o_{13}	5, 12	20, 15	5, 11	(8, 9)
	o_{14}	(6, 5)	16, 7	3, 7	9, 6

Open-Loop Stackelberg with 2 as Leader (points to (11,6))
 Open-Loop Stackelberg with 1 as Leader (points to (6,5))
 Open-Loop Nash (points to (8,9))

Fig. 2. Open-loop bimatrix game for Example 2.1.

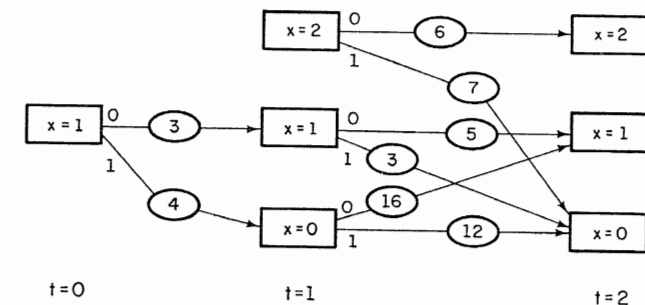
Table 2

	c_{i1}	c_{i2}	c_{i3}	c_{i4}	c_{i5}	c_{i6}	c_{i7}	c_{i8}	c_{i9}	c_{i10}	c_{i11}	c_{i12}	c_{i13}	c_{i14}	c_{i15}	c_{i16}
$u_i(0,1)$	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
$u_i(1,2)$	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
$u_i(1,1)$	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
$u_i(1,0)$	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1

and the Stackelberg strategy when Player 1 is the leader is (o_{14}, o_{21}) , leading to the trajectory $x(1) = 0$ and $x(2) = 0$ and the costs $J_{1s1}^o = 6$ and $J_{2s1}^o = 5$. Furthermore, the Nash solution, which is the common element in D_{1o} and D_{2o} , is (o_{13}, o_{24}) ; its trajectory is $x(1) = 1$ and $x(2) = 2$, and the costs are $J_{1N}^o = 8$ and $J_{2N}^o = 9$. Consider, next, the closed-loop controls.

(ii) *Closed-loop controls.* Assume that the players are restricted to announce, before the start of the game, control laws $u_1(t, x)$ and $u_2(t, x)$ that are functions of the time and the state of the game. The actual values of their controls can then be determined only while the game is played, once the actual value of the state at each time is known. Such controls are called closed-loop controls. In this game, there are 16 such choices for each player, and the sets of closed-loop admissible controls are $U_1^c = \{c_{1j}; j = 1, \dots, 16\}$ and $U_2^c = \{c_{2j}; j = 1, \dots, 16\}$, where c_{ij} are as in Table 2 for $i = 1, 2$.

For every c_{2j} that Player 2 may choose, Player 1 will have to solve an optimization problem, and his corresponding optimal closed-loop control can be easily obtained (e.g., via dynamic programming). For example, if Player 2 chooses c_{26} , then the optimization problem for Player 1 is shown in Fig. 3, and it is easily seen that c_{14} is his optimal

Fig. 3. Optimization problem for Player 1 when $u_2(t, x) = c_{26}$.

closed-loop control. If this procedure is repeated for all $u_2(x, t) \in U_2^c$, the rational reaction set D_{1c} for Player 1 is obtained as follows:

$$D_{1c} = \{(c_{115}, c_{21}), (c_{18}, c_{22}), (c_{113}, c_{23}), (c_{16}, c_{24}), (c_{111}, c_{25}), (c_{14}, c_{26}), \\ (c_{19}, c_{27}), (c_{12}, c_{28}), (c_{115}, c_{29}), (c_{116}, c_{210}), (c_{15}, c_{211}), (c_{16}, c_{212}), \\ (c_{111}, c_{213}), (c_{112}, c_{214}), (c_{11}, c_{215}), (c_{12}, c_{216})\}.$$

Following the same procedure, the set D_{2c} is obtained as follows:

$$D_{2c} = \{(c_{11}, c_{211}), (c_{12}, c_{211}), (c_{13}, c_{211}), (c_{14}, c_{211}), (c_{15}, c_{211}), (c_{16}, c_{211}), \\ (c_{17}, c_{211}), (c_{18}, c_{211}), (c_{19}, c_{211}), (c_{110}, c_{23}), (c_{111}, c_{211}), \\ (c_{112}, c_{23}), (c_{113}, c_{211}), (c_{114}, c_{23}), (c_{115}, c_{211}), (c_{116}, c_{23})\}.$$

There are two closed-loop Stackelberg controls with Player 2 as leader, (c_{15}, c_{211}) and (c_{16}, c_{212}) , both leading to the same costs $J_{1s2}^c = 7$ and $J_{2s2}^c = 2$ and to the same trajectory $x(1) = 2$ and $x(2) = 1$. Taking $D_{1c} \cap D_{2c}$, there is only one closed-loop Nash control pair, (c_{15}, c_{211}) , giving the same costs $J_{1N}^c = 7$ and $J_{2N}^c = 2$ and the same⁶ trajectory $x(1) = 2$ and $x(2) = 1$.

Thus, as in the Nash solution, open-loop and closed-loop Stackelberg solutions are generally different. However, since the Stackelberg and Nash controls are selected from the same spaces, whether $U_1^o \times U_2^o$ or $U_1^c \times U_2^c$, it is always true that the leader is better off in the Stackelberg solution than in the Nash solution.

(iii) *Feedback strategies.* The procedure described above for calculating the closed-loop Stackelberg controls is quite lengthy, especially in games with a large number of states, stages and controls, as is usually the case. One approach for simplifying this computation is to use the following dynamic programming technique, which is similar to the one described in Ref. 5 for the Nash solution. At time $t = 1$, there are three possible states. From every state, the transition to the next stage $t = 2$ is a 2×2 matrix game whose Nash and Stackelberg controls (with 2 as leader) are calculated and marked by N and S_2 in Fig. 1. Eliminating all other trajectories except for those marked by S_2 , we see that, at stage $t = 0$, there are two choices of controls for each player, resulting in the 2×2 matrix game shown

⁶ The coincidence of one closed-loop Stackelberg solution with Player 2 as leader and the closed-loop Nash solution in this example is only accidental. In fact, it can be easily checked that, when Player 1 is the leader, the closed-loop Stackelberg solution is different from the closed-loop Nash solution.

		Player 2	
		0	1
Player 1	0	(6, 4)	11, 0
	1	16, 7	3, 7

(a) Stackelberg

(b) Nash

Fig. 4. Stackelberg and Nash feedback strategies via dynamic programming.

in Fig. 4a. The Stackelberg strategy computed from this bimatrix game is encircled on Fig. 4a and is identified as (c_{14}, c_{26}) . Similarly (Ref. 5), eliminating all trajectories except for those marked by N , we see that the Nash controls calculated from the resulting bimatrix game at $t = 0$ and shown in Fig. 4b are identified as (c_{15}, c_{211}) . Therefore, we readily note that this dynamic programming technique, when used for calculating the Stackelberg strategies, did not lead to the closed-loop Stackelberg controls obtained in (ii), while when used for the Nash controls it did indeed lead to the closed-loop Nash controls. It is therefore necessary to differentiate between these two types of controls. For this reason, we shall call the controls obtained via dynamic programming approach feedback strategies; and the subscripts or superscript f will be used to denote related quantities. Thus, the Stackelberg feedback strategies for this example are (c_{14}, c_{26}) with trajectory $x(1) = 1$, $x(2) = 0$ and costs $J_{1s2}^f = 6$ and $J_{2s2}^f = 4$, and the Nash feedback strategies are (c_{15}, c_{211}) with trajectory $x(1) = 2$ and $x(2) = 1$ and costs $J_{1N}^f = 7$ and $J_{2N}^f = 2$ which are identical to the closed-loop Nash controls. Another conclusion which can be drawn from the above computations is that the property that the leader is better off in the Stackelberg solution than in the Nash solution is no longer true when feedback strategies are considered (note that, in this example, $J_{2s2}^f > J_{2N}^f$). Naturally, this is due to the fact that the trajectories eliminated at $t = 1$ are different in both cases.

Although this example is quite simple, it illustrates the difficulties encountered in determining the closed-loop Stackelberg controls in nonzero-sum dynamic games. The principle of optimality (Ref. 6), which in effect is the tool that simplifies the computational procedures, does not generalize to the Stackelberg solution as it does to the Nash solution. In fact, it can be easily checked that neither the open-loop nor the closed-loop Stackelberg solutions in the previous example have the Stackelberg property for the game resulting from the transition from $t = 1$ to $t = 2$ (i.e., starting at $t = 1$). At $x(1) = 2$, which is on the trajectory of the closed-loop Stackelberg solution for the game starting at $t = 0$, the remaining Stackelberg solution is $(1, 0)$, leading

to $J_1 = 2$ and $J_2 = 5$, while the closed-loop Stackelberg control for the game starting at $t = 1$ gives the controls $(0, 1)$, leading to $J_1 = 6$ and $J_2 = 3$. These results will now be generalized, and the equivalence between the closed-loop Nash controls and the Nash feedback strategies will be proven.

3. Stackelberg Solutions in Differential Games

The differences among the open-loop, closed-loop, and feedback controls are due to the fact that they are selected from different sets of admissible controls. A definition of these sets in differential games is therefore necessary. Let $[t_0, t_f]$ be the interval of time over which the game is defined. Let $\tau \in [t_0, t_f]$ and let $U_{1\tau}^o$ and $U_{2\tau}^o$ denote the sets of all admissible open-loop (for example, measurable functions) controls on $[\tau, t_f]$ for Players 1 and 2, respectively. Now, define the set

$$X_{\xi, \tau} = \{x(t): \dot{x} = f(x, t, u_1, u_2), t \in [\tau, t_f], x(\tau) = \xi, u_1 \in U_{1\tau}^o, u_2 \in U_{2\tau}^o\},$$

and consider the following performance indices defined on $[\tau, t_f]$:

$$J_i(\xi, \tau, u_1, u_2) = K_i(x(t_f)) + \int_{\tau}^{t_f} L_i(x, t, u_1, u_2) dt, \quad i = 1, 2, \quad (4)$$

with $u_1 \in U_{1\tau}^o$, $u_2 \in U_{2\tau}^o$ and $x(t)$ satisfying

$$\dot{x} = f(x, t, u_1, u_2) \quad x(\tau) = \xi, \quad t \in [\tau, t_f]. \quad (5)$$

Clearly, the open-loop admissible control sets for the game starting at t_0 are $U_{1t_0}^o$ and $U_{2t_0}^o$, and the necessary conditions for the existence of Stackelberg strategies in these sets are easily obtained as follows. The rational reaction set D_{1o} , when Player 2 is the leader, is composed of $(u_1(t), u_2(t)) \in U_{1t_0}^o \times U_{2t_0}^o$ satisfying the following necessary conditions:

$$\dot{x} = f(x, t, u_1(t), u_2(t)), \quad x(t_0) = x_0, \quad (6)$$

$$p' = -\partial H_1 / \partial x, \quad p'(t_f) = \partial K_1(x(t_f)) / \partial x(t_f), \quad (7)$$

$$0 = \partial H_1 / \partial u_1, \quad (8)$$

where

$$H_1(x, t, u_1, u_2, p) \triangleq L_1(x, t, u_1, u_2) + p'f(x, t, u_1, u_2), \quad (9)$$

and the pair $(u_{1s2}(t), u_{2s2}(t)) \in D_{1o}$ that minimizes $J_2(x_0, t_0, u_1, u_2)$

subject to (6)–(9) as constraints can be shown, simply by using variational techniques, to satisfy the following necessary conditions⁷:

$$\dot{\lambda}_1' = -\partial H_2 / \partial x, \quad \lambda_1'(t_f) = \partial K_2(x(t_f)) / \partial x(t_f) - \lambda_2'(t_f) [\partial^2 K_1(x(t_f)) / \partial x(t_f)^2], \quad (10)$$

$$\dot{\lambda}_2' = -\partial H_2 / \partial p, \quad \lambda_2(t_0) = 0, \quad (11)$$

$$0 = \partial H_2 / \partial u_1 = \partial H_2 / \partial u_2, \quad (12)$$

where

$$H_2(x, t, p_1, u_1, u_2, \lambda_1, \lambda_2, \lambda_3) \triangleq L_2(x, t, u_1, u_2) + \lambda_1' f(x, t, u_1, u_2) + \lambda_2' (-\partial H_1 / \partial x)' + \lambda_3' (\partial H_1 / \partial u_1)'. \quad (13)$$

When $K_1(x(t_f)) = 0$ and $K_2(x(t_f)) = 0$, these conditions become similar to those obtained in Ref. 3.

The closed-loop admissible control sets $U_{1t_0}^c$ and $U_{2t_0}^c$ at t_0 , from which the players select their closed-loop controls before the start of the game are defined as follows:

$$U_{it_0}^c = \{u_i(t, x(t)): [t_0, t_f] \times X_{\xi, t_0} \rightarrow U_{it_0}^o\}, \quad i = 1, 2.$$

These are the sets of all functions of t and $x(t)$ defined on $[t_0, t_f]$ such that the corresponding solution $x(t)$ of (5) with τ replaced by t_0 , when substituted back in $u_i(t, x(t))$, produces a function of time which belongs to the space $U_{it_0}^o$. The necessary conditions for the existence of a closed-loop Stackelberg solution are not easily obtained by using variational techniques as in the case of the open-loop solution. The closed-loop rational reaction set D_{1c} when Player 2 is the leader is composed of the pairs $(u_1(t, x(t)), u_2(t, x(t))) \in U_{1t_0}^c \times U_{2t_0}^c$ satisfying the following necessary conditions:

$$\dot{x} = f(x, t, u_1(t, x(t)), u_2(t, x(t))), \quad x(t_0) = x_0, \quad (14)$$

$$p' = -\partial H_1 / \partial x - \partial H_1 / \partial u_2 [\partial u_2(t, x(t)) / \partial x(t)], \quad p'(t_f) = \partial K_1(x(t_f)) / \partial x(t_f), \quad (15)$$

$$0 = \partial H_1 / \partial u_1, \quad (16)$$

where H_1 is as defined in (9). Because of the term including $\partial u_2(t, x(t)) / \partial x(t)$ in (15), it can be easily shown that variational techniques fail to produce a candidate in $U_{1t_0}^c \times U_{2t_0}^c$ which minimizes $J_2(x_0, t_0, u_1, u_2)$ when subject to (14)–(16) as constraints. Other

⁷ Note that $\partial H_2 / \partial x = (\nabla_x H_2)'$ is a row vector. The same notation is used for all partial derivatives of a scalar function with respect to a vector.

possible approaches for performing this minimization are not investigated in this paper and remain unexplored for future research.

We now consider the case of feedback strategies. Let the interval $[t_0, t_f]$ be divided into N equal subintervals of time. Consider an arbitrary interval $[\tau_j, \tau_{j+1}]$, and assume that the Stackelberg feedback strategies (in the sense that is being defined) $u_{1s2}^f(t, x(t))$ and $u_{2s2}^f(t, x(t))$ for the game defined over the interval $[\tau_{j+1}, t_f]$ have been obtained. Let $J_i(x(\tau_{j+1}), \tau_{j+1}, u_{1s2}^f, u_{2s2}^f)$, $i = 1, 2$, be the costs corresponding to these strategies. Furthermore, let $U_{1\tau_j}^f$ and $U_{2\tau_j}^f$ denote the subsets of $U_{1\tau_j}^c$ and $U_{2\tau_j}^c$ obtained by eliminating all controls that do not coincide with $u_{1s2}^f(t, x(t))$ and $u_{2s2}^f(t, x(t))$, respectively, over $[\tau_{j+1}, t_f]$. Now, consider the game defined over $[\tau_j, t_f]$, where the state equation is as in (5) and the performance indices (4) are reduced to

$$J_i(x(\tau_j), \tau_j, u_1, u_2) = J_i(x(\tau_{j+1}), \tau_{j+1}, u_{1s2}^f, u_{2s2}^f) + \int_{\tau_j}^{\tau_{j+1}} L_i(x, t, u_1, u_2) dt, \quad i = 1, 2, \quad (17)$$

where $u_1 \in U_{1\tau_j}^f$ and $u_2 \in U_{2\tau_j}^f$. Let $(u_{1s2}^f(t, x(t)), u_{2s2}^f(t, x(t))) \in U_{1\tau_j}^f \times U_{2\tau_j}^f$ be the Stackelberg strategies⁸ for the game defined by (17) and (5) with τ replaced by τ_j . Now, if there exist such strategies for all $\tau_j \in [t_0, t_f]$ when this procedure is repeated backward in time until $U_{1t_0}^f$ and $U_{2t_0}^f$ are obtained, and if their limit as $|\tau_{j+1} - \tau_j| \rightarrow 0$ for all j (or as $N \rightarrow \infty$) exist, then the resulting strategies in $U_{1t_0}^f \times U_{2t_0}^f$ are called Stackelberg feedback strategies.

In a similar way (simply by replacing the word Stackelberg by the word Nash) as above, the Nash feedback strategies $u_{1N}^f(t, x(t))$ and $u_{2N}^f(t, x(t))$ are defined. We now prove the equivalence between the closed-loop Nash controls and the Nash feedback strategies.⁹

Proposition 3.1. The closed-loop Nash control pair $(u_{1N}^c(t, x(t)), u_{2N}^c(t, x(t)))$ are equal to the Nash feedback strategies $(u_{1N}^f(t, x(t)), u_{2N}^f(t, x(t)))$.

Proof. The proof of this proposition is straightforward. Let (ξ, τ)

⁸ It should be clear that these strategies defined over $[\tau_j, t_f]$ and those obtained originally for $[\tau_{j+1}, t_f]$ coincide over $[\tau_{j+1}, t_f]$; hence, in order to avoid proliferation of notation, they are denoted by the same expressions $u_{is2}^f(t, x(t))$.

⁹ In this proposition, we assume that the Nash closed-loop and feedback strategies exist and are unique.

be any point in the state space. Every pair $(\hat{u}_1(t, x(t)), \hat{u}_2(t, x(t))) \in D_{1c}$ has the property that, if $\hat{u}_2(t, x(t))$ is restricted to the time interval $[\tau, t_f]$, then $\hat{u}_1(t, x(t))$, which is obtained by solving an ordinary optimization (rather than a game) problem where it is known that dynamic programming or any other method lead to the same optimal control, when restricted to $[\tau, t_f]$, is also optimal for the optimization problem resulting from (4)–(5) when $u_2 = \hat{u}_2(t, x(t))$. Similarly, every pair $(\hat{u}_1(t, x(t)), \hat{u}_2(t, x(t))) \in D_{2c}$ has the same property for $\hat{u}_2(t, x(t))$ when $\hat{u}_1(t, x(t))$ is restricted to $[\tau, t_f]$. Therefore, since $(u_{1N}^c(t, x(t)), u_{2N}^c(t, x(t))) \in D_{1c} \cap D_{2c}$, it follows that these properties hold simultaneously for both controls and, hence, it is also a Nash feedback strategy.

This proposition justifies the simultaneous use in Ref. 5 of the closed-loop Nash controls and the Nash feedback strategies as being identical solutions. However, because the closed-loop Stackelberg control pair lies on the rational reaction set D_{1c} of the follower and not generally in the intersection of D_{1c} and D_{2c} , it cannot be concluded that it coincides with the Stackelberg feedback strategies. In other words, Proposition 3.1 simply says that, at any time during the course of play and from any allowable state at that instant, if the players recalculate their closed-loop Nash controls, these controls will be the same as the remaining part of the controls calculated initially. This, however, is not true in the case of the closed-loop Stackelberg controls. In fact, when dynamic programming is used, several controls in closed-loop form are eliminated from consideration at t_0 because they do not possess this optimality (Nash or Stackelberg) property from all other possible starting points (ξ, τ) . Thus, because of Proposition 3.1, in the case of the Nash solution, the closed-loop Nash controls will not be among those controls that are eliminated at t_0 ; while, in the case of the Stackelberg solution, the closed-loop Stackelberg controls most likely will. Furthermore, the closed-loop controls eliminated at t_0 in the Nash solution are not the same as those eliminated at t_0 in the Stackelberg solution (i.e., $U_{1t_0}^f$ is not the same in both cases; for example, see Fig. 4) and, hence, it is no longer possible to conclude that the Stackelberg feedback solution is more beneficial to the leader than the Nash feedback solution.

In order to illustrate the dynamic programming technique described earlier, the necessary conditions for the existence of Stackelberg feedback strategies for a class of discrete games, where dynamic programming is more easily applied, are obtained in the following section.

4. Stackelberg Feedback Strategies in Discrete Games

Consider the multistage discrete game defined by

$$x(l+1) = f(x(l), l, u_1(l), u_2(l)), \quad x(0) = x_0, \quad l = 0, \dots, N-1, \quad (18)$$

where the state $x(l)$ and the decision (control) variables $u_1(l)$ and $u_2(l)$ are n -dimensional, m_1 -dimensional, and m_2 -dimensional vectors of real numbers for all $l = 0, 1, \dots, N-1$. Let the cost functionals defined over stages k, \dots, N be of the form

$$J_i(x(k), k, u_1, u_2) = K_i(x(N)) + \sum_{l=k}^{N-1} L_i(x(l), l, u_1(l), u_2(l)), \quad (19)$$

where $u_i = (u_i(k), \dots, u_i(N-1))$, $i = 1, 2$. Suppose that Player 2 is the leader, and assume that the transition from the k th to the $(k+1)$ th stage is under consideration. Let u_{1s2}^f and u_{2s2}^f be the Stackelberg feedback strategies for the game starting at stage $k+1$ and ending at stage N , and let $V_i(x(k+1), k+1) = J_i(x(k+1), k+1, u_{1s2}^f, u_{2s2}^f)$, $i = 1, 2$, be the costs corresponding to these strategies and obtained from

$$V_i(x(k+1), k+1) = K_i(x(N)) + \sum_{l=k+1}^{N-1} L_i(x(l), l, u_{1s2}^f(l, x(l)), u_{2s2}^f(l, x(l))), \quad i = 1, 2, \quad (20)$$

where $x(l)$ is obtained from

$$x(l+1) = f(x(l), l, u_{1s2}^f(l, x(l)), u_{2s2}^f(l, x(l))), \quad l = k+1, \dots, N-1. \quad (21)$$

The cost functionals for the game defined over stages k, \dots, N can therefore be written as

$$J_i(x(k), k, u_1(k), u_2(k)) = V_i(x(k+1), k+1) + L_i(x(k), k, u_1(k), u_2(k)). \quad (22)$$

Assuming that no constraints exist on the controls, we see that, for a fixed $u_2(k)$, the follower (Player 1) determines his optimal $u_1(k)$ (assuming that it exists) as a function of $u_2(k)$ and $x(k)$ from

$$\begin{aligned} & \partial J_1(x(k), k, u_1(k), u_2(k)) / \partial u_1(k) \\ &= [\partial V_1(x(k+1), k+1) / \partial x(k+1)] [\partial f / \partial u_1(k)] + \partial L_1 / \partial u_1(k) = 0. \end{aligned} \quad (23)$$

The leader, therefore, must minimize $J_2(x(k), k, u_1(k), u_2(k))$ subject to

(23) as constraint. The necessary conditions for this minimization are

$$\begin{aligned} & \partial L_2 / \partial u_i(k) + [\partial V_2(x(k+1), k+1) / \partial x(k+1)] [\partial f / \partial u_i(k)] + (\partial / \partial u_i(k)) \\ & \times [\lambda'(k) [\partial L_1 / \partial u_1(k)]' + \lambda'(k) [\partial f / \partial u_1(k)]' [\partial V_1(x(k+1), k+1) / \partial x(k+1)]'] = 0, \\ & i = 1, 2, \end{aligned} \quad (24)$$

where $\lambda(k)$ is an m_1 -dimensional Lagrange multiplier. When (23)–(24) are solved, $u_{1s2}^f(k, x(k))$ and $V_i(x(k), k)$ for the game from k to N are obtained. This procedure is then repeated until the starting stage is reached.¹⁰ The boundary conditions for (23)–(24) are given at the terminal stage by $V_i(x(N), N) = K_i(x(N))$; $i = 1, 2$. Note that, when stage $k = 0$ is reached, all feedback strategies defined over the stages 1 to $N-1$ will have been eliminated except for those that are feedback Stackelberg strategies for the game defined on stages 1 to $N-1$.

5. Conclusions

Several properties of the closed-loop Stackelberg controls have been investigated and the difficulties (conceptual and computational) encountered in their determination have been pointed out. Unlike the closed-loop Nash controls, it has been shown that the closed-loop Stackelberg controls cannot be obtained by applying dynamic programming techniques. The solution obtained via dynamic programming, called Stackelberg feedback strategies, has the property that, at any instant of time during the course of play and from any allowable state at that instant of time, it provides the leader with the best choice of control (in the sense of Stackelberg), regardless of previous decisions and with the assumption that Stackelberg feedback strategies will be used for the remainder of the interval of play. On the other hand, if the starting time is fixed, the leader's closed-loop Stackelberg control is the best control law (among all other admissible closed-loop controls) that he can announce prior to the start of the game, but it does not have this same desirable property from any other starting time. Since the leader is virtually the only decision maker in the Stackelberg solution, the decision of choosing between a closed-loop or a feedback strategy depends generally on whether t_0 is known or not and whether the system parameters are completely certain or not.

¹⁰ In discrete games, the Stackelberg feedback strategies also have the property that the players may announce their controls (the leader first) stage by stage, once the current value of the state vector at each stage is known. Such development of information is called successive (Ref. 7).

References

1. VON STACKELBERG, H., *The Theory of the Market Economy*, Oxford University, Oxford, England, 1952.
2. SIMAAN, M., and CRUZ, J. B., JR., *On the Stackelberg Strategy in Nonzero-Sum Games*, Journal of Optimization Theory and Applications, Vol. 11, No. 5, 1973.
3. CHEN, C. I., and CRUZ, J. B., JR., *Stackelberg Solution for Two-Person Games with Biased Information Patterns*, IEEE Transactions on Automatic Control, Vol. AC-17, No. 6, 1972.
4. STARR, A. W., and HO, Y. C., *Nonzero-Sum Differential Games*, Journal of Optimization Theory and Applications, Vol. 3, No. 3, 1969.
5. STARR, A. W., and HO, Y. C., *Further Properties of Nonzero-Sum Differential Games*, Journal of Optimization Theory and Applications, Vol. 3, No. 4, 1969.
6. BELLMAN, R., *Dynamic Programming*, Princeton University Press, Princeton, New Jersey, 1957.
7. PROPOI, A. I., *Minimax Problems of Control Under Successively Acquired Information*, Automation and Remote Control, Vol. 31, No. 1, 1970.

Algorithms for Discounted Stochastic Games¹S. S. RAO,² R. CHANDRASEKARAN,³ AND K. P. K. NAIR⁴

Communicated by R. A. Howard

Abstract. In this paper, a two-person zero-sum discounted stochastic game with a finite state space is considered. The movement of the game from state to state is jointly controlled by the two players with a finite number of alternatives available to each player in each of the states. We present two convergent algorithms for arriving at minimax strategies for the players and the value of the game. The two algorithms are compared with respect to computational efficiency. Finally, a possible extension to nonzero sum stochastic game is suggested.

1. Introduction

In a stochastic game, the play proceeds in a sequence of steps or transitions assumed to take place at every unit time interval and, at each transition, the play is said to be in some state i chosen from a finite set of states. In each state, a finite number of alternatives is available to each of the players. While in each state there is a zero-sum reward and the game moves to other states probabilistically, both the reward and transition probabilities depend on the actions taken by the players in the state. Shapley (Ref. 1) introduced this concept of stochastic game but with a stop probability in each state, derived a

¹ This research was supported in part by funds allocated to the Department of Operations Research, School of Management, Case Western Reserve University under Contract No. DAHC 19-68-C-0007 (Project Themis) with the U.S. Army Research Office, Durham, North Carolina. The authors thank the referees for their valuable suggestions.

² Associate Professor, Department of Operations Research, Case Western Reserve University, Cleveland, Ohio.

³ Associate Professor, Department of Operations Research, Case Western Reserve University, Cleveland, Ohio.

⁴ Associate Professor, Department of Business Administration, University of New Brunswick, Fredericton, New Brunswick, Canada.