A general theory for Stackelberg games with partial state information *

Tamer BAŞAR
Coordinated Science Laboratory and Department of Electrical Engineering, University of Illinois, Urbana, IL 61801, U.S.A.

Received 1 March 1981.
Revised 11 August 1981
Communicated by Y.C. Ho

This paper presents a general method for derivation of a tight lower bound on the Stackelberg cost of the leader in general two-person deterministic dynamic games with partial dynamic state information. The method converts the original dynamic Stackelberg problem into two open-loop optimization problems whose solutions can readily be obtained using the standard techniques of optimization and optimal control theory. When applied to the class of linear-quadratic dynamic games with partial dynamic information, defined on general Hilbert spaces, each one of these open-loop optimization problems becomes a quadratic programming problem with linear constraints, thus allowing for an explicit computation of the Stackelberg cost value. The paper also includes a specific example, illustrating application of these results on a discrete-time linear-quadratic dynamic game wherein the leader has access to partial state information.

1. Introduction

The Stackelberg solution concept, first introduced by H. von Stackelberg [1] for static games, and then extended and applied to dynamic games in the papers [2–4], has recently attracted considerable attention in the literature after the development of an indirect method to obtain the solution of dynamic games with closed-loop information structure. The essence of this indirect method introduced in [5] and [6] for the Stackelberg solution of two-person deterministic dynamic games is the following: First find a two-person team problem whose optimal team cost provides a tight lower bound for the leader's Stackelberg cost in the dynamic game, and then determine a particular closed-loop representation of the leader's optimal feedback solution in the team problem, which will force the follower to the strategy that minimizes the team cost, even though he is actually minimizing his own cost functional.

Within the context of linear-quadratic dynamic games defined in discrete time, two different team problems have been introduced in [6], depending on whether the follower acts at the last stage of the game or not. For the latter case the related team problem is the one which is determined completely by the leader's cost function so that, under certain conditions on the parameters of the game, the leader can force the follower to a strategy which jointly minimizes his (the leader's) own cost function. An appropriate strategy for the leader to accomplish this is a linear one-step memory representation of his feedback team strategy, which can be determined recursively. Other, more complicated (nonlinear and nondifferentiable), one-step memory representations are also possible [7], and these sometimes extend the region of applicability (in the parameter space) of this indirect approach.

In the former case, on the other hand (i.e. when the follower also acts at the last stage of the game), the related team problem has a 'reduced' cost function which is obtained from the leader's cost function by taking into account the optimal response of the follower at the last stage. This optimal response is incorporated in the follower's cost function so that we now also have a 'reduced' cost function for the follower. In this new game the follower does not act at the last stage, and therefore the problem becomes similar to that discussed earlier. For the same class of (linear-quadratic) dynamic game problems, Tolwinski has obtained in [7] a different class of 'reduced team' problems (or rather the minimum value of these reduced team costs) for the case when the follower has

* Part of this work was performed while the author was with the Department of Applied Mathematics, Marmara Research Institute, Gebze-Kocaeli, Turkey; and partly it was sponsored by the U.S. Air Force under Grant AFOSR-78-3633 and the National Science Foundation under Grant ECS-79-19396.

North-Holland Publishing Company
Large Scale Systems 3 (1982) 47–56

extra degrees of freedom in influencing the state variable, which cannot be 'detected' by the leader through his observation of the state.

The indirect method of [6] was then applied to continuous-time linear-quadratic differential games [8–10] and to stochastic two-person games [11,12] and later extended to many (3)-player deterministic [13] and stochastic [14] games. For an interpretation of these results from the viewpoint of incentives, and with applications in economics and social choice theory, we refer the reader to [15].

In the present paper, we first discuss (in the next section) derivation of the Stackelberg solution for general dynamic games as a closed-loop representation of the team solution (that necessarily involves memory, whenever a 'complete detectability condition' is satisfied, in which case the Stackelberg cost (of the leader) coincides with the minimum value of the leader's cost function. We then present a general method which leads to a tight lower bound on the Stackelberg cost of the leader whenever the complete detectability condition is not satisfied. This indirect method is valid for a sufficiently large class of dynamic games (such as linear, nonlinear, discrete-time, continuous-time, etc.) and when the dynamic information available to the leader is not necessarily full state information, and it involves the solution of two open-loop optimization problems (cf. Sections 3 and 4). When this method is applied to the special class of linear quadratic dynamic games with partial dynamic information, the Stackelberg cost value (of the leader) is obtained through the solution of two quadratic programming problems with linear constraints, defined on general Hilbert spaces (cf. Section 5).

2. Stackelberg solution of dynamic games with perfect state information

Consider a two-person dynamic game in normal form, described by the cost functionals $J_1(\gamma_1, \gamma_2)$ and $J_2(\gamma_1, \gamma_2)$ where the strategies γ_1 and γ_2 belong to a priori determined strategy speces Γ_1 and Γ_2, respectively. If Player 1 (P1) is the leader and P2 is the follower, an extended definition of the Stackelberg solution which also accounts for nonunique responses of the follower is as follows [6]:

For each $\gamma_1 \in \Gamma_1$, first introduce the *rational reaction set* of the follower by

$$R(\gamma_1) = \{\gamma_2^0 \in \Gamma_2 : J_2(\gamma_1, \gamma_2^0) \leq J_2(\gamma_1, \gamma_2), \quad \forall \gamma_2 \in \Gamma_2\}. \tag{1}$$

Then a strategy $\gamma_1^* \in \Gamma_1$ is said to be a Stackelberg strategy of the leader if

$$\sup J_1(\gamma_1^*, \gamma_2) \leq \sup J_1(\gamma_1, \gamma_2), \quad \forall \gamma_1 \in \Gamma_1, \tag{2}$$

where the supremum is taken over $\gamma_2 \in R(\gamma_1^*)$ on the left-hand side (LHS) of (2) and over $\gamma_2 \in R(\gamma_1)$ on the right-hand side (RHS). The quantity on the LHS is known as the *Stackelberg cost* of the leader, and is equivalently defined as

$$J_1^* = \inf_{\Gamma_1} \sup_{R(\gamma_1)} J_1(\gamma_1, \gamma_2) = \sup_{R(\gamma_1^*)} J_1(\gamma_1^*, \gamma_2). \tag{3}$$

It is a well-established fact in the literature that any direct approach towards the solution of this problem meets with formidable difficulties, whenever the leader has access to dynamic information [3–6]. An alternative proposed in [6] to solve this problem involves two steps: (1) Determine (through some indirect methods) the Stackelberg cost J_1^* of the leader, and (2) find a strategy for the leader that leads to realization of that cost level by also taking into account rational responses of the follower.

One natural lower bound for J_1^* is clearly the infimum of J_1 over the product set $\Gamma_1 \times \Gamma_2$, i.e.

$$J_1^* \geq \inf_{\Gamma_1 \times \Gamma_2} J_1(\gamma_1, \gamma_2). \tag{4}$$

Let us now assume that this infimum is actually achieved (as a minimum), and let $\Gamma_i^t \subset \Gamma_i$ $(i = 1,2)$ be equivalence classes of strategies γ_i^t with the property

$$J^{1^t} = \inf_{\Gamma_1 \times \Gamma_2} J_1(\gamma_1, \gamma_2) = J_1(\gamma_1^t, \gamma_2^t), \quad \forall \gamma_i^t \in \Gamma_i^t, \quad i = 1,2. \tag{5}$$

These equivalence classes are, in general, infinite sets if the dynamic game under consideration is an infinite game. Then we have the following proposition.

Proposition 2.1. *If there exists a* $\gamma_1^* \in \Gamma_1^t$ *such that*

$$\left\{\gamma_2^0 \in \Gamma_2 : \inf_{\Gamma_2} J_2(\gamma_1^*, \gamma_2) = J_2(\gamma_1^*, \gamma_2^0)\right\} \subset \Gamma_2^t,$$

then γ_1^* constitutes a Stackelberg strategy for the leader.

If the requirement of Proposition 2.1 is satisfied, then this implies that the leader can force the follower to play in such a way so as to globally minimize his own cost functional, even though he (the follower) intends to minimize his (possibly quite different) cost functional. Let us now investigate the possibilities for the leader to adopt such a powerful position. Firstly, let us introduce some notation and terminology:

Let η_1 and η_2 denote the information available to P1 and P2, respectively, during the course of the game, and let u and v denote their control (decision) vectors (for P1 and P2, respectively), taking values in some abstract vector spaces. Then, we clearly have the relations

$$u=\gamma_1(\eta_1), \qquad v=\gamma_2(\eta_2). \tag{6}$$

Since, in general, $\eta_1=\eta_1(u, v, \xi)$ and $\eta_2=\eta_2(u, v, \xi)$, where ξ is some primitive variable, (6) can be written, under resonable assumptions, as

$$u=\tilde{\gamma}_1(\xi), \qquad v=\tilde{\gamma}_2(\xi) \tag{7}$$

which we may call *open-loop* relations, as opposed to *closed-loop* relations (6). Now what makes Γ_i^t (defined earlier by (5)) in general an infinite set, is the redundancy of information embedded in Γ_i. For a minimization problem which is free of such informational redundancy, such as an open-loop optimization problem, one may expect at most a finite number of elements in each corresponding equivalence class; here, we assume that there exists only one in the form (7) (which is justified if J_1 is strictly convex in terms of the decision vectors). Then there is only one strategy in Γ_i^t that depends only on ξ, which we denote by $\tilde{\gamma}_i^t$. Furthermore, for the class of deterministic dynamic games, which we shall henceforth be dealing with, we may take ξ to be the initial state of the game (denoted x_0) which is known by both parties.

The assumptions that we have made so far are reasonable and justifiable. Now we make two assumptions which are not that reasonable and which may fail for some important classes of deterministic dynamic games:

Assumption A. *The leader can detect* exactly *the control value of the follower through his information, i.e. the inverse image of $\eta_1(u,\cdot, x_0)$ exists for every u and x_0.*

Before stating the second assumption, we first introduce the following terminology: If $(t_0, t_f]$[1] denotes the interval on which the dynamic game is defined, and s is chosen such that $t_0<s<t_f$, let γ_{1s} (respectively, γ_{1-s}) denote the restriction of a strategy $\gamma_1 \in \Gamma_1$ to the interval $(t_0, s]$ (respectively, $(s, t_f]$) with the corresponding restricted strategy set being Γ_{1s} (respectively, Γ_{1-s}). Further write $\gamma_1=[\gamma_{1s}, \gamma_{1-s}]$.

Assumption B. *For the dynamic game defined on $(t_0, t_f]$ there exists an s $(t_0<s<t_f)$ such that* P2 *does not act throughout the interval $(s, t_f]$, and there exists a $\hat{\gamma}_{1-s} \in \Gamma_{1-s}$ such that, with $\hat{\gamma}_1 \triangleq [\tilde{\gamma}_{1s}^t, \hat{\gamma}_{1-s}]$,*

$$\inf_{\Gamma_2} J_2(\hat{\gamma}_1, \gamma_2) > J_2(\tilde{\gamma}_1^t, \tilde{\gamma}_2^t). \tag{8}$$

(Note that Assumption B implies that the leader has the ability to make the follower's cost function sufficiently large, by his actions, after he completely 'detects' the follower's control value.)

Proposition 2.2. *Under Assumptions* A *and* B, *there exists a Stackelberg strategy for the leader satisfying the hypothesis of Proposition* 2.1, *namely*

$$\gamma_1^*(\eta_1)=\begin{cases}\gamma_1^t(\eta_1) & \text{if } v=\tilde{\gamma}_2^t(x_0),\\ \hat{\gamma}_1(\eta_1) & \text{otherwise},\end{cases} \tag{9}$$

where γ_1^t is any element of Γ_1^t, and $\hat{\gamma}_1$ is defined by (8).

Proof. Firstly, note that $\gamma_1^* \in \Gamma_1^t$, by Assumption A and the fact that it minimizes $J_1(\gamma_1, \tilde{\gamma}_2^t)$ over Γ_1. Secondly, it forces the follower to play $v=\tilde{\gamma}_2^t$, since otherwise the cost incurred to him under the strategy $u=\hat{\gamma}^1(\eta_1)$ is larger, by Assumption B.

Before discussing the implications of this result, let us consider a specific example to illustrate the underlying idea.

Example 2.3: Consider the 2-stage scalar deterministic dynamic game described (in extensive form) by the state equations

$$\begin{aligned} x_1&=x_0-v,\\ x_2&=x_1-u \end{aligned} \tag{10}$$

and cost functionals[2]

$$J_1=(x_2)^2+2u^2+\beta v^2, \quad \beta \geqslant 0, \tag{11a}$$

[1] Here $(t_0, t_f]$ stands for both the continuous and discrete time interval.

[2] Here, by an abuse of notation, we denote the cost functionals of the game in extensive form again by J_1 and J_2.

$$J_2 = (x_2)^2 + v^2 \tag{11b}$$

where P1 (the leader) controls u as a function of x_0 and x_1, i.e. $u = \gamma_1(x_1, x_0)$, and P2 controls v as a function of x_0, i.e. $v = \gamma_2(x_0)$; hence, $\gamma_1 \in \Gamma_1$ is any measurable mapping of $\mathbb{R} \times \mathbb{R}$ into $\mathbb{R}$ and $\gamma_2 \in \Gamma_2$ is any measurable mapping of $\mathbb{R}$ into $\mathbb{R}$.

The minimization problem leads, in this case, to the equivalence classes

$$\Gamma_1^t = \{\gamma_1 \in \Gamma_1 : \gamma_1(x_1^t, x_0) = [\beta/(3\beta+2)]\, x_0\}, \tag{12a}$$

$$\Gamma_2^t = \{\gamma_2^t(x_0) = [2/(3\beta+2)]\, x_0\} \tag{12b}$$

where

$$x_1^t \triangleq [3\beta/(3\beta+2)]\, x_0. \tag{12c}$$

Note that Γ_1^t is an infinite set, whereas Γ_2^t is a singleton. The corresponding (absolute minimum) value of J_1 is

$$J_1^t = [2\beta/(3\beta+2)]\, x_0^2.$$

Now, using the notation introduced earlier, the optimum open-loop strategies $(\tilde{\gamma}_1^t, \tilde{\gamma}_2^t)$ associated with this problem are also unique and are given by

$$\tilde{\gamma}_1^t(x_0) = [\beta/(3\beta+2)]\, x_0, \tag{13a}$$

$$\tilde{\gamma}_2^t(x_0) = [2/(3\beta+2)]\, x_0. \tag{13b}$$

Furthermore, Assumption A is clearly satisfied, and Assumption B is also satisfied, which can be seen by simply taking $\hat{\gamma}_1(x_0) = -kx_0$ for $k > 0$. Hence, the hypotheses of Proposition 2.2 are satisfied, and we can therefore declare (9) as a Stackelberg strategy for the leader, which can be written, in view of the relation $v = x_0 - x_1$, as

$$\gamma_1^*(x_1, x_0) = \begin{cases} [\beta/(3\beta+2)]\, x_0 & \text{if } x_1 = x_1^t, \\ -kx_0 & \text{otherwise,} \end{cases} \tag{14}$$

where x_1^t is defined by (12c) and k is any positive number.[3] It should be noted that here we have taken the unique open-loop element $\tilde{\gamma}_1^t$ of Γ_1^t as a particular strategy γ_1^t; other possibilities (such as $\gamma_1^t(x_1, x_0) = \frac{1}{3}x_1$) also exist—there is in fact an infinite number of such choices.

To recapitulate,

(1) For the dynamic game described by (10)–(11) and under the closed-loop information pattern, a Stackelberg solution exists for all values of β.

(2) The Stackelberg cost of the leader is equal to the global minimum of his cost function (obtained by cooperative action of the follower), even though the follower seeks to minimize his own (different) cost functional.

(3) The leader's Stackelberg strategy is a *representation* of his optimal feedback strategy in the related team problem described by (5), and it necessarily involves *memory* (a feedback strategy for P1, which is independent of x_0, cannot be a Stackelberg strategy).

(4) The Stackelberg strategy of P1 is *nonunique* (parameterized, in this case, by a positive scalar k), but the optimal response of P2 to all these strategies of P1 is unique (and independent of k).

(5) The Stackelberg strategy of the leader, as given by (14), is not even continuous; this, however, does not rule out the possibility for existence of continuous (and differentiable) Stackelberg strategies.

(6) The strategy (14) may be viewed as a *threat strategy* on part of the leader, since he essentially threatens the follower to worsen his cost if he does not play the optimal team strategy (13b).

The general solution presented in Proposition 2.2 may lack some desirable properties; such as continuity, differentiability, as exemplified above within the context of Example 2,3, but it nevertheless provides a verification of the possibility that the easily computable bound J_1^t can in fact be realized as the Stackelberg cost J_1^*. Hence, once it is known that the bound in (4) is tight, under the given information structure, then one can refer back to Proposition 2.1 and seek to determine a $\gamma^{1*} \in \Gamma_1^t$ with some more appealing properties than those possessed by (14). Carrying out such a procedure for the dynamic game of Example 2.3, if we restrict our analysis to linear strategies for P1, we first observe that any such strategy in Γ_1^t can be written as (cf. [6])

$$\gamma_1(x_1, x_0) = \tfrac{1}{3}x_1 + p\left(x_1 - \frac{3\beta}{3\beta+2}x_0\right) \tag{15}$$

where p is a scalar parameter. To determine the value of p for which (15) constitutes a Stackelberg strategy for the leader, we have to substitute (15) into J_2, minimize the resulting expression over $\gamma_2 \in \Gamma_2$ by also utilizing (10), and compare the argument of this minimization problem with the strategy (13b). Such an analysis readily leads to

[3] This result is in fact valid whenever $k > -1 + \sqrt{8/13}$.

the unique value

$$p = \tfrac{2}{3} - 1/\beta \tag{16}$$

provided that $\beta \neq 0$. Hence, the linear strategy (15), together with the specific value of p as given above, also constitutes a Stackelberg strategy for the leader, provided that $\beta \neq 0$. For the case $\beta = 0$, it can be shown by mimicking the analysis of [16, p. 23] that no continuously differentiable Stackelberg strategy exists for the leader; so, in this case, we have to utilize nondiferentiable strategies like (14).

A natural question that comes into mind now is whether J_1^t always constitutes a tight lower bound for J_1^*; the reply is "no" since Assumptions A and B that lead to Proposition 2.2 may, at times, not be satisfied. One such case occurs in discrete-time finite stage games when the follower also acts at the last stage; for such games the leader cannot detect (through his state observation) the follower's action at the last stage and therefore cannot enforce his team solution. However, this problem is also tractable, as discussed in [6], since it is possible to define a 'reduced' team problem for the leader (by taking into account possible rational responses of the follower at the last stage) whose optimal cost may provide a tight lower bound for J_1^*. For an illustration of this approach, within the context of an extended version of Example 2.3, we refer the reader to the conference paper [22].

There is still a large class of deterministic dynamic game problems, however, for which a reduction to a new game satisfying Assumptions A and B is not possible, in particular if the leader does not have access to full state information. For such problems a new technique has to be developed in order to determine the Stackelberg cost J_1^*, and this is the topic of the remaining sections. Before dealing with the general problem, we first consider, in the next section, a 2-dimensional 2-stage dynamic game to illustrate the difficulties encountered in obtaining J_1^* when the leader has access to partial dynamic information, and to motivate and introduce our new approach to this clas of problems.

3. A dynamic game with partial state information

Consider the 2-stage 2-player dynamic game characterized by the 2-dimensional state equation

$$x_1 = x_0 - v, \qquad x_0 = \xi(1, 1)',$$
$$x_2 = x_1 - (1,1)'u, \tag{17}$$

and cost functionals

$$J_1 = x_2' x_2 + u^2 + v' \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} v, \tag{18a}$$

$$J_2 = x_2' \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} x_2 + v' \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} v. \tag{18b}$$

Here x_i is the 2-dimensional state vector, ξ is a known scalar parameter, v is the 2-dimensional control vector of P2 (the follower) who acts at stage 0, and u is the scalar control variable of P1 (the leader). The value of the initial state x_0, as given above, is known by both players, and in addition P1 has access to the following partial state information at stage 1:

$$y = (1, 1)\, x_1. \tag{19}$$

Therefore, P1's strategies will be of the form $u = \gamma_1(y)$ where $\gamma_1 \in \Gamma_1$ is any measurable mapping of $\mathbb{R}$ into $\mathbb{R}$, whereas P2's strategies will be constants, i.e. $v = \gamma_2 \in \mathbb{R}^2 = \Gamma_2$.

Let us first investigate whether Assumptions A and B are satisfied for this problem. In view of the discussion thet led to Proposition 2.2. in Section 2, we first observe that the leader can actually threaten the follower to make his cost function arbitrarily large (by simply picking $\gamma_1(y) = k$, where k is arbitrarily large); therefore Assumption B is satisfied for this problem. But for this threat to be implementable, P1 should be able to detect, at stage 1, what P2 has actually done at stage 0 since, otherwise, a strategy of the form (9) does not exist —and this is not the case here. The leader has a single channel which does not permit him to observe the exact value of v; i.e. Assumption A, henceforth referred to as the *complete detectability condition*, is not satisfied here. All P1 can observe is the value of y (which he may wish to force to a given or precalculated desired value), but there are several choices for v which would lead to the same value for y. In fact, it can be shown that (see [22]) the leader cannot force the follower's actions to lead to the team value $y^t = \frac{21}{13}\,\xi$ of y, and therefore, the minimum team cost cannot be achieved in this problem, i.e.

$$\tfrac{7}{13}\xi^2 = J_1^t < J_1^*. \tag{20}$$

Another point worth emphasizing here is that we

cannot know a priori whether there exists a Stackelberg strategy for the leader in this problem, but the Stackelberg cost J_1^* is always a well-defined finite quantity. Therefore, our task now is to determine J_1^*.

An indirect approach to determine J_1^*

Let us assume, at the outset, that there exists a Stackelberg strategy for the leader, denoted by $\gamma_1^* \in \Gamma_1$, and an optimal response of the follower, denoted by $v^* \in \Gamma_2$, leading to the Stackelberg cost $J_1(\gamma_1^*, v^*) = J_1^*$; furthermore, let us denote the realized value of y, after utilization of $v = v^*$, by y^*. Since

$$v^* = \arg \min_{v \in \mathbb{R}^2} J_2(\gamma_1^*(y), v), \tag{21}$$

and since the only dependence of $u = \gamma_1^*(y)$ on v is through y, the optimization problem (21) faced by the follower is equivalent to the constrained optimization problem

$$v^* = \arg \min_{v \in S} J_2(\gamma_1^*(y^*), v) \tag{22}$$

where

$$S = \{v \in \mathbb{R}^2 : y = 2\xi - (1, 1)\, v = y^*\}. \tag{23}$$

Letting $\gamma_1^*(y^*) = u^*$ (which is a constant, yet to be determined), we solve this minimization problem and obtain v^* uniquely as a function of u^* and y^*:

$$v^* = \tfrac{1}{6}(u^* - 4y^* + 7\xi,\ -u^* - 2y^* + 5\xi)'. \tag{24}$$

Now, by substituting this into J_1, we arrive at the expression

$$\begin{aligned} J_1(u^*, v^*) = F(u^*, y^*) &= \left(\tfrac{2}{3}y^* - \tfrac{7}{6}u^* - \tfrac{1}{6}\xi\right)^2 \\ &+ \left(\tfrac{1}{3}y^* - \tfrac{5}{6}u^* + \tfrac{1}{6}\xi\right)^2 + u^{*2} \\ &+ 2\left(\tfrac{1}{6}u^* - \tfrac{2}{3}y^* + \tfrac{7}{6}\xi\right)^2 \\ &+ \left(\tfrac{1}{6}u^* + \tfrac{1}{3}y^* - \tfrac{5}{6}\xi\right)^2, \end{aligned} \tag{25}$$

and the minimum value of $F(u^*, y^*)$ over all u^* and y^* should equal J_1^*. Carrying out the minimization we obtain

$$y^* \cong 1.6393\xi, \qquad u^* \cong 0.5409\xi, \qquad J_1^* \cong 0.5409\xi^2. \tag{26}$$

Therefore, this dynamic game indeed admits a Stackelberg strategy for the leader, which is

$$\gamma_1^*(y) = \begin{cases} 0.5409\xi & \text{if } y = y^* \cong 1.6393\xi, \\ k \gg 0 & \text{otherwise,} \end{cases} \tag{27}$$

and the Stackelberg cost is

$$J_1^* \cong 0.5409\xi^2. \tag{28}$$

The specific range of values for k can actually be determined by substituting $\gamma_1 = k$ into J_2, minimizing the resulting expression over $v \in \mathbb{R}^2$ and comparing this minimum value with $J_2(\gamma_1^*, v^*)$, where $v^* \cong \xi(0.1640, 0.1967)'$.

Several remarks are now in order:

(1) The dynamic game of this section, with partial state information, admits a Stackelberg strategy for the leader, which is, however, not an element of Γ_1^t.

(2) The Stackelberg cost of the leader is higher than the global minimum value of his cost functional.

(3) The leader's Stackelberg strategy necessarily involves memory and it may again be viewed as a threat strategy; but now the leader can threaten the follower only in terms of the single (scalar) variable that he observes and not through the control vector of P2—this being the main reason behind the strict inequality $J_1^* > J_1^t$.

(4) (27) does not characterize the complete class of strategies that lead to the Stackelberg cost; but any Stackelberg strategy must satisfy the side condition $\gamma_1^*(y^*) \cong 0.5409\xi$. It is quite possible that some more appealing (such as continuous, differentiable,...) strategies could constitute a Stackelberg solution for the problem, realizing the bound J_1^*, but this can only be discovered by trying out specific structural forms satisfying the foregoing side condition; one such form being the linear strategy $\gamma_1(y) = 0.5409\xi + p(y - y^*)$, which in this case happens to provide a Stackelberg strategy for a specific value of p.

The next section is devoted to a generalization of this indirect method (of determining J_1^*) to a sufficiently broad class of dynamic games, and subsequently, in Section 5, the method is applied to linear-quadratic dynamic games defined on general Hilbert spaces.

4. A general method to determine J_1^* for dynamic games with partial state information

Let us return to our initial formulation of Section 2, and introduce a two-person dynamic game through the cost functionals $J_1\gamma_1, \gamma_2)$, $J_2(\gamma_1, \gamma_2)$ and strategy spaces Γ_1, Γ_2. This is a deterministic

game and therefore both players have access to the initial state vector x_0 which we may take as given a priori. The players also acquire information about the evolution of the game, but for P2 (the follower) we may take his information structure open-loop since, in a deterministic Stackelberg game, dynamic state information does not bring the follower any additional advantage. For P1 (the leader), however, dynamic information is in general helpful, and we denote it in this case by ϵ. Further, denote the value of this information (which in fact takes values in a finite or infinite dimensional vector space), after both players have chosen their strategies, by y, so that we now have the loop relations

$$u=\gamma_1(y), \qquad v=\gamma_2,$$
$$y=\eta(u,v) \tag{29}$$

where dependence on the state (if any) is suppressed since it can in turn be expressed in terms of the control vectors. The vectors u, v, and y are assumed to belong, respectively, to the vector spaces U, V, and Y on which no additional structure is imposed.[4] It should be noted that, under an assumption of causality which is implicitly required in the definition of a dynamic game, there exists a function $\bar{\eta}: V \to Y$ so that (29) may be written more conveniently as

$$u=\gamma_1(y), \qquad v=\gamma_2,$$
$$y=\bar{\eta}(v). \tag{30}$$

(See [17] for a discussion on this point.)

Let us now consider the following two steps:

Step 1. Minimize $J_2(u, v)$ over $S \subset V$, where

$$S(u,y) \triangleq \{v \in V: y=\eta(u,v)\}, \quad u \in U, \quad y \in Y. \tag{31}$$

Define the (partial) optimal response set

$$R(u,y)=\{v^0 \in S: J_2(u,v^0) \leqslant J_2(u,v), \forall v \in S\}. \tag{32}$$

(Note that here u does not functionally depend on v; it is any element of U which comprises only open-loop strategies.)

Step 2. If $R(u, y)$ is a singleton, denote the corresponding unique map by $T: U \times Y \to V$, and consider minimization of $J_1(u, T(u,y))$ over $(u, y) \in U \times Y$ and subject to $y=\eta(u, T(u, y))$. Denote any minimizing solution pair by (u^*, y^*).

If $R(u,y)$ is not a singleton, define (u^*, y^*) through the inequality

$$J_1^* \triangleq \sup_{v \in R(u^*,y^*)} J_1(u^*, v) \leqslant \sup_{v \in R(u,y)} J_1(u,v) \quad \forall (u,y) \in U \times Y. \tag{33}$$

Theorem 4.1. *Let there exist a pair of* $(u^*, y^*) \in U \times Y$ *satisfying the requirements at Step* 2, *and further let there exist a* $\hat{\gamma}_1 \in \Gamma_1$ *with the property*

$$\inf_{v \in V} J_2(\hat{\gamma}_1, v) > J_2(u^*, v^*) \quad \forall v^* \in R(u^*, y^*),$$

such that the composite strategy

$$\gamma_1^*(y)=\begin{cases} u^* & \text{if } y=y^*, \\ \hat{\gamma}_1(y) & \text{otherwise}, \end{cases} \tag{34}$$

is an element of Γ_1. *Then* J_1^* *defined by* (33) *provides a tight lower bound on the Stackelberg cost of the leader.*[5]

Proof. Let us prove this result for the case of unique follower response; the other more general case is a straightforward extension.

Simply note that, under the hypothesis of the theorem, the leader, by applying the strategy (34), can force the follower to the strategy

$$v^*=T(u^*, y^*),$$

under which the realized cost value is J_1^*. Clearly the leader cannot do any better, since a strategy in the structural form (34) enables the leader to utilize his maximum power through his information $y=\eta(u, v)$, and the pair (u^*, y^*) in (33) is obtained by minimizing J_1 subject to the freedom allotted to P2 in the nondetectable (by the leader) region of V.

Remark. A significant aspect of the result established above is that the original dynamic Stackelberg game is reduced to two static (open-loop) optimization problems which can be solved by utilizing the standard techniques of optimization

[4] Y is taken here, without any loss of generality, to be the full range space of η.

[5] In the case of nonunique responses of the follower, the realized cost value for P1 might be lower than J_1^* if P2 does not utilize the supremizing strategy in (33). In the case of unique follower response, however, J_1^* is the realized cost value.

(or optimal control, if the original problem is formulated in the state space). That is in principle, by carrying out the two steps outlined prior to Theorem 4.1, the Stackelberg cost value can readily be determined. A corresponding Stackelberg strategy for the leader is (34); there may, of course, be plenty others, some of them with more appealing properties in view of our discussion in Section 3.

Two special cases

(1) When the leader has access to only open-loop information, η becomes the null-function and therefore S coincides with V, and consequently $R(u, y) = R(u)$ becomes the optimal response set of the follower in the original Stackelberg game. Step 2, then, directly determines the open-loop Stackelberg strategy of the leader in this open-loop game. Hence Steps 1 and 2 reduce to the well-known method of obtaining the open-loop Stackelberg strategy, whenever the leader's information is open-loop.

(2) If the leader has access to perfect closed-loop information and the complete detectability condition is satisfied, then η is an invertible transformation and S becomes a singleton. This implies that $R(u, y)$ also becomes a singleton determined completely by u and y, and consequently (33) becomes equivalent to

$$J_1^* = \inf_{(u,y)} \sup_{v \in R(u,y)} J_1(u, v) = \inf_{(u,v) \in U \times V} J_1(u, v);$$

that is, the leader achieves the global minimum value of J_1. Clearly, this conclusion is in agreement with our previous result on this perfect information case, i.e. Proposition 2.2.

5. Linear-quadratic dynamic games with partial dynamic information

In this section we apply Theorem 4.1 and the two steps of derivation preceding it to the general class of linear-quadratic dynamic games with partial dynamic information for the leader and no dynamic information for the follower (note that this latter assumption creates no loss of generality, as discussed earlier). The class of games under consideration is defined on general Hilbert spaces and therefore our formulation covers both continuous-time and discrete time problems.

Given a two-person linear-quadratic dynamic game in extensive form, we can eliminate the intermediate state variables and arrive (also in view of our discussion that led to (30) from (29)) at the following formulation (where we now use u_1, u_2 instead of u, v for the control variables of the players):

$$J_i(u_1, u_2) = \sum_{k,j=1,2} \langle u_k, A^i_{kj} u_j \rangle + \sum_{j=1,2} \langle u_j, l^i_j \rangle, \quad i = 1,2, \tag{35}$$

$$y = N u_2. \tag{36}$$

Here u_i belongs to U_i which is a Hilbert space with an appropriate inner product $\langle \cdot, \cdot \rangle$,[6] y is the observation of P1 and belongs to another Hilbert space Y, N is a bounded linear operator with full range so that, for any strongly positive linear operator $B: Y \to Y$ (denoted $B > 0$), $NBN^* > 0$ where N^* denotes the adjoint of N,[7] A^i_{kj} is a bounded linear aperator from U_j into U_k, with $A^i_{jj} > 0$ $\forall i, j = 1,2$, and $A^i_{ji} = 0$ for $i \neq j$, $l^i_j \in U_j$ is a known function of the initial state of the game (i.e. a deterministic quantity, since the initial state is also known), and finally

$$A^i_{ii} - A^i_{ij} \left(A^i_{jj} \right)^{-1} A^i_{ij}{}^* > 0, \quad i, j = 1,2, \quad i \neq j, \tag{37}$$

which basically makes both J_1 and J_2 strictly convex on $U_1 \times U_2$. To complete the description of the game, we have to specify the strategy space of P1, which is taken as a class of appropriate mappings γ_1 mapping Y into U_1; this, however, will not be needed in the sequel since Steps 1 and 2 basically involve 'open-loop' optimization problems. It is important, though, to remember that P1 utilizes the value of y in the actual selection of his control.

Let us now proceed with the derivation of J_1^* for this problem, by following the two steps outlined in Section 2.

Step 1. The optimization problem here is a strictly convex quadratic programming problem with linear equality constraints, which is known to admit a unique solution in U_2 (see [18, 19]). If $u_2^0 \in U_2$ denotes this solution, then there exists a

[6] For the sake of simplicity in notation, we do not differentiate between inner products defined on different Hilbert spaces.

[7] In this section, superscript 'star' is reserved for adjoint—not for optimum—with the sole exception of J_1^*.

$\lambda \in Y$ such that the first Gateaux variation of

$$L(u_2) = J_2 + \langle \lambda, Nu_2 \rangle$$

vanishes at $u_2 = u_2^0$; this is also a sufficient condition because of strict convexity. Carrying out this minimization, we obtain, also by utilizing the linear constraint (36), the unique solution

$$u_2^0(y, u_1) = By + Cu_1 + Dl_2^2 \tag{38}$$

where $B: Y \to U_2$, $C: U_1 \to U_2$, $D: U_2 \to U_2$ are bounded linear operators defined by

$$B = \left(A_{22}^2\right)^{-1} N^* \left[N\left(A_{22}^2\right)^{-1} N^*\right]^{-1}, \tag{39a}$$

$$C = -\tfrac{1}{2}[I - BN]\left(A_{22}^2\right)^{-1} A_{21}^2, \tag{39b}$$

$$D = -\tfrac{1}{2}[I - BN]\left(A_{22}^2\right)^{-1}. \tag{39c}$$

This then determines $R(u, y)$ completely, which is a singleton in this case.

Step 2. Now the problem is minimization of $J_1(u_1, u_2)$ over $U_1 \times Y$ and subject to the linear constraint (38). Substitution of (38) into $J_1(u_1, u_2)$ for u_2, leads to a quadratic function $F(u_1, y)$ which is strictly convex on $U_1 \times Y$. Therefore it admits a unique solution, which can be obtained by simply taking the Gateaux variation of F with respect to u_1 and y, separately, and setting these expressions equal to zero. We delete details of this standard procedure, and only give here the solution which we denote by (u_1^0, y^0):

$$y^0 = K^{-1} l,$$

$$u_1^0 = -K_1^{-1}\left(K_2 y^0 + \tilde{l}_1\right)$$

where $K_1: U_1 \to U_1$, $K_2: Y \to U_1$, $K: Y \to Y$ are bounded linear operators, with $K_1 > 0$, $K > 0$ defined by

$$K_1 = 2A_{11}^1 + A_{12}^1 C + C^* A_{12}^{1*} + 2C^* A_{22}^1 C,$$

$$K_2 = A_{12}^1 B + 2C^* A_{22}^1 B,$$

$$K = 2B^* A_{22}^1 B - K_2^* K_1^{-1} K_2,$$

and $l \in Y$, $\tilde{l}_1 \in U_1$ are respectively defined by

$$l = K_2^* K_1^{-1} \tilde{l}_1 - B^* l_2^1 - 2B^* A_{22}^1 D l_2^2,$$

$$\tilde{l}_1 = \left(A_{12}^1 + 2C^* A_{22}^1\right) D l_2^2 + l_1^1 + C^* l_2^1.$$

Hence the first hypothesis of Theorem 4.1 is fulfilled. The second hypothesis is also clearly fulfilled since we can find elements $\hat{\gamma}_1$ even in U_1 that make J_2 arbitrarily large, because of the strict convexity assumption. In discrete-time finite stage problems, for example, this would correspond to choosing components of the control vector at the last stage arbitrarily large (as one possibility); and in continuous-time problems this would correspond to making the control vector arbitrarily large (in norm) in a sufficiently small subinterval that also includes the terminal time. Consequently the Stackelberg cost J_1^* of the leader in this class of dynamic games is determined as

$$J_1^* = J_1\left(u_1^0, u_2^0\left(y^0, u_1^0\right)\right)$$

and a possible strategy for the leader that forces the follower to play $u_2 = u_2^0$, so that this cost level is attained, is

$$\gamma_1^0 = \begin{cases} u_1^0 & \text{if } y = y^0, \\ \hat{\gamma}_1(y) & \text{otherwise,} \end{cases}$$

where $\hat{\gamma}_1(y)$ is determined as discussed above. Other, structurally more appealing, strategies would also be possible; the only required condition on a candidate Stackelberg strategy γ_1 would be

$$\gamma_1(y^0) = u^0,$$

i.e. *any Stackelberg strategy is a closed-loop representation* [16,20] *of the open-loop policy* u_1^0 *on the equilibrium information set* (y^0) *of the leader*. In the linear class, for example, candidate strategies are given as

$$\gamma_1(y) = u_1^0 + \mathfrak{L}(y - y^0)$$

where $\mathfrak{L}: Y \to U_1$ is any linear bounded operator that is compatible with the 'control-information dependence' requirements (such as causality) of the dynamic game under consideration. Whether such an $\mathfrak{L}$ exists for a given linear-quadratic game, as well as possibilities for other structural forms, can be investigated only if we have some further structure given on the abstract dynamic game of this section; we do not pursue this point any further here.[8] For a specific application of the foregoing results and the ideas introduced in Section 4 to continuous-time differential games when the leader has access to sampled state information, the reader is referred to [21].

[8] For further elaboration of this point see the recent paper [24] which discusses existence and derivation of linear Stackelberg strategies under partial dynamic information.

6. Concluding remarks

We have presented, in this paper, a new indirect method for derivation of a tight lower bound on the Stackelberg cost value of the leader in deterministic dynamic games with partial state information. This indirect method involves solution of two open-loop optimization problems and leads, as a byproduct, to a Stackelberg strategy for the leader. This Stackelberg strategy, however, may not be that appealing since it lacks features like continuity, differentiability, etc. But a strategy that possesses such properties may be found by trying out several (appealing) structural forms under the restriction that they satisfy a certain side condition —this side condition being the open-loop value of the Stackelberg strategy on the information set of the leader. Even if a Stackelberg strategy with the sought desired features cannot be found, the bound yielded by our indirect method serves to compare performances of some simple suboptimal strategies for the leader against the best performance achievable. This seems to be a promising avenue for further research.

It is possible to extend the indirect method of this paper to many-player dynamic game problems with different levels of hierarchy and with partial state information and to stochastic dynamic game problems wherein the leader has access to partial redundant information on the follower's actions. For one class of such extensions see [23].

References

[1] H. von Stackelberg, Marktform und Gleichgewicht (Springer, Vienna, 1934).

[2] C.I. Chen and J.B. Cruz, Jr., Stackelberg solution for two-person games with biased information patterns, IEEE Trans. Automat. Control 17 (1972) 791–798.

[3] M. Simaan and J.B. Cruz, Jr., On the Stackelberg strategy in nonzero-sum Games, J. Optimization Theory Appl. 11 (1973) 533–555.

[4] M. Simaan and J.B. Cruz, Jr., Additional aspects of the Stackelberg strategy in nonzero-sum games, J. Optimization Theory Appl. 11 (1973) 613–626.

[5] T. Başar and H. Selbuz, A new approach for derivation of closed-loop Stackelberg strategies, Proc. 1978 IEEE Conf. Decision and Control, San Diego, CA (1979) 1113–1118.

[6] T. Başar and H. Selbuz, Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems, IEEE Trans. Automat. Control 24 (1979) 166–179.

[7] B. Tolwinski, Closed-loop Stackelberg solution to multistage linear-quadratic game, J. Optimization Theory Appl. 35 (1981).

[8] T. Başar and G.J. Olsder, Team-optimal closed-loop Stackelberg strategies in linear-quadratic differential games, Proc. 17th Allerton Conf. Communication, Control, Computing, October 1979.

[9] T. Başar and G.J. Olsder, Team-optimal closed-loop Stackelberg strategies in hierarchical control problems, Automatica 16 (1980) 409–414.

[10] G.P. Papavassilopoulos and J.B. Cruz, Jr., Sufficient conditions for Stackelberg and Nash strategies with memory, J. Optimization Theory Appl. 31 (2) (1980) 233–260.

[11] T. Başar, Hierarchical decision making under uncertainty, in: P.T. Liu (Ed.), Dynamic Optimization and Mathematical Economics (Plenum Press, New York, 1980) 205–221.

[12] Y.C. Ho, P.B. Luh and R. Muralidharan, Information structure, Stackelberg games and incentive controllability, IEEE Trans Automatic Control 26 (1981) 454–460.

[13] T. Başar, Equilibrium strategies in dynamic games with multi levels of hierarchy, Proc. 2nd IFAC Symposium on Large Scale Systems, Toulouse, France, 1980; ibid. Automatica 17 (1981).

[14] T. Başar, Stochastic multi-criteria decision problems with multi levels of hierarchy, IEEE Trans. Automatic Control 26 (1981) 549–553.

[15] Y.C. Ho, P.B. Luh and G.J. Olsder, A control-theoretic view on incentives, Proc. 4th Internat. Conf. Analysis and Optimization of Systems (Springer, New York, 1980).

[16] T. Başar, Information structures and equilibria in dynamic games, in: M. Aoki and A. Marzollo (Eds.), New Trends in Dynamic System Theory and Economics (Academic Press, New York, 1979) 3–55.

[17] H.S. Witsenhausen, Alternatives to the tree model for extensive games, in: J.D. Grote (Ed.), The Theory and Applications of Differential Games (Reidel, 1975) 77–84.

[18] D.G. Luenberger, Optimization by Vector Space Methods (Wiley, New York, 1969).

[19] A.V. Balakrishnan, Applied Functional Analysis (Springer, New York, 1976).

[20] T. Başar, Optimum coordination of linear interconnected systems, Large Scale Systems 1 (1980) 12–27.

[21] T. Başar, A new method for the Stackelberg solution of differential games with sampled-data state information, Proc 7th IFAC Congress, Kyoto, Japan, 1981.

[22] T. Başar, Memory strategies and a general theory for Stackelberg games with partial state information, Proc. 4th Internat. Conf. on Analysis and Optimization of Systems, Versailles, France (Springer, New York, 1980).

[23] T. Başar, Performance bounds for hierarchical systems under partial dynamic information, J. Optimization Theory Appl., to appear.

[24] Y.P. Zheny and T. Başar, Existence and derivation of optimal affine incentive schemes for Stackelberg games with partial information: A geometric approach, CSL, Internal Rept., Univ. of Illinois, Urbana, IL, 1981.