Stackelberg Strategies and Incentives in Multiperson Deterministic Decision Problems

YING-PING ZHENG, TAMER BASAR, FELLOW, IEEE, AND JOSE B. CRUZ, JR., FELLOW, IEEE

Abstract—In this paper discrete and continuous-time two-person decision problems with a hierarchical decision structure are studied and applicability and appropriateness of a function–space approach in the derivation of causal real-time implementable optimal Stackelberg (incentive) strategies under various information patterns are discussed. Results on existence and derivation of incentive strategies for dynamic games formulated in abstract inner-product spaces, in the absence of any causality restriction on the leader's policies, are first presented and then these results are extended (and specialized) in two major directions: 1) discrete-time dynamic games with informational advantage to the leader at each stage of the decision process, which involves partial observation of the follower's decisions; and derivation of multistage incentive strategies for the leader under a feedback Stackelberg solution adapted to the feedback information pattern; and 2) derivation of causal, physically realizable optimum affine Stackelberg policies for both discrete and continuous-time problems, in terms of the gradients of the cost functionals evaluated at the optimum (achievable) operating point (which is in some cases the globally minimizing solution of the leader's cost functional). The paper is concluded with some applications of the theory to important special cases, some extensions to infinite-horizon problems, and some numerical examples that further illustrate these results.

I. Introduction and a General Description of the Stackelberg Problem

A. General Introduction

THE PRESENCE of multiple decisionmakers is a common phenomenon in many large-scale decision problems, especially if they involve humanistic and socioeconomic elements. The decisionmakers may have noncommensurable, and at times conflicting, preferences, or they may have basically the same goal but may wish to decentralize the decisionmaking process in order to alleviate the heavy burden of acquiring, transmitting, and processing the excessive amount of information needed for a centralized control (Athans [1], Başar and Cruz [5]). In either case, the decisionmakers (or players, in the terminology of game theory) would have different objective functions, and acquire possibly different information in the decisionmaking process. Furthermore, there would be an order in which the decisionmakers act and/or announce their policies, and this order would either be fixed (predetermined) or determined as a consequence of the players' actions. All these factors contribute to the concept of solution to be adopted for a general multiperson decisionmaking problem, and have to be taken into account before the derivation of the solution process.

There is a growing variety of solution concepts in dynamic game theory (such as team solution, Pareto solution, Nash solution, etc.; see Başar and Cruz [5], Başar and Olsder [7]), and among these the Stackelberg (or leader–follower) strategies (Cruz [12], [13]) have recently attracted more and more attention, in both the control and economics literatures. This concept was first introduced by Von Stackelberg [26] for a class of static decision problems arising in economics. Then its dynamic version was presented in a control theoretic framework by Chen and Cruz [11] and Simaan and Cruz [24], [25]. This solution concept is especially suitable for hierarchical multilevel decision problems wherein the decisionmakers hold nonsymmetric roles in the decisionmaking process. One of the players, called the leader, occupies a higher decision level, and this superior position enables him to announce his strategy in advance and enforce it on the other players. By taking into account the optimal responses of the followers, which may be determined as the solution of some other multiperson decision problem under a specific solution concept relevant to that problem (see [18], [3]), the leader seeks the policy that leads to a most favorable outcome for him.

Such situations arise in many real world problems. In a large organization, the headquarters decisionmaker (the leader) cannot dictate every subdivision's (the follower's) task in fine detail; in its stead, it simply announces and executes appropriate strategies (policies), such as the resource allocation strategy, penalty or reward policy, the profit-sharing policy, etc., so as to induce the subdivisions to work in accordance with the interests of the entire organization ([2], [15], [16], [19], [22]). Some recent investigations have been devoted to the study and construction of such leader–follower strategies in special types of organizations. For example, a standard and efficient way for a government (the leader) to solve the water pollution prob-

Manuscript received May 21, 1983; revised June 15, 1983. The work of T. Başar and J. B. Cruz, Jr. was supported in part by the Joint Services Electronics Program under Contract N00014-79-C-0424 and in part by the Office of Naval Research under Contract N00014-82-K-0469.

Y. P. Zheng was with Coordinated Science Laboratory, University of Illinois, 1101 W. Springfield Avenue, Urbana, IL 61801. He is now with the Institute of Automation, Academia Sinica, Beijing, People's Republic of China.

T. Başar and J. B. Cruz, Jr. are with the Decision and Control Laboratory, Coordinated Science Laboratory, University of Illinois, 1101 W. Springfield Avenue, Urbana, IL 61801.

0018-9472/84/0100-0010$01.00 ©1984 IEEE

lem is to design some subsidy programs or penalty policies to encourage or induce the chemical plants (the followers) to act cooperatively. A utility company (the leader) may use a price strategy (or a pricing strategy) to induce the customers (the followers) to consume the utility resource more reasonably ([18], [20]). In a market with both free competition and government adjustment, the government (the leader) may design a strategy of adjusting the effective income of the potential buyers of the commodity so as to induce the competing duopolistic firms to cooperate and achieve a Pareto-optimal solution [23]. All of these problems can be studied in the framework of Stackelberg dynamic game theory, thus making this new field very promising in applications.

B. General Description of the Stackelberg Problem

To be more precise in our description of a Stackelberg game and the related solution concepts, let us now consider a two-person dynamic game problem with a hierarchical decision structure under which player 1 acts as the leader and player 2 as the follower. The state $x(\cdot)$ of the underlying decision process evolves according to either (in continuous time)

$$\dot{x}(t) = f(t, x(t), u(t), v(t)), \qquad t \in [0, T) \tag{1}$$

or (in discrete time)

$$x(k+1) = f(k, x(k), u(k), v(k)), \qquad k = 0, 1, \cdots, N-1, \tag{2}$$

where $u(\cdot)$ is the leader's decision variable and $v(\cdot)$ is the follower's, and they are either time-functions or time-series, belonging to the corresponding Hilbert spaces.

$$u(\cdot) \in L_2^{m_1}[0, T), \qquad v(\cdot) \in L_2^{m_2}[0, T), \quad \text{in continuous time;}$$

or

$$u(\cdot) \in l_2^{m_1}[0, N-1], \qquad v(\cdot) \in l_2^{m_2}[0, N-1], \quad \text{in discrete time.}$$

Generically, let us denote the decision variables of the leader and the follower by u and v, respectively, and the decision spaces by U and V. Furthermore, let $X = L_2^n[0, T)$ or $l_2^n[0, N-1]$ denote the state space for the process, where $x(\cdot)$ belongs, and let $Y_1 \subseteq X \times V$ and $Y_2 \subseteq X \times U$ denote the information (observation) spaces of the leader and the follower, respectively. A permissible policy (strategy) $\gamma_i \in \Gamma_i$ for player i is a Borel-measurable mapping from his observation space into his decision space, satisfying some additional regularity conditions like causality, Lipschitz continuity, etc. that will be delineated later in proper contexts. One underlying assumption here is that, with $x_0 \in \mathbb{R}^n$ fixed, to each $(\gamma_1, \gamma_2) \in \Gamma_1 \times \Gamma_2$, there corresponds a unique state trajectory $x(\cdot) \in X$ and a unique pair of cost values $\{J_1(\gamma_1, \gamma_2), J_2(\gamma_1, \gamma_2)\}$.

The Stackelberg game problem involves, in a nutshell, determination of a leader's policy $\gamma_1^* \in \Gamma_1$ satisfying

$$J_1(\gamma_1^*, T(\gamma_1^*)) = \min_{\gamma_1} J_1(\gamma_1, T(\gamma_1)) \tag{3}$$

where T: $\Gamma_1 \to \Gamma_2$ is the unique rational response mapping of the follower, i.e.,

$$T(\gamma_1) = \arg\min_{\gamma_2 \in \Gamma_2} J_2(\gamma_1, \gamma_2) \tag{4}$$

where we tacitly assumed existence of a unique solution to (4).

Even though this definition is valid for all types of information available to the players (i.e., for arbitrary Y_1 and Y_2), the derivation of the solution will depend to a great extent on the underlying information structure, as to be elucidated in the sequel.

1) Open-Loop Information Structure: The players' information comprises only the *a priori* information, e.g., the structural parameters of the problem and the initial conditions. In this case, strategies and the decision variables coincide, and are chosen as time-functions from the beginning.

Necessary conditions for the open-loop solution of Stackelberg dynamic game problems can be obtained without any conceptual difficulties, although it is rather difficult to solve them analytically or even numerically [13], [7, ch. 7].

2) Closed-Loop Information Structure: Here the leader is assumed to acquire state information with memory, i.e., elements of Y_1 are given by $y_1(t) = \{x(\tau), \tau \leqslant t\}$ or $y_1(k) = \{x(i), i = 0, \cdots, k\}$, thus leading to policies in the form $u(t) = \gamma_1(t; x(\tau), \tau \leqslant t)$ or $u_1(k) = \gamma_1(k; x(k), x(k-1), \cdots, x(0))$. The follower, on the other hand, could acquire closed-loop or open-loop information. Any direct approach towards the solution of such dynamic Stackelberg games meets with formidable difficulties, since the optimization problem (4) is "structurally" dependent on the structure of leader's strategy, that is, the follower faces an optimization problem "parameterized" by the structure of γ_1. Such "nonclassical" optimal control problems and indirect ways of obtaining the solution have been discussed in many papers; see [21], [8], [9], [6], [3], [27], and it has been shown that in certain cases the leader can achieve the global minimum value of his cost function J_1. This feature has also been established by an "indirect method" [4] under two conditions: 1) the leader can detect the follower's action (detectability); and 2) by choosing an appropriate strategy the leader is able to threaten the follower by severe punishment in case of any deviation from the desired solution trajectory (enforceability). It has been shown, moreover, that the closed-loop information is rich enough to allow for the solution to satisfy additional design specifications. One such specification involves a "robustness" feature; that is, in case of a deviation from an optimal path, not to punish the follower indefinitely at all future stages, but rather use an effective threat policy which would carry a punitive action role for only a few (two or three) stages. This aspect of the problem and its solution has been discussed in some recent papers in the literature, see e.g., [27].

3) Feedback Strategies and the Feedback Stackelberg Solution Concept: A subclass of closed-loop strategies comprises those policies that depend only on the current value of the state without memory. That is

$$y_1(t) = \{x(t)\} \quad \text{or} \quad y_1(k) = \{x(k)\}.$$

Under such feedback information pattern for the leader, the Stackelberg solution is still very difficult to obtain, and in fact, in most cases, it will not even exist if the initial state $x(0)$ is taken a variable and a solution is sought for all $x(0) \in \mathbb{R}^n$. A way to circumvent this difficulty, in the case of discrete-time problems, is to require that any subprocess-to-go is also an optimal process in the Stackelberg sense [25]. This permits the adoption of a dynamic programming type approach which involves the solution of static Stackelberg games at each stage (and in retrograde time). In comparison with the Stackelberg solution, the feedback Stackelberg solution gives only a suboptimal solution; though it has the advantage of being simpler in structure, computationally feasible and implementable. Furthermore, it has better robustness properties against noise and disturbance, since the leader can update on his policy at each stage of the decision process.

4) Incentive Strategies: As we have indicated in case 2), the leader may expect to achieve the global optimum of his cost function (the team solution), as though the follower was cooperating with him, provided that he has memory, can detect follower's actions and can announce and implement enforceable policies. In order to investigate the Stackelberg problem from this viewpoint, we include the decision $v(\cdot)$ of the follower directly in the information $y_1(\cdot)$ available to the leader, thereby allowing the leader to adopt a strategy in the form

$$u(\cdot) = \gamma_1(x(\cdot), \qquad v(\cdot)) \tag{5}$$

which explicitly displays the dependence of the leader's decision variable u on the follower's, v. Such a dependence is not necessarily instantaneous, and may involve delays; furthermore, y_1 may carry only partial information on v, such as the one obtained through the present and past values of the state. Whatever the nature of the dependence is, such a structure (as in (5)) is called an incentive strategy, because it displays the extent of the leader's power in enforcing a certain action on the follower through a punishment or reward scheme and by utilizing the information acquired through y_1 ([18], [28]).

C. Outline of the Following Sections

This paper is devoted to an extensive discussion and derivation of closed-loop Stackelberg strategies and incentive policies in dynamic decision problems of the types introduced above, and an elaboration on their properties. In the next section we first discuss the incentive decision problem when the leader's permissible strategies are of the form (5), in abstract inner-product spaces, and present some general results on the existence and derivation of linear incentive policies. These results are then extended in Sections III and IV in two different directions. In Section III we treat the discrete-time Stackelberg problem with dynamic informational advantage to the leader at each stage of the game, and under the feedback Stackelberg solution concept. General conditions are obtained for existence of a solution and for this solution to coincide with the global Stackelberg solution. In Section IV, we extend the results of Section II to derivation of *causal* incentive schemes and construction of real-time closed-loop Stackelberg strategies from a normal-form description, in both discrete and continuous time. Some applications to important special cases with illustrative numerical examples are given in Section V.

II. Some General Results on Existence and Derivation of Optimal Incentive Strategies

In this section we consider an abstract reformulation of the dynamic game problem of Section I-B with the leader allowed to have a partial measurement of the follower's decision variable v. Towards this end, let U and V be Hilbert spaces, with elements u and v, respectively, and the cost functional J_i of player i $(i = 1, 2)$ be a mapping from $U \times V$ into $\mathbb{R}$. In this reformulation, the dynamic nature of the decision process is suppressed, $\Gamma_2 \equiv V$, and Γ_1 is the class of all Borel-measurable mappings from Y into U, where Y is a Hilbert space comprising observations of the form

$$y = Nv$$

where $N: V \to Y$ is a linear operator with full range in Y. The case when N is invertible is known as the perfect information case; otherwise we say that the leader has only partial information on the actions of the follower.

A. Perfect Information Case

Let $(u^d, v^d) \in U \times V$ be a desirable solution from the point of view of the leader—this point could, for example, be chosen as the global minimizer of the leader's cost function $J_1(u, v)$ over $U \times V$, if such a solution exists. Then, an optimal incentive policy for the leader is one that forces the follower to choose the decision v^d, by making the incurred cost corresponding to $v \neq v^d$ sufficiently large; in other words, for a given incentive strategy γ_1 to be implementable it should satisfy the strict inequality

$$J_2(u = \gamma_1(v), v) > J_2(u^d, v^d), \qquad \text{for all } v \neq v^d, \qquad v \in V \tag{6}$$

together with the side condition

$$\gamma_1(v^d) = u^d. \tag{7}$$

To formalize this concept, we introduce the set

$$\Omega_d = \{(u, v) \in U \times V: J_2(u, v) \leqslant J_2(u^d, v^d)\} \tag{8}$$

and immediately arrive at the following result.

Proposition 1: A desired decision pair $(u^d, v^d) \in U \times V$ can be induced by an incentive strategy $\gamma_1 \in \Gamma_1$, if to each $v \in V$, $v \neq v^d$, there corresponds a $u = \gamma_1(v) \in U$ such that $(u, v) \notin \Omega_d$. A strategy that accomplishes this is the so-called (discontinuous) threat policy given by

$$\gamma_1(v) = \begin{cases} u^d, & \text{if } v = v^d \\ \text{any } u \text{ such that } (u, v) \notin \Omega_d, & \text{if } v \neq v^d. \end{cases} \tag{9}$$

Remark 1: The preceding proposition provides a sufficient condition for existence of an optimal incentive strategy. This condition is also necessary if we make an additional behavioral assumption on the follower, which is that on the boundary of Ω_d (which is his indifference curve) he chooses points that are detrimental to the leader.

The next proposition shows that the hypothesis of Proposition 1 is satisfied for an important class of problems.

Proposition 2: If $J_2(u, v)$ is continuous and strictly convex on $U \times V$, any desired decision pair $(u^d, v^d) \in U \times V$ is inducible by an appropriate incentive strategy.

Proof: Since $J_2(u, v)$ is continuous and strictly convex, the set Ω_d is closed and strictly convex. We now prove the proposition by contradiction. Assume that there exists a $(u^d, v^d) \in U \times V$ which cannot be induced by an appropriate incentive strategy. That is there exists a $\bar{v} \in V$, $\bar{v} \neq v^d$, such that $(\bar{u}, \bar{v}) \in \Omega_d$ for every $\bar{u} \in U$. Let $v_\alpha = \alpha v^d + (1 - \alpha)\bar{v}$, $0 < \alpha < 1$; then $(\bar{u}, v_\alpha) = \alpha(u^d, v^d) + (1 - \alpha)(u_\alpha, \bar{v}) \in \Omega_d$, where $u_\alpha = (1/1 - \alpha)(\bar{u} - \alpha u^d) \in U$. When $\alpha \to 1$, $(\bar{u}, v_\alpha) \to (\bar{u}, v^d)$ and hence the limit point $(\bar{u}, v^d)$ belongs to Ω_d for every $\bar{u} \in U$. In particular, if $\bar{u}$ is chosen as $u^d + u_0$ and $u^d - u_0 (u_0 \in U)$, the convex combination $(u^d, v^d) = (1/2)(u^d + u_0, v^d) + 1/2(u^d - u_0, v^d)$ should be an inner point of the strictly convex set Ω_d. This is contradictory to the fact that (u^d, v^d) is a boundary point of Ω_d and this completes the proof.

Incentive policies that induce the pair (u^d, v^d), under the hypotheses of Proposition 2 are not only of the type (9), but could also be continuous and even continuously differentiable. However, if we further restrict the class of incentive strategies to affine ones (because of their simple structure), we have to impose an additional restriction on J_2, as elucidated in the Proposition 3 below, whose proof can be found in [28].

Proposition 3: For an incentive Stackelberg game, let $J_2(u, v)$ be strictly convex and Fréchet differentiable on $U \times V$, and its gradient with respect to u, evaluated at the desired decision point $(u^d, v^d) \in U \times V$, does not vanish, i.e.,

$$\nabla_u J_2(u^d, v^d) \neq 0. \tag{10}$$

Then, the desired decision pair can be induced by an affine incentive strategy

$$\gamma_1(v) = u^d - Q(v - v^d) \tag{11}$$

where Q: $V \to U$ is a linear operator whose adjoint Q^*: $U \to V$ satisfies the equation

$$\nabla_v J_2(u^d, v^d) = Q^* \nabla_u J_2(u^d, v^d) \tag{12}$$

which admits at least one solution under (10).

It should be noted that whenever a global minimum to $J_1(u, v)$ exists on $U \times V$ (say, (u^t, v^t)), by letting $(u^d, v^d) = (u^t, v^t)$ above in (11) and (12), the leader can force the follower to minimize collectively the leader's cost functional $J_1(u, v)$.

B. Partial Dynamic Information

If the leader does not have access to v, he cannot necessarily enforce an arbitrary decision pair $(u^d, v^d) \in U \times V$ on the follower, and, in particular, (u^t, v^t) is in general not achievable. In fact, achievable solution pairs will be elements of the product space $U \times Y$, with the best achievable performance for the leader being [4].

$$\min_{U \times Y} \tilde{J}_1(u, y) \tag{13}$$

where

$$\tilde{J}_1(u, y) = J_1(u, v^*(u, y)) \tag{14}$$

$$v^*(u, y) = \arg\left\{ \min_{v \in V} J_2(u, v) \text{ subject to } Nv = y \right\}. \tag{15}$$

Here we have tacitly assumed that in (15) the argument is unique for every $(u, y) \in U \times Y$, which in fact holds whenever $J_2(u, v)$ is strictly convex on $U \times V$ [28, Lemma 2]. Further introducing

$$\tilde{J}_2(u, y) = J_2(u, v^*(u, y))$$

it can be shown [28] that strict convexity of $J_2(u, v)$ implies strict convexity of $\tilde{J}_2(u, y)$ on $U \times Y$, and hence the incentive problem with partial dynamic information becomes equivalent to one with perfect information, with $(u, y) \in U \times Y$ being the decision variables and $\tilde{J}_i(u, y)$, $i = 1, 2$, the cost functionals. Propositions 1–3 apply directly to this transformed, or so-called "projected" problem, provided that the desirable solution pair (u^d, y^d) is chosen out of $U \times Y$. In this context, a direct application of Proposition 3 leads to affine optimal incentive policies

$$\gamma_1(y) = u^d - Q(y - y^d) \tag{16}$$

where Q^*: $U \to Y$ satisfies

$$\nabla_y \tilde{J}_2(u^d, y^d) = Q^* \nabla_u \tilde{J}_2(u^d, y^d) \tag{17}$$

provided that $\tilde{J}_2(u, y)$ is Fréchet differentiable on $U \times Y$ and $\nabla_u \tilde{J}_2(u^d, y^d) \neq 0$.

Obviously, the operator Q^* in either (12) or (17) is not uniquely defined. Thus, there exist several candidates for the solution of the incentive problem at our disposal to satisfy some additional requirements. Some possible ways of constructing the operator Q^*, with application examples and other details on these approaches can be found in [28]. Yet another possible selection criterion based on sensitivity considerations has been presented and discussed in [10].

III. The Feedback Stackelberg Game with Stagewise Information Advantage to the Leader

As one application of the general results presented in the previous section, we consider here a feedback dynamic game in discrete-time, as described by the state evolution (2) and with player i's cost function given by $J_i^0(x_0, u, v)$, where

$$
\begin{aligned}
&J_i^k\left(x(k), u_k^{N-1}, v_k^{N-1}\right) \\
&\quad = \sum_{j=k}^{N-1} g_i(j, x(j), u(j), v(j)) + g_i(N, x(N)) \\
&u_k^{N-1} = \{u(k), u(k+1), \cdots, u(N-1)\}, \\
&v_k^{N-1} = \{v(k), v(k+1), \cdots, v(N-1)\} \\
&u = u_0^{N-1}, \qquad v = v_0^{N-1}, \\
&u(k) \in \mathbb{R}^{m_1}, \qquad v(k) \in \mathbb{R}^{m_2}, \qquad x(k) \in \mathbb{R}^n, \\
&\qquad k = 0, 1, \cdots, N-1. \quad (18)
\end{aligned}
$$

We endow the leader with such an information pattern that permits him to use incentive strategies under partial observation of the follower's current actions; that is, letting

$$
\begin{aligned}
y_1(k) = y(k) = N_k(x(k), \quad v(k)) \in Y_k \subseteq \mathbb{R}^p; \\
N_k: \mathbb{R}^n \times \mathbb{R}^{m_2} \to \mathbb{R}^p \qquad (19)
\end{aligned}
$$

we assume that permissible policies for the leader are Borel measurable mappings

$$\gamma_1(k; \cdot): \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^{m_1} \qquad (20)$$

so that

$$u(k) = \gamma_1[k; x(k), y(k)]. \qquad (21)$$

For the follower, on the other hand, we assume that only feedback state information is available, i.e.,

$$v(k) = \gamma_2[k; x(k)]. \qquad (22)$$

What we envisage here is a decision making process wherein the leader is dominant only stagewise, not only by announcing his policy ahead of the follower but also by incorporating partial information on the follower's current action in his incentive strategy. More precisely, the rules that underlie the game are as follows: At each stage $k = 0, 1, \cdots, N-1$, the leader announces his strategy $u(k) = \gamma_1[k; x(k), y(k)]$ first, to which the follower reacts by minimizing his stagewise cost function. This then determines the values of $y(k)$, $u(k)$, $v(k)$ and $x(k+1)$ in terms of $x(k)$, and transition to the next stage takes place. Of course, while making decisions at each stage, the players will have to anticipate their future moves and arrive at their policies accordingly. At each stage a dynamic Stackelberg game (incentive) problem of the type discussed in Section II is solved, with the leader, not only announcing his policy ahead of the follower, but also having informational advantage (partial information on the follower's decision). We call such a game a "feedback Stackelberg game with informational advantage to the leader" and the associated solution concept the "feedback Stackelberg solution with informational advantage to the leader" (FSIA). Note that this solution concept coincides with the standard feedback Stackelberg concept (cf. [24], [25]) in the case $N_k(x(k), v(k))$ is independent of $v(k)$.

We now discuss derivation of the FSIA for the finite horizon multistage decision process formulated in this section.

Let us consider the last step decision problem starting from $x(N-1)$, with only $u(N-1)$ and $v(N-1)$ to be determined (the problem $(N-1)$). Following the discussion of Section II-B, the best response of the follower to fixed values of $x(N-1)$, $u(N-1)$ and $y(N-1) = N_{N-1}(x(N-1), v(N-1))$ will be

$$
\begin{aligned}
&v(N-1) \\
&\quad = \arg\Big\{ \min_{v(N-1)} J_2^{N-1}(x(N-1), u(N-1), v(N-1)), \\
&\qquad v_{N-1} \in N_{N-1}^{-1}(x(N-1), y(N-1)) \Big\} \\
&\triangleq v_{N-1}^*(N-1, u(N-1), y(N-1)), \qquad (23)
\end{aligned}
$$

where $N_{N-1}^{-1}(x, y) = \{v \in \mathbb{R}^{m_2}\colon\ N_{N-1}(x, v) = y\}$, thus, leading to the "projected" cost functionals

$$
\begin{aligned}
&\tilde{J}_i^{N-1}[x(N-1), u(N-1), y(N-1)] \\
&\quad \triangleq J_i^{N-1}\big[x(N-1), u(N-1), v_{N-1}^*(x(N-1), \\
&\qquad \times u(N-1), y(N-1))\big] \qquad (i = 1, 2). \qquad (24)
\end{aligned}
$$

Therefore, the lowest cost value the leader can hope to attain is

$$
\begin{aligned}
I_1^{N-1}[x(N-1)] &\triangleq \min_{u(N-1),\, y(N-1)} \tilde{J}_1^{N-1} \\
&\quad \cdot [x(N-1), u(N-1), y(N-1)]. \qquad (25)
\end{aligned}
$$

Let us assume that, for each $x(N-1) \in \mathbb{R}^n$, there exists a unique solution $(u^t(N-1), y^t(N-1))$ to (25). (If the solution to (25) is not unique, we adopt one of the possible solutions according to some other consideration of preference (for the leader), see [4] for a discussion on this point.)

Now introduce the counterpart of set (8), in this context, which will depend explicitly on $x(N-1)$:

$$
\begin{aligned}
&\Omega_{N-1}(x(N-1)) = \big\{u, y) \in \mathbb{R}^{m_1} \times Y_{N-1} | \tilde{J}_2^{N-1} \\
&\quad \cdot (x(N-1), u, y) \leqslant \tilde{J}_2^{N-1}(x(N-1) u^t(N-1), \\
&\quad \cdot y^t(N-1))\big\} \qquad (26)
\end{aligned}
$$

and let $\Omega_{N-1}^c(x(N-1))$ denote its complement. Then, we have

Definition 1: For problem $(N-1)$, a state $x(N-1)$ is called *incentive controllable* if either Y_{N-1} is a singleton or for any $y \in Y_{N-1}$, $y \neq y^t(N-1)$, there exists $u \in \mathbb{R}^{m_1}$ such that $(u, y) \in \Omega_{N-1}^c(x(N-1))$. Furthermore, if all states $x(N-1) \in \mathbb{R}^n$ are incentive controllable, then the problem $(N-1)$ is called *completely incentive controllable*.

Now, an existence result follows immediately from Proposition 2.1:

Proposition 4: Assume that problem $(N-1)$ is completely incentive controllable. Then for each x_{N-1} there exists an incentive strategy $u(n-1) = \gamma_1[N-1; x(N-1), y(N-1)]$ which forces the follower to take the decision $v(N-1) = v_{N-1}^*(x(N-1), u^t(N-1), y^t(N-1))$, with the realized cost value for the leader being the minimum value of $\tilde{J}_1^{N-1}$, i.e., $I_1^{N-1}(x(N-1))$.

Remark 2: If $y(N-1)=v(N-1)$, or $v(N-1)=N_{N-1}^{-1}(x(N-1), y(N-1))$ exists uniquely, the leader has complete access to the follower's decision and the problem becomes one with perfect dynamic information (see Section II-A). In this case the attainable lower bound for an incentive controllable state x_{N-1} is exactly the team solution

$$I_1^{N-1}(x(N-1)) = \min_{u(N-1), v(N-1)} \cdot J_1^{N-1}(x(N-1), u(N-1), v(N-1)) \quad (27)$$

which is obviously the absolute lower bound.

The result of Proposition 4 can now be applied recursively by simply replacing J_i^k with the cost-to-go function $\bar{J}_i^k$ to be introduced below and by appropriately redefining $I_i^k(x(k))$. Towards this end, let

$$\bar{J}_i^k(x(k), u(k), v(k)) \triangleq g_i^k(k, x(k), u(k), \cdot v(k)) + I_i^{k+1}(x(k+1)) \quad (28)$$

where

$$x(k+1) = f(k, x(k), u(k), v(k)) \quad (29)$$

and

$$I_1^{k+1}(x(k+1)) = \min_{u,y} \tilde{J}_1^{k+1}(x(k+1), u, y) \equiv \tilde{J}_1^{k+1}(x(k+1), u^t(k+1), y^t(k+1)) \quad (30)$$

$$I_2^{k+1}(x(k+1)) = \tilde{J}_2^{k+1}(x(k+1), u^t(k+1), y^t(k+1)). \quad (31)$$

Construct

$$v_k^*(x(k), u(k), y(k)) = \arg\left\{\min_v \bar{J}_2^k(x(k), u(k), v), \cdot v \in N_k^{-1}(x(k), y(k))\right\} \quad (32)$$

$$\tilde{J}_i^k(x(k), u(k), y(k)) = \bar{J}_i^k(x(k), u(k), v_k^*(x(k), u(k), y(k)) \quad (i=1,2) \quad (33)$$

$$(u^t(k; x(k)), y^t(k; x(k))) = \arg\left\{\min_{u,y} \tilde{J}_1^k(x(k), u, y)\right\} \quad (34)$$

$$I_i^k(x(k)) = \tilde{J}_i^k(x(k), u^t(k), y^t(k)) \quad (i=1,2) \quad (35)$$

$$\Omega_k(x(k)) = \left\{(u, y) \in \mathbb{R}^{m_2} \times Y_k | \tilde{J}_2^k(x(k), u, y) \leqslant \tilde{J}_2^k(x(k), u^t(k), y^t(k))\right\}. \quad (36)$$

Then, the problem considered at stage k has the projected cost functions $\tilde{J}_1^k$ and $\tilde{J}_2^k$, with a lower bound on the former given by $I_1^k(x(k))$. The following (recursive) definition now paves the way for Proposition 5, the generalization of Proposition 4.

Definition 2: The $(N-k)$-stage problem (k) is called *completely incentive controllable*, if 1) the corresponding problem $(k+1)$ is completely incentive controllable; and 2) the equivalent one-stage incentive problem (28)–(29) is completely incentive controllable in the sense of Definition 1.

Proposition 5: For a completely incentive controllable problem (k), and for each starting state x_k, there exists an optimal incentive strategy

$$u^*(k) = \gamma_1^*[k; x(k), y(k)] = \begin{cases} u^t(k; x(k)), & \text{when } y(k) = y^t(k; x(k)) \\ u \in \mathbb{R}^{m_1} \text{ such that } (u, y) \in \Omega_k^c(x(k)), & \text{when } y(k) \neq y^t(k; x(k)), \end{cases} \quad (37)$$

that forces the follower to take the decision $v(k) = v_k^*(x(k), u^t(k), y^t(k))$, with the realized cost value for the leader being $I_1^k(x(k)$. This constitutes a FSIA solution for the dynamic game problem considered in this section.

Remark 3: Equations (28)–(35) constitute the recurrence relations between $I_i^k(x(k)$ and $I_i^{k+1}(x(k+1))$, $(i=1,2)$. This is the generalized optimality principle for the feedback Stackelberg game problem with informational advantage to the leader, under the assumption of the complete incentive controllability.

We now put some more structure on the underlying spaces and functionals, in order to obtain some specific results. The first set of such restraints and the main result that ensues are the following.

Proposition 6: The feedback Stackelberg game is completely incentive controllable if for each $x(k) \in \mathbb{R}^n$, and $k = 0, \cdots, N-1$, Y_k is a vector space and $\tilde{J}_2^k(x(k), u, y)$ is continuous and strictly convex in the pair $(u, y) \in \mathbb{R}^{m_1} x Y_k$.

Proof: Verification of this result involves a repeated application of Proposition 2 in a routine way, and is therefore omitted.

Corollary 1: When we construct the sequence $\{\hat{J}_i^k(x(k), u(k), y(k))\}$ according to relations (28)–(35), and recursively from $k = N-1$ backwards, if all $\tilde{J}_2^k(x(k), u(k), y(k))$ are continuous, strictly convex in $u(k)$ and $y(k)$ for all $x(k) \in \mathbb{R}^n$ and $k \geqslant 0$, then the problem always admits a FSIA solution, with one such optimal incentive strategy given by (37).

The conditions of this corollary (and of Proposition 6) are actually satisfied for a class of problems of practical importance. Consider, for example, the following set of sufficient conditions:

1) $g_2(N; x)$ is convex in $x \in \mathbb{R}^n$;
2) $g_2(k; x, u, v)$ is decomposable in the form; $g_2(k; x, u, v) = P_2(k; x) + q_2(k; u, v)$, where $P_2(k; x)$ is convex in x and $q_2(k; u, v)$ is strictly convex in (u, v);
3) $f(k; x, u, v)$ and $N_k(x, v)$ are affine in their arguments;
4) $u^t(k; x)$ and $y^t(k; x)$ are affine in x;
5) $v_k^*(x, u, y)$ is affine in x, u and y.

These guarantee satisfaction of the hypotheses of the corollary. One such special class is the linear-quadratic problem where

$$g_2(N; x) = \langle x, Q_N x\rangle \qquad (Q_N \geqslant 0)$$

$$g_2(k; x, u, v) = \langle x, Q_k x\rangle + \langle u, R_k u\rangle + \langle v, S_k v\rangle$$

$$(Q_k \geqslant 0, R_k > 0, S_k > 0)$$

where $\langle \cdot , \cdot \rangle$ denotes appropriate inner products in vector spaces;

$$f(k; x, u, v) = A_k x + B_k u + C_k v$$

$$N_k(x, v) = N_k v$$

where A_k, B_k, C_k, N_k are matrices of appropriate dimensions, with N_k being of full rank.

It readily follows from Proposition 3 that in this case the FSIA solutions are not only of the type (37), but can also be taken to be affine, in which case

$$u^t(k) = L_1(k)x(k), \qquad y^t(k) = L_2(k)x(k)$$

$$v_k^*(x, u, y) = M_1(k)x + M_2(k)u(k) + M_3(k)y(k)$$

$$u^*(k) = \gamma_1^*[k; x(k), y(k)]$$

$$= L_1(k)x(k) - Q_1(k)[y(k) - L_2(k)x(k)]$$

with capital letters denoting matrices of appropriate dimensions, and $Q_1(k)$ being a gain matrix whose transpose satisfies a gradient equation of the type (12), for each $k \geq 0$. Explicit expressions for these matrices can be obtained by basically solving (28)–(35), recursively, and by noting that $\tilde{J}_i^k$ and I_i^k are quadratic functionals for each $k \geqslant 0$.

Remark 4: The preceding results find natural extensions to the class of problems wherein the control and measurement spaces are arbitrary (infinite dimensional) Banach spaces, instead of being finite dimensional. Particularly, for the linear-quadratic problem discussed above, the same affine structure prevails provided that we interpret the inner-products appropriately and replace all matrices with linear operators. Such a result, then, would be applicable to continuous-time dynamic games in which the decision makers have access to sampled information and the feedback solution is defined in between different sampled subintervals.

Remark 5: Under the conditions of Proposition 4 and Corollary 3.1, and when the leader has perfect access to the follower's decision variable at each stage, the affine FSIA solution has also a robust feature in the sense that its truncated version constitutes a FSIA solution to a dynamic feedback game of shorter duration, defined on the interval $[k, N-1]$, for any $k > 0$. This result is a direct consequence of the fact that the trajectory corresponding to the original FSIA solution satisfies the principle of optimality (being the team solution from the leader's point of view) and the leader's affine FSIA strategy employs only current state information.

IV. Derivation of Causal Stackelberg Solutions to Discrete and Continuous-Time Dynamic Games

In this section we turn our attention to the global Stackelberg solution in both discrete and continuous-time dynamic games of the type introduced in §II-B, and under the closed-loop information pattern. Here, the leader will not have any stagewise informational advantage over the follower, but he will still dominate the decision process by announcing his strategy ahead of time and enforcing it on the follower, in accordance with the solution concept (3)–(4). Furthermore, because of its appealing features, we restrict attention to those strategies for the leader that are linear in the dynamic part of the information, and also assume, without any loss of generality, that the follower employs only open-loop policies (which does not lead to any degradation in his performance (see, e.g., [7]).

Let J_1 and J_2 be appropriate cost functionals for players 1 and 2, which, for fixed initial state $x_0 \in \mathbb{R}^n$, can always be rewritten (by elimination of the state variable) as functions of solely the decision variables $(u, v) \in U \times V$ (see Section I-B for notation). Since every discrete or continuous-time dynamic game can be expressed in this form, the analyses and results of Section II are directly applicable here provided that the corresponding optimal incentive strategy for the leader is permissible, i.e., it is causal and satisfies the additional structural restrictions that may be imposed on elements of Γ_1. Specifically, let us assume that:

1) Through the closed-loop state information, the leader is able to infer perfectly the past values of $v(\cdot)$, the decision function of the follower.

2) $J_2(u, v)$ is Fréchet-differentiable and strictly convex on $U \times V$.

3) A global minimum to $J_1(u, v)$ exists on $U \times V$, which we denote by $(u^\dagger, v^\dagger) \in U \times V$, and which is adopted as a desirable solution by the leader.

4) At this solution point,

$$\nabla_u J_2(u^\dagger, v^\dagger) \neq 0. \tag{38}$$

Then, we know from Proposition 3 that, in the absence of causality, every optimal affine Stackelberg solution can be written as

$$u = \gamma_1(v) = u^\dagger - Q(v - v^\dagger) \tag{39}$$

where the adjoint of Q, $Q^*\colon U \to V$, satisfies

$$\nabla_v J_2(u^\dagger, v^\dagger) = Q^* \nabla_u J_2(u^\dagger, v^\dagger). \tag{40}$$

Now, the real question here is whether we can find an operator Q whose adjoint satisfies (40) above, and which is further causal and leads to a policy γ_1, as given by (39), belonging to a given closed-loop policy space Γ_1. We show below that, under the closed-loop pattern and for both discrete and continuous-time problems satisfying appropriate structural assumptions, such a linear operator can be contructed.

Towards this end, we first introduce some notation. Let the inner-product of two elements $f(\cdot)$ and $g(\cdot)$ in $L_2^m[0, T)$ be defined by

$$\langle f, g\rangle \triangleq \int_0^T f'(t)g(t)\,dt = \int_0^T g'(t)f(t)\,dt \tag{41}$$

and further introduce the notation

$$\langle f, g\rangle_{(t_1, t_2)} \triangleq \int_{t_1}^{t_2} f'(t)g(t)\,dt = \int_{t_1}^{t_2} g'(t)f(t)\,dt \tag{42}$$

where $0 \leqslant t_1 \leqslant t_2 \leqslant T$. Similarly, for discrete-time processes, the inner-product of $f(\cdot), g(\cdot) \in l_2^m[0, N]$ is defined by

$$\langle f, g\rangle \triangleq \sum_{i=0}^{N} f'(i)g(i) = \sum_{i=0}^{N} g'(i)f(i) \tag{43}$$

and furthermore

$$\langle f, g\rangle_{(k,l)} = \sum_{i=k}^{l} f'(i)g(i) = \sum_{i=k}^{l} g'(i)f(i) \tag{44}$$

where k, l are integers, $0 \leqslant k \leqslant l \leqslant N$. Now introduce

$$\phi(\cdot) \triangleq \nabla_u J_2(u, v), \quad \Psi(\cdot) \triangleq \nabla_v J_2(u, v) \tag{45}$$

where the gradients are evaluated at some specific values of $u \in U$ and $v \in V$ that will be clear from the context.

To reveal a property of $\phi(\cdot)$ and $\Psi(\cdot)$ which is vital in the construction of affine incentive strategies (cf. (39)), let us consider the variation in J_2 resulting from, for example, an infinitesimal variation $\delta u(\cdot)$ in $u(\cdot)$:

$$\delta J_2 = \langle \nabla_u J_2(u, v), \quad \delta u(\cdot)\rangle = \int_0^T \phi'(t)\delta u(t)\, dt$$

$$\text{or} = \sum_{i=0}^{N} \phi'(i)\delta u(i). \tag{46}$$

Thus, the value of $\phi(\cdot)$ at time t simply represents the local sensitivity of J_2 with respect to $u(t)$, in other words, the ability of the leader to influence J_2 by changing his decision variable $u(t)$ at time t. Likewise, the time function $\Psi(\cdot) = \nabla_v J_2(u, v)$ represents the follower's ability to influence his cost functional J_2 by changing $v(t)$, the value of $v(\cdot)$ at time t. Hence, they can be referred to as "sensitivity functions" representing the sensitivity of J_2 to the players' actions, which may be taken as a measure of the players' control ability in the related optimization problems.

Of course, when we speak of "changing" or "influence" as above, we use these terms in the meaning of "infinitesimal variations" or the "first order approximation." Thus, they make sense only in a small neighborhood of a specific point $(u, v) \in U \times V$.

Hence, in the absence of a causality restriction, the results of Section II admit an explicit "physical interpretation." The only condition for existence of an affine incentive solution to the Stackelberg dynamic game is that the sensitivity of J_2 with respect to $u(\cdot)$ should not be zero (cf. Proposition 3, and also (38)). That is, whenever the leader is able to influence the follower's cost-functional (infinitesimally), he can always force the follower to choose the prescribed value for his decision variable.

Now, when the leader is faced with the additional constraint that his control at time t cannot depend on the "future values" of $v(\cdot)$, the operator Q in the incentive strategy (39) should be a causal operator (or equivalently, Q^* satisfying (40) should be anti-causal). Moreover, if the leader needs a nonzero time duration ϵ to infer the necessary information on $v(t)$ from the current observation, the control $u(t)$ can only depend on the value of $v(\tau)$ for $\tau \leqslant t - \epsilon$; such an operator Q will be called *ϵ-strong causal* (and Q^*-ϵ-strong anticausal).

Let us first introduce

$$\Phi(t) \triangleq \langle \phi, \phi\rangle_{(t,T)} \text{ or } \langle \phi, \phi\rangle_{(t,N)} \tag{47}$$

and

$$\psi(t) = \langle \Psi, \Psi\rangle_{(t,T)} \text{ or } \langle \Psi, \Psi\rangle_{(t,N)}. \tag{48}$$

If $\Phi(t) = 0$, i.e., $\phi(\tau) = 0$, $\tau \geqslant t$ (recall that, for functions $f(t)$, $g(t)$ in Hilbert space $L_2^m[t_0, t_f]$, $f(\cdot) = g(\cdot)$ means $f(t) = g(t)$ for almost all $t \in [0, T)$, except perhaps on some set of measure zero; this fact has to be noted throughout the paper), then the leader cannot control the situation during $\tau \in [t, T)$. If, concurrently, $\psi(t) \neq 0$, the follower can change the value of $J_2(u, v)$ by infinitesimal variations in $v(\cdot)$. Thus, it is intuitively evident that the leader may not be able to enforce any desired decision pair (u^d, v^d) by a causal incentive strategy, because he cannot respond effectively to the variation in the follower's decision, even though he may be able to detect it.

To put the above intuitive reasoning into precise form, we first prove for continuous-time systems the following result.

Lemma 1: For any $\phi(\cdot) \in U = L_2^{m_1}[t_0, t_f]$ and $\Psi(\cdot) \in V = L_2^{m_2}[t_0, t_f]$, a set of sufficient conditions for existence of an anticausal bounded linear operator Q^*: $U \to V$ satisfying (40) which can be rewritten as $Q^*\phi = \Psi$, is the following:

a) For all $t \in [0, T)$, $\Psi(t) \neq 0$ implies $\Phi(t) \neq 0$. (Let t_ϕ be the smallest time such that $\Phi(t_\phi) = 0$, and t_Ψ be the smallest time making $\psi(t) = 0$; then the condition says $t_\phi \geqslant t_\Psi$.)

b) When $t_\phi = t_\Psi$, the following integral exists and remains bounded

$$\int_0^{t_\phi} \frac{\Psi'(t)\Psi(t)}{\Phi(t)}\, dt. \tag{49}$$

(This second condition means that the follower's "control ability," measured in terms of $\Psi(t)$, cannot be much stronger than the leader's, at the point $t_\phi = t_\Psi$ when they concurrently lose their control ability.)

Proof: The lemma can be proved simply by giving one of the possible solutions for operator Q^*, which is

$$Q^*[f(\tau)](t) \triangleq \begin{cases} \displaystyle\int_t^{t_\phi} \frac{\Psi(t)\phi'(\tau)f(\tau)}{\Phi(t)}\, d\tau & (t < t_\phi) \\ 0 & (t \geqslant t_\phi). \end{cases} \tag{50}$$

Thus, it is an integral operator with kernel

$$R(t, \tau) = \begin{cases} \dfrac{\Psi(t)\phi'(\tau)}{\Phi(t)} & (\tau, t < t_\phi, \tau > t) \\ 0 & (\text{otherwise}) \end{cases} \tag{51}$$

(see, e.g., [29, p. 67]). Note that

$$\begin{aligned}\|R\|^2 &\triangleq \int_0^T\int_0^T \operatorname{Tr}\left[R(t,\tau)R'(t,\tau)\right] d\tau\, dt \\ &= \int_t^{t_\phi}\int_t^{t_\phi} \operatorname{Tr} \frac{\Psi(t)\phi'(t)\phi(\tau)\Psi'(t)}{\Phi^2(t)} d\tau\, dt \\ &= \int_0^{t_\phi} \frac{\Psi'(t)\Psi(t)}{\Phi^2(t)} \int_t^{t_\phi} \phi'(\tau)\phi(\tau)\, d\tau\, dt \\ &= \int_t^{t_\phi} \frac{\Psi'(t)\Psi(t)}{\Phi(t)} dt < \infty \qquad (52)\end{aligned}$$

and therefore Q^* is well-defined and bounded, with $\|Q^*\|^2 \leqslant \|R\|^2$. It is anticausal, since the value of $Q^*[f(\tau)]$ at time t depends only on the values of $f(\tau)$ for $\tau \in [t, t_\phi)$. Finally it is straightforward to verify that $Q^*[\phi(\cdot)](t) = \Psi(t)$, except perhaps at times t belonging to a set of measure zero.

Remark 6: The operator Q, being the adjoint of the anticausal operator Q^*, is a causal operator. The adjoint of (50) can readily be computed to be

$$\begin{aligned}Q[g(t)](\tau) &= \int_0^T R'(t,\tau) g(t)\, dt \\ &= \int_0^\tau \frac{\phi(\tau)\Psi'(t)}{\Phi(t)} g(t)\, dt. \qquad (53)\end{aligned}$$

The lemma can easily be generalized to the case when Q^*: $U \to V$ is required to be an ϵ-strong anticausal operator. The sufficient conditions become, in this case the following:

i) whenever $\langle \Psi, \Psi \rangle_{(t,T)} \neq 0$, we must have $\Phi(t+\epsilon) = \langle \phi, \phi \rangle_{(t+\epsilon,T)} \neq 0$, that is $t_\phi \geqslant t_\Psi + \epsilon$. This implies that $\Psi(t) = 0$ for all $t \geqslant T - \epsilon$; and

ii) when $t_\phi = t_\Psi + \epsilon$, the following integral exists and remains bounded:

$$\int_0^{t_\phi - \epsilon} \frac{\Psi'(t)\Psi(t)}{\Phi(t+\epsilon)} dt. \qquad (54)$$

In this case the kernel corresponding to (51) is

$$R(t,\tau) = \begin{cases} \dfrac{\Psi(t)\phi'(\tau)}{\Phi(t+\epsilon)} & (t < t_\phi - \epsilon,\ t + \epsilon < \tau < t_\phi) \\ 0 & (\text{otherwise}). \end{cases} \qquad (55)$$

Moreover, the counterparts of (50) and (53), in this case, are, respectively,

$$Q^*[f(\tau)](t) = \begin{cases} \displaystyle\int_{t+\epsilon}^{t_\phi} \frac{\Psi(t)\phi'(\tau)}{\Phi(t+\epsilon)} f(\tau)\, d\tau & (t < t_\phi - \epsilon) \\ 0 & (\text{otherwise}) \end{cases} \qquad (56)$$

and

$$Q[g(t)](\tau) = \begin{cases} \displaystyle\int_0^{\tau-\epsilon} \frac{\phi(\tau)\psi'(t)}{\Phi(t+\epsilon)} g(t)\, dt & (\epsilon \leqslant \tau < t_\phi) \\ 0 & (\tau < \epsilon \text{ or } \epsilon \geqslant t_\phi). \end{cases} \qquad (57)$$

The counterpart of Lemma 1 for the discrete-time case is much simpler, since Condition b is then implied by Condition a.

Lemma 2: For any $\phi(\cdot) \in U = l_2^{m_1}[0, N-1]$, $\Psi(\cdot) \in V = l_2^{m_2}[0, N-1]$, a sufficient condition for existence of a one-step-strong anticausal linear operator Q^*: $U \to V$ such that $Q^*\phi = \Psi$ is that

iii) whenever $\psi(k) = \sum_{i=k}^{N-1}\Psi'(i)\Psi(i) \neq 0$, we must have $\Phi(k+1) = \sum_{i=k+1}^{N-1}\phi'(i)\phi(i) \neq 0$; that is, $t_\phi \geqslant t_\Psi + 1$.[1]

The proof of this lemma is similar to that of Lemma 1 and is therefore omitted. The corresponding linear operators are

$$Q^*[f(i)](k) = \begin{cases} \displaystyle\sum_{i=k+1}^{N-1} \frac{\Psi(k)\phi'(i)}{\Phi(k+1)} f(i) & (k < t_\phi - 1) \\ 0 & (k \geqslant t_\phi - 1) \end{cases} \qquad (58)$$

and

$$Q[g(k)](i) = \begin{cases} \displaystyle\sum_{k=0}^{i-1} \frac{\phi(i)\Psi'(k)}{\Phi(k+1)} g(k) & (1 \leqslant i < t_\phi) \\ 0 & (i = 0 \text{ or } i \geqslant t_\phi) \end{cases} \qquad (59)$$

which are one-step strong anticausal and one-step strong causal, respectively. The general conclusion we derive from these two lemmas is the following:

Proposition 7: For the general Stackelberg dynamic game problem (Section I-B) with $U = L_2^{m_1}[0, T]$, $V = L_2^{m_2}[0, T]$ or $U = l_{l_2}^{m_1}[0, N-1]$, $V = l_2^{m_2}[0, N-1]$, in addition to the assumptions 1)–4) made in this section, let conditions a) and b) of Lemma 1 (or correspondingly, condition iii of Lemma 2) be satisfied, and let the leader have perfect access to the past values of the follower's control variable (by possibly inferring these values perfectly through the observation of the state). Then the operator Q defining the affine incentive strategy

$$u = \gamma_1(v) = u^\dagger - Q[v - v^\dagger] \qquad (60)$$

where

$$(u^\dagger, v^\dagger) = \underset{(u,v) \in U \times V}{\arg\min}\ J_1(u, v), \qquad (61)$$

can be chosen as a causal operator (correspondingly, one-step strong-causal operator). One of its possible forms is

[1] Here t_ϕ and t_Ψ are the discrete-time counterparts of those introduced in Lemma 1.

given by (53) (or, correspondingly, by (59)), and it provides a global Stackelberg solution to the problem.

Remark 7: For the discrete-time case, a very simple and useful version of the sufficient conditions is that at the last decision stage

$$\phi(N) \neq 0 \quad \text{and} \quad \Psi(N) = 0.$$

The causal incentive solution obtained above offers us a possible way of constructing the "closed-loop" solution to the Stackelberg dynamic game problem, provided that $v(\cdot)$ can be reconstructed from the observed state information in real-time. If, for example, there is a causal operator H such that

$$v(\cdot) = Hx(\cdot) \tag{62}$$

then we have the closed-loop solution

$$u(\cdot) = u^{\dagger}(\cdot) - Q[Hx(\cdot) - v^{\dagger}(\cdot)] \tag{63}$$

which is physically realizable. More specific derivations along this line are provided in the next section.

We should note that when the ϵ-strong causality conditions i) and ii) are taken instead of a) and b), the statement of Proposition 6 can be modified in a straightforward manner, which then says that affine ϵ-strong causal solutions exist. These may be used in realizing the optimum Stackelberg strategy of the leader, with an ϵ-delay in the reconstruction of $v(\cdot)$ from the state observation $x(\cdot)$.

Finally, we should remark that the results of this section, in particular those of Lemmas 1, 2, and Proposition 7, can be extended to the case when the leader has only partial state information and/or partial dynamic information on the follower's actions, without much difficulty and with only minor modifications. This extension involves, basically, the derivation of an achievable desirable solution $(u^{\dagger}, y^{\dagger})$ to replace (61) (cf. Section II-B), "projected" cost functional $\tilde{J}_2(u, y)$ for the follower, and rewording of Lemmas 1–2 and Proposition 7 in terms of this new notation. We do not pursue this point here; see, however, the specific problem solved in Section V-B.

V. Applications and Examples

In this section, the concepts and results presented in Sections II and IV will be applied to some special cases of practical interest. Some numerical examples will be given to show the applicability of the theory and the general approach.

A. Causal Stackelberg Solution to Discrete-Time Linear Quadratic Dynamic Game Problems

One of the important subclasses of problems widely discussed in the literature (see e.g., [9], [27]) is the discrete-time dynamic Stackelberg game with linear-state equation

$$x(k+1) = A(k)x(k) + B(k)u(k) + C(k)v(k) \qquad (k = 0, 1, \cdots, N-1) \tag{64}$$

and quadratic cost-functionals

$$J_i = x'(N)Q(N)x(N) + \sum_{j=0}^{N-1} \{x'(j)Q_i(j)x(j) + u'(j)R_i(j)u(j) + v'(j)S_i(j)v(j)\} \tag{65}$$

where $i = 1, 2$, refer to the leader and the follower, respectively.

The approach presented in the previous section can be used in obtaining a causal solution to this problem under the closed-loop information pattern. Here we give only a numerical example to illustrate the method.

Example 1:

$$x(k+1) = x(k) + u(k) + v(k) \qquad (k = 0, 1, \cdots, N-2)$$

$$x(N) = x(N-1) + u(N-1)$$

$$J_1 = x^2(N) + \sum_{k=0}^{N-1} (x^2(k) + 2u^2(k) + v^2(k))$$

$$J_2 = x^2(N) + \sum_{k=0}^{N-1} (x^2(k) + u^2(k) + 3v^2(k)).$$

In accordance with the method presented in Sections II and IV, first the team solution of minimizing J_1 is obtained from the standard Riccati recurrence relations, which involves the value function $x'(k)P(k)x(k)$, the corresponding team optimal controls $u^{\dagger}(k)$ and $v^{\dagger}(k)$, and optimal trajectory $x^{\dagger}(\cdot)$. Then the gradients $\phi(k) = \nabla_{u(k)} J_2^{\dagger}$ and $\Psi(k) = \nabla_{v(k)} J_2^{\dagger}$ at the desired team solution are derived from the dynamic equations as

$$\phi(k) = \sum_{j=k+1}^{N} x^{\dagger}(j) + u^{\dagger}(k) \qquad (k = 0, 1, \cdots, N-1)$$

$$\Psi(k) = \sum_{j=k+1}^{N} x^{\dagger}(j) + 3v^{\dagger}(k) \qquad (k = 0, 1, \cdots, N-2),$$

$$\psi(N-1) = 0,$$

Note that $\Phi(k) = \sum_{i=k}^{N-1} \phi^2(i)$, and from (60) and (59) the optimal Stackelberg strategy for the leader is

$$\gamma(i; x) = u^{\dagger}(i) + \phi(i) \sum_{k=0}^{i-1} \frac{\Psi(k)[v(k) - v^{\dagger}(k)]}{\Phi(k+1)}, \qquad u(0) = u^{\dagger}(0).$$

$$\equiv P(i)x(i) + \phi(i) \sum_{k=0}^{i-1} \frac{\Psi(k)}{\Phi(k+1)} \cdot [x(k+1) - x(k) - u^{\dagger}(k) - v^{\dagger}(k)], \qquad u(0) = u^{\dagger}(0).$$

The values of these coefficients for the case $N = 4$ are

TABLE I

k	0	1	2	3	4
$P(k)$	1.4576075	1.4592593	1.4761905	1.6666667	1.000000
$u^\dagger(k)/x^\dagger(k)$	-0.2288037	-0.2296296	-0.2380953	-0.3333333	—
$v^\dagger(k)/x^\dagger(k)$	-0.4576074	-0.4592593	-0.4761905	0	—
$x^\dagger(k)/x^\dagger(k-1)$		0.3135888	0.3111111	0.2857143	0.6666667
$x^\dagger(k)/x(o)$	1	0.3135888	0.097561	0.0278746	0.018583
$\phi(k)/x(o)$	0.2288037	0.0720093	0.0232288	0.0092915	—
$\Psi(k)/x(o)$	-0.9152148	-0.2880372	-0.0929152	0	—
$\Phi(k)/x^2(o)$	0.0581624	0.0058112	0.006259	0.0000863	—

listed in Table I, with the corresponding policies being

$$u(1) = u^\dagger(1) - 11.340855\big(v(0) - v^\dagger(0)\big)$$
$$(2) = u^\dagger(2) - 3.6583393 \cdot \big(v(0) - v^\dagger(0) - 1.068982(v(1) - v^\dagger(1)\big)$$
$$(3) = u^\dagger(3) - 1.4633326 \cdot \big(v(0) - v^\dagger(0) - 0.4275919(v(1) - v^\dagger(1)\big) - 10.003726\big(v(2) - v^\dagger(2)\big).$$

Here

$$v(0) - v^\dagger(0) = x(1) - x(0) - u^\dagger(0) - v^\dagger(0)$$
$$v(1) - v^\dagger(1) = x(2) - x(1) - u^\dagger(1) - v^\dagger(1)$$
$$v(2) - v^\dagger(2) = x(3) - x(2) - u^\dagger(2) - v^\dagger(2).$$

This is a causal closed-loop Stackelberg solution which achieves the globally optimal team solution.

B. The Linear Quadratic Infinite-Time Stackelberg Problem

Consider the continuous-time problem formulated by

$$\dot{x} = Ax + Bu + Cv \qquad x(t) \in \mathbb{R}^n, \qquad u(t) \in \mathbb{R}^{m_1}, \qquad v(t) \in \mathbb{R}^{m_2} \qquad x(0) = x_0, \qquad t \geqslant 0; \tag{66}$$

$$J_i = \int_0^\infty (x'Q_ix + u'R_iu + v'S_iv)\,dt \qquad (i = 1,2) \tag{67}$$

where B and C have full-column rank m_1 and m_2, respectively, $(A,(B\,\vdots\,C))$ is controllable, $(Q_1^{1/2}, A)$ is observable, $Q_i \geqslant 0$, $R_1 > 0$, $S_i > 0$, $R_2 \geqslant 0$. The team solution that minimizes J_1 is

$$u^\dagger = -R_1^{-1}B'Px^\dagger, \qquad v^\dagger = -S_1^{-1}C'Px^\dagger, \qquad J_1^\dagger = x_0'Px_0 \tag{68}$$

where P is the unique positive definite solution of the algebraic Riccati equation

$$P\big[BR_1^{-1}B' + CS_1^{-1}C'\big]P - A'P - PA - Q_1 = 0 \tag{69}$$

and the optimal trajectory $x^\dagger$ satisfies

$$\dot{x}^\dagger = A_cx^\dagger \triangleq \big[A - \big(BR_1^{-1}B' + CS_1^{-1}C'\big)P\big]x^\dagger, \qquad x^\dagger(0) = x_0. \tag{70}$$

We now attempt to solve this problem under two different causal-functional dependences for the leader's policy, viz $\gamma_1\colon V \to U$; $\gamma_1(v) = u^\dagger - Q(v - v^\dagger)$ and $\gamma_1\colon X \to U$; $\gamma_1(x) = u^\dagger - Q(x - x^\dagger)$, where Q is, in each case, a linear causal operator.

In the former case, we first calculate the gradients of J_2 with respect to u and v (see Appendix A) and arrive at

$$\phi(t) = \nabla_uJ_2(u^\dagger, v^\dagger) = 2Mx^\dagger(t) \tag{71}$$
$$\Psi(t) = \Delta_vJ_2(u^\dagger, v^\dagger) = 2Nx^\dagger(t) \tag{72}$$

where

$$M = \big(B'I_0 - R_2R_1^{-1}B'P\big) \tag{73}$$
$$N = \big(C'I_0 - S_2S_1^{-1}C'P\big) \tag{74}$$

and I_0 is the solution of the matrix equation

$$A'I_0 + I_0A_c + Q_2 = 0. \tag{75}$$

If a constant matrix gain solution is desired, then we must have

$$Q'M = N. \tag{76}$$

Unless Range $M' \supset$ Range N', such a Q does not exist, and hence the problem does not admit a solution. However, if we also allow dependence on the initial state x_0, affine causal solutions exist provided that for all $x_0 \neq 0$, $Me^{A_ct}x_0 \neq 0$, which is equivalent to the requirement that (M, A_c) be observable. In this case an optimal affine incentive scheme is

$$u(t) = u^\dagger - Q(v - v^\dagger)$$
$$= u^\dagger(t) - \int_0^t \frac{\phi(t)\Psi'(\sigma)[v(\sigma) - v^\dagger(\sigma)]}{\int_0^\infty \phi'(s)\phi(s)\, ds} d\sigma. \quad (77)$$

Next, we seek a solution in the form $u = u^\dagger - Q(x - x^\dagger)$, where Q is causal. By using the approach outlined in Section II-B and taking the entire trajectory x as the leader's observed information, we have, uniquely,

$$v^*(u, x) = C^+(\dot{x} - Ax - Bu) \quad (78)$$

where $C^+ \triangleq (C'C)^{-1}C'$ is the pseudo-inverse of C. Note that in this case the absolute lower bound given by (68) is attainable, since the operator N of Section II-B is invertible (C being a matrix of full-column rank). Now, projecting the problem into $U \times X$ where (u, x) belongs, we obtain

$$\tilde{J}_2(u, x) = \langle x, Q_2 x\rangle + \langle u, R_2 x\rangle + \langle(\dot{x} - Ax - Bu), \bar{C}(\dot{x} - Ax - Bu)\rangle \quad (79)$$

where $\bar{C} \triangleq C^{+\prime}S_2C^+$. The gradients $\nabla_u \tilde{J}_2$ and $\nabla_x \tilde{J}_2$ at $(u^\dagger, x^\dagger)$ can be evaluated as (see Appendix B)

$$\phi(t) = \nabla_u \tilde{J}_2(u^\dagger, x^\dagger) = 2Mx^\dagger(t) \quad (80)$$
$$\Psi(t) = \nabla_x \tilde{J}_2(u^\dagger, x^\dagger) = 2Nx^\dagger(t) \quad (81)$$

where

$$M = (B'C^{+\prime}S_2S_1^{-1}C' - R_2R_1^{-1}B')P \quad (82)$$
$$N = Q_2 + C^{+\prime}S_2S_1^{-1}C'PA_c + A'C^{+\prime}S_2S_1^{-1}C'P. \quad (83)$$

The conclusion we arrive at here is almost the same as in the case (73)–(74). When Range $M' \supset$ Range N', there exists a constant gain solution $u = u^\dagger - Q(x - x^\dagger)$ with Q satisfying $Q'M = N$. Otherwise, provided that (M, A_c) is observable, there exists an affine causal solution, depending on $x_0 \neq 0$, given by

$$u(t) = u^\dagger - Q(x - x^\dagger)$$
$$= u^\dagger(t) - \int_0^t \frac{\phi(t)\Psi'(\sigma)(x(\sigma) - x^\dagger(\sigma))}{\int_\sigma^\infty \phi'(s)\phi(s)\, ds} d\sigma, \quad (84)$$

where $\phi(\cdot)$ and $\Psi(\cdot)$ are given by (80) and (81), respectively.

We now provide a numerical example to illustrate these results.

Example 2:

$$\dot{x} = 2x + u + v, \qquad x(0) = x_0 \qquad t \in [0, \infty)$$
$$J_1 = \int_0^\infty (6x^2 + u^2 + v^2)\, dt$$
$$J_2 = \int_0^\infty (qx^2 + rv^2)\, dt, \qquad q > 0, r > 0.$$

The team solution is

$$u^\dagger = -3x^\dagger, \qquad v^\dagger = -3x^\dagger, \qquad x^\dagger(t) = e^{-4t}x_0,$$

since from $2P^2 - 4P - 6 = 0$ we have $P = 3$ and $A_c = -4$. From (75), $I_0 = (1/2)q$. From (73)–(74), $M = (1/2)q$, $N = (1/2)q - 3r$. That is, $\phi(t) = qx^\dagger(t)$, $\Psi(t) = (q - 6r)x^\dagger(t)$. The optimal Stackelberg strategy is

$$u(t) = u^\dagger(t) - Q[v - v^\dagger](t)$$

where the operator Q is either $Q = N/M = 1 - 6r/q$ or

$$Q[g(t)](\tau) = \phi(\tau)\int_0^\tau \frac{\Psi(t)g(t)}{\int_t^\infty \phi^2(s)\, ds} dt$$
$$= 8\left(1 - \frac{6r}{q}\right)\int_0^\tau e^{4(t-\tau)}g(t)\, dt.$$

This solution can be implemented by a first-order block with transfer function

$$\overline{W}(s) = \frac{U(s) - U^\dagger(s)}{V(s) - V^\dagger(s)} = 8\left(1 - \frac{6r}{q}\right)\frac{1}{s+4};$$

where s is the Laplace variable.

On the other hand, from (82) and (83)

$$M = 3r/2, \qquad N = (q - 6r)/2.$$

Thus, the optimum Stackelberg strategy in case of x-dependence is

$$u(t) = u^\dagger(t) - Q[x - x^\dagger](t),$$

where the operator Q is either $Q = N/M = q/3r - 2$ or

$$Q[g(t)](\tau) = 3re^{-4\tau}x_0\int_0^\tau \frac{(q - 6r)e^{-4t}x_0 g(t)}{\int_t^\infty 9r^2x_0^2e^{-8s}\, ds} dt$$
$$= \int_0^\tau \left(\frac{q}{3r} - 2\right)e^{4(t-\tau)}8g(t)\, dt$$

which may be implemented in the frequency domain by

$$W(s) = \frac{U(s) - U^\dagger(s)}{X(s) - X^\dagger(s)} = 8\left(\frac{q}{3r} - 2\right)\frac{1}{s+4}.$$

C. The Linear Quadratic Finite-Time Closed-Loop Stackelberg Problem

In this subsection we provide an example illustrating the results obtained in Section IV when $t_\phi < \infty$.

Example 3:

Consider the problem with the specifics

$$\dot{x} = 2x + u + b(t)v, \qquad t \in [0,1],$$
$$J_1 = 4x^2(1) + \int_0^1 (6x^2 + u^2 + v^2)\, dt,$$
$$J_2 = 2x^2(1) + \int_0^1 (qx^2 + rv^2)\, dt \qquad (q > 0, r > 0).$$

where the time-varying gain $b(t)$ is a continuous bounded function; furthermore, when $t \to 1$, $b(t) \to 0$ with the order of magnitude being

$$b(t) = b_0\epsilon^\alpha + 0(\epsilon^\alpha), \qquad \epsilon = 1 - t, \qquad \alpha > 0.$$

The team solution for J_1 is

$$u^\dagger(t) = -2P(t)x^\dagger(t), \qquad v^\dagger = -2b(t)P(t)x^\dagger(t)$$

$$\dot{x}^\dagger = \left(2 - 2(1+b)^2\right)Px^\dagger \Rightarrow x^\dagger(\tau)$$

$$= \left[\exp\left(-\int_0^\tau 2P(b^2 + 2b)\,dt\right)\right]x_0$$

$$\frac{dP}{dt} + 4P - 2P^2(1+b^2) + 3 = 0, \qquad P(1) = 2.$$

Both $x^\dagger(t)$ and $P(t)$ are bounded continuous functions on $[0,1]$.

$$\nabla_u J_2^\dagger = \phi(\tau) = 2q\int_\tau^1 e^{2(t-\tau)}x^\dagger(t)\,dt + 4x^\dagger(1)e^{2(1-\tau)}$$

$$\nabla_v^\dagger J_2 = \Psi(\tau) = 2qb(\tau)\int_\tau^1 e^{2(t-\tau)}x^\dagger(t)\,dt + 4x^\dagger(1)e^{2(1-\tau)}b(\tau) + 2rv^\dagger(\tau).$$

When $\tau \to 1$, $\phi(\tau) \to \phi(1) = 4x^\dagger(1)$, thus $t_\phi = 1$. Furthermore,

$$\Phi(t) = \int_t^1 \phi^2(\tau)\,d\tau = 4x^\dagger(1)(1-t) + 0(1-t).$$

Therefore, condition (49) is satisfied:

$$\int_t^1 \frac{\Psi^2(\tau)}{\Phi(\tau)}\,d\tau < \infty$$

and by Lemma 1, the operator

$$Q[g(\cdot)](\tau) = \int_0^\tau \frac{\phi(\tau)\Psi(t)}{\Phi(t)}g(t)\,dt$$

is linear, bounded, and can be used in the construction of the Stackelberg strategy $u = u^\dagger - Q(x - x^\dagger)$.

VI. Concluding Remarks

In this paper we have discussed derivation of closed-loop Stackelberg strategies and incentive policies for a general class of dynamic decision problems with a hierarchical decision structure, in both discrete and continuous time. The first set of results involve discrete-time dynamic games in which the leader has informational advantage over the follower, in the sense that he can observe the follower's actions at each stage (before he acts) either perfectly or partially. Under a feedback Stackelberg solution concept that takes this informational advantage into account, we have studied derivation of optimal affine policies. Furthermore, we have investigated the conditions under which such a solution coincides with the global Stackelberg solution (cf. Section III).

A second set of results presented in this paper has involved an analysis of existence and derivation of causal real-time implementable global Stackelberg solutions in dynamic games wherein the leader is allowed to use memory policies. In this context, we have treated both discrete-time and continuous-time problems, and using a function space approach we have solved certain special cases both analytically and numerically (Sections II, IV, and V).

Appendix A

Derivation of (71)–(72):
Since $x(t) = e^{At}x_0 + \int_0^t e^{A(t-\tau)}(Bu(\tau) + Cv(\tau)\,d\tau$,

$$\langle \delta x, Q_2 x\rangle = \int_0^\infty \delta x'(t)Q_2x(t)\,dt$$

$$= \int_0^\infty \int_0^t (\delta v'(\tau)C' + \delta u'(\tau)B')e^{A'(t-\tau)}Q_2x(t)\,d\tau\,dt$$

$$= \int_0^\infty d\tau \int_\tau^\infty dt\left[\delta v'(\tau)C' + \delta u'(\tau)B'\right]e^{A'(t-\tau)}Q_2x(t).$$

Therefore, for variations δu and δv we have

$$\frac{1}{2}\delta J_2 = \langle \delta x, Q_2x\rangle + \langle \delta u, R_2u\rangle + \langle \delta v, S_2v\rangle$$

$$= \langle \delta u, R_2u + \int_\tau^\infty B'e^{A'(t-\tau)}Q_2x(t)\,dt\rangle + \langle \delta v, S_2v + \int_\tau^\infty C'e^{A'(t-\tau)}Q_2x(t)\,dt\rangle$$

$$\frac{1}{2}\nabla_u J_2 = R_2u + \int_\tau^\infty B'e^{A'(t-\tau)}Q_2x(t)\,dt$$

$$\frac{1}{2}\nabla_v J_2 = S_2v + \int_\tau^\infty C'e^{A'(t-\tau)}Q_2x(t)\,dt.$$

When $u = u^\dagger$, $v = v^\dagger$,

$$\frac{1}{2}\phi(\tau) = \frac{1}{2}\nabla_u J_2(u^\dagger, v^\dagger) = R_2u^\dagger + B'e^{-A'\tau}\left[\int_\tau^\infty e^{A't}Q_2e^{A_ct}\,dt\right]x_0$$

$$\frac{1}{2}\Psi(\tau) = \frac{1}{2}\nabla_v J_2(u^\dagger, v^\dagger) = S_2v^\dagger + C'e^{-A'\tau}\left[\int_\tau^\infty e^{A't}Q_2e^{A_ct}\,dt\right]x_0.$$

Since

$$I_\tau = \int_\tau^\infty e^{A't}Q_2e^{A_ct}\,dt$$

$$= e^{A}e^{A_ct}A_c^{-1}|_\tau^\infty - \int_\tau^\infty A'e^{A't}Q_2e^{A_ct}A_c^{-1}\,dt$$

$$= A'^{-1}e^{A't}Q_2e^{A_ct}|_\tau^\infty - \int_\tau^\infty A'^{-1}e^{A't}Q_2e^{A_ct}A_c\,dt,$$

it follows that

$$A'I_\tau + I_\tau A_c + e^{A'\tau}Q_2^{A_c\tau} = 0.$$

Let I_0 satisfy

$$A'I_0 + I_0A_c + Q_2 = 0;$$

then

$$I_\tau = e^{A'\tau}I_0e^{A_c\tau}.$$

Therefore,

$$\frac{1}{2}\phi(t) = -R_2R_1^{-1}B'Px^\dagger(t) + B'I_0x^\dagger(t)$$

$$\frac{1}{2}\Psi(t) = -S_2S_1^{-1}C'Px^\dagger(t) + C'I_0x^\dagger(t)$$

and relations (73) and (74) follow.

APPENDIX B

Derivation of Gradients (80)–(81): Note that

$$\tilde{J}_2(u,x) = \langle x, Q_2 x\rangle + \langle u, R_2 u\rangle + \langle(\dot{x} - Ax - Bu), \bar{C}(\dot{x} - Ax - Bu)\rangle,$$

and consider only those x and u with their values and variations δx and δu satisfying

$$x(\infty) = \dot{x}(\infty) = 0$$
$$\delta x(0) = \delta x(\infty) = \delta\dot{x}(\infty) = 0$$
$$u(\infty) = \delta u(\infty) = 0.$$

We have, for variations δx and δu:

$$\delta\langle\dot{x}, \bar{C}\dot{x}\rangle = 2\langle\delta\dot{x}, \bar{C}\dot{x}\rangle = 2\int_0^\infty \delta\dot{x}(t)'\bar{C}\dot{x}(t)\,dt$$
$$= 2\delta x'\bar{C}\dot{x}|_0^\infty - 2\int_0^\infty \delta x'\bar{C}\ddot{x}\,dt$$
$$= -2\int_0^\infty \delta x'\bar{C}\ddot{x}\,dt$$

$$\nabla_x\langle\dot{x}, \bar{C}\dot{x}\rangle = -2\bar{C}\ddot{x}$$

$$\delta\langle\dot{x}, \bar{C}Bu\rangle = \int_0^\infty \delta\dot{x}'\bar{C}Bu\,dt + \int_0^\infty \dot{x}'\bar{C}B\delta u\,dt$$
$$= \int_0^\infty \dot{x}'\bar{C}B\delta u\,dt + \delta x'\bar{C}Bu|_0^\infty - \int_0^\infty \delta x'\bar{C}B\dot{u}\,dt$$

$$\nabla_x\langle\dot{x}, \bar{C}Bu\rangle = -\bar{C}B\dot{u}$$
$$\nabla_u\langle\dot{x}, \bar{C}Bu\rangle = B'\bar{C}\dot{x}.$$

Therefore,

$$\frac{1}{2}\nabla_u\tilde{J}_2 = R_2u - B'\bar{C}\dot{x} + B'\bar{C}Ax + B'\bar{C}Bu$$
$$= R_2u - B'\bar{C}Cv$$

$$\frac{1}{2}\nabla_x\tilde{J}_2 = Q_2x - \bar{C}\ddot{x} + A'\bar{C}Ax + \bar{C}A\dot{x}$$
$$-A'\bar{C}\dot{x} + \bar{C}B\dot{u} + A'\bar{C}Bu$$
$$= Q_2x - \bar{C}C\dot{v} - A'\bar{C}Cv$$
$$= Q_2x - C^{+\prime}S_2\dot{v} - A'C^{+\prime}S_2v.$$

At point $(u^\dagger, x^\dagger)$, these expressions become equal to

$$\frac{1}{2}\phi(t) = \frac{1}{2}\nabla_u\tilde{J}_2(u^\dagger, x^\dagger)$$
$$= \left(-R_2R_1^{-1}B'Px^\dagger + B'\bar{C}CS_1^{-1}C'Px^\dagger\right)(t)$$
$$= \left(B'\bar{C}CS_1^{-1}C' - R_2R_1^{-1}B'\right)Px^\dagger(t)$$
$$\triangleq M(t)x^\dagger(t)$$

$$\frac{1}{2}\Psi(t) = \frac{1}{2}\nabla_x\tilde{J}_2(u^\dagger, x^\dagger)$$
$$= \left(Q_2x^\dagger + C^{+\prime}S_2S_1^{-1}C'P\dot{x}^\dagger + A'C^{+\prime}S_2S_1^{-1}C'Px^\dagger\right)(t)$$
$$= \left(Q_2x^\dagger + C^{+\prime}S_2S_1^{-1}C'PA_cx^\dagger + A'C^{+\prime}S_2S_1^{-1}C'Px^\dagger\right)(t)$$
$$= \left(Q_2 + C^{+\prime}S_2S_1^{-1}C'PA_c + A'C^{+\prime}S_2S_1^{-1}C'P\right)x^\dagger(t) = N(t)x^\dagger(t).$$

This then completes the verification.

REFERENCES

[1] M. Athans, "On large-scale systems and decentralized control," *IEEE Trans. Automat. Contr.*, vol. AC-23, 1978.

[2] K. J. Arrow and R. Radner, "Allocation of resources in large teams," *Econometrica*, vol. 47, 1979.

[3] T. Başar, "Equilibrium strategies in dynamic games with multilevels of hierarchy," *Automatica*, vol. 17, pp. 749–754, 1981.

[4] ____, "A general theory for Stackelberg games with partial state information," *Large Scale Systems*, vol. 3, pp. 47–56, 1982.

[5] T. Başar and J. B. Cruz, Jr., "Concepts and methods in multi-person Coordination and Control," in *Optimization and Control of Dynamic Operational Research Models*, S. G. Tzafestas, Ed. Amsterdam: North Holland, 1982, Ch. 11, pp. 351–387.

[6] T. Başar and G. J. Olsder, "Team-optimal closed-loop Stackelberg strategies in hierarchical control problems," *Automatica*, vol. 16, pp. 409–414, 1980.

[7] ____, *Dynamic Noncooperative Game Theory.* London: Academic, 1982.

[8] T. Başar and H. Selbuz, "A new approach for derivation of closed-loop Stackelberg strategies," *Proc. 19th IEEE Conf. Decision and Control*, San Diego, CA, 1979, pp. 1113–1118.

[9] ____, "Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 166–179, 1979.

[10] D. Cansever and T. Başar, "A minimum sensitivity approach to incentive design problems," in *Proc. 21st IEEE Conf. Decision and Control*, Orlando, FL, 1982, pp. 158–163.

[11] C. I. Chen and J. B. Cruz, Jr., "Stackelberg solutions for two-person games with biased information patterns," *IEEE Trans. Automat. Contr.*, vol. AC-17, pp. 791–798, 1972.

[12] J. B. Cruz, Jr., "Survey of Nash and Stackelberg equilibrium strategies in dynamic games," *Annals of Economic and Social Management*, vol. 4, pp. 339–344, 1975.

[13] ____, "Leader–follower strategies for multilevel systems," *IEEE Trans. Automat. Contr.*, vol. AC-23, pp. 244–255, 1978.

[14] B. F. Gardner, Jr. and J. B. Cruz, Jr., "Feedback Stackelberg strategies for a two-player game," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 270–271, 1977.

[15] T. Groves, "Incentives in teams," *Econometrica*, vol. 41, pp. 617–631, 1973.

[16] T. Groves and M. Loeb, "Incentives in a divisionalized firm," *Management Sci.*, vol. 25, pp. 221–230, 1979.

[17] Y. C. Ho, P. B. Luh, and R. Muralidharan, "Information structure, Stackelberg games, and incentive controllability," *IEEE Trans. Automat. Contr.*, vol. AC-26, pp. 454–460, 1981.

[18] Y. C. Ho, P. B. Luh, and G. J. Olsder, "A control-theoretic view on incentives," *Automatica*, vol. 18, pp. 167–179, 1982.

[19] L. P. Jennergren, "On the design of incentives in business firms—A survey of some research," *Management Sci.*, vol. 26, pp. 180–201, 1980.

[20] P. B. Luh, Y. C. Ho, and R. Muralidharan, "Load adaptive pricing: an emerging tool for electric utilities," *IEEE Trans. Automat. Contr.*, vol. AC-27, pp. 320–328, 1982.

[21] G. P. Papavassilopoulos and J. B. Cruz, Jr., "Nonclassical control problems and Stackelberg games," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 155–166, 1979.

[22] R. Radner, "Monitoring cooperative agreements in a repeated principal–agent relationship," *Econometrica*, vol. 49, pp. 1127–1148, 1981.

[23] M. Salman and J. B. Cruz, Jr., "An incentive model of duopoly with government coordination," *Automatica*, vol. 17, pp. 821–829, 1981.

[24] M. Simaan and J. B. Cruz, Jr., "On the Stackelberg strategy in nonzero-sum games," *J. Optimiz. Theory Appl.*, vol. 11, pp. 533–555, 1973.

[25] ____, "Additional aspects of the Stackelberg strategy in nonzero-sum games," *J. Optimiz. Theory Appl.*, vol. 11, pp. 613–626, 1973.

[26] H. Von Stackelberg, *Marktform und Gleichgewicht.* Vienna: Springer-Verlag. Eng. transl., *The Theory of the Market Economy.* Oxford: Oxford University Press, 1952.

[27] B. Tolwinski, "Closed-loop Stackelberg solution to multi-stage linear-quadratic game," *J. Optimization Theory and Applications*, vol. 34, pp. 485–501, 1981.

[28] Y. P. Zheng and T. Başar, "Existence and derivation of optimal affine incentive schemes for Stackelberg games with partial information: A geometric approach." *Int. J. Control*, vol. 35, no. 6, pp. 997–1011, 1982.

[29] A. V. Balakrishnan, *Applied Functional Analysis.* New York: Springer-Verlag, 1976.

An Improved Method for Solving Multiple Criteria Problems Involving Discrete Alternatives

M. MURAT KÖKSALAN, MARK H. KARWAN, AND STANLEY ZIONTS

***Abstract*—An approach is presented for solving a discrete multiple criteria problem. The approach asks pairwise comparisons of a decisionmaker. Under mild assumptions, the method obtains the most preferred alternative. The required number of pairwise comparisons is generally modest. Our experience with the method indicates that for reasonable underlying utility functions, a heuristic stopping rule generally yields the most preferred alternative after several comparisons, usually fewer than 20.**

I. Introduction

IN THIS PAPER we consider discrete alternative multiple criteria decisionmaking for the case in which all criteria are cardinal. We consider a problem of choosing among alternatives where a quantity or score on a natural scale exists to represent the performance of each alternative on each criterion. The selection of the best alternative is complex because of the presence of multiple criteria. The choice would be simple in two cases:

- if one of the alternatives is at least as good as every other alternative with respect to every criterion under consideration; or
- if we know the underlying utility function (or value function) of the decisionmaker as a function of the considered criteria.

In the first case the alternative that is superior in each criterion is clearly the most preferred alternative. In the second case the most preferred alternative may be found by substituting the scores of each alternative into the utility function. The alternative that maximizes the utility function is the most preferred alternative. However, both of these cases are rather rare in practice.

A number of methods have been suggested for the solution of this problem. Many of these methods construct a composite function to approximate an assumed underlying utility function. Some of these methods may be classified as methods of conjoint analysis (see Green and Srinivasan [1]) which require holistic evaluations among alternatives by the decisionmaker in constructing a composite function. The burden placed on the decisionmaker in these evaluations is generally substantial. In the decision analysis approach, utility functions are first constructed for each criterion. These functions are then combined into a composite function approximating the utility function (see Keeney and Raiffa [3]). The utility function is then used to compare alternatives. There is a variety of approaches that do not use a composite function. Rivett [7] uses multidimensional scaling techniques and obtains a graph of alternatives in which the most preferred and the least preferred alternatives are at opposite ends. The method requires considerable input from the decisionmaker. The methods based on outranking relations (e.g., Roy [8], Siskos [9]) also do not require a composite function. The approach suggested by Zionts [12] assumes an underlying linear utility function and finds the best alternative by asking a number of comparisons between pairs of alternatives. Korhonen *et al.* [5] extend this approach to the case of an underlying quasiconcave nondecreasing utility function.

The approach we suggest is based on that of Korhonen *et al.* which is discussed further in Section II. In Section III we present some theoretical results on which our method is based. Other aspects of the approach are presented in

Manuscript received March 16, 1983; revised July 1983.

M. M. Köksalan and M. H. Karwan are with the Department of Industrial Engineering, and S. Zionts is with the School of Management, the State University of New York, Buffalo, NY 14214.

0018-9472/84/0100-0024$01.00