

Solutions and Properties of Multi-stage Stackelberg Games*

PETER B. LUH,† SHI-CHUNG CHANG† and TSU-SHUAN CHANG‡

Key Words—Game theory; Stackelberg games; hierarchical decision making.

Abstract—The inducible region concept is used to solve multi-stage, deterministic Stackelberg games, and to study properties of the solutions. We first delineate the inducible region by using a process similar to dynamic programming. A desired outcome is then selected from the region. To construct a Stackelberg strategy, we move forwards in time to cover various contingent situations. This approach gives a distinct picture about the nature of the problem. In particular, two aspects of the principle of the optimality are discussed. A design procedure is presented to construct a relatively robust Stackelberg strategy. The issue of credibility is also addressed.

1. Introduction

STACKELBERG games deal with multi-person hierarchical optimization problems. For the two-level case, one particular decision maker, known as the leader, is in a position to declare strategies and implement controls to induce the behavior of other decision makers (known as the followers). The leader's major task for such a problem is the design of his strategy, by taking into account followers' rational reactions, to minimize his cost. Many real world problems possess, more or less, this feature. Examples include resource allocation, organization theory, electricity pricing (Luh, Ho and Muralidharan, 1982), and command and control in the military context.

Among recent results on this topic, many achieve the leader's absolute minimum cost—the team cost (Basar and Selbuz, 1979; Basar, 1982; Tolwinski, 1981a; Ho, Luh and Olsder, 1982; Chang and Ho, 1983). This says that under the conditions provided, the leader is able to induce the followers to behave as if they were also minimizing his cost function. While it is intuitively clear that the team cost may not always be achievable, another methodology began to emerge (Tolwinski, 1981b, 1983), and the concept of inducible region was then formally introduced in Chang and Luh (1982a, b). [For a historical discussion on the development, see Section 6 of Chang and Luh (1983).] The basic idea can be explained as follows. For a given leader's strategy, the followers are assumed to react rationally. The resulting leader's and followers' decisions can be thought of as an 'outcome' of the problem. If the leader chooses a different strategy, the outcome will be changed accordingly. The *inducible region* is then defined as the collection of these outcomes for all the leader's strategies. Consequently, the optimal outcome the leader can induce is the best one in the inducible region, and a procedure to solve a Stackelberg problem is to go through the following steps: (1) delineate the inducible region, (2) select the best outcome from the inducible region, and (3) construct a Stackelberg strategy to induce the occurrence of the desired outcome. Results on Stackelberg games obtained by using the inducible region approach have been reported in Chang and Luh (1983) for single-stage, two-level problems and in Luh, Chang and Ning (1984) for three-level problems.

In this paper, we use the inducible region concept to solve two-person, multi-stage, deterministic Stackelberg games, and to investigate properties of solutions. We start in Section 2 with the formulation of the problem. In Section 3, we delineate the inducible region. Due to the causality constraint in dynamic games, the procedure to delineate it is to work backwards in time, and is thus called the *backward sweep* (in contrast to the *forward sweep* to be introduced later). In Section 4, we point out several practical considerations in the design of a Stackelberg strategy. They then suggest that we shall examine more closely the meaning of the principle of optimality in the Stackelberg game context. It is observed that the principle of optimality has two aspects, one along the optimal path and one off it. They are discussed, respectively, in Sections 5 and 6. A design procedure, the *forward sweep*, is then proposed in Section 7 for the construction of a relatively robust and credible Stackelberg strategy, the 'optimal Stackelberg strategy'. The relationship among the principles of optimality, the optimal Stackelberg strategy, and credibility are elaborated in Section 8. Comparison with Tolwinski's (1983) results and concluding remarks are given in Section 9.

2. Problem formulation

Consider a two-person Stackelberg game with DM0 as the leader and DM1 the follower. Let the strategy of DM1 be denoted as γ_1 , with $\gamma_1 \in \Gamma_1$, and the cost function be denoted as $J_1(\gamma_0, \gamma_1)$. The Stackelberg solution concept can be explained as follows. DM0 is assumed to announce his strategy γ_0 in advance and commit himself to it. For a given γ_0 , DM1 is assumed to select a $\gamma_1 \in R(\gamma_0)$, where

$$R(\gamma_0) = \{\gamma_1 \in \Gamma_1 | J_1(\gamma_0, \gamma_1) \leq J_1(\gamma_0, \tilde{\gamma}_1) \quad \forall \tilde{\gamma}_1 \in \Gamma_1\} \quad (1)$$

is DM1's rational reaction set. It is assumed that $R(\gamma_0) \neq \emptyset$ for simplicity. DM0's major task for this problem is the design of his strategy, by taking into account DM1's rational reaction, to minimize J_0 . That is, DM0 wants to find a 'Stackelberg strategy' γ_0^* such that

$$\sup J_0(\gamma_0^*, \gamma_1) \leq \sup J_0(\gamma_0, \tilde{\gamma}_1) \quad \forall \gamma_0 \in \Gamma_0, \quad (2) \\ \gamma_1^* \in R(\gamma_0^*) \quad \gamma_1^* \in R(\gamma_0)$$

The supremes in the above inequality take into account the possible nonunique reactions of DM1. The value on the left-hand side of (2) is DM0's Stackelberg cost, denoted as J_0^* .

Consider now the extensive form of a deterministic, T-stage game. The strategy of DM1 at stage t is denoted as γ_{1t} , and its corresponding decision is denoted as u_{1t} , with $u_{1t} \in U_{1t}$, $t = 0, 1, \dots, T-1$. For notational simplicity, we also define the following terms

$$u_t^1 = (u_{1t}, u_{2t}, \dots, u_{Nt}, \gamma_{1t})$$

$$u_t^0 = (u_{0t}, u_{1t}, \dots, u_{Nt}, \gamma_{1t})$$

$$u_t = (u_{0t}, u_{01}, \dots, u_{Nt}, \gamma_{1t})$$

Similar notations will be used for other variables, functions, and sets. Without loss of generality, we assume that the system dynamics has been substituted out [Ho (1980), p. 651]. Therefore there is no more state variable, and at stage t , the game is characterized by (x_t, u_t^0, u_t^1) , with x_t being the initial state. The sequence of actions is shown in Fig. 1. At each stage, DM1 acts

* Received 12 October, 1982; revised 21 April, 1983; revised 30 August, 1983. The original version of this paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by associate editor T. Başar under the direction of editor A. Sage.

† Department of Electrical Engineering and Computer Science, University of Connecticut, Storrs, CT 06268, U.S.A.

‡ Department of Electrical Engineering, State University of New York at Stony Brook, NY 11794, U.S.A.

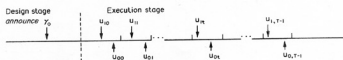


FIG. 1. Sequence of actions.

first, and DM0 acts next.[†] DM0 is assumed to have complete information and perfect memory. That is to say at stage t , DM0 observes u_{1t} precisely, and recalls (x_0, u_0^t, u_1^t) perfectly.[‡] We assume that DM1 has open-loop information.[§] Consequently we have

$$\gamma_{0t}: X_0 \times U_0^t \times U_1^{t-1} \rightarrow U_{0t}$$

$$\gamma_{1t}: X_0 \rightarrow U_{1t}$$

where γ_0 is assumed to be measurable with respect to its information set.[¶] Finally, given a tuple (γ_0, γ_1) and x_0 , we shall rewrite the cost function $J_t(\gamma_0, \gamma_1)$ as $J_t(u_0, u_1, x_0)$.

3. Delineation of the inducible region

As mentioned in Section 1, the inducible region is the collection of the outcomes for all DM0's strategies. For the multi-stage problem just formulated, the inducible region can be stated more precisely as follows. For a given γ_0 , DM1's reaction might not be unique. According to (2), DM0 assumes in his strategy designing stage that DM1 chooses a particular γ_1^t where

$$\gamma_1^t \in \arg \sup_{\gamma_1 \in R_{DM1}} J_0(\gamma_0, \gamma_1) \quad (3)$$

We assume for simplicity that such a γ_1^t exists. The corresponding outcome (u_0^t, u_1^t) , i.e. the one that satisfies

$$u_{1t} = \gamma_{1t}(x_0)$$

and

$$u_{0t} = \gamma_{0t}(u_0^{t-1}, u_1^{t-1}, x_0), \quad \text{for } 0 \leq t \leq T-1 \quad (4)$$

is thus called the outcome from DM0's viewpoint for this γ_0 . As in Chang and Luh (1983), the inducible region is then defined as

$$IR = \{(u_0^t, u_1^t) \mid \text{there exists a } \gamma_0 \in \Gamma_0 \text{ s.t. } (u_0^t, u_1^t) \text{ is the outcome from DM0's viewpoint}\} \quad (5)$$

From the above definition, DM0 cannot let any point outside IR be the outcome of the game. Consequently, we have the following theorem.

Theorem 1.

(a) The leader's Stackelberg cost is given by

$$J_0^* = \inf_{(u_0, u_1) \in IR} J_0(u_0, u_1, x_0) \quad (6)$$

(b) A Stackelberg strategy γ_0^* exists iff (6) has a solution in IR.

[†] The results can be extended, without major difficulties, to cases where DM0 acts first, or DM0 and DM1 act simultaneously. For details, please see Chang (1983).

[‡] For a problem with incomplete information, we first have to go through a process that converts it to a one with complete information (Basar, 1982), and then apply the method discussed here. For details, please see Chang (1983).

[§] Once γ_0 is announced, DM1 faces an optimal control problem. Under the deterministic formulation, open-loop information and closed-loop information have the same result. Thus we can assume, without loss of generality, that DM1 has open-loop information.

[¶] In this paper, we consider pure strategies only.

From the above theorem, we can see that an essential problem is to delineate IR. To do this, we shall incorporate the inducible region concept for intermediate subgames. Let $IR_t(u_0^t, u_1^t, x_0)$ represent the inducible region of a subgame starting at stage t with (u_0^t, u_1^t, x_0) as the past history. That is, if we consider a new Stackelberg game starting at stage t with (u_0^t, u_1^t, x_0) as given, then $IR_t(u_0^t, u_1^t, x_0)$ is its inducible region. A pair of (u_0, u_1) is said to be stage- t -inducible if $(u_0, u_1) \in IR_t(u_0^t, u_1^t, x_0)$. Let $D_t(x_0)$ denote the collection of all stage- t -inducible pairs of (u_0, u_1) , i.e.

$$D_t(x_0) = \{(u_0, u_1) \mid (u_0, u_1) \in IR_t(u_0^t, u_1^t, x_0), 0 \leq t \leq T-1\} \quad (7)$$

For the convenience of derivation, we also define

$$D_T(x_0) = \{(u_0, u_1) \mid (u_0, u_1) \in U_0 \times U_1\} \quad (8)$$

as the set of all the possible combinations of u_0 and u_1 . Due to the causality constraint, the key step in delineating IR is the design of a recursive algorithm that generates IR, from IR_{T-1} backwards in time, so that the inducible region for the whole game, IR, can be obtained at the end of the recursion.

Consider a subgame starting at stage t with (u_0^t, u_1^t, x_0) specified, $0 \leq t \leq T-1$. Assume that $D_{t+1}(x_0)$ is available, either from the result of stage $t+1$ for $0 \leq t \leq T-2$, or from (8) for $t = T-1$. It can be shown that a pair (u_0^t, u_1^t) is in $IR_t(u_0^t, u_1^t, x_0)$ if and only if the following two conditions are satisfied:

1. $(u_0^t, u_1^t) \in D_{t+1}(x_0)$. For $t = T-1$, this must be the case. For $0 \leq t \leq T-2$, if $(u_0^t, u_1^t) \notin D_{t+1}(x_0)$, then DM1 will not choose u_1^{t+1} later on, and thus $(u_0^t, u_1^t) \notin IR_t(u_0^t, u_1^t, x_0)$, a contradiction.
2. (u_0^t, u_1^t) is preferred by DM1 at stage t . Otherwise, DM1 will not select u_0^t at stage t . This implies that DM0 is able to make DM1 worse off if he deviates at stage t .

Motivated by the second condition, we define the following term:

$$\begin{aligned} M_t(u_0^t, u_1^t, x_0) &\hat{=} \inf_{u_0} \sup_{u_1} M_{t+1}(u_0^{t+1}, u_1^{t+1}, x_0) \\ &= \inf \sup \inf \sup \dots \\ &\quad \inf \sup J_t(u_0, u_1, x_0), \\ &\quad \text{for } 0 \leq t \leq T-1 \end{aligned} \quad (P-2)$$

with $M_T(u_0^T, u_1^T, x_0) \hat{=} J_T(u_0, u_1, x_0)$ as the terminal condition. In conjunction with the first condition, we then obtain[†]

$$\begin{aligned} IR_t(u_0^t, u_1^t, x_0) &= \{(u_0^t, u_1^t) \mid (u_0^t, u_1^t) \in D_{t+1}(x_0) \text{ and} \\ &J_t(u_0, u_1, x_0) < M_t(u_0^t, u_1^t, x_0)\} \cup S_{2t}(u_0^t, u_1^t, x_0) \end{aligned} \quad (9)$$

where

$$S_{2t}(u_0^t, u_1^t, x_0) \hat{=} \{(u_0^t, u_1^t) \mid (u_0^t, u_1^t)$$

that maximizes J_0 among the solutions of (P-2)[‡].

The intuitive interpretation is that no matter what DM0's future decision sequence u_0^t is, DM1 can always obtain a cost no greater than M_t by minimizing J_t . Thus any pair (u_0, u_1) with $J_t > M_t$ is not stage- t -inducible because DM1 will not choose such a (u_0, u_1) . On the other hand, any pair in $D_{t+1}(x_0)$ with $J_t < M_t$ is stage- t -inducible, since DM0 can penalize DM1 at least to M_t by maximizing J_t if DM1 deviates at stage t . Finally,

S_2 is the set of inducible outcomes on the $J_1 = M_t$ boundary [according to (3)]. Detailed derivations are provided in Chang (1983). The set of stage- t -inducible outcomes is then given by

$$D_t(x_0) = \{(u_0, u_1) | (u_0^t, u_1^t) \in IR_t(u_0^t, u_1^t, x_0)\}. \quad (10)$$

By repeating the above process, we can obtain at stage 0 the inducible region for the entire game, i.e. $IR_0(x_0) = D_0(x_0) = IR$. The optimal inducible outcome (u_0^0, u_1^0) can then be found by solving the parameter optimization problem (6).

The above procedure for delineating IR is to go from the terminal stage backwards in time, thus it is called the backward sweep. It should be noted that in the backward sweep, the set D_t shrinks at each stage, i.e. $D_t \subset D_{t+1}$, $0 \leq t \leq T-1$. This is caused by the deletion of outcomes when $M_t < M_{t+1}$.

Example 1. Let us consider the two-stage game tree of Fig. 2. At each stage, U_t is the set $\{0, 1\}$, where 0 represents the upward branch and 1 represents the downward branch. The decision sequence is denoted by $(u_0, u_{00}; u_{11}, u_{01})$, and the fixed initial state x_0 is dropped for all the expressions.

First for stage 1 (the terminal stage), consider a subgame with the past history $(u_{10} = 0, u_{00} = 0)$. We have $M_1(u_{10} = 0, u_{00} = 0) = 8$ from $(P-2, -1)$ and $IR_1(u_{10} = 0, u_{00} = 0) = \{(u_{11}, u_{01}) | (0, 0), (0, 1)\}$ from (9). M_1 and IR_1 for different past histories can be derived in a similar way. The elements in D_1 are marked by \times in Fig. 2. Now consider stage 0. From $(P-2)$ and (9), we have $M_0 = 6$, and $IR = IR_0$ consists of outcomes marked by Δ . Minimizing J_0 over IR, we see that the desired outcome is $(u_0^0, u_1^0; u_{11}^0, u_{01}^0) = (1, 0, 0, 0)$, with $J_0^0 = 0$, $J_1^0 = 3$.

4. Strategy construction

In the previous section, we presented a recursive algorithm to delineate the inducible region. The issue of how to construct a strategy to induce the desired outcome remains untouched. Consider first the 'maximum penalty strategy' γ_0^M specified by

$$u_{0t} = \begin{cases} u_0^{t-1} & \text{if } u_1^{t-1} = (u_1^t)^{t-1} \\ 1 & \text{otherwise,} \end{cases}$$

† Comparing to equation (26) of Tolwinski (1983), we see that the condition $(u_0, u_1) \in D_{t+1}(x_0)$ corresponds to $J_t < M_t$ for $s \geq t+1$.

where γ_0^M is essentially the solution to the supremization subproblem in $(P-2)$. Precise definition of γ_0^M can be found in Chang (1983). In other words, if DMI behaves desirably, DM0 selects the desired u_0^0 . Otherwise, DM0 punishes DMI to the most extent by selecting the u_0 that maximizes J_1 . It can be easily seen that the maximum penalty strategy can induce any desired outcome in IR. γ_0^M for Example 1 is shown in Fig. 2 by dotted lines.

In addition to γ_0^M , it is well known that in many cases DM0 has a lot of flexibility in designing his Stackelberg strategy to induce his desired outcome (Ho, Luh and Oldser, 1982). As an example, the strategy shown in Fig. 3 is also a Stackelberg strategy for Example 1. In fact, the only conditions for a strategy to be a Stackelberg strategy are

$$(C1) \text{ if } u_1 = u_1^t, \text{ then } u_0 = u_0^t;$$

$$(C2) \text{ if } u_1 \neq u_1^t, \text{ the resultant } J_1 \text{ should be greater than } J_1(u_0^t, u_1^t, x_0).$$

It is possible that the resultant J_1 equals $J_1(u_0^t, u_1^t, x_0)$ if the corresponding J_0 is less than or equal to J_0^t , see (3).

The condition (C2) is the 'declarability condition'† which says that DMI must be penalized upon deviations so that the desired outcome (u_0^0, u_1^0) is credible. If we stick to the ideal case that (i) both decision makers are completely rational and will not make any mistake, and (ii) DM0 will commit himself to the announced strategy, then all the Stackelberg strategies are equivalent since they yield the same costs. Thus although DM0 might have much flexibility in designing his Stackelberg strategy, it is not worthwhile to further discuss strategy construction since γ_0^M has already been obtained. However, it is of practical interest to consider the following issues for multi-stage games:
1. DMI might deviate unintentionally. A quick examination of Fig. 2 shows that γ_0^M may not be a good strategy if DMI deviates. For instance, when deviation happens at stage 1, then DM1 pays a high cost ($J_0 = 7$) in punishing DMI by adopting γ_0^M . It can be seen easily that by letting $u_{01} = 0$ instead of 1, DM0 still penalizes DMI for his deviation ($J_1 = 5 > J_1^0 = 3$), but this is much better for DM0 ($J_0 = 3 < J_0^M = 7$).

† The word 'declarability' was suggested by Professor Y. C. Ho (1983).

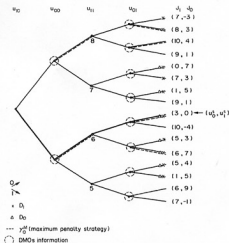


FIG. 2. Example 1.

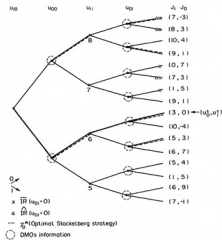


FIG. 3. Optimal Stackelberg strategy for Example 1.

2. DM1 might not believe that DM0 will carry out the announced strategy. We shall again use Example 1 to illustrate this point. Firstly, along the optimal path (u_0^*, u_0^*) : $(u_1^*, u_1^*) = (1, 0, 0, 0)$, it is doubtful that DM0 will stick to $u_0^* = 0$ after the desired $(1, 0, 0, 0)$ has been realized. Since in this case, he can select $u_0 = 1$ and get $J_0 = -4$ rather than end up with $J_0 = 0$. Note that the path associated with $u_0 = 1$, i.e. $(1, 0, 0, 1)$, cannot be announced at the beginning as the desired outcome since it gives $J_1 = 10$ and thus is not inducible. Worrying about being taken advantage by DM0 and getting $J_1 = 10$, DM1 might tend to select $u_1 = 1$. Secondly, off the optimal path, suppose that DM1 already deviated at stage 0 with $u_0 = 0$. He might wonder whether DM0 will really implement γ_0^* or do something smarter (e.g. $u_0 = 0$; $u_1 = 0$ if $u_1 = 0$ and $u_1 = 1$ if $u_1 = 1$ (as shown in Fig. 3)).

From the above discussions, we see that the construction of a closed-loop Stackelberg strategy is no longer straightforward when DM0 wants to have a relatively 'robust' and 'credible' strategy. In order to clarify some of the issues, and to provide a design method for the construction of a 'good' Stackelberg strategy, we shall next examine the principle of optimality in the Stackelberg game context.

5. The principle of optimality for Stackelberg games

In a Stackelberg game, the principle of optimality has two aspects, one along the optimal path (about the desired outcome) and one off the optimal path (about deviations). In this section, we shall discuss the one along the optimal path.

Principle of optimality along the optimal path (POP-I):

A desired outcome (u_0^*, u_1^*) is said to satisfy POP-I if it has the following property: for every $0 \leq t \leq T-1$, when $(u_0^*)^t, (u_1^*)^{t+1}$ has been realized, $(u_0^*)^t, (u_1^*)^{t+1}$ remains to be a desired outcome with regard to the new game resulting from previous decisions.

From this definition, we can see that if POP-I is satisfied, then DM0 has no incentive to break his commitment along the desired path. It is well known that POP-I does not hold for general cases (Simaan and Cruz, 1973). It is also well known that for the special case when the team outcome is achievable, then POP-I holds (Cruz, 1978; Papavasiliopoulos and Cruz, 1979; Başar and Selbuz, 1979). However, there is no satisfactory explanation for why POP-I does not hold in general. Although the so-called 'feedback Stackelberg strategy' has been introduced to be a solution satisfying POP-I, it has a very different solution concept (Simaan and Cruz, 1973).

To fully understand POP-I, we shall consider a subgame starting at the middle of stage t with $(u_0^*, u_1^*)^t, x_0$ specified.† In parallel to (3)–(5), we can define the inducible region for this subgame. It can then be seen easily that the inducible region is given by

$$\begin{aligned} \hat{R}_t(u_0^*, u_1^*)^{t+1}, x_0 &= \{(u_0^*, u_1^*)^t \mid u_0^t \in U_0, \text{ and} \\ & (u_0^*, u_1^*)^t \in R_{t+1}(u_0^*, u_1^*)^{t+1}, x_0\}, \\ & \text{for } 0 \leq t \leq T-2 \\ & = U_0, T-1, \text{ for } t = T-1. \end{aligned} \quad (11)$$

Before going into details, let us consider the example of Section 4 to fix some ideas.

Example 1 (Continued). It has been pointed out that given $(u_0^*, u_0^*; u_1^*) = (1, 0, 0, 0)$, $u_0 = 0$ is not the best choice for DM0. In terms of inducible region, this can be explained as follows. Given $(1, 0, 0, 0)$, the inducible region described by (10) is $\hat{R}_0(u_0 = 1, u_0 = 0; u_1 = 0) = \{0, 1\}$. The corresponding J_0^* are 0 and -4 , respectively. Thus if DM0 has a chance to select u_0 , again (i.e. discard the strategy he announced previously), the new

desired decision will be $u_0 = 1 \neq u_0^*$. Note that because the sequence $(1, 0, 0, 1) \notin \hat{R}_0$, DM0 cannot select it to be the desired outcome for the entire game at the very beginning. However, in the consideration of the new game starting at the middle of stage 1, it emerges to be inducible. Since it yields a lower J_0 , DM0 will prefer it if he is given a chance to reselect an outcome.

Now consider the general case. For a given past history $(u_0^*, u_1^*)^{t-1}, x_0$, we shall say that a sequence $(u_0^*, u_1^*)^t$ emerges to be inducible at the middle of stage t if $(u_0^*, u_1^*)^t \in \hat{R}_t(u_0^*, u_1^*)^{t-1}, x_0$ but $(u_0^*, u_1^*)^t \notin \hat{R}_t(u_0^*, u_1^*)^{t-1}, x_0$. From the understanding of the above example, it can be seen that POP-I will be satisfied if and only if all the newly emerged sequences along (u_0^*, u_1^*) are not better than (u_0^*, u_1^*) for DM0. That is

$$J_0(u_0^*, u_1^*, x_0) \leq \min_{(u_0, u_1) \in \hat{R}_t(u_0^*, u_1^*)^{t-1}, x_0} J_0(u_0, u_1, x_0)$$

where

$$\begin{aligned} A = \{ & (u_0, u_1) \mid (u_0, u_1) \notin \hat{R}_t(u_0^*, u_1^*)^{t-1}, \\ & = \{(u_0^*)^t, (u_1^*)^{t+1}\} \text{ and } (u_0^*, u_1^*)^t \\ & \in \hat{R}_t(u_0^*, u_1^*)^{t-1}, x_0\} \text{ for some } t \geq 0. \end{aligned} \quad (12)$$

We can see that POP-I does not hold for general problems. The reason is that DM0 generally has more choices as new sequences emerge to be inducible, and it's very likely that one of these sequences is better than the original desired outcome. If this happens, then POP-I fails. We can also explain why POP-I holds when the team outcome is achievable, which is in IR by assumption. Since $J_0(u_0^*, u_1^*, x_0) \leq J_0(u_0, u_1, x_0)$ for any (u_0, u_1) , (u_0^*, u_1^*) remains to be preferable to any newly emerged sequences and therefore POP-I is satisfied. However, the converse is not true and can be easily seen by changing $J_1(1, 0, 0, 0)$ in example 1 from -4 to any positive real number, say 2.

6. Principle of optimality off the optimal path (POP-II)

To consider the case where DM1 deviates, we have the following definition.

Principle of optimality off the optimal path (POP-II):

A Stackelberg strategy γ_0^* is said to satisfy POP-II if it has the following property: for every $0 \leq t \leq T-1$, when the follower deviates at stage t and the leader is given a chance to reannounce his strategy, γ_0^* remains to be a Stackelberg strategy with regard to the new game resulting from previous decisions.

To see the conditions under which POP-II is satisfied, we assume first that DM1 deviates only once, and it happens at stage 0. Let γ_0^* be the announced Stackelberg strategy that satisfies both (C1) and (C2) with (u_0^*, u_1^*) as the desired outcome. Suppose that (u_0^*, u_1^*) is observed. Under the assumption that DM0 carries out γ_0^* and DM1 does not deviate afterwards (i.e. given u_0 and γ_0^*), DM1 selects u_1 to minimize J_1 , the remaining decision sequence is determined. Let it be denoted as (u_0^*, u_1^*) . The question is, if DM0 is given a chance to reannounce his strategy after the deviation is observed, will he stick to the old γ_0^* ? The answer is generally negative. The reason is that in this case, DM0 faces a new problem, and the original declarability condition (C2) is no longer in effect. Thus the inducible region for the new game is simply $\hat{R}_0(u_0, u_1, x_0)$ (10), and DM0 would select (u_0^*, u_1^*) as the newly desired outcome, where

$$(u_0^*, u_1^*) = \arg \min_{(u_0, u_1) \in \hat{R}_0(u_0, u_1, x_0)} J_0(u_0, u_1, x_0). \quad (13)$$

We see that POP-II is satisfied at (u_0, u_1, x_0) if we can let $(u_0^*, u_1^*) = (u_0, u_1)$. However, generally $(u_0^*, u_1^*) \neq (u_0, u_1)$ since the latter has to satisfy the declarability condition. POP-II at other points can be treated recursively in a similar way.

Unlike POP-I, which concentrates on (u_0^*, u_1^*) , POP-II deals exclusively with sequences off the optimal path. For a general problem, POP-I fails because the emergence of better inducible sequences (along the optimal path) that were uninducible before. On the other hand, POP-II fails because the emergence of better inducible sequences (off the optimal path) that were previously

†From the formulation of the problem, DM0 makes his decision at the middle of a stage. If at the middle of a stage DM0 is given a chance to reannounce his strategy, he has at least two options. He can either carry out the old γ_0 , and then design (u_0^*, u_1^*) for the subgame starting at stage $t+1$, or he can design (u_0^*, u_1^*) at the middle of stage t and implement it right away. Since the latter gives DM0 more choices, it yields lower J_0 . We shall thus consider the subgame starting at the middle of stage t .

uninducible, or were inducible but could not be included in the original γ_0^* because of the violation of (C2). The satisfaction of POP-II does not guarantee the satisfaction of POP-I, and POP-I does not imply POP-II either. An example is the strategy shown in Fig. 3. The fact is that POP-I and POP-II together are, in a loose sense, equivalent to the Bellman's principle of optimality in the optimal control set-up.

The failure of POP-I and POP-II for general problems immediately implies that the dynamic programming method cannot be applied to Stackelberg games. The proposed inducible region approach, which includes the backwards sweep, a parameter optimization problem (6), and the forward sweep to be introduced later, is therefore believed to be a very important methodology. The failure of POP-I and POP-II also explains the often observed phenomenon: the backoff of a policy by an authority. Usual excuses for it are the unexpected change of the situation, and the availability of new and better information, etc. Our result, however, shows that even without any of these complications, the leader might intend to backoff. Thus it is an inherent property in the leader-follower setup. Sometimes it has to rely on a third party (e.g. the court) to enforce the leader to carry out his promise. On the other hand, the design of a robust and credible strategy is therefore a very crucial issue. This will be discussed in the next section.

7. The optimal Stackelberg strategy

From the understanding of POP-I and POP-II, we now present a design procedure for the construction of a Stackelberg strategy that remains to be optimal in the presence of DM1's deviations. This strategy will be called the *Optimal Stackelberg Strategy*, and denoted by γ_0^* .

The basic philosophy is to punish DM1 in the way that hurts DM0 the least upon any deviation. Consider first the situation where DM1 deviates only once, and it happens at stage 0. Suppose $u_{10} = \bar{u}_{10} \neq u_{10}^*$. Constrained by the declarability condition (C2), the remaining decision sequence should be in the following set

$$\bar{R}_0(\bar{u}_{10}, x_0) = \{(u_0, u_1^*) \mid (u_0, u_1) \in \bar{R}(\bar{u}_{10}, x_0), \text{ and } J_1(u_0, u_1, x_0) \geq J_1(\bar{u}_{10}, u_1^*, x_0)\} \quad (14)$$

Thus the best DM0 can select is (\bar{u}_0, \bar{u}_1) , where

$$(\bar{u}_0, \bar{u}_1) = \arg \min_{(u_0, u_1) \in \bar{R}_0(\bar{u}_{10}, x_0)} J_0(u_0, \bar{u}_{10}, u_1, x_0) \quad (15)$$

The sequence (\bar{u}_0, \bar{u}_1) then specifies a portion of γ_0^* . To consider multiple deviations, we apply the above method recursively as follows. Given a past history (u_0^*, u_1^*, x_0) and the desired outcome (u_0^*, u_1^*) of the previous iteration, suppose now $u_{11} = \bar{u}_{11} \neq u_{11}^*$, DM0 selects the best (u_0^*, u_1^{*+1}) from the following set

$$\begin{aligned} \bar{R}_1(u_0^*, u_1^{*+1}, x_0) \\ = \{(u_0, u_1^{*+1}) \mid (u_0, u_1^{*+1}) \in \bar{R}(u_0^*, u_1^{*+1}, x_0) \text{ and } \\ J_1(u_0, u_1, x_0) \geq J_1(u_0^*, u_1^*, x_0)\} \quad (16) \end{aligned}$$

The difference between (12) and (14) is that in (12), we are designing a Stackelberg strategy for the new game resulting from previous decisions, and the original (C2) is not in effect. On the other hand, we are now designing a Stackelberg strategy for the original problem, thus (C2) is required. Note also that the (C2) used in delineating \bar{R}_1 in (16) is associated with the latest desired outcome, with its J_1 being greater than or equal to J_1^* . The new desired sequence (u_0^*, u_1^{*+1}) then specifies another portion of γ_0^* . By going through all the possible deviations, a complete γ_0^* can then be constructed. Note that although γ_0^* depends on (u_0^*, u_1^{*+1}, x_0) , it can actually be announced ahead of time for all possible deviations, since the declarability condition is always satisfied. Intuitively, we have a complete contingent plan to be announced at the beginning of the game. Let us now consider Example 1 again.

† The equality holds when $J_0(u_0, u_1, x_0) \leq J_0(\bar{u}_0, \bar{u}_1, x_0)$ as mentioned in Section 4.

Example 1 (Continued). To construct the optimal Stackelberg strategy, we first consider the case that DM1 deviates only once at stage 0 with $u_{10} = 0$. $\bar{R}_0(u_{10} = 0)$ is shown in Fig. 3 with elements marked by \times . $\bar{R}(u_{10} = 0)$ is shown in the same figure with elements marked by Δ . The new desired sequence is then $(u_0^*, u_1^*) = (0, 0.0)$, with $J_1 = 7$ and $J_0 = -3$. Thus, a portion of γ_0^* for $u_{10} = 0$ can be specified. We then move forwards to consider the situations where DM1 deviates at stage 1. We have $\bar{u}_{11} = 1$ for $(u_{10}, u_{11}) = (0, 0, 1)$ and $\bar{u}_{11} = 0$ for $(1, 0, 1)$. This completes the specification of γ_0^* as shown in Fig. 3 by dotted lines. It is interesting to observe that if DM1 deviates only at stage 0, he is penalized for the deviation ($J_1 = 7 > J_1 = 3$), however DM0 actually gains from DM1's mistake ($J_0 = -3 < J_0^* = 0$).

Recall that in the derivation of the inducible region, the backward sweep is used, and the set D_1 shrinks at each stage. In the designing of γ_0^* , we delineate \bar{R}_1 by moving forwards in time to consider various contingent situations. This process is thus called *forward sweep*. In the forward sweep, two things happen: *new sequences emerge as discussed in Section 5, and some existing sequences are deleted due to the violation of (C2)*. One interesting phenomenon is that since some of the originally uninducible sequences might become inducible through DM1's deviations, DM0 might even be better off. However, there is no way for DM0 to induce the occurrence of such deviations. Otherwise DM0 will select that sequence as the desired outcome at the very beginning. Note that although γ_0^* might still not be unique, the set of possible γ_0^* is much smaller than the set of all γ_0^* . It is also easy to see that if a γ_0^* satisfies POP-II, then it must be an optimal Stackelberg strategy. However, the converse is not true.

8. Credibility and optimality

In a Stackelberg game, DM0 is required to carry out his announced strategy in advance and it is assumed to carry out his announced strategy later on. Whether the latter is the case is the issue of credibility, and has been addressed in Ho and Olsder (1981) and Ho and Tolwinski (1982). With the concept of POP-I, POP-II and the optimal Stackelberg strategy, the following remarks can be made.

1. If a desired outcome satisfies POP-I, there will be no incentive for DM0 to break his commitment along the optimal path.
2. A Stackelberg strategy satisfying POP-II has the nice property of 'one stone, two birds' in that it penalizes DM1's deviation and minimizes DM0's cost at the same time. There will be no incentive for DM0 to renege his strategy upon any DM1's deviation.
3. If a strategy satisfies POP-I and POP-II simultaneously, it is rather credible. However, to be fully credible, a Stackelberg strategy has to satisfy POP-II and 'POP-I everywhere'. The latter means that POP-I holds not only for (u_0^*, u_1^*) , but also for all the desired outcomes considered in the contingent plan. Such an example can be found by changing $J_0(1, 0, 0, 1)$ of Fig. 3 from -4 to 2 , then the γ_0^* shown is a fully credible strategy.
4. For a given problem, there might not exist a Stackelberg strategy that satisfies POP-II and 'POP-I everywhere'. Then the optimal Stackelberg strategy of Section 7 will be a reasonable choice for DM0.

From the above discussion, we can see that under the deterministic formulation, the understanding of POP-I, POP-II and the forward sweep procedure in constructing an optimal Stackelberg strategy renders much insight to the credibility issue. It also provides a systematic way of designing a relatively robust and credible Stackelberg strategy.

9. Conclusions

In this paper, the inducible region concept is used to solve multi-stage Stackelberg games. The methodology developed includes several steps. First, the backward sweep is used to delineate the inducible region. The optimal inducible outcome is then selected from the delineated region. Finally, the forward sweep is used to construct a closed-loop Stackelberg strategy. Although a simple two-stage game tree was used to illustrate key ideas, the results are applicable to infinite games as well.

Those who are familiar with Tolwinski's (1983) results might see mathematical resemblance between his results and the result of Section 3 in this paper where the inducible region is delineated. As mentioned in Section 1, a historical discussion on the

development of the inducible region approach has been provided in Section 6 of Chang and Luh (1981). Furthermore, what makes our results distinct is the incorporation of the inducible region concept for intermediate subgames. Many qualitative aspects of multi-stage Stackelberg games can now be explored clearly with the help of the backward stage, the forward stage, and the delaying and emerging of inducible regions. In particular, POP-I, POP-II, and the issue of credibility are addressed. A simple procedure for the construction of a relatively robust and credible, and hence good Stackelberg strategy is also presented.

Acknowledgments—The authors appreciate useful discussions with Professors Y. C. Ho, T. Başar, B. Tóthvári, and Mr. Y. P. Chang. The research work in this paper was supported by the National Science Foundation under grants ECS8305982 and ECS8205673.

References

- Basar, T. and H. Sillbur (1978). Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems. *IEEE Trans. Aut. Control*, AC-24, 386.
- Basar, T. (1982). A general theory for Stackelberg games with perfect state information. *Large Scale Systems*, 3, 47.
- Chang, T.-S. and P. B. Luh (1982a). The concept of inducible region in Stackelberg games. *Proceedings of the 1982 American Nuclear Conference*, Arlington, Virginia, pp. 138-140.
- Chang, T.-S. and P. B. Luh (1982b). A complete solution for two-person, single-stage, zero-sum, Stackelberg games. *Proceedings of the 21st IEEE Conference on Decision and Control*, Orlando, Florida, pp. 176-180.
- Chang, T.-S. and Y.-C. Ho (1983). Stackelberg games: non-zero-sum multi-agent incentive problems. *IEEE Trans. Aut. Control*, AC-28, 677.
- Chang, T.-S. and P. B. Luh (1981). Derivation of necessary and sufficient conditions for single-stage Stackelberg games via the inducible region concept. *IEEE Trans. Aut. Control*, to be published.
- Chang, S.-C. (1981). Master Thesis, Department of Electrical Engineering and Computer Science, University of Connecticut.
- Chu, J., J. B. (1978). Leader-follower strategies for multilevel systems. *IEEE Trans. Aut. Control*, AC-23, 204.
- Ho, Y.-C. (1980). Team decision theory and information structure. *Proceedings of IEEE*, 68, 644.
- Ho, Y.-C., P. B. Luh and R. Marudharan (1981). Information structure, Stackelberg games and incentive controllability. *IEEE Trans. Aut. Control*, AC-26, 454.
- Ho, Y.-C. and G. J. Olsder (1981). Aspect of the Stackelberg game problem—existence, bifurc, and hierarchy. *Proceedings of the IFAC 8th Triennial World Congress*, Kyoto, Japan, Vol. 13, pp. 134-138.
- Ho, Y.-C., P. B. Luh and G. J. Olsder (1982). A control theoretic view on incentive. *Automatica*, 18, 369.
- Ho, Y.-C. and B. Tóthvári (1982). Credibility and rationality of players' strategy in multilevel games. *Proceedings of the 20th IEEE Conference on Decision and Control*, Orlando, Florida, pp. 609-613.
- Ho, Y.-C. (1983). On incentive problems. *Systems & Control Letters*, 3, 63-68.
- Lin, P. B., Y.-C. Ho and R. Marudharan (1982). Load adaptive pricing: an emerging tool for electric utilities. *IEEE Trans. Aut. Control*, AC-27, 323.
- Luh, P. B., T.-S. Chang and T. Ning (1984). Three-level hierarchical decision problems. *IEEE Trans. Aut. Control*, to be published.
- Papavasiliopoulos, G. P. and J. B. Cruz, Jr (1978). Nonzero-sum control problems and Stackelberg games. *IEEE Trans. Aut. Control*, AC-24, 123.
- Simaan, M. and J. B. Cruz, Jr (1975). Additional aspects of the Stackelberg strategy in nonzero-sum games. *J. Dynamic Theory & Appl.*, 13, 815.
- Tóthvári, B. (1981a). Closed-loop Stackelberg solution to multi-stage linear-quadratic game. *J. Dynamic Theory & Appl.*, 10, 481.
- Tóthvári, B. (1981b). Equilibrium relations for a class of hierarchical games. In J. Gauthier and M. N. Sargent (Eds.), *Applications of Systems Theory to Economics, Management and Technology*. PWN, Warsaw, pp. 385-400.
- Tóthvári, B. (1983). A Stackelberg solution of dynamic games. *IEEE Trans. Aut. Control*, AC-28, 45.