Delft University of Technology

Delft Center for Systems and Control

Technical report 10-062

Stackelberg equilibria for discrete-time dynamic games – Part II: Stochastic games with deterministic information structure*

K. Staňková and B. De Schutter

If you want to cite this report, please use the following reference instead:
K. Staňková and B. De Schutter, "Stackelberg equilibria for discrete-time dynamic games – Part II: Stochastic games with deterministic information structure," *Proceedings of the 2011 IEEE International Conference on Networking, Sensing and Control*, Delft, The Netherlands, pp. 255–260, Apr. 2011.

Delft Center for Systems and Control
Delft University of Technology
Mekelweg 2, 2628 CD Delft
The Netherlands
phone: +31-15-278.51.19 (secretary)
fax: +31-15-278.66.79
URL: http://www.dcsc.tudelft.nl

Stackelberg Equilibria for Discrete-Time Dynamic Games Part II: Stochastic Games with Deterministic Information Structure

Kateřina Staňková, Bart De Schutter

***Abstract*— We consider a two-person discrete-time dynamic game with a prespecified fixed duration. Each player maximizes her profit over the game horizon, taking decisions of the other player into account. Our goal is to find the Stackelberg equilibria for such a game. After having discussed deterministic Stackelberg games in the companion paper (Stackelberg Equilibria for Discrete-Time Dynamic Games – Part I: Deterministic Games), in this paper we focus on stochastic games with a deterministic information structure. While for the stochastic game with open-loop structure the solution procedure is straightforward and already reported in the literature, the problem with the closed-loop problem information structure for stochastic games remains a challenge. After discussing a rather standard approach to solve the open-loop stochastic game, we propose an approach to find (sub)optimal solutions of the closed-loop game. Moreover, we discuss solution approach for generalized games in which the leader has access to the follower's past actions, the so-called inverse Stackelberg games.**

Keywords: stochastic games, discrete-time infinite dynamic games, Stackelberg games, information structure, team problems

I. Introduction & literature overview

This paper extends results of the companion paper (Stackelberg Equilibria for Discrete-Time Dynamic Games – Part I: Deterministic Games) into the realm of stochastic games with a deterministic information structure. In such a game there is a noise in the state equation, but the players do not have a biased perception of the states of the game. The game is referred to as infinite, because the decision spaces of the players comprise an infinite number of alternatives. We focus on the noncooperative variant of this game [1], [2], in which the goals of individual players might be conflicting. More specifically, we deal with Stackelberg problems [1]–[6] as opposed to Nash problems [7], [8].

The open-loop Stackelberg solution concept in the infinite discrete-time dynamic games was first treated in [9]. Some other references that discuss the open-loop and the feedback Stackelberg solutions in discrete-time infinite dynamic games are [3], [10], [11]. Applications of this concept in microeconomics can be found in [12]. In this paper we show a standard method how to solve open-loop stochastic games.

This paper extends the theory of Stackelberg equilibria for stochastic games presented in [2], [11] under a closed-loop information pattern. Not much attention has been given to such problems. The difficulties encountered when dealing with the closed-loop stochastic variant of the game were first pointed out in [13] and are stated in [2], without proposing a solution strategy. In [14] the feedback strategy solutions in linear-quadratic stochastic dynamic games with noisy observations are discussed. However, stochastic games with a closed-loop information structure remain a challenge. In this paper we suggest a methodology to find a suboptimal solution of such games. Moreover, we propose a solution procedure for a generalized variant of the game in which the leader has access to the follower's past actions.

This paper is composed as follows. In Section II basic notions used in this paper are introduced. In Section III the open-loop variant of the stochastic game is dealt with. In Section IV we study the closed-loop variant of the game. In Section V the generalized variant of the stochastic game, also referred to as the inverse (or reverse) Stackelberg game, is considered. In Section VI the conclusions and possibilities for future research are discussed.

K. Staňková and B. De Schutter are with the Delft Center for Systems & Control, Delft University of Technology, The Netherlands. `k.stankova@tudelft.nl`, `b.deschutter@tudelft.nl`

II. Preliminaries

Let us consider a two-player Stackelberg game, in which P_1 is the leader and P_2 is the follower. Let $\mathscr{K} = \{1,\dots,K\}$ denote the stages of the game, where K is the maximal number of moves the player P_i ($i \in \{1,2\}$) is allowed to make. Let $X^{(1)} \subset \mathbb{R}$ and $X^{(2)} \subset \mathbb{R}$ be nonempty closed intervals and let $X \stackrel{\text{def}}{=} X^{(1)} \times X^{(2)}$ be the state space of the game. The state of the game for the k-th stage is then referred to as $x_k = \big(x_k^{(1)}, x_k^{(2)}\big)^T$, where $x_k^{(1)}$ is P_1's state and $x_k^{(2)}$ is P_2's state. Let $U_k^{(1)} \subset \mathbb{R}$ (for each $k \in \mathscr{K}$) be a closed interval called P_1's decision space. Its elements are P_1's permissible decisions $u_k^{(1)}$ at stage k, announced by P_1 at the beginning of the stage. Let $U_k^{(2)} \subset \mathbb{R}$ (for each $k \in \mathscr{K}$) be P_2's decision space. Its elements are P_2's permissible decisions $u_k^{(2)}$ at stage k.

In this paper we will extend the game to the stochastic infinite discrete-time dynamic game by adding an additional player, the so-called "nature", into the set of players. The nature acts randomly, but according to an a priori known probability law. The state of the game evolves according to the difference equation

$$x_{k+1} \sim F_k\big(x_k, u_k^{(1)}, u_k^{(2)}, \theta_k\big), \quad k \in \mathscr{K}, \tag{1}$$

with $F_k : X \times U_k^{(1)} \times U_k^{(2)} \times \Theta \to X$. Here $\theta_k \in \Theta$ is the decision variable of the nature at stage k. The initial state x_1 is then also a random variable and the joint probability distribution function of $(x_1, \theta_1, \dots, \theta_K)$ is assumed to be known. For the sake of simplicity, in this paper we will consider a situation with

$$F_k\big(x_k, u_k^{(1)}, u_k^{(2)}\big) \stackrel{\text{def}}{=} f_k\big(x_k, u_k^{(1)}, u_k^{(2)}\big) + \theta_k.$$

Then (1) becomes

$$x_{k+1} \sim f_k\big(x_k, u_k^{(1)}, u_k^{(2)}\big) + \theta_k, \quad k \in \mathscr{K}, \tag{2}$$

with state function

$$f_k : X \times U_k^{(1)} \times U_k^{(2)} \to X,$$

where $\theta_1, \ldots, \theta_K$ is a sequence of statistically i.i.d. Gaussian vectors with values in $\mathbb{R}^2$ and with $\mathrm{cov}(\theta_k) > 0$, $k \in \mathscr{K}$.

Let the information gained and recalled by P_i ($i \in \{1,2\}$) at stage k of the game be determined by $\eta_k^{(i)}$, an a priori known selection from $\big(x_1^{(1)}, \ldots, x_k^{(1)};\ x_1^{(2)}, \ldots,\ x_k^{(2)}\big)$. Specifications of $\eta_k^{(i)}$ for all $k \in \mathscr{K}$ characterize the information structure of the game for P_i. Let $N_k^{(i)} \stackrel{\text{def}}{=} \big\{\big(\eta_k^{(i)}\big)^T\big\}$. Let $\Gamma_k^{(i)}$ be a prespecified class of measurable mappings $\gamma_k^{(i)} : N_k^{(i)} \to U_k^{(i)}$, called P_i's permissible strategies at stage k. The aggregate mapping $\gamma^{(i)} = \big(\gamma_1^{(i)}, \gamma_2^{(i)}, \ldots, \gamma_K^{(i)}\big)$ is P_i's strategy, and the class $\Gamma^{(i)}$ of all such mappings $\gamma^{(i)}$ so that $\gamma_k^{(i)} \in \Gamma_k^{(i)}$, $k \in \mathscr{K}$, is P_i's strategy set. We will refer to a P_2's strategy based on the P_1's strategy as the P_2's response to the P_1's strategy.

Definition 2.1: (Information structure) *In a 2-person discrete time dynamic game, we say that P_i's ($i \in \{1,2\}$) information has (for all $k \in \mathscr{K}$)*

(a) *an open-loop structure if $\eta_k^{(i)} = x_1$,*
(b) *a closed-loop structure if $\eta_k^{(i)} = \big(x_1, \ldots, x_k\big)$,*
(c) *a feedback structure if $\eta_k^{(i)} = x_k$.*

In this paper we will consider cases (a) and (b), case (c) is a subject of the future research.

Let $L^{(i)} : \big(X \times U_1^{(1)} \times U_1^{(2)}\big) \times \big(X \times U_2^{(1)} \times U_2^{(2)}\big) \times \ldots \times \big(X \times U_k^{(1)} \times U_k^{(2)}\big) \to \mathbb{R}$ be called P_i's profit function. Each player maximizes $L^{(i)}$, taking into account possible actions of the other player.

A. Assumptions on state and profit functions

In order to simplify the analysis, we will mostly assume (unless stated differently), the following: The functions $f_k\big(\cdot, u_k^{(1)}, u_k^{(2)}\big)$, $f_k\big(x_k, \cdot, u_k^{(2)}\big)$, and $f_k\big(\cdot, u_k^{(1)}, \cdot\big)$ are continuously differentiable on $\mathbb{R}_+^2$, $U_k^{(1)}$, and $U_k^{(2)}$, respectively, The profit functions are assumed to be stage-additive, i.e., there exists $g_k^{(i)} : X \times X \times U_k^{(1)} \times U_k^{(2)} \to \mathbb{R}$, for all $k \in \mathscr{K}$, so that[1]

$$L^{(i)}\big(u^{(1)}, u^{(2)}\big) = \sum_{k=1}^{K} g_k^{(i)}\big(x_k, u_k^{(1)}, u_k^{(2)}, x_{k+1}\big), \tag{3}$$

$i \in \{1,2\}$, and continuously differentiable on $U^{(1)} \times U^{(2)}$. Furthermore, $g_k^{(i)}\big(\cdot, u_k^{(1)}, u_k^{(2)}, \cdot\big)$ and $g_k^{(i)}\big(x_k, u_k^{(1)}, \cdot, x_{k+1}\big)$ are continuously differentiable on $\mathbb{R}^2$ and $U_k^{(2)}$, respectively.

B. Stochastic game formulation

An extensive form description of the game contains the set of players, the index set defining the stages of the game, the state space and the decision spaces, the state equation, the observation sets, the state-observation equation, the information structure of the game, the information spaces, the strategy sets, and the profit functionals.

Similarly as it is done in [2] for a general N-person discrete-time game, we can transform the game into a normal-form game. For each fixed initial state x_1 and for each pair $\big(\gamma^{(1)}, \gamma^{(2)}\big)$, where $\gamma^{(i)} \in \Gamma^{(i)}$, $i \in \{1,2\}$, the extensive form description leads to a unique vector

$$u_k^{(i)} \stackrel{\text{def}}{=} \gamma_k^{(i)}\big(\eta_k^{(i)}\big), \quad i \in \{1,2\}, \quad k \in \mathscr{K}, \tag{4}$$

because of the causal nature of the information structure and because the state evolves according to a stochastic difference equation (2). Then, substitution of (4) into $L^{(1)}$ and $L^{(2)}$ leads to a pair of functions reflecting the corresponding profits of the players. This further implies existence of a composite mapping

$$J^{(i)} : \Gamma^{(1)} \times \Gamma^{(2)} \to \mathbb{R}$$

for each $i \in \{1,2\}$, which is the strategy-dependent profit function. Then the permissible strategy spaces $\big(\Gamma^{(1)}, \Gamma^{(2)}\big)$ together with

$$\big(J^{(1)}(\gamma^{(1)}, \gamma^{(2)}), J^{(2)}(\gamma^{(1)}, \gamma^{(2)})\big)$$

constitute the normal form description of the game for each fixed initial state x_1. If the game can be expressed in the normal form, techniques used for finding solutions of finite games can be adopted [2].

III. Open-loop game

This section summarizes results presented in [2], [14]. Consider that the system evolves according to

$$x_{k+1} \sim f_k\big(x_k, u_k^{(1)}, u_k^{(2)}\big) + \theta_k, \quad k \in \mathscr{K}, \tag{5}$$

where $\theta_1, \ldots, \theta_K$ is a sequence of statistically i.i.d. Gaussian vectors with values in $\mathbb{R}^2$ and with $\mathrm{cov}(\theta_k) > 0$, $k \in \mathscr{K}$. The profit functionals for P_1 and P_2 are given by (3).

A. Open-loop information pattern for both players

If the information has an open-loop structure for both players and is known a priori, the solution can be found by converting the game into equivalent static normal form and by consequent utilizing methods used for finding the deterministic open-loop Stackelberg solution. Indeed, we can recursively substitute (5) into (3) and take expected values of $L^{(i)}$ over the random variables $\theta_1, \ldots, \theta_K$ to obtain functionals $L^{(i)}$ depending only on the players' decisions $u^{(1)}$ and $u^{(2)}$ and on x_1. The game can be then treated as a static game. The solution depends on the statistical moments of the random disturbances in the state equation – unless the system equation is linear and the profit functions are quadratic [11]. In such a case the solution may be independent of the disturbances.

[1] For the sake of notation convenience we refer to the profit functions as $L^{(i)}\big(u^{(1)}, u^{(2)}\big)$, with $u^{(i)} \stackrel{\text{def}}{=} \big(u_1^{(i)}, \ldots, u_K^{(i)}\big)^T$.

B. Open-loop information pattern for the leader and closed-loop information pattern for the follower

For the deterministic Stackelberg games dealt with in the companion paper the optimal open-loop Stackelberg equilibrium strategy of P_1 does not change if P_2 has an additional state information. The optimal strategy of P_2 then becomes any closed-loop representation of the open-loop response on the equilibrium trajectory associated with the open-loop Stackelberg solution; hence, the optimal strategy of P_2 is nonunique.

However, for the stochastic Stackelberg game with an open-loop information structure for both P_1 and P_2 the solution does not coincide with the solution of the game in which P_1 has an open-loop information structure and P_2 has a closed-loop information structure and the latter has to be obtained independently of the former. The steps to solve the game with the open-loop information structure for P_1 and closed-loop information structure for P_2 are (see [2] for its derivation):

1) For any $\big(u_k^{(1)} \in U_k^{(1)} | k \in \mathscr{K}\big)$ maximize

$$\mathbb{E}\Big(\sum_{k=1}^{K} g_k^{(2)}\big(x_k, u_k^{(1)}, u_k^{(2)}, x_{k+1}\big) \,|\, u_k^{(2)} = \gamma_k^{(2)}\big(x_1,\ldots,x_k\big)\Big)$$

subject to (2) and over $\Gamma^{(2)}$. We will denote the maximizing strategy of P_2 by $\gamma^{(2)\mathrm{o}}$. Following [15], [16], we can see that with $\eta_k^{(2)} \overset{\mathrm{def}}{=} (x_l, l \leq k)$ any $\gamma^{(2)\mathrm{o}}$ satisfies the dynamic programming principle

$$\begin{aligned}
V(k,x) = \max_{u_k^{(2)} \in U_k^{(2)}} &\mathbb{E}\big[g_k^{(2)}\big(x_k, u_k^{(1)}, u_k^{(2)}, f_k\big(x_k, u_k^{(1)}, u_k^{(2)}\big) \\
&+ \theta_k\big) + V\big(k+1, f_k\big(x_k, u_k^{(1)}, u_k^{(2)}\big) + \theta_k\big)\big] \\
= \mathbb{E}\big[&g_k^{(2)}\big(x_k, u_k^{(1)}, \gamma_k^{(2)\mathrm{o}}\big(\eta_k^{(2)}\big), \\
&f_k\big(x_k, u_k^{(1)}, \gamma_k^{(2)\mathrm{o}}\big(\eta_k^{(2)}\big)\big) + \theta_k\big) \\
&+ V\big(k+1, f_k\big(x_k, u_k^{(1)}, \gamma_k^{(2)\mathrm{o}}\big(\eta_k^{(2)}\big)\big) + \theta_k\big)\big], \qquad (6)
\end{aligned}$$

with

$$V(k,x) \overset{\mathrm{def}}{=} \max_{u_k^{(2)} \in U_k^{(2)}} \mathbb{E}\Big[\sum_{i=k}^{K} g_i^{(2)}\big(x_i, u_k^{(1)}, u_k^{(2)}, x_{i+1}\big)\Big].$$

2) Maximize

$$\begin{aligned}
J^{(1)}\big(\gamma^{(1)}, \gamma^{(2)\mathrm{o}}\big) = \mathbb{E}\big[&L^{(1)}\big(u^{(1)}, u^{(2)}\big) | u_k^{(1)} = \gamma_k^{(1)}(x_1), \\
&u_k^{(2)} = \gamma_k^{(2)\mathrm{o}}\big(\eta_k^{(2)}\big), k \in \mathscr{K}\big]
\end{aligned}$$

over $\Gamma^{(1)}$, subject to (6) and the state equation (2) with $u_k^{(2)}$ replaced by $\gamma^{(1)}$-dependent $\gamma_k^{(2)\mathrm{o}}\big(\eta_k^{(2)}\big)$.[2] The solution of this optimization problem constitutes the Stackelberg strategy of the leader in the stochastic dynamic game under consideration, provided that the maximization of V provides a unique solution $\big(\gamma_1^{(2)\mathrm{o}}, \ldots, \gamma_K^{(2)\mathrm{o}}\big)$ for each $\big(u_1^{(1)}, \ldots, u_K^{(1)}\big)$.

[2]Note that $\gamma_k^{(2)\mathrm{o}}(\cdot)$ is dependent on $\gamma^{(1)}$, but this dependence cannot be in general written in the closed-form.

Steps 1) and 2) provide a straightforward procedure for finding Stackelberg equilibrium of the game defined by (2), (3).

IV. CLOSED-LOOP STOCHASTIC GAME

If P_1 has access to dynamic information, but does not have direct access to decisions of P_2, a direct approach to find the solution cannot be applied, as any optimal response of P_2 to a strategy announced by P_1 cannot be expressed analytically in terms of strategy of P_1, even in the case of linear-quadratic games. Moreover, the indirect method proposed in the companion paper solving the deterministic game cannot be used here, since for stochastic game each strategy has a unique representation as opposed to the infinitely many closed-loop representations of a given strategy in a deterministic system. Consequently, derivation of closed-loop solutions of stochastic dynamic games remains a challenge. If, however, we make some structural assumptions on possible strategies of P_1 - which means seeking suboptimal solutions instead of optimal solutions, then the problem may become tractable. In particular, if, under the structural assumptions on $\Gamma^{(1)}$, restricting $\Gamma^{(1)}$ to a certain function set, the class of the permissible strategies of P_1 can be described by a finite number of parameters, and if P_2's optimal response can be determined as a function of these parameters, then the original game may transform into the static one in which P_1 selects her strategy from an Euclidian space of the corresponding dimension; such a static game is, in general, solvable [17], although more likely numerically than analytically. The following example illustrates such an approach and shows that even with very simple stochastic dynamic games and very restrictive assumptions on $\Gamma^{(1)}$, the corresponding (sub-optimal) solution methodology is not trivial.

Example 4.1 (Two-stage stochastic game): Let

$$\begin{aligned}
&L^{(1)} = -{x_3}^2 - 2\big(u_2^{(1)}\big)^2 - \big(u_1^{(2)}\big)^2, \\
&L^{(2)} = -{x_3}^2 - \big(u_1^{(2)}\big)^2, \\
&x_2 = x_1 - u_1^{(2)} + \theta_1, \quad x_3 = x_2 - u_2^{(1)} + \theta_2, \qquad (7)
\end{aligned}$$

where θ_1 and θ_2 are i.i.d. random variables with zero mean and variances σ_1, σ_2 and x_1 is known a priori, and assume that the strategies of P_1 are linear in x_2, i.e., we restrict ourselves to the strategies

$$\Gamma^{(1)} = \{\gamma^{(1)} | \gamma^{(1)}\big(x_1, x_2\big) = \rho_2 x_1 + \rho_1 x_2\}. \qquad (8)$$

Player P_2 acts at the stage 2 and has access to x_1 and x_2, player P_2 acts at stage 1 and has only access to x_1. Player P_2 maximizes

$$J^{(2)} = \mathbb{E}\{-\big(x_2 - \gamma^{(1)}\big(x_2, x_1\big)\big)^2 - \big(\gamma^{(2)}\big(x_1\big)\big)^2\},$$

over $\gamma^{(2)} \in \Gamma^{(2)}$ in order to determine her optimal response to any $\gamma^{(1)} \in \Gamma^{(1)}$ chosen by P_1.

With $\Gamma^{(1)}$ defined by (8), the problem is equivalent to the problem of finding optimal ρ_1 and ρ_2, which are x_1-

dependent. Under such restriction $J^{(2)}$ has a unique maximum, leading to the optimal strategy for P_2

$$\gamma^{(2),\mathrm{o}} = \frac{(1-\rho_1-\rho_2)(1-\rho_1)x_1}{\rho_1^2 - 2\rho_1 + 2}.$$

By substituting $\gamma^{(1)}$ and $\gamma^{(2),\mathrm{o}}$ into $J^{(1)}\big(\gamma^{(1)},\gamma^{(2)}\big)$, together with the corresponding values of x_3 and x_2, we obtain

$$\begin{aligned} F(\rho_1,\rho_2) = &-\frac{\big(1-\rho_1-\rho_2\big)^2 x_1^2}{1+\big(1-\rho_1\big)^2} - \big((1-\rho_1)^2+\rho_1^2\big)\sigma_1 \\ &- \frac{2\big(\rho_1+2\rho_2-\rho_1\rho_2\big)^2 x_1^2}{\big(1+\big(1-\rho_1\big)^2\big)^2} - \sigma_2, \end{aligned}$$

which has to be maximized over ρ_1 and ρ_2 for fixed x_1. Note that:

- F is continuous in both ρ_1 and ρ_2, $F(\rho_1,\rho_2) \le 0$ for all $\rho_1,\rho_2 \in \mathbb{R}$, and $F(\rho_1,\rho_2) \to -\infty$ if $|\rho_1|$, $|\rho_2| \to \infty$. Consequently (by the Weierstrass theorem [18]), there exists at least one pair (ρ_1^*,ρ_2^*) maximizing F for any given (x_1,σ_1).
- The optimal (ρ_1^*,ρ_2^*) depends on $\big(x_1,\sigma_1\big)$, but not on σ_2. Hence, for each fixed $\big(x_1,\sigma_1\big)$ the function $F(\rho_1,\rho_2)$ can be maximized using classical optimization methods [19].
- $\gamma^{(1)}\big(x_1,x_2\big) = \rho_2^* x_1 + \rho_1^* x_2$ is a suboptimal strategy, i.e., it does not lead to the team maximum of the game for P_1.

V. Generalized stochastic Stackelberg game

Let us now assume that P_1 has information about the past states of the game and the past decisions of P_2. With such extended information structure a given strategy of P_1 will have multiple representations, thus it might be possible to enforce her team maximum

$$J^{(1)}\big(\gamma^{(1)\S},\gamma^{(2)\S}\big) = \max_{\gamma^{(1)}\in\Gamma^{(1)}} \max_{\gamma^{(2)}\in\Gamma^{(2)}} J^{(1)}\big(\gamma^{(1)},\gamma^{(2)}\big).$$

The goal of P_1 is to find an optimal representation of the team-optimal strategy, i.e., the strategy which enforces the team maximum. Such a strategy may be dependent on decisions made by P_2.

Example 5.1: Let

$$\begin{aligned} &L^{(1)} = -{x_3}^2 - 2\big(u_2^{(1)}\big)^2 - \big(u_1^{(2)}\big)^2, \\ &L^{(2)} = -{x_3}^2 - \big(u_1^{(2)}\big)^2, \\ &x_2 = x_1 - u_1^{(2)} + \theta_1, \quad x_3 = x_2 - u_2^{(1)} + \theta_2. \end{aligned} \tag{9}$$

Let P_1 has access to $u^{(2)}$, in addition to x_1 and x_2. By this way, the game is generalized into a so-called inverse Stackelberg game [20]–[22].

The best outcome that P_1 can achieve is the team maximum

$$J^{(1)}\big(\gamma^{(1)\S},\gamma^{(2)\S}\big) = \max_{\gamma^{(1)}\in\Gamma^{(1)}} \max_{\gamma^{(2)}\in\Gamma^{(2)}} J^{(1)}\big(\gamma^{(1)},\gamma^{(2)}\big),$$

with

$$\begin{aligned} J^{(1)}\big(\gamma^{(1)},\gamma^{(2)}\big) = \mathbb{E}\big\{ &-\big(x_2-\gamma^{(1)}(x_2,x_1)-\theta_2\big)^2 \\ &+ 2\big(\gamma^{(1)}\big(x_2,x_1\big)\big)^2 - \big(\gamma^{(2)}\big(x_1\big)\big)^2\big\}. \end{aligned}$$

Note that the profit shows dependence on $u_1^{(2)} = \gamma^{(2)}\big(x_1\big)$ not only directly, but also through x_2. This team problem is in fact a linear-quadratic stochastic control problem [14] and the team-optimum strategies are

$$\gamma^{(1)\S} = \frac{x_2}{3}, \quad \gamma^{(2)\S} = \frac{5x_1}{14}, \tag{10}$$

which is the unique maximizing pair on $\Gamma^{(1)}\times\Gamma^{(2)}$. It is not, however, unique in the extended strategy space of P_1, as (for example) the following parametrized strategy also constitutes an optimal solution, with $\kappa\in\mathbb{R}$:

$$\begin{aligned} \gamma_\kappa^{(1)}\big(x_2,x_1,u_1^{(2)}\big) &= \frac{x_2}{3} + \kappa\big(u_1^{(2)} - \frac{5}{14}x_1\big), \\ \gamma_\kappa^{(2)} &= \frac{5x_1}{14}, \end{aligned} \tag{11}$$

which characterizes the class of optimal strategies linear in x_2, x_1, and $u^{(2)}$, all leading to the same maximum expected value for P_1. We will refer to these strategies as "representations of $\gamma^{(1)\S}$" under the team-optimum solution $\big(\gamma^{(1)\S},\gamma^{(2)\S}\big)$. Among this family of strategy pairs we are looking for the one with the additional property: If P_1 maximizes her expected profit function, then the strategy in $\Gamma^{(2)}$ that leads to this maximum is still $\gamma^{(2)\S}$, resp. $\gamma_\kappa^{(2)}$. The corresponding strategy for P_1 would then correspond to the global inverse Stackelberg solution, yielding the (team) maximum profit for P_1.

Let us now focus on linear representations (11), which leads to the quadratic maximization problem:

$$\mathbb{E}\big(-\big(x_2-\gamma_\alpha^{(1)}(x_2,x_1,u_1^{(2)})-\theta_2\big)^2 + \big(u_1^{(2)}\big)^2\big),$$

with $x_2 = x_1 - u_1^{(2)} + \theta_1$. Since x_1 is independent of θ_1 and θ_2, which have both zero mean, this problem is equivalent to the following deterministic optimization problem:

$$\max_{u_1^{(2)}\in\mathbb{R}} \Big(-\big(\frac{2}{3}\big(u_1^{(2)}-x_1\big) - \alpha\big(u_1^{(2)} - \frac{5}{14}x_1\big)\big)^2 - \big(u_1^{(2)}\big)^2\Big)$$

The solution to this problem is the pair

$$\begin{aligned} \gamma^{(1)\dagger}\big(x_2,x_1,u^{(2)}\big) &= \frac{x_2}{3} + \frac{8}{27}\big(u^{(2)} - \frac{5}{14}x_1\big), \\ \gamma^{(2)\dagger}\big(x_1\big) &= \frac{5}{14}x_1, \end{aligned}$$

which constitutes a global inverse Stackelberg solution. This is, in fact, the unique solution within the class of linear strategies.

Remark 5.1: In Example 5.1 P_1 has access to all information that P_2 has access to. However, in stochastic inverse Stackelberg problems the information may not always be nested for the leader. If in this example P_1 had access to x_2 and $u_1^{(2)}$, then the problem would be of a nonnested information structure. To such problems the methodology shown in the example does not apply, since the dynamic information for P_1 no longer exhibits redundancy. While such problems have not been so far studied in detail, they were discussed in [23], [24] and are also a subject of our future research. For stochastic inverse Stackelberg problems with nested information the approach used in Example 5.1 can be generalized as follows.

Consider a two-person stochastic inverse Stackelberg problem with the profit functions $L^{(1)}\left(u^{(1)},u^{(2)},\theta\right)$ and $L^{(2)}\left(u^{(1)},u^{(2)},\theta\right)$ for P_1 and P_2, respectively, where θ is some random vector with a known probability distribution function. Let $y^{(1)}=h^{(1)}(\theta)$ be the estimate of P_1 on Θ and $y^{(2)}=h^{(2)}(\theta)$ be the estimate of P_2 on Θ, with the property, that what P_2 knows is also known by P_1, i.e., the sigma-field generated by $y^{(1)}$ includes the sigma-field generated by $y^{(2)}$. Let $\Gamma^{(i)}$ be the set of all measurable strategies of the form $u^{(i)}=\gamma^{(i)}\left(y^{(i)}\right)$, $i=1,2$, and $\hat{\Gamma}^{(1)}$ be the set of all measurable policies of the form $u^{(1)}=\gamma^{(1)}\left(y^{(1)},u^{(2)}\right)$. Let us introduce the pair of policies

$$\left(\gamma^{(1),*},\gamma^{(2),*}\right)\stackrel{\text{def}}{=}\arg\max_{\gamma^{(1)}\in\Gamma^{(1)},\gamma^{(2)}\in\Gamma^{(2)}}\mathbb{E}_{\theta}\left\{L^{(1)}\left(\gamma^{(1)}\left(h^{(1)}(\theta)\right),\right.\right.$$
$$\left.\left.\gamma^{(2)}\left(h^{(2)}(\theta)\right),\theta\right)\right\},$$

assuming that the underlying team problem admits a maximizing solution. Then an optimal representation of P_1's strategy $\gamma^{(1),*}$ is $\hat{\gamma}^{(1)}\in\hat{\Gamma}^{(1)}$ satisfying

$$\hat{\gamma}^{(1)}\left(h^{(1)}(\theta),\gamma^{(2),*}\left(h^{(2)}(\theta)\right)\right)=\gamma^{(1),*}\left(h^{(1)}(\theta)\right)\quad\text{a.s.}\quad(12)$$

The following result now follows.

Proposition 5.1: For the stochastic inverse Stackelberg problem with nested information as formulated above, the pair $\left(\hat{\gamma}^{(1)},\gamma^{(2),*}\right)$ constitutes the team maximum outcome for P_1. Equivalently, if a strategy $\hat{\gamma}^{(1)}\in\hat{\Gamma}^{(1)}$ satisfies (12), the optimum of the underlying stochastic decision problem for the leader exists.

VI. Conclusions & Future research

In this paper we have introduced a stochastic variant of discrete-time infinite dynamic games and have dealt with finding their Stackelberg equilibrium solutions. Such solutions depend on the information patterns of the games and vary with the characteristics of the individual problems.

We have reviewed classical approaches used to solve the games with open-loop information structure and proposed an approach to find a suboptimum of the games with closed-loop information structure. Moreover, we have shown how to find a solution to a generalized variant of the game with the closed-loop information structure, in which the leader has access to the follower's past actions, the so-called inverse Stackelberg game.

The main subjects of future research are finding solutions of the generalized stochastic game with a nonnested information structure and study of the stochastic games with feedback information structure.

Acknowledgments

This research has been supported by the European 7th framework STREP project "Hierarchical and distributed model predictive control of large-scale systems (HD-MPC)", contract number INFSO-ICT-223854, and by the European 7th Framework Network of Excellence "Highly-complex and networked control systems (HYCON2)". The authors would like to thank dr. Ján Buša from the Department of Mathematics, Faculty of Electrical Engineering & Computer Science, The Technical University of Košice, for his helpful insights.

References

[1] A. Perea y Monsuwé, *Information Structures in Non-Cooperative Games*. Maastricht, The Netherlands: Unigraphic, 1997.
[2] T. Başar and G. J. Olsder, *Dynamic Noncooperative Game Theory*. Philadelphia, Pennsylvania: SIAM, 1999.
[3] F. Kydland, "Noncooperative and dominant player solutions in discrete dynamic games," *International Journal of Economic Review*, vol. 16, pp. 321–335, 1975.
[4] A. Bagchi, *Stackelberg Differential Games in Economical Models*. Berlin, Germany: Springer-Verlag, 1984.
[5] H. Peters, *Game Theory: A Multi-Leveled Approach*. Dordrecht, The Netherlands: Springer, 2008.
[6] K. Staňková, "On Stackelberg and Inverse Stackelberg Games & Their Applications in the Optimal Toll Design Problem, the Energy Market Liberalization Problem, and in the Theory of Incentives," Ph.D. dissertation, Delft University of Technology, Delft, The Netherlands, 2009.
[7] J. Nash, "Noncooperative games," *Annals of Mathematics*, vol. 54, pp. 286–295, 1951.
[8] G. Olsder, "Adaptive Nash strategies for repeated games resulting in Pareto solutions," Delft University of Technology, Department of Mathematics and Informatics, Reports of the Department of Mathematics and Informatics 86-09, 1986.
[9] M. Simaan and J. B. Cruz, "Additional aspects of the Stackelberg strategy in the nonzero sum games," *Journal of Optimization Theory and Applications*, vol. 11, pp. 613–626, 1973.
[10] J. B. Cruz, "Leader-follower strategies for multilevel systems," *IEEE Transactions on Automatic Control*, vol. AC-23, pp. 244–255, 1978.
[11] T. Başar, "Information structures and equilibria in dynamic games," in *New Trends in Dynamic System Theory and Economics*, M. Aoki and A. Marzollo, Eds. New York and London: Academic Press, 1979, pp. 3–55.
[12] K. Okuguchi, *Expactations and Stability in Oligopoly Models*, ser. Lecture Notes in Economics and Mathematical Systems. Berlin: Springer-Verlag, 1976.
[13] D. Castanon, "Equilibria in stochastic dynamic games of stackelberg type," Ph.D. dissertation, M.I.T. Electronics Systems Laboratory, Cambridge, MA, 1976.
[14] T. Başar, "Stochastic stagewise stackelberg strategies for linear-quadratic systems," in *Stochastic Control Theory and Stochastic Differential Systems*, ser. Lecture Notes in Control and Information Sciences, M. Kohlmann and W. Vogel, Eds. New York: Springer-Verlag, 1979, ch. 16, pp. 264–276.
[15] R. Bellman, *Dynamic Programming*. Princeton, New Jersey: Princeton University Press, 1957.
[16] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons, Inc., 1994.
[17] M. Bardi, T. E. S. Raghavan, and T. Parthasarathy, Eds., *Stochastic and Differential Games: Theory and Numerical Methods*, ser. Annals of International Society of Dynamic Games. Boston: Birkhäuser, 1999, vol. 4.
[18] W. Rudin, *Functional analysis*. New York: McGraw-Hill, 1973.
[19] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*. London, UK: Academic Press, 2004.
[20] G. J. Olsder, "Phenomena in inverse Stackelberg problems," Mathematisches Forschungsinstitut Oberwolfach, Oberwolfach, Germany, Tech. Rep. 11/2005, 2005.

[21] K. Staňková and G. J. Olsder, "Inverse Stackelberg games versus adverse-selection principal-agent model theory," in *Proceedings of the 12th International Symposium on Dynamic Games and Applications*, Sophia-Antipolis, France, 2006.

[22] K. Staňková, M. C. J. Bliemer, and G. J. Olsder, "Inverse Stackelberg games and their application to dynamic bilevel optimal toll design problem," in *Proc. 12th International Symposium on Dynamic Games and Applications*, 2006.

[23] T. Başar, "Affine incentive schemes for stochastic systems with dynamic information," *SIAM Journal on Scientific and Statistical Computing*, vol. 22, pp. 199–210, 1984.

[24] ——, "Stochastic incentive problems with partial dynamic information and multiple levels of hierarchy," *European Journal of Political Economy*, pp. 203–217, 1989.