Contributions to Game Theory and Management, XII, 273–281

Optimal Incentive Strategy in a Discounted Stochastic Stackelberg Game*

Dmitry B. Rokhlin and Gennady A. Ougolnitsky

I. I. Vorovich Institute of Mathematics,
Mechanics and Computer Sciences of Southern Federal University,
8a, Milchakova, Rostov-on-Don, Russia
E-mail: dbrohlin@sfedu.ru
E-mail: gaugolnickiy@sfedu.ru

Abstract We consider a game where manager's (leader's) aim is to maximize the gain of a large corporation by the distribution of funds between m producers (followers). The manager selects a tuple of m non-negative incentive functions, and the producers play a discounted stochastic game, which results in a Nash equilibrium. Manager's aim is to maximize her related payoff over the class of admissible incentive functions. It is shown that this problem is reduced to a Markov decision process.

1. Introduction

The problem of incentives plays a key role in economics and management. Its mathematical formalization is proposed by the theory of incentives (Laffont and Martimort, 2002), mechanism design (Myerson, 1983), the theory of control in organizational systems (Novikov, 2013). However, the majority of the respective problem formulations are static.

A natural dynamic incentive model is provided by the dynamical inverse Stackelberg games, where the leader strategy depends on the followers' actions (an incentive mechanism). A general review can be found in (Olsder, 2009). In the paper (Rokhlin and Ougolnitsky, 2018) (inspired by Novikov and Shokhina, 2003) we formalized the incentive problem as a stochastic inverse Stackelberg game and obtained a simple description of leader's optimal strategy in the case of a single follower. In the present paper we extend this result for the case of multiple followers.

Consider a game where manager's (leader's) aim is to maximize the gain of a large corporation by the distribution of funds between m producers (followers). To each follower the leader reports a non-negative stimulating (incentive) function $c_i(x, a)$, depending on the state of the system x (e.g., the market price of the produced good) and the actions $a = (a^i, \dots, a^m)$ of the producers (e.g., the production levels). At each stage of the game the producers select their actions a_t^i independently and get the rewards $r_i(x_t, a_t) = c_i(x_t, a_t) - g_i(x_t, a_t)$, where g_i are the production costs. The manager, or the corporation, one-stage gain equals to $f(x_t, a_t) - \sum_{i=1}^m c_i(x_t, a_t)$, where f can be regarded as the sales revenue. The stochastic "law of motion" of the state variable x_t is governed by a transition kernel q: informally, $\mathsf{P}(x_{t+1} \in B | x_t, a_t) = q(B | x_t, a_t)$.

Each player's gain is estimated over the infinite horizon with the common discount factor β. So,

$$\mathsf{E} \sum_{t=0}^{\infty} \beta^t \left(f(x_t, a_t) - \sum_{i=1}^m c_i(x_t, a_t) \right) \to \max$$

* The research is supported by the Russian Science Foundation, project 17-19-01038.

is the objective functional of the leader, and

$$\mathsf{E}\sum_{t=0}^{\infty}\beta^t\left(c_i(x_t,a_t)-g_i(x_t,a_t)\right)\to\max$$

are the objective functionals of the followers. For each tuple $(c_1,\dots,c_m)$ the pool of producers responds by a Nash equilibrium in the corresponding discounted stochastic game. The leader performs the optimization over the functions c_i from an appropriate class. From the previous work (Novikov and Shokhina, 2003; Rokhlin and Ougolnitsky, 2018) it is known that it is optimal for the leader to economically motivate the followers to implement the strategies $a_t^i=\overline{u}_i(x_t)$, where $\overline{u}=(\overline{u}_1,\dots,\overline{u}_n)$ is an optimal stationary deterministic Markov strategy in the Markov decision problem

$$\mathsf{E}\sum_{t=0}^{\infty}\beta^t\left(f(x_t,a_t)-\sum_{i=1}^{m}g_i(x_t,a_t)\right)\to\max.$$

Passing to the case of multiple followers, draw some technical difficulties related to the existence of a stationary Markov equilibrium. To overcome these technical issues we modify the class of incentive functions, considered in (Rokhlin and Ougolnitsky, 2018), to make them continuous in actions. Furthermore, we confine ourselves to the games with a coarser transition kernel (He and Sun, 2017). Other related assumptions on the transition kernel q, providing the existence of a stationary Markov equilibrium (see Jaśkiewicz and Nowak, 2018), would be suffice.

In Section 2. we give a general formal description of a discounted stochastic game and a Markov decision process. In Section 3. we use this formalism to precisely describe an ε-optimal strategy of the leader and her value function in our model, formulated as a Stackelberg game: see Theorem 4. In two final remarks we compare this theorem with the results of (Rokhlin and Ougolnitsky, 2018), and mention that the technical coarser transition kernel condition can be dropped by passing to a correlated equilibrium.

2. Basics of discounted stochastic games

Let $(\Omega,\mathscr{F})$ be a measurable space, and let (Y,τ) be a topological space. A function $F:\Omega\times Y\to\mathbb{R}$ is called a *Caratheodory function* if the function $F(\cdot,y)$ is $\mathscr{F}$-measurable for each $y\in Y$ and the function $F(\omega,\cdot)$ is τ-continuous for each $\omega\in\Omega$ (Aliprantis and Border, 2006Definition 4.50). If (Y,τ) is a separable metrizable space, then such function F is jointly measurable Aliprantis and Border, 2006Lemma 4.51. Denote by $\mathscr{C}_b(\Omega\times Y)$ the set of uniformly bounded Caratheodory functions. Also, recall that a standard Borel space is a measurable space isomorphic to a Borel subset of a Polish space (separable, completely metrizable topological space) Srivastava, 1998.

Let $I=\{1,\dots,m\}$ be the set of players. The *discounted stochastic game* is determined by

- A standard Borel state space $(X,\mathscr{B}(X))$ with its Borel σ-algebra $\mathscr{B}(X)$.
- Separable metrizable spaces $(\overline{A}_i,\tau_i)$, $i\in I$ of players' actions.
- Compact-valued mappings $x\mapsto A_i(x)\subset\overline{A}_i$. A set $A_i(x)$ describes the admissible actions of i-th player in the state $x\in X$. It is assumed that the multivalued mappings $x\mapsto A_i(x)$ are measurable (Hu and Papageorgiou, 1997Chapter 2,

Definition 1.1), that is $\{x \in X : A_i(x) \cap U \neq \emptyset\} \in \mathscr{B}(X)$ for any open set $U \subset \overline{A}_i$.
- Reward functions $r_i \in \mathscr{C}_b(X \times \overline{A})$, where $\overline{A} = \overline{A}_1 \times \cdots \times A_m$ is endowed with the product topology τ.
- A transition probability $Q(\cdot|\cdot)$ from $X \times A$ to X (Bogachev, 2007Definition 10.7.1), that is
  - the function $(x, a) \mapsto Q(B|x, a)$ is $\mathscr{B}(X) \times \mathscr{B}(A)$-measurable for every $B \in \mathscr{B}(X)$,
  - the function $B \mapsto Q(B|x, a)$ is a probability measure on $\mathscr{B}(X)$ for every $(x, a) \in X \times A$.

  It is assumed that the function $a \mapsto \int w(y)Q(dy|x, a)$ is continuous for any $x \in X$ and any bounded Borel measurable function w on X.
- A discount factor $\beta \in [0, 1)$.

We assume that the players use stationary Markov strategies, which can be identified with the transitions probabilities σ_i from X to $\overline{A}_i$ such that $\sigma_i(x)(A_i(x)) = 1$. For $x \in X$ each tuple $\sigma = (\sigma_i)_{i \in I}$ induces the probability measure

$$\begin{aligned} \mathsf{P}_{x,\sigma}(dx_0 da_0 \dots dx_t da_t) = \delta_x(dx_0) \prod_{i \in I} \sigma_i(x_0)(da_0^i) \times \\ \times Q(dx_1|x_0, a_0) \dots Q(dx_t|x_{t-1}, a_{t-1}) \prod_{i \in I} \sigma_i(x_t)(da_t^i) \end{aligned} \tag{1}$$

on the space of sequences $(x_t, a_t)_{t \in \mathbb{Z}_+}$, $(x_t, a_t) \in X \times \overline{A}$ endowed with the product σ-algebra.

The expected discounted payoff of the player i equals to

$$J_i(x, \sigma) = \mathsf{E}_{x,\sigma} \sum_{t=0}^{\infty} \beta^t r_i(x_t, a_t). \tag{2}$$

A tuple $\sigma^* = (\sigma_i^*)_{i \in I}$ is called a Nash equilibrium if

$$J_i(x, \sigma^*) \geq J_i(x, \sigma_{-i}^*, \sigma_i), \quad i \in I$$

for any strategies σ_i and any $x \in X$. We use the standard notation $\sigma_{-i} = (\sigma_j)_{j \in (I \setminus i)}$.

Formally, a *Markov decision process* is a stochastic game involving a single player: $m = 1$. Omitting the index “1”, in this case we can write the objective function (2) as follows:

$$J(x, \sigma) = \mathsf{E}_{x,\sigma} \sum_{t=0}^{\infty} \beta^t r(x_t, a_t).$$

Denote by $v(x) = \sup_\sigma J(x, \sigma)$ the related value function.

Theorem 1. *For the described Markov decision process the following assertions hold true:*

(i) *v is the unique solution of the Bellman equation*

$$v(x) = \sup_{a \in A(x)} \left\{ r(x, a) + \beta \int_X v(y) Q(dy|x, a) \right\} \tag{3}$$

in the space of bounded Borel measurable functions on X.

(ii) *There exists an optimal strategy $\sigma^*(x)(dy) = \delta_{u^*(x)}(dy)$, which can be identified with a Borel measurable selector*

$$u^*(x) \in \arg\max_{a\in A(x)} \left\{ r(x,a) + \beta \int_X v(y) Q(dy|x,a) \right\}.$$

(iii) *If σ^* is an optimal strategy: $v(x) = J(x,\sigma^*)$, then*

$$v(x) = \int_{A(x)} \left(r(x,a) + \beta \int_X v(y) Q(dy|x,a) \right) \sigma^*(x)(da) \tag{4}$$

The proof of (i), (ii) can be found e.g. in (Himmelberg et al., 1976; Hu and Papageorgiou, 2000, Section VII.2)). The relation (4) is known as the dynamic programming principle or the "fundamental equation": (Dynkin and Yushkevich, 1979Chapter 6).

If $(\sigma_i^*)_{i\in I}$ is a Nash equilibrium in an m-player game, $m > 1$, then each σ_i^* is an optimal solution of the problem

$$J_i(x, \sigma^*_{-i}, \sigma_i) \to \max_{\sigma_i}. \tag{5}$$

A simple calculation shows that

$$J_i(x,\sigma_i,\sigma^*_{-i}) = \mathsf{E}_{s,\sigma_i,\sigma^*_{-i}} \sum_{t=0}^{\infty} \beta^t r_i(x_t, a_t^i; \sigma^*_{-i}),$$

where

$$r_i(x, a^i; \sigma^*_{-i}) = \int_{\overline{A}_{-i}} r_i(x, a^i, a^{-i})\, \sigma^*_{-i}(da^{-i}) \tag{6}$$

and the expectation $\mathsf{E}_{s,\sigma_i,\sigma^*_{-i}}$ is taken with respect to the measure generated by the transition probabilities

$$Q_{\sigma^*_{-i}}(B|x,a^i) = \int_{\overline{A}_{-i}} Q(B|x,a^i,a^{-i}) \sigma^*_{-i}(x)(da^{-i}) \tag{7}$$

and by the strategies σ_i on the space of sequences $(x_t, a_t^i)_{t\in\mathbb{Z}_+}$, $(x_t, a_t^i) \in X \times \overline{A}_i$ in the same way as (1).

It follows that (5) is a Markov decision process, satisfying the assumptions of Theorem 1. Let $V_{\sigma^*_{-i}}(x) = \sup_{\sigma_i} J(x,\sigma_i,\sigma^*_{-i})$ be the related value function. Since σ_i^* is an optimal solution: $V_{\sigma^*_{-i}}(x) = J(x,\sigma_i^*,\sigma^*_{-i})$, from the optimality principle (4) and the Bellman equation (3) we get

$$\begin{aligned} V_{\sigma^*_{-i}}(x) &= \int_{A(x)} \left(r_i(x,a^i;\sigma^*_{-i}) + \beta \int_X V_{\sigma^*_{-i}}(y) Q_{\sigma^*_{-i}}(dy|x,a^i) \right) \sigma_i^*(x)(da^i) \\ &\geq r_i(x,a^i;\sigma^*_{-i}) + \beta \int_X V_{\sigma^*_{-i}}(y) Q_{\sigma^*_{-i}}(dy|x,a^i), \quad a^i \in A_i(x). \end{aligned} \tag{8}$$

For fixed x and $(\sigma_i^*)_{i\in I}$ consider the one-shot game $\Gamma(x,\sigma^*)$ on $A_1(x) \times \cdots \times A_m(x)$, where the payoff of i-th player equals to

$$H_i(x,a) = r_i(x,a^i;\sigma^*_{-i}) + \beta \int_X V_{\sigma^*_{-i}}(y) Q_{\sigma^*_{-i}}(dy|x,a^i).$$

From (8) it follows that $(\sigma_i^*(x))_{i\in I}$ is a mixed strategy Nash equilibrium in the game $\Gamma(x,\sigma^*)$. So, we have proved the following well-known result: see (Jaśkiewicz and Nowak, 2018, He and Sun, 2017) for similar statements.

Theorem 2. *For a discounted stochastic game with payoff functionals (2) and stationary Markov strategies* $(\sigma_i^*)_{i\in I}$ *the following conditions are equivalent.*

(i) $(\sigma_i^*)_{i\in I}$ *is a Nash equilibrium.*
(ii) *Each* σ_i^*, $i \in I$ *is an optimal solution of the Markov decision process with the objective functional (5), transition kernel (7), and reward (6).*
(iii) $(\sigma_i^*(x))_{i\in I}$ *is a mixed strategy Nash equilibrium in the game* $\Gamma(x,\sigma^*)$ *for each* $x \in X$.

There are several additional assumptions that ensure the existence of a stationary Markov equilibrium: see the survey (Jaśkiewicz and Nowak, 2018). We rely on the results of (He, 2014; He and Sun, 2017).

Assume that

(A) $Q(\cdot|x,a)$ is absolutely continuous with respect to a probability measure λ on $(X,\mathscr{B}(X))$ for each $(x,a)\in X\times\overline{A}$. The related density $(x,a,y)\mapsto q(y|x,a)$ is assumed to be $\mathscr{B}(X)\times\mathscr{B}(\overline{A})\times\mathscr{B}(X)$-measurable. Here $\mathscr{B}(\overline{A})$ is the Borel σ-algebra of the topological space $(\overline{A},\tau)$.
(B) For each $x\in X$ the mapping $a\mapsto Q(\cdot|x,a)$ is continuous in the total variation norm:

$$\lim_{a_n\to a}\sup_{B\in\mathscr{B}(X)}|Q(B|x,a_n)-Q(B|x,a)|=0.$$

Consider a probability space $(\Omega,\mathscr{F},\mathsf{P})$. A sub-$\sigma$-algebra $\mathscr{G}\subset\mathscr{F}$ is said to be *setwise coarser* He, 2014 than $\mathscr{F}$ if for every $D\in\mathscr{F}$ with $\mathsf{P}(D)>0$ there exists a set $D_0\subset D$, $D_0\in\mathscr{F}$ such that $\mathsf{P}(D_0\triangle D_1)>0$ for any $D_1\in\{D'\cap D : D'\in\mathscr{G}\}$ (equivalently, $\mathscr{F}$ has no $\mathscr{G}$-atom under P: see (He and Sun, 2017; He and Sun, 2018) for this terminology). A stochastic game has a *coarser transition kernel* if there exists a σ-algebra $\mathscr{G}\in\mathscr{B}(X)$ such that $q(\cdot|x,a)$ is $\mathscr{G}$-measurable for all $(x,a)\in X\times\overline{A}$ and $\mathscr{G}$ is setwise coarser than $\mathscr{B}(X)$.

Theorem 3 (He and Sun, 2017, Theorem 1). *Assume that the assumptions (A), (B) are satisfied and the game has a coarser transition kernel. Then there exists a Nash equilibrium* $(\sigma_i^*)_{i\in I}$.

In (He and Sun, 2017) Theorem 3 was formulated and proved under a more general assumption that there exists a *decomposable coarser transition kernel.*

3. The result

We use the notation of Section 2. and assume that the conditions of Theorem 3 are satisfied. A formal description of the game in question is given as follows.

(I) The manager selects a tuple $c=(c_i)_{i\in I}$ of non-negative stimulating functions $c_i\in\mathscr{C}_b(X\times\overline{A})$.
(II) The pool of $m\geq 1$ producers with the reward functions

$$r_i(x,a)=c_i(x,a)-g_i(x,a),\quad 0\leq g_i\in\mathscr{C}_b(X\times\overline{A})$$

and payoffs

$$J_i(x,\sigma,c) = \mathsf{E}_{x,\sigma} \sum_{t=0}^{\infty} \beta^t (c_i(x,a) - g_i(x,a))$$

play the described stochastic game which results in a (stationary Markov) Nash equilibrium $(\sigma_i^*)_{i\in I}$.

(III) The manager gets the payoff

$$J_L(x,\sigma^*,c) = \mathsf{E}_{x,\sigma^*} \sum_{t=0}^{\infty} \beta^t \left(f(x_t,a_t) - \sum_{i=1}^{N} c_i(x_t,a_t) \right), \tag{9}$$

where $f \in \mathscr{C}_b(X \times \overline{A})$.

Denote by $T(c)$ the set of Nash equilibriums for a given stimulating function c. The leader aim is to maximize her payoff for a worst Nash equilibrium:

$$G(x,c) = \inf_{\sigma^* \in T(c)} J_L(x,\sigma^*,c).$$

The problem of this sort is known as a weak Stackelberg game: see (Breton et al., 1988). Let us call

$$V_L(x) = \sup\{G(x,c) : c_i \in \mathscr{C}_b(X \times \overline{A}),\ i \in I\}$$

the value of the leader. A tuple $\overline{c}_\varepsilon$ is called an ε-Stackelberg solution, if

$$V_L(x) - \varepsilon \le G(x,\overline{c}_\varepsilon), \qquad x \in X.$$

A pair $(\overline{c}_\varepsilon,\sigma^*)$, $\sigma^* \in T(\overline{c}_\varepsilon)$ is called an ε-Stackelberg equilibrium.

Consider an auxiliary Markov decision process:

$$J(x,\sigma) = \mathsf{E}_{x,\sigma} \sum_{t=0}^{\infty} \beta^t \left(f(x_t,a_t) - \sum_{i=1}^{N} g_i(x_t,a_t) \right) \to \max_{(\sigma_i)_{i\in I}}. \tag{10}$$

This problem is attributed to the leader, who performs the maximization over the tuples $\sigma = (\sigma_i)_{i\in I}$. By Theorem 1 this problem has an optimal solution of the form $\overline{\sigma} = (\delta_{\overline{u}_i(x)})_{i\in I}$, which can be identified with a Borel measurable selector

$$(\overline{u}_i(x))_{i\in I} \in \arg\max_{a\in A(x)} \left\{ f(x,a) - \sum_{i=1}^{m} g_i(x,a) + \beta \int_X V(y) Q(dy|x,a) \right\}.$$

Here $V(x) = \sup_\sigma J(x,\sigma)$ is the value function of the problem (10). It would coincide with the optimal payoff of the leader if she was engaged in the production without resorting to the services of producers.

We will assume that any producer suffers no cost if his production level is zero:

$$g_i(x,0,a^-) = 0. \tag{11}$$

Theorem 4. *Under the assumptions of Theorem 3 the following assertions hold true.*

(i) $V_L(x) = V(x)$.
(ii) *The tuple* $\overline{c}_\varepsilon = (c_{i,\varepsilon})_{i\in I}$,

$$c_{i,\varepsilon}(x,a) = g_i(x,a) + \frac{\varepsilon}{m}(1-\beta)(1-|a^i - \overline{u}_i(x)|)^+, \quad y^+ := \max\{y,0\}$$

is an ε*-Stackelberg solution. The related Nash equilibrium is unique:* $T(\overline{c}_\varepsilon) = \{(\delta_{\overline{u}_i(x)})_{i\in I}\}$.

Proof. For any $\sigma^* \in T(c)$, $0 \le c \in \mathscr{C}_b(X\times A)$ we have

$$J_i(x,\sigma^*,c) \ge J_i(x,(\delta_0^i,\sigma^*_{-i}),c) \ge 0$$

in view of (11). It follows that

$$\begin{aligned} J_L(x,\sigma^*,c) &\le J_L(x,\sigma^*,c) + \sum_{i=1}^m J_i(x,\sigma^*,c) \\ &= \mathsf{E}_{x,\sigma^*}\sum_{t=0}^\infty \beta^t\Big(f(x_t,a_t) - \sum_{i=1}^m g_i(x_t,a_t)\Big) = J(x,\sigma^*) \le V(x). \end{aligned} \tag{12}$$

Hence, $V_L(x) \le V(x)$.

Let $\sigma^* \in T(\overline{c}_\varepsilon)$. As was mentioned in Section 2., each component of the tuple $(\sigma_i^*)_{i\in I}$ is an optimal solution of the Markov decision process

$$\begin{aligned} J_i(x,(\sigma_i,\sigma^*_{-i}),c) &= \mathsf{E}_{x,\sigma_i,\sigma^*_{-i}}\sum_{t=0}^\infty \beta^t(c_i(x_t,a_t^i;\sigma^*_{-i}) - g_i(x_t,a_t^i;\sigma^*_{-i})) \\ &= (1-\beta)\frac{\varepsilon}{m}\mathsf{E}_{x,\sigma_i,\sigma^*_{-i}}\sum_{t=0}^\infty \beta^t(1-|a_t^i - \overline{u}_i(x_t)|)^+ \to \max_{\sigma_i}. \end{aligned}$$

Clearly,

$$J_i(x,(\sigma_i,\sigma^*_{-i}),\overline{c}_\varepsilon) \le (1-\beta)\frac{\varepsilon}{m}\mathsf{E}_{x,\sigma_i,\sigma^*_{-i}}\sum_{t=0}^\infty \beta^t \le \frac{\varepsilon}{m},$$

$$J_i(x,(\delta_{\overline{u}_i},\sigma^*_{-i}),\overline{c}_\varepsilon) = \frac{\varepsilon}{m}.$$

Hence,

$$V_{\sigma^*_{-i}}(x,\overline{c}_\varepsilon) := \sup_{\sigma_i} J_i(x,(\sigma_i,\sigma^*_{-i}),\overline{c}_\varepsilon) = \frac{\varepsilon}{m}.$$

By Theorem 2 for each x the tuple $(\sigma_i^*(x))_{i\in I}$ is a mixed strategy Nash equilibrium in the one-shot game on $A_1(x)\times\cdots\times A_m(x)$, where the payoff of i-th player equals to

$$\begin{aligned} H_i(x,a,\overline{c}_\varepsilon) &= c_i(x_t,a_t^i;\sigma^*_{-i}) - g_i(x_t,a_t^i;\sigma^*_{-i}) + \beta\int_X V_{\sigma^*_{-i}}(y)Q_{\sigma^*_{-i}}(dy|x,a^i) \\ &= (1-|a^i - \overline{u}_i(x)|)^+ + \frac{\varepsilon}{m}. \end{aligned}$$

In this trivial game the tuple $(\sigma_i^*(x))_{i\in I} = (\delta_{\overline{u}_i(x)})_{i\in I}$ of pure strategies is the unique Nash equilibrium.

Finally, substituting $\overline{c}_\varepsilon$ in (9), we get

$$G(x,\overline{c}_\varepsilon) = J_L(x,(\delta_{\overline{u}_i})_{i\in I},\overline{c}_\varepsilon) = \mathsf{E}_{x,(\delta_{\overline{u}_i})_{i\in I}} \sum_{t=0}^{\infty} \beta^t \left(f(x_t,a_t) - \sum_{i=1}^{m} g_i(x_t,a_t) \right)$$
$$- (1-\beta)\frac{\varepsilon}{m}\mathsf{E}_{x,(\delta_{\overline{u}_i})_{i\in I}} \sum_{t=0}^{\infty} \beta^t \sum_{i=1}^{m} (1-|a_t^i - \overline{u}_i(x_t)|)^+ = V(x) - \varepsilon.$$

By (12) it follows that $\overline{c}_\varepsilon$ is an ε-Stackelberg solution, and $V_L(x) = V(x)$.

Remark 1. In the case of a single follower ($m = 1$) similar results were obtained in (Rokhlin and Ougolnitsky, 2018). However, the incentive premium in (Rokhlin and Ougolnitsky, 2018) was discontinuous in a:

$$c_\varepsilon(x,a) - g(x,a) = \varepsilon(1-\beta)I_{\{a=\overline{u}(x)\}} \tag{13}$$

(we drop the index "1"). Furthermore, an analogue of Theorem 4 was proved either under a strong assumption that in the auxiliary problem (10) there exists a continuous optimal strategy $\overline{u}$ (by the way, this requires a topology on the state space), or by working with the notion of (ε,η)-Stackelberg solution and with the class of universally measurable stimulating functions c. The assumption of the present paper that the production costs g_i, the revenue function f and the stimulating functions c_i belong to the class $\mathscr{C}_b(X \times \overline{A})$ leads to more natural and simple results. On the other hand, in (Rokhlin and Ougolnitsky, 2018Theorem 2) it was shown that for the incentive premium (13) the follower can deviate from $\overline{u}$ only at the expense of "large losses". Thus, such premium has its own merits.

In the case of finite state and action spaces, where arise no measure theoretic difficulties, Theorem 4 remains valid for discontinuous incentive premium

$$c_{i,\varepsilon}(x,a) - g_i(x,a) = \frac{\varepsilon}{m}(1-\beta)I_{\{a^i=\overline{u}_i(x)\}}.$$

Remark 2. In He and Sun, 2017 it is mentioned that Theorem 3 implies the existence of a *stationary Markov correlated equilibrium* under the assumptions (A), (B). Closely following (He and Sun, 2017), we succintly describe this point as follows. Consider the extended state space $X' = X \times [0,1]$ endowed with the product σ-algebra $\mathscr{B}(X) \otimes \mathscr{B}([0,1])$ and the product measure $\lambda' = \lambda \otimes \eta$, where η is the Lebesgue measure on $\mathscr{B}([0,1])$. In the related model at each stage all players obtain a signal $z_t \in [0,1]$. These signals are independent random variables, which are uniformly distributed on $[0,1]$. The transition probability takes the form

$$Q'(B \times C|x,z,a) = Q(B|x,a)\eta(C).$$

For the density $q'(\cdot|x,z,a)$ of $Q'(\cdot|x,z,a)$ with respect to λ' we have

$$q'(y,u|x,z,a) = q(y|x,a).$$

The function $q'(\cdot|x,z,a) = q(\cdot|x,a)$ is measurable with respect to the σ-algebra $\mathscr{G}' = \mathscr{B}(X)\otimes\{\emptyset,L\}$ for all x,a, and this σ-algebra is setwise coarser than $\mathscr{B}(X)\otimes\mathscr{B}([0,1])$. Hence, the new model satisfies the condition of coarser transition kernel and possesses a stationary Markov equilibrium. By definition, this means the existence of a correlated equilibrium in the original model, satisfying the assumptions (A), (B).

So, if in the scheme (I) – (III), describing the Stackelberg game, at stage (II) we replace a Nash equilibrium by a correlated equilibrium, then all assertions of Theorem 4 remain valid, if the condition of coarser transition kernel is dropped. The proof in fact does not change, since it is insensitive to the state space.

4. Conclusion

The present paper is related to the development of the theory of incentives in a dynamical stochastic formulation. Essentially, it generalizes the results of (Rokhlin and Ougolnitsky, 2018) for the case of multiple followers. Overall, the leader should assume that she does not rely on producers' services and attribute their costs to herself. After determining optimal production strategies from the corresponding Markov decision process she should economically motivate the producers to follow these policies. The closely related problems of multiple leaders and continuous time deserve further study.

References

Aliprantis, C. D., Border, K. C. (2006). Infinite dimensional analysis: a hitchhiker's guide. Springer-Verlag, Berlin.

Bogachev, V. (2007). Measure theory. Volume II. Springer-Verlag, Berlin.

Breton, M., Alj, A., Haurie, A. (1988). *Sequential Stackelberg equilibria in two-person games.* J. Optim. Theory Appl., **59(1)**, 71–97.

Dynkin, E. B., Yushkevich, A. A. (1979). Markov control processes and their applications. Springer-Verlag, New York.

He, W. (2014). Theory of correspondences and games. Ph.D. Thesis, National University of Singapore.

He, W., Sun, Y. (2017). *Stationary Markov perfect equilibria in discounted stochastic games.* J. Econ. Theory, **169**, 35–61.

He, W., Sun, Y. (2018). *Conditional expectation of correspondences and economic applications.* Econ. Theory, **66(2)**, 265–299.

Himmelberg, C. J., Parthasarathy, T., VanVleck, F. S. (1976). *Optimal plans for dynamic programming problems.* Math. Oper. Res., **1**(4), 390–394.

Hu, S., Papageorgiou N. S. (1997). Handbook of multivalued analysis. Volume 1: Theory. Kluwer, Dordrecht.

Hu, S., Papageorgiou N. S. (2000). Handbook of multivalued analysis. Volume 2: Applications. Kluwer, Dordrecht.

Jaśkiewicz, A., Nowak, A. S. (2018). *Non-zero-sum stochastic games.* In Basar, T., Zaccour, G. (Eds.), Handbook of Dynamic Game Theory. Springer, Cham.Springer.

Laffont, J.-J., Martimort, D. (2002). The theory of incentives: the principal-agent model. Princeton University Press, Princeton, NJ.

Myerson, R. B. (1983). *Mechanism design by an informed principal.* Econometrica, **51(6)**, 1767–1797.

Novikov, D. A. (2013). Theory of control in organizations. Nova Science Publishers, New York.

Novikov, D. A., Shokhina, T. E. (2003). Incentive mechanisms in dynamic active systems. Autom. Remote Control, **64**, 1912–1921.

Olsder, G. J. (2009). *Phenomena in inverse Stackelberg games, part 2: Dynamic problems.* J. Optim. Theory Appl., **143(3)**, 601–618.

Rokhlin, D. B., Ougolnitsky, G. A. (2018). *Stackelberg equilibrium in a dynamic stimulation model with complete information.* Autom. Remote Control, **79**, 701–712.

Srivastava, S. M. (1998). A course on Borel sets. Springer, New York.