

Principles of Stable Cooperation in Stochastic Games

Elena M. Parilina

St. Petersburg University,

Faculty of Applied Mathematics and Control Processes,

Bibliotekhnaya pl. 2, St. Petersburg, 198504, Russia

E-mail: elena.parilina@gmail.ru

WWW home page: <http://www.apmath.spbu.ru/ru/staff/parilina/>

Abstract The paper considers stochastic games in the class of stationary strategies. The cooperative form of this class of stochastic games is constructed. The cooperative solution is found. Conditions of dynamic stability for stochastic games are obtained. Principles of dynamic stability include three conditions: subgame consistency, strategic stability and irrational behavior proof condition of the cooperative agreement. Also the paper considers the example for which the cooperative agreement is found and the conditions of dynamic stability are checked.

Keywords: cooperative stochastic game, stationary strategies, time consistency, subgame consistency, payoff distribution procedure, strategic stability, irrational behavior proof condition

1. Introduction

A stochastic game is a dynamic game process. If cooperation is possible in the game an important property of cooperative agreement is stability in dynamics. The work of Petrosyan and Zenkevich, 2009, considers three principles of stable cooperation: time-consistency (dynamic consistency), strategic stability and irrational behavior proof condition.

L.A. Petrosyan was the first to introduce the concept of dynamic consistency for differential games (Petrosyan, 1977). This condition appeared to be typical also for stochastic games (Petrosyan, 2006). The paper considers stochastic games in stationary strategies with the finite number of states any of which can be realized at every game stage. With this type of the definition of stochastic games the consistency of cooperative agreement should take place in every position (state) of the game. In other words, the claim of subgame consistency is laid to the cooperative agreement. Subgame consistency of the cooperative agreement lets the players expect receiving the allocation according to the same optimality principle in every stochastic subgame.

The condition of strategic stability is guaranteed by the existence of Nash equilibrium in the regularized game with the payoffs that players expect to receive as a result of the cooperative agreement. The regularization of the game is constructed on the basis of the initial stochastic game with the use of payoff distribution procedure (Petrosyan and Danilov, 1979). The conditions of strategic stability for stochastic games were also considered by Grauer and Petrosyan, 2002.

The irrational behavior proof condition (Yeung, 2006) lets guarantee that if some player (group of players) cancels the agreement at some stage of the game and players play individually from this stage to the end of the game they receive not

less than if each player plays by himself during the whole game. This condition also lets secure the cooperative agreement from the force majeure circumstances.

The definition of stochastic games was introduced by Shapley, 1953a. At present a lot of papers deal with the study of stochastic games (Petrosyan et al., 2004, Petrosyan and Baranova, 2006, Herings and Peeters, 2004). Stochastic games have wide application in the field of telecommunication system modeling (Parilina, 2010, Altman et al., 2003), in economics (Amir, 2003), in the problem of tax evasion (Raghavan, 2006).

2. Stochastic games in stationary strategies

Stochastic game begins with the chance turn, i.e. with the choice of the initial state of the game which the game process begins with. The state of the stochastic game is determined as simultaneous normal form game of n players. One of the finite number of states is realized at each stage of the stochastic game. In the state some action profile is realized depending on which transition to the next states is accomplished with some probability. The payoff of the players is discounted when the game goes on. There are some notations:

- The set of players is $N = \{1, \dots, n\}$.
- The set of states is $\{\Gamma^j\}_{j=1}^t$ where $\Gamma^j = \langle N, X_1^j, \dots, X_n^j, K_1^j, \dots, K_n^j \rangle$ is state j , set N is equal for all Γ^j , $j = 1, \dots, t$, X_i^j is the finite set of pure strategies of player i in Γ^j , $K_i^j(x_1^j, \dots, x_n^j) = K_i^j(x^j)$ is a payoff function of player i in state Γ^j , $j = 1, \dots, t$.
- The probability that state Γ^k is realized if at the previous stage (in state Γ^j) action profile $x^j = (x_1^j, \dots, x_n^j)$ has realized, is $p(j, k; x^j)$. It is obvious that $p(j, k; x^j) \geq 0$ and $\sum_{k=1}^t p(j, k; x^j) = 1$ for each $x^j \in X^j = \prod_{i \in N} X_i^j$ and for any $j, k = 1, \dots, t$.
- The discount factor is $\delta \in (0, 1)$.
- The vector of the initial distribution on states $\Gamma^1, \dots, \Gamma^t$ is $\pi^0 = (\pi_1^0, \dots, \pi_t^0)$, where π_j^0 is the probability that state Γ^j is realized at the first stage of the game, $\sum_{j=1}^t \pi_j^0 = 1$;
- The set of player i 's stationary strategies is $\Xi_i = \{\eta_i\}$. Using stationary strategies the player's choice of the strategy in each state from set $\{\Gamma^1, \dots, \Gamma^t\}$ at any stage depends only on which state is realized at this stage, i.e. $\eta_i : \Gamma^j \mapsto x_i^j \in X_i^j, j = 1, \dots, t$.

Definition 1. Call the set

$$G = \left\langle N, \{\Gamma^j\}_{j=1}^t, \{\Xi_i\}_{i \in N}, \delta, \pi^0, \{p(j, k; x^j)\}_{j=\overline{1,t}, k=\overline{1,t}, x^j \in \prod_{i=1}^n X_i^j} \right\rangle \quad (1)$$

finite stochastic game in stationary strategies.

Definition 2. Call stochastic game (12) with vector $\pi^0 = (0, \dots, 0, 1, 0, \dots, 0)$ (with 1 in the j th component), i.e. the game beginning with state Γ^j , the finite stochastic subgame in stationary strategies and denote as G^j , $j = 1, \dots, t$.

Remark 1. Obviously, player i 's stationary strategy in game G is player i 's stationary strategy in any subgame G^1, \dots, G^t .

Payoff in finite stochastic game is a random variable. So we have to determine the utility function of the payoff. Consider the mathematical expectation of the player's payoff as the utility of his payoff in stochastic game G . Let $\bar{E}_i(\eta)$ be the expected payoff of player i in game G and $E_i^j(\eta)$ be the expected payoff of player i in subgame G^j when strategy profile η is realized in stochastic game G (subgame G^j). Form vector $E_i(\eta) = (E_i^1(\eta), \dots, E_i^t(\eta))$.

For the expected payoff of player i in subgame G^j the following recurrent equation takes place:

$$E_i^j(\eta) = K_i^j(x^j) + \delta \sum_{k=1}^t p(j, k; x^j) E_i^k(\eta) \quad (2)$$

under condition that $\eta(\Gamma^j) = x^j$, i.e. $\eta(\cdot) = (\eta_1(\cdot), \dots, \eta_n(\cdot))$ where $\eta_i(\Gamma^j) = x_i^j \in X_i^j$, $x^j = (x_1^j, \dots, x_n^j)$ for each $j = 1, \dots, t$, $i \in N$.

Since stochastic game G is considered in the class of stationary strategies defined above and the set of states $\{\Gamma^1, \dots, \Gamma^t\}$ is finite then it is sufficiently to consider t of subgames G^1, \dots, G^t accordingly beginning with the states $\Gamma^1, \dots, \Gamma^t$.

Hereinafter, let $\eta(\cdot) = (\eta_1(\cdot), \dots, \eta_n(\cdot))$ be the stationary strategy profile such as $\eta_i(\Gamma^j) = x_i^j \in X_i^j$ where $j = 1, \dots, t$, $i \in N$. We restrict our consideration to the set of player i 's pure stationary strategies in stochastic game G . Denote it as \tilde{E}_i .

The matrix of transition probabilities in stochastic game G under the realization of stationary strategy profile $\eta(\cdot)$ looks like:

$$\Pi(\eta) = \begin{pmatrix} p(1, 1; x^1) & \dots & p(1, t; x^1) \\ p(2, 1; x^2) & \dots & p(2, t; x^2) \\ \dots & \dots & \dots \\ p(t, 1; x^t) & \dots & p(t, t; x^t) \end{pmatrix}. \quad (3)$$

We can rewrite equation (2) in matrix form using (3) as follows:

$$E_i(\eta) = K_i(\eta) + \delta \Pi(\eta) E_i(\eta), \quad (4)$$

where $K_i(\eta) = (K_i^1(x^1), \dots, K_i^t(x^t))$, and $K_i^j(x^j)$ is the player i 's payoff in state Γ^j on condition that strategy profile $x^j \in X^j$ is realized in this state.

Equation (3) is equivalent to the following one:

$$E_i(\eta) = (\mathbb{I} - \delta \Pi(\eta))^{-1} K_i(\eta), \quad (5)$$

where \mathbb{I} is an identity $t \times t$ matrix.

Remark 2. Matrix $(\mathbb{I} - \delta \Pi(\eta))^{-1}$ always exists for $\delta \in (0, 1)$. It is not difficult to prove this statement. It is known that all the eigenvalues of stochastic matrix $\Pi(\eta)$ are in the interval $[-1, 1]$. For the existence of matrix $(\mathbb{I} - \delta \Pi(\eta))^{-1}$ it is necessary and sufficient that the determinant of matrix $(\Pi(\eta) - \frac{1}{\delta} \mathbb{I})$ be not equal to zero. So matrix $(\Pi(\eta) - \frac{1}{\delta} \mathbb{I})$ must not have the eigenvalue equal to $\frac{1}{\delta}$. The last condition takes place because $\frac{1}{\delta} > 1$, so this number cannot be the eigenvalue of stochastic matrix $\Pi(\eta)$.

The expected payoff of player i in the stochastic game G we can find in the following way:

$$\bar{E}_i(\eta) = \pi^0 E_i(\eta). \quad (6)$$

3. Cooperation in stochastic games

Suppose now, that the players from N decide to cooperate to receive the maximum total payoff. Denote the pure strategy profile maximizing the sum of the expected players' payoffs in stochastic game G as $\bar{\eta}(\cdot) = (\bar{\eta}_1(\cdot), \dots, \bar{\eta}_n(\cdot))$, i.e.

$$\max_{\eta \in \prod_{i \in N} \Xi_i} \sum_{i \in N} \bar{E}_i(\eta) = \sum_{i \in N} \bar{E}_i(\bar{\eta}). \tag{7}$$

Problem (6) may have more than one decision. Call strategy profile $\bar{\eta}(\cdot)$ satisfying (6) as cooperative decision.

The coalitional form of noncooperative game is usually given by the pair $\langle N, V \rangle$, where N is the set of players and V is a real-valued function, called the characteristic function of the game, defined on the set 2^N (the set of all subsets of N), and satisfying two properties (1) $V(\emptyset) = 0$, and (2) (superadditivity) for any disjoint coalitions $S, T \subset N$, $S \cap T = \emptyset$, the next inequality is satisfied: $V(S) + V(T) \leq V(S \cup T)$. The value $V(S)$ is a real number for each coalition $S \subset N$, which may be interpreted as the worth or power of coalition S when its members play together as a unit. Condition (2) says that the value of two disjoint coalitions is at least as great when they play together as when they work apart. The assumption of superadditivity is not needed for some of the theory of coalitional games, but it seems to be a natural condition.

Define the characteristic function $\bar{V}(S)$ in stochastic game G via characteristic function $V^j(S)$ of stochastic subgames G^j , $j = 1, \dots, t$, as follows:

$$\bar{V}(S) = \pi^0 V(S) \tag{8}$$

for any coalition $S \subset N$ where $V(S) = (V^1(S), \dots, V^t(S))$, $V^j(S)$ is the value of the characteristic function of stochastic subgame G^j derived for coalition S .

The task is to determine the characteristic function $V^j(S)$ for any coalition S .

Firstly, consider $S = N$. Bellman equation (Bellman, 1957) for the value $V(N)$ can be written as follows:

$$V(N) = \max_{\eta \in \prod_{i \in N} \Xi_i} \left[\sum_{i \in N} K_i(\eta) + \delta \Pi(\eta) V(N) \right] = \sum_{i \in N} K_i(\bar{\eta}) + \delta \Pi(\bar{\eta}) V(N),$$

where $\bar{\eta}(\cdot)$ is the pure strategy profile satisfying condition (6).

The value $V(N)$ is got from the previous equation:

$$V(N) = (\mathbb{I} - \delta \Pi(\bar{\eta}))^{-1} \sum_{i \in N} K_i(\bar{\eta}). \tag{9}$$

Secondly, consider $S \subset N$, $S \neq \emptyset$. To define the value of characteristic function $V^j(S)$ for this coalition, $j = 1, \dots, t$, for each subgame G^j , define an auxiliary zero-sum stochastic game G_S^j where coalition $S \subset N$ plays as a maximizing player and coalition $N \setminus S$ plays as a minimizing player. Define the value of function $V^j(S)$ for subgame G^j as a lower value of antagonistic stochastic game G_S^j in pure stationary strategies (in fact, the lower value of matrix game):

$$V^j(S) = \max_{\eta_S} \min_{\eta_{N \setminus S}} \sum_{i \in S} E_i^j(\eta_S, \eta_{N \setminus S}), \tag{10}$$

where $(\eta_S(\cdot), \eta_{N \setminus S}(\cdot))$ is a strategy profile in pure stationary strategies and $\eta_S(\cdot) = (\eta_{i_1}(\cdot), \dots, \eta_{i_k}(\cdot))$ is a vector of stationary strategies of players $i_1, \dots, i_k \in S$, $i_1 \cup \dots \cup i_k = S$, $\eta_S(\cdot) \in \prod_{j=1}^k \tilde{\Xi}_{i_j}$, the set of pure stationary strategies of coalition $S \subset N$, and $\eta_{N \setminus S}(\cdot)$ is a vector of stationary strategies of players $i_{k+1}, \dots, i_n \in N \setminus S$, $i_{k+1} \cup \dots \cup i_n = N \setminus S$, $\eta_{N \setminus S}(\cdot) \in \prod_{j=k+1}^n \tilde{\Xi}_{i_j}$, the set of pure stationary strategies of coalition $N \setminus S$.

Finally, consider $S = \emptyset$ and get the value of characteristic function:

$$V^j(\emptyset) = 0. \quad (11)$$

Remark 3. Characteristic functions $\bar{V}(S)$ determined by (10) and $V^j(S)$ determined by (7)–(11) are superadditive.

Definition 3. Cooperative stochastic subgame G_{co}^j is a set $\langle N, V^j(\cdot) \rangle$, where N is the set of players, and $V^j : 2^N \rightarrow R$ is the characteristic function calculated by (7) – (11).

Definition 4. Cooperative stochastic game G_{co} is a set $\langle N, \bar{V}(\cdot) \rangle$, where N is the set of players and $\bar{V} : 2^N \rightarrow R$ is the characteristic function calculated by (10).

Definition 5. Vector $\alpha^j = (\alpha_1^j, \dots, \alpha_n^j)$ satisfying the two following conditions:

$$1) \sum_{i \in N} \alpha_i^j = V^j(N),$$

$$2) \alpha_i^j \geq V^j(\{i\}) \text{ for any } i \in N,$$

is called the allocation in subgame G_{co}^j ($j = 1, \dots, t$). Denote the set of allocations in cooperative subgame G_{co}^j as A^j , $j = 1, \dots, t$.

Definition 6. Vector $\bar{\alpha} = (\bar{\alpha}_1, \dots, \bar{\alpha}_n)$, where $\bar{\alpha}_i = \pi^0 \alpha_i$, $\alpha_i = (\alpha_i^1, \dots, \alpha_i^t)$, and $(\alpha_1^j, \dots, \alpha_n^j) = \alpha^j \in I^j$ is called an allocation in cooperative stochastic game G_{co} . Denote the set of allocations in cooperative stochastic game G_{co} as \bar{I} .

Suppose that the set of allocations in any subgame G_{co}^j , $j = 1, \dots, t$, is nonempty. So the set of allocations in cooperative stochastic game G_{co} is also nonempty.

4. Principles of stable cooperation

4.1. Subgame consistency of cooperative agreement

Suppose that players cooperate in stochastic game and for every subgame G_{co}^j choose allocation $\alpha^j = (\alpha_1^j, \dots, \alpha_n^j) \in I^j$. The problem is how to realize payments to the players at each stage of the stochastic game for getting the expected payoff α_i^j for player i in stochastic subgame G^j . If players receive payoffs according to their payoff functions in the states they hardly ever get the components of the chosen allocation in mathematical expectation sense. To find the way out of the situation we should suggest the method of redistribution of total players' payoff in every state realized in the stochastic game process. This method was proposed by Petrosyan and Danilov, 1979, for differential games.

There are two principles of constructing the real payments to the players in the dynamic game adapted to the theory of stochastic games:

1. The sum of payments to the players in every state is equal to the sum of players' payoffs in strategy profile realized in this state according to cooperative decision $\bar{\eta}(\cdot)$.
2. The expected sum of payments to player i in each subgame G^j is equal to the i th component of an allocation in subgame G_{co}^j that players have chosen before the beginning of the game.

Taking into account that in stochastic game (12) the number of subgames is equal to the number of possible states we should find vector $\beta_i = (\beta_i^1, \dots, \beta_i^t)$ for every $i \in N$, where β_i^j is a payment to player i in state Γ^j , $j = 1, \dots, t$. And these payments have to satisfy the two above principles and if so then these payments can be called payoff distribution procedure (PDP) (see Petrosyan and Danilov, 1979).

Find the conditions that the new payoffs of the players are satisfied the principles of PDP in the terms of stochastic games.

1. The first principle is equivalent to the following equation:

$$\sum_{i \in N} \beta_i^j = \sum_{i \in N} K_i^j(\bar{x}^j), \quad (12)$$

where \bar{x}^j is an action profile realized under cooperative decision $\bar{\eta}(\cdot)$ in state Γ^j , $j = 1, \dots, t$.

2. To find the condition of the second principle we need to work out the expected total payoff of player i in stochastic subgame with new payments β_i^j in state Γ^j , $j = 1, \dots, t$. Denote this value as B_i and write the recurrent equation for this value:

$$B_i^j = \beta_i^j + \delta \sum_{k=1}^t p(j, k; x^j) B_i^k,$$

or in vector form:

$$B_i = \beta_i + \delta \Pi(\bar{\eta}) B_i, \quad (13)$$

where $B_i = (B_i^1, \dots, B_i^t)$. Equation (13) is equivalent to the following one:

$$B_i = (\mathbb{I} - \delta \Pi(\eta))^{-1} \beta_i. \quad (14)$$

In respect to the second principle of PDP and equation (14) we obtain the following equation for β_i :

$$\alpha_i = (\mathbb{I} - \delta \Pi(\eta))^{-1} \beta_i, \quad (15)$$

where $\alpha_i = (\alpha_i^1, \dots, \alpha_i^t)$, $(\alpha_1^j, \dots, \alpha_n^j) = \alpha^j \in I^j$. Equation (15) can be rewritten in an equivalent form:

$$\beta_i = (\mathbb{I} - \delta \Pi(\bar{\eta})) \alpha_i. \quad (16)$$

It is easy to show that β_i found from (16) satisfies (12). As $\sum_{i \in N} \beta_i^j$ is equal to $(\mathbb{I} - \delta \Pi(\bar{\eta})) \sum_{i \in N} \alpha_i = (\mathbb{I} - \delta \Pi(\bar{\eta})) V(N)$, and $V(N)$ is found from (7) then equation (12) holds.

Remark 4. Equation (16) equals the following functional equation

$$\alpha_i = \beta_i + \delta II(\bar{\eta})\alpha_i. \quad (17)$$

The second item in the right part of equation (17) is the expected value of the component of the allocation in subgame beginning with the next stage. Suppose that the allocation for each subgame is chosen from the same optimality principle that has been chosen by the players at the beginning of the game.

Obviously, if players keep to cooperative decision $\bar{\eta}(\cdot)$ the expected payoff of player i in stochastic game with new payments in some action profiles (which are realized in the states under cooperative decision $\bar{\eta}$) is equal to the expected value of the correspondent component of the allocation in cooperative stochastic game G_{co} .

Now for every allocation $\bar{\alpha} = (\bar{\alpha}_1, \dots, \bar{\alpha}_n)$, where $\bar{\alpha}_i = \pi^0 \alpha_i$, $\alpha_i = (\alpha_i^1, \dots, \alpha_i^t)$, $(\alpha_1^j, \dots, \alpha_n^j) = \alpha^j \in I^j$ we can determine the regularization of stochastic game G by the following definition.

Definition 7. Noncooperative stochastic game G_α (subgame G_α^j , $j = 1, \dots, t$) is called α -regularization of stochastic game G (subgame G^j), if for any player $i \in N$ in state I^j payoff function $K_i^{\alpha,j}(x^j)$ is defined as follows:

$$K_i^{\alpha,j}(x^j) = \begin{cases} \beta_i^j, & \text{if } x^j = \bar{x}^j; \\ K_i^j(x^j), & \text{if } x^j \neq \bar{x}^j, \end{cases} \quad (18)$$

where PDP $\beta = (\beta_1, \dots, \beta_n)$ (Petrosyan and Baranova, 2006) is found from (??).

The procedure of regularization of the stochastic game G (subgame G^j) suggests a method of construction of real payments to the players in every state and one can insist that players are interested in the redistribution of their payoffs because getting $\beta_i^1, \dots, \beta_i^t$ in states I^1, \dots, I^t respectively player i receives the same sum (in terms of mathematical expectation) in game G_α (G_α^j) as he has planned to receive in the cooperative stochastic game G_{co} (G_{co}^j) and the expected sum of the remained payments will belong to the same optimality principle which has been chosen by the players at the beginning of the game. In this case we can say that *subgame consistency* (*dynamic consistency*) of the chosen cooperative agreement takes place.

4.2. Strategic stability of cooperative agreement

Introduce the additional notations. Set $\Gamma(k)$ as the state realized at stage k of stochastic game G . It is obvious that $\Gamma(k) \in \{\Gamma^1, \dots, \Gamma^t\}$. Write $x(k)$ the strategy profile realized in state $\Gamma(k)$. Set the subgame of stochastic game G_α from Definition 3 beginning from state $\Gamma(k)$ as $G_\alpha^{\Gamma(k)}$.

Call the sequence $((\Gamma(1), x(1)), (\Gamma(2), x(2)), \dots, (\Gamma(k-1), x(k-1)))$ the history of stage k and denote it as $h(k)$. Let T be $\{(\Gamma^1, \bar{x}^1), (\Gamma^2, \bar{x}^2), \dots, (\Gamma^t, \bar{x}^t)\}$.

Stochastic game G and G_α are games with perfect information in the sense that at each stage k ($k = 1, 2, \dots$) all players from N know state $\Gamma(k)$ and the history of stage k .

Definition 8. We call behavior strategy profile $\varphi^*(\cdot) = (\varphi_1^*(\cdot), \dots, \varphi_n^*(\cdot))$ strong transferable equilibrium in regularized game G_α if for any coalition $S \subset N$, $S \neq \emptyset$, the inequality

$$\sum_{i \in S} \bar{E}_i^\alpha(\varphi^*) \geq \sum_{i \in S} \bar{E}_i^\alpha(\varphi^* \parallel \varphi_S) \quad (19)$$

is true for any behavior strategy of coalition S : $\varphi_S(\cdot) = \{\varphi_i(\cdot)\}_{i \in S} \in \prod_{i \in S} \Phi_i$, and $\bar{E}_i^\alpha(\cdot)$ is the expected payoff of player i in α -regularization of stochastic game G .

Theorem 1. If in α -regularization of stochastic game G with α such that $\bar{\alpha} = \pi^0 \alpha$, the following inequality holds for any coalition $S \subset N$, $S \neq \emptyset$:

$$\sum_{i \in S} \beta_i \geq (\mathbb{I} - \delta \Pi(\bar{\eta})) F(S), \quad (20)$$

where $F(S) = (F^1(S), \dots, F^t(S))$,

$$F^j(S) = \max_{\substack{x_S^j \in \prod_{i \in S} X_i^j \\ x_S^j \neq \bar{x}_S^j}} \left\{ \sum_{i \in S} K_i^j(\bar{x}^j \parallel x_S^j) + \delta \sum_{l=1}^t p(j, l; \bar{x}^j \parallel x_S^j) V^l(S) \right\},$$

then in regularized game G_α there exists a strong transferable equilibrium with payoffs $(\bar{\alpha}_1, \dots, \bar{\alpha}_n)$.

Proof. Consider the behavior strategy profile $\hat{\varphi}(\cdot) = (\hat{\varphi}_1(\cdot), \dots, \hat{\varphi}_n(\cdot))$ in game G_α :

$$\hat{\varphi}_i(h(k)) = \begin{cases} \bar{x}_i^j, & \text{if } \Gamma(k) = \Gamma^j, j = \overline{1, t}, h(k) \subset T; \\ \hat{x}_i^j(S), & \text{if } \Gamma(k) = \Gamma^j, j = \overline{1, t}, \exists l \in [1, k-1] \\ & \text{and } S \subset N, i \notin S: h(l) \subset T, \\ & \text{and } (\Gamma(l), x(l)) \notin T, \\ & \text{but } (\Gamma(l), (x(l) \parallel \bar{x}_S(l))) \in T, \\ \text{anyone} & \text{in other cases,} \end{cases} \quad (21)$$

where $\hat{x}_i^j(S)$ is the player i 's pure strategy in state Γ^j which with strategies $x_p^j(S)$, $p \neq i$, $p \notin S$ forms the strategy of coalition $\{N \setminus S\}$ in antagonistic game against coalition S in subgame G^{Γ^j} .

The proof of the theorem repeats the proof of folk theorems (see Dutta, 1995) using the structure of strategy (21). Prove that $\hat{\varphi}(\cdot) = (\hat{\varphi}_1(\cdot), \dots, \hat{\varphi}_n(\cdot))$ determined in (21) is a strong transferable equilibrium in stochastic game G_α .

From definition (21) it follows that on condition that all players keep cooperative decision $\bar{\eta}(\cdot)$ the expected payoff of coalition S in subgame G_α^j , $j = 1, \dots, t$, is equal to the following one:

$$E_S^j(\hat{\varphi}(\cdot)) = \sum_{i \in S} E_i^j(\hat{\varphi}(\cdot)) = \sum_{i \in S} E_i^j(\bar{\eta}(\cdot)).$$

Let $E_S(\hat{\varphi}(\cdot))$ be equal to vector $(E_S^1(\hat{\varphi}(\cdot)), \dots, E_S^t(\hat{\varphi}(\cdot)))$ then for any coalition $S \subset N$, $S \neq \emptyset$ the next equality takes place:

$$E_S(\hat{\varphi}) = (\mathbb{I} - \delta \Pi(\bar{\eta}))^{-1} \sum_{i \in S} \beta_i. \quad (22)$$

Consider strategy-profile $(\widehat{\varphi}(\cdot) \parallel \varphi_S(\cdot))$, $S \subset N$, $S \neq \emptyset$, when some coalition S deviates from strategy $\widehat{\varphi}_S(\cdot)$. Let stage k be such that there exists number $l \in [1, k-1]$ such that history $h(l) \subset T$ and state $(\Gamma(l), x(l)) \notin T$ but $(\Gamma(l), (x(l) \parallel \bar{x}_S(l))) \in T$. Without loss of generality suggest that $\Gamma(k) = \Gamma^j$. Determine the payoff of coalition S in game G_α in strategy profile $(\widehat{\varphi}(\cdot) \parallel \varphi_S(\cdot))$ by formula $\sum_{i \in S} \overline{E}_i^\alpha(\widehat{\varphi} \parallel \varphi_S) = \pi^0 \sum_{i \in S} E_i^\alpha(\widehat{\varphi} \parallel \varphi_S)$, where

$$\begin{aligned} \sum_{i \in S} E_i^\alpha(\widehat{\varphi} \parallel \varphi_S) &= \sum_{i \in S} E_i^{\alpha, [1, k-1]}(\widehat{\varphi} \parallel \varphi_S) \\ &\quad + \delta^{k-1} \Pi^{k-1}(\widehat{\varphi} \parallel \varphi_S) \sum_{i \in S} E_i^{\alpha, [k, \infty)}(\widehat{\varphi} \parallel \varphi_S), \end{aligned} \quad (23)$$

where the first term in the right side of equation (23) is the expected payoff of coalition S at the first $k-1$ stages of game G_α , $\sum_{i \in S} E_i^{\alpha, [k, \infty)}(\widehat{\varphi} \parallel \varphi_S)$ in the second term is the expected payoff of coalition S in the subgame of game G_α beginning from stage k . Since there were no deviations of any coalition from the cooperative decision $\overline{\eta}(\cdot)$ up to stage $k-1$ inclusive as it was shown before the following equalities holds for the elements of the right side of (23):

$$\begin{aligned} \sum_{i \in S} E_i^{\alpha, [1, k-1]}(\widehat{\varphi} \parallel \varphi_S) &= \sum_{i \in S} E_i^{\alpha, [1, k-1]}(\overline{\eta}), \\ \Pi^{k-1}(\widehat{\varphi} \parallel \varphi_S) &= \Pi^{k-1}(\overline{\eta}). \end{aligned}$$

In the second term of the right side of (23) as $E_i^{\alpha, [k, \infty)}(\widehat{\varphi} \parallel \varphi_S)$ we mean vector $(E_i^{\alpha, 1}(\widehat{\varphi} \parallel \varphi_S), \dots, E_i^{\alpha, t}(\widehat{\varphi} \parallel \varphi_S))$ where $E_i^{\alpha, j}(\widehat{\varphi} \parallel \varphi_S)$ is the expected payoff of player $i \in S$ in regularized subgame G_α^j beginning with state Γ^j .

Find the expected payoff of coalition S in subgame G_α^j beginning with stage k and state $\Gamma(k)$ is equal to Γ^j . The following formula takes place:

$$\sum_{i \in S} E_i^{\alpha, j}(\widehat{\varphi} \parallel \varphi_S) = \sum_{i \in S} K_i^j(\bar{x}^j \parallel x_S^j) + \delta \sum_{l=1}^t p(j, l; \bar{x}^j \parallel x_S^j) V^l(S), \quad (24)$$

because players from coalition $N \setminus S$ will punish coalition S playing the antagonistic game against coalition S beginning from stage $k+1$ according to the definition of strategy profile $\widehat{\varphi}(\cdot)$. In (24) the value of characteristic function $V^j(S)$ is determined by (9).

Since the expected payoffs of coalition S in strategy profiles $\widehat{\varphi}(\cdot)$ and $(\widehat{\varphi}(\cdot) \parallel \varphi_S(\cdot))$ are equal up to stage $k-1$, then as a result of deviation coalition S can guarantee the increase of payoff only at the expense of the part of game G_α beginning with stage k , i.e. at the expense of the expected payoff in subgame G_α^j , $j = 1, \dots, t$. Coalition S in strategy profile $(\widehat{\varphi}(\cdot) \parallel \varphi_S(\cdot))$ can guarantee the following expected payoff from stage k :

$$\max_{\substack{x_S^j \in \prod_{i \in S} X_i^j \\ x_S^j \neq \bar{x}_S^j}} \left\{ \sum_{i \in S} K_i^j(\bar{x}^j \parallel x_S^j) + \delta \sum_{l=1}^t p(j, l; \bar{x}^j \parallel x_S^j) V^l(S) \right\}. \quad (25)$$

According to the definition of PDP the expected payoff of coalition S in regularized subgame G_α^j in strategy profile $\widehat{\varphi}(\cdot)$ can be found from the equation:

$$\sum_{i \in S} E_i^\alpha(\widehat{\varphi}) = (\mathbb{I} - \delta \Pi(\overline{\eta}))^{-1} \sum_{i \in S} \beta_i, \tag{26}$$

where $E_i^\alpha(\widehat{\varphi}(\cdot)) = (E_i^{\alpha,1}(\widehat{\varphi}(\cdot)), \dots, E_i^{\alpha,t}(\widehat{\varphi}(\cdot)))$. Taking into account inequality (20) from (25), (26) and reasoning presented above we can obtain inequality

$$E_S^\alpha(\widehat{\varphi}(\cdot)) \geq E_S^\alpha(\widehat{\varphi}(\cdot) \parallel \varphi_S(\cdot)).$$

Thus behavior strategy profile (21) is the strong transferable equilibrium in α -regularization of game G . The expected payoff of player i in game G_α in strategy-profile $\widehat{\varphi}(\cdot)$ is equal to $\overline{\alpha}_i$ where $\overline{\alpha}_i = \pi^0 \alpha_i$ and vector $\alpha_i = (\alpha_i^1, \dots, \alpha_i^t)$ consists of i th components of allocations $\alpha^1, \dots, \alpha^t$ derived for cooperative subgames G^1, \dots, G^t accordingly.

Corollary 1. *If α -regularization of game G for any player $i \in N$ the next inequality takes place:*

$$\beta_i \geq (\mathbb{I} - \delta \Pi(\overline{\eta})) W_i,$$

where $W_i = (W_i^1, \dots, W_i^t)$,

$$W_i^j = \max_{\substack{x_i^j \in X_i^j \\ x_i^j \neq \overline{x}_i^j}} \left\{ K_i^j(\overline{x}^j \parallel x_i^j) + \delta \sum_{l=1}^t p(j, l; \overline{x}^j \parallel x_i^j) V^l(\{i\}) \right\},$$

then in α -regularization of stochastic game G there exists Nash equilibrium with players' payoffs $(\overline{\alpha}_1, \dots, \overline{\alpha}_n)$.

4.3. Condition of irrational behavior proofness

To protect the players against losses in cases when cooperation breaks up at some stage of the game it is necessary that the following equation takes place for every $i \in N$ and any $k = 1, 2, \dots$

$$V(\{i\}) \leq E_i^{\alpha, [1, k]} + \delta^k \Pi^k(\overline{\eta}) V(\{i\}), \tag{27}$$

where $E_i^{\alpha, [1, k]}$ is the mathematical expectation of player i 's payoff at the first k stages of regularized game G_α .

We suppose that before the beginning of the next game stage players know if the cooperation has broken up or not, i.e. information delay is not supposed in such a problem definition. In the left side of inequality (27) there is the value of characteristic function $V(\{i\}) = (V^1(\{i\}), \dots, V^t(\{i\}))$ derived for player i where $V^j(\{i\})$ is the value of characteristic function of player i in subgame G^j . In the right side of inequality (27) the first term is equal to the expected value of player i 's payoff if at the first k stages of the game players keep to cooperative decision $\overline{\eta}(\cdot)$, the second term is the expected payoff of player i beginning with stage $k + 1$ if player i plays independently from this stage.

Proposition 1. *In stochastic game G_α for condition of irrational behavior proofness it is sufficient that the following inequality takes place for any $i \in N$:*

$$(\mathbb{I} - \delta \Pi(\overline{\eta}))(\alpha_i - V(\{i\})) \geq 0, \tag{28}$$

where $\alpha_i = (\alpha_i^1, \dots, \alpha_i^t)$ and α_i^j is the i th component of allocation $\alpha^j \in I^j$.

Proof. Show condition (28) is sufficient for inequality (27) for any $k = 1, 2, \dots$. We use the mathematical induction method for the proof.

Rewrite inequality (27) for $k = 1$:

$$V(\{i\}) \leq \beta_i + \delta \Pi(\bar{\eta}) V(\{i\}). \quad (29)$$

Transform inequality (28) considering definition α_i using PDP (15) and get inequality (29).

Suppose that from the truth of inequality (28) the truth of inequality (27) for $k = l$ follows. Rewrite inequality (27) for $k = l$:

$$V(\{i\}) \leq \beta_i + \dots + \delta^{l-1} \Pi^{l-1}(\bar{\eta}) \beta_i + \delta^l \Pi^l(\bar{\eta}) V(\{i\}). \quad (30)$$

Proof the statement for $k = l + 1$. Inequality (27) for $k = l + 1$ is as follows:

$$V(\{i\}) \leq \beta_i + \dots + \delta^l \Pi^l(\bar{\eta}) \beta_i + \delta^{l+1} \Pi^{l+1}(\bar{\eta}) V(\{i\}). \quad (31)$$

We should proof that if (28) is true then inequality (27) takes place for $k = l + 1$. After the transformation the right part of inequality (31) will have the following form:

$$\beta_i + \delta \Pi(\bar{\eta}) \{ \beta_i + \delta \Pi(\bar{\eta}) \beta_i + \dots + \delta^{l-1} \Pi^{l-1}(\bar{\eta}) \beta_i + \delta^l \Pi^l(\bar{\eta}) V(\{i\}) \}$$

Taking into account inequality (30) the expression in braces is not less than $V(\{i\})$, so the right part of inequality (31) is not less than $\beta_i + \delta \Pi(\bar{\eta}) V(\{i\})$. Considering the definition of PDP (16) and inequality (28) we get the truth of inequality (27) for $k = l + 1$. So the statement is proved.

5. Example

Consider the following stochastic game G :

1. The set of players is $N = \{1, 2\}$.
2. The set of states is $\{\Gamma^1, \Gamma^2\}$, where $\Gamma^j = \langle N, X_1^j, X_2^j, K_1^j, K_2^j \rangle$, $j = 1, 2$, $X_1^j = \{x_{11}^j, x_{12}^j\}$ is the set of actions of player 1, and $X_2^j = \{x_{21}^j, x_{22}^j\}$ is the set of actions of player 2. For state Γ^1 players' payoffs are determined as follows:

$$\begin{pmatrix} (2; 2) & (1; 12) \\ (11; 3) & (5; 4) \end{pmatrix}.$$

And for state Γ^2 players' payoffs are determined as follows:

$$\begin{pmatrix} (3; 1) & (2; 7) \\ (8; 2) & (5; 3) \end{pmatrix}.$$

3. Transition probabilities from state Γ^1 look like

$$\begin{pmatrix} (0.6; 0.4) & (0.7; 0.3) \\ (0.3; 0.7) & (0.6; 0.4) \end{pmatrix},$$

where element (k, l) of the matrix consists of the transition probability from state Γ^1 to states Γ^1, Γ^2 accordingly on condition that player 1 chooses k th action and player 2 chooses l th action in state Γ^1 .

Transition probabilities from state Γ^2 look like

$$\begin{pmatrix} (0.8; 0.2) & (0.3; 0.7) \\ (0.3; 0.7) & (0.2; 0.8) \end{pmatrix}.$$

4. The discount factor is $\delta = 0.99$.
5. The vector of the initial distribution on the set of states is $\pi^0 = (1/2, 1/2)$.

Determine the cooperative form G_{co} of game G described above. Firstly, calculate the cooperative decision $\bar{\eta} = (\bar{\eta}_1, \bar{\eta}_2)$ in stationary strategies using (5) and (6). We obtain the unique stationary strategy profile $\bar{\eta}_1(\Gamma^1) = x_{11}^1$, $\bar{\eta}_1(\Gamma^2) = x_{12}^2$, $\bar{\eta}_2(\Gamma^1) = x_{22}^1$, $\bar{\eta}_2(\Gamma^2) = x_{21}^2$.

Secondly, work out the values of characteristic function $V(\cdot) = (V^1(\cdot), V^2(\cdot))$ for all possible coalitions using (7)-(11):

$$V(\emptyset) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad V(\{1\}) = \begin{pmatrix} 500.00 \\ 500.00 \end{pmatrix}, \quad V(\{2\}) = \begin{pmatrix} 334.44 \\ 332.78 \end{pmatrix}, \quad V(\{1, 2\}) = \begin{pmatrix} 1152.48 \\ 1147.52 \end{pmatrix}.$$

Using (10) calculate the values of characteristic function $\bar{V}(\cdot)$ for all possible coalitions:

$$\bar{V}(\emptyset) = 0.00, \quad \bar{V}(\{1\}) = 500.00, \quad \bar{V}(\{2\}) = 333.61, \quad \bar{V}(\{1, 2\}) = 1150.00.$$

So, we determine the cooperative stochastic subgame G_{co}^j as the set $\langle N, V^j(\cdot) \rangle$, $j = 1, 2$, and cooperative stochastic game G_{co} as the set $\langle N, \bar{V}(\cdot) \rangle$.

Finally, suppose that players choose for example Shapley value (Shapley, 1953b) as allocation of their total payoff in cooperative stochastic game G_{co} and in all subgames G_{co}^j , $j = 1, 2$.

Shapley values calculated for subgames look like:

$$\alpha_1 = \begin{pmatrix} 659.02 \\ 657.37 \end{pmatrix}, \quad \alpha_2 = \begin{pmatrix} 493.46 \\ 490.15 \end{pmatrix},$$

where $\alpha_i = (\alpha_i^1, \alpha_i^2)$, and α_i^j is the i th component of Shapley value of subgame G_{co}^j using characteristic function $V^j(\cdot)$, $j = 1, 2$, $i \in N$. Then taking into account vector of initial distribution π^0 determine the allocation $\bar{\alpha}$ in G_{co} by Definition 8:

$$\bar{\alpha} = (\bar{\alpha}_1, \bar{\alpha}_2) = (658.20, 491.80).$$

Verify the principles of stable cooperation. To satisfy the principle of subgame consistency we should calculate the PDP for the allocation $\bar{\alpha}$ equals to $\pi^0 \alpha$ using (16):

$$\beta_1 = \begin{pmatrix} 7.08 \\ 6.08 \end{pmatrix}, \quad \beta_2 = \begin{pmatrix} 5.92 \\ 3.92 \end{pmatrix}.$$

As we can see not all the components of PDP are equal to the corresponding payoffs of players in the states. So if players get their payoffs determined by the initial rules of the game they can't receive the components of the chosen allocation (in the mathematical expectation sense). It tells about subgame inconsistency of chosen cooperative agreement. Realize α -regularization of initial stochastic game G using PDP and Definition 3. In α -regularization of game G players' payoffs in state Γ^1 are as follows:

$$\begin{pmatrix} (2; 2) & (7.08; 5.92) \\ (11; 3) & (5; 4) \end{pmatrix},$$

and in state Γ^2 players' payoffs look like:

$$\begin{pmatrix} (3; 1) & (2; 7) \\ (5.92; 3.92) & (5; 3) \end{pmatrix}.$$

Check the second principle of stable cooperation which is strategic stability of cooperative agreement. For this purpose verify the truth of inequality (20) from Theorem 1. Compute $F(S)$ for all $S \subset N$ and obtain that $F(\{1\}) = (500.00, 498.00)$, $F(\{2\}) = (332.44, 332.78)$. Then inequality (20) is equivalent to the following ones:

$$\beta_1 = \begin{pmatrix} 7.08 \\ 6.08 \end{pmatrix} \geq \begin{pmatrix} 5.59 \\ 4.39 \end{pmatrix} = (\mathbb{I} - \delta\Pi(\bar{\eta}))F(\{1\}),$$

$$\beta_2 = \begin{pmatrix} 5.92 \\ 3.92 \end{pmatrix} \geq \begin{pmatrix} 3.22 \\ 3.43 \end{pmatrix} = (\mathbb{I} - \delta\Pi(\bar{\eta}))F(\{2\}).$$

They are true and we can say about strategic stability of the players' chosen cooperative agreement.

Verify the third principle condition of irrational behavior proofness. The sufficient condition for this principle from Proposition 1 holds as you can see here:

$$(\mathbb{I} - \delta\Pi(\bar{\eta}))(\alpha_1 - V(\{1\})) = \begin{pmatrix} 2.08 \\ 1.08 \end{pmatrix} \geq 0,$$

$$(\mathbb{I} - \delta\Pi(\bar{\eta}))(\alpha_2 - V(\{2\})) = \begin{pmatrix} 2.08 \\ 1.08 \end{pmatrix} \geq 0.$$

For this numerical example we have made the α -regularization of the initial game G , and checked the principles of stable cooperation, they all are satisfied.

References

- Altman, E., El-Azouzi, R. and Jimenez, T. (2003). *Slotted Aloha as a stochastic game with partial information*. Proceedings of Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networks (WiOpt '03), Sophia Antipolis, France.
- Amir, R. (2003). *Stochastic games in economics: The latter-theoretic approach*. Stochastic Games and Applications in A. Neyman and S. Sorin (eds.), NATO Science Series C, Mathematical and Physical Sciences, Vol. 570, Chap. 29, 443–453.
- Bellman, R. (1957). *Dynamic programming*. Princeton University Press, Princeton, NJ.
- Dutta, P. (1995). *A folk theorem for stochastic games*. Journal of Economic Theory **66**, 1–32.
- Grauer, L. V. and Petrosjan, L. A. (2002). *Strong Nash Equilibrium in Multistage Games*. International Game Theory Review, **4(3)**, 255–264.
- Herings, P. J.-J. and Peeters, R. J. A. P. (2004). *Stationary Equilibria in Stochastic Games: Structure, Selection, and Computation*. Journal of Economic Theory, Vol. 118, No. 1, 32–60.
- Parilina, E. M. (2010). *Cooperative data transmission game in wireless network*. Upravlenie bol'simi sistemami, **31.1**, 93–110 (in Russian).
- Petrosyan, L. A. (1977). *Stability of the solutions in differential games with several players*. Vestnik Leningrad. Univ. Mat. Mekh. Astronom, **19(4)**, 46–52 (in Russian).
- Petrosjan, L. A. (2006). *Cooperative Stochastic Games*. Advances in dynamic games, Annals of the International Society of Dynamic Games, Vol. 8, Part VI, 139–146.
- Petrosjan, L. A. and Baranova, E. M. (2006). *Cooperative Stochastic Games in Stationary Strategies*. Game Theory and Applications, **XI**, 7–17.
- Petrosjan, L. A. and Danilov, N. N. (1979). *Stability of the solutions in nonantagonistic differential games with transferable payoffs*. Vestnik Leningrad. Univ. Mat. Mekh. Astronom. **1**, 52–59.
- Petrosyan, L. A. and Zenkevich, N. A. (2009). *Principles of dynamic stability*. Upravlenie bol'simi sistemami, **26.1**, 100–120 (in Russian).

- Petrosyan, L. A., Baranova, E. M., Shevkoplyas, E. V. (2004). *Multistage Cooperative Games with Random Duration*. Optimal control and differential games. Proceedings of Institute of mathematics and mechanics, **10(2)**, 116–130 (in Russian).
- Raghavan, T. E. S. (2006). *A Stochastic Game Model of Tax Evasion*. Advances in dynamic games, Annals of the International Society of Dynamic Games, 2006, Vol. 8, Part VI, 397–420.
- Shapley, L. S. (1953a). *Stochastic Games*. Proceedings of National Academy of Sciences of the USA, **39**, 1095–1100.
- Shapley, L. S. (1953b). *A Value for n -person Games*. Contributions to the Theory of Games-II, by H.W. Kuhn and A.W. Tucker, editors. (Annals of Mathematical Studies, Vol. 28), Princeton University Press, 307–317.
- Yeung, D. W. K. (2006). *An irrational-behavior-proof condition in cooperative differential games*. International Game Theory Review, **8(4)**, 739–744.