

УДК 519.83

ББК 22.18

# ПРИБЛИЖЕННОЕ РАВНОВЕСИЕ В КОНЕЧНО ПОВТОРЯЮЩЕЙСЯ ИГРЕ «ДИЛЕММА ЗАКЛЮЧЕННОГО»

АЛЁНА М. ПИСАРЕВА

ЕЛЕНА М. ПАРИЛИНА\*

Санкт-Петербургский государственный университет  
199034, Санкт-Петербург, Университетская наб., 7-9  
e-mail: st055836@student.spbu.ru, e.parilina@spbu.ru

В статье изучается конечно повторяющаяся «Дилемма заключенного». Для поддержания кооперации в игре предлагается новая ситуация в стратегиях поведения, при которой отклонение игроков наказывается не до конца игры, а на период заданной продолжительности в зависимости от шага игры. Доказано существование приближенного равновесия в данных стратегиях, и найдено максимальное значение выигрыша игрока при отклонении от данной ситуации приближенного равновесия.

*Ключевые слова:* повторяющиеся игры, дилемма заключенного, приближенное равновесие, стратегия наказания.

*Поступила в редакцию:* 09.01.24 *После доработки:* 02.03.24 *Принята к публикации:* 01.06.24

## 1. Введение

Повторяющиеся игры являются предметом исследования специалистов в области теории игр, поскольку являются упрощенной моделью динамических игр, для которых удастся получить интересные

---

©2024 А.М. Писарева, Е.М. Парилина

\* Работа второго автора поддержана грантом Российского научного фонда № 22-11-00051, <https://rscf.ru/project/22-11-00051/>

результаты, включая равновесия, в явном виде. Важную роль в этой области играют народные теоремы, которые описывают равновесия по Нэшу при разных предположениях о поведении игроков [5, 7, 18, 19]. Идеи, лежащие в основе народных теорем, лучше всего можно понять, рассмотрев сначала самый простой вариант народной теоремы для бесконечно повторяющихся игр, когда игроки имеют полную информацию о том, что выбирали игроки на предыдущих шагах: если игроки достаточно высоко оценивают будущие выигрыши (достаточно терпеливы или, что тоже самое, имеют большой дисконтирующий фактор), тогда стратегии, приносящие любую комбинацию выигрышей, текущая ценность которых не ниже, чем выигрыши в равновесии по Нэшу в одношаговой игре, могут формировать равновесие в бесконечно повторяющейся игре [12]. Равновесность такого набора стратегий удастся обеспечить за счет реализации стратегий наказания.

В нашей работе рассматривается конечно повторяющаяся игра «Дилемма заключенного», которая привлекает внимание исследователей как простейшая модель, в которой сотрудничество или кооперация обоих игроков не является равновесием по Нэшу, и единственное равновесие по Нэшу состоит в том, что оба игрока отказываются от сотрудничества. Как говорилось выше, народная теорема утверждает, что в бесконечно повторяющейся версии «Дилеммы заключенного», при условии, что игроки достаточно терпеливы (коэффициент дисконтирования их выигрышей стремится к единице), существует равновесие по Нэшу, такое, что оба игрока взаимодействуют, и это взаимодействие обеспечивается стратегической поддержкой, т.е. является равновесием [8, 13]. В работе [9] доказывается существование стратегий, обеспечивающих равновесие, устойчивое относительно отклонения группы игроков от кооперации в случае повторяющихся и многошаговых игр многих лиц. Трудности поддержки кооперации возникают, если игра имеет конечное число повторений. Для этого случая можно доказать, что единственным абсолютным равновесием по Нэшу в повторяющейся игре является разыгрывание равновесия по Нэшу одношаговой игры на каждом шаге [12].

Поскольку кооперативное поведение игроков не удастся стратегически поддержать в конечно повторяющихся играх, то исследователи

определяют  $\varepsilon$ -равновесие или приближенное равновесие [14], которое представляет собой набор стратегий, при индивидуальном отклонении от которого игроки не могут увеличить свой выигрыш на величину, большую чем  $\varepsilon$ . В равновесии по Нэшу ни у одного игрока нет стимула изменить свое поведение, а в приближенном равновесии по Нэшу это требование ослаблено, чтобы допустить возможность того, что у игрока может быть небольшой стимул к отклонению (получение  $\varepsilon$ ) [6, 16]. Если обратиться к конечной  $T$ -повторяющейся игре «Дилемма заключенного», где выигрыш усредняется за  $T$  периодов, то профиль стратегий типа «око за око» (tit-for-tat) или «абсолютное наказание» (grim trigger) (см. [7]) хотя и не является равновесием по Нэшу в повторяющейся игре, но является  $\varepsilon$ -равновесием для некоторого положительного  $\varepsilon$ . Допустимые значения отклонения зависят от матрицы выигрышей в одношаговой игре и от количества повторений этой игры [17]. Новые равновесия в стратегиях наказания для повторяющихся игр «Дилемма заключенного», разыгрываемых на сетях, предложены в работе [10], а в статье [11] поддержка кооперативного поведения осуществляется за счет перераспределения выплат между игроками.

Абсолютное наказание, которое длится с шага, следующего за отклонением и до конца игры (grim trigger strategy profile), нельзя назвать хорошей мерой поддержки кооперации с экономической точки зрения. В повторяющейся игре «Дилемма заключенного» постоянный выбор стратегии предательства (некооперации) после того, как какой-то игрок однажды отклонился от кооперации, можно назвать хорошим подходом для определения наказания противника, полностью отказывающегося от сотрудничества, но это не лучший способ определения наказания в игре с теми игроками, которые могут время от времени предавать [4]. В связи с этим, определение стратегий наказания, которые позволяют наказывать отклонившихся игроков не навсегда, а на определенный временной промежуток, кажутся разумными. В этой статье делается шаг в этом направлении и определяется новая ситуация равновесия в стратегиях с ограниченным по времени наказанием.

В данной работе предлагаются новые ситуации приближенного равновесия, поддерживающие кооперативное поведение игроков в ко-

нечно повторяющейся игре «Дилемма заключенного». Делаются предположения о поведении игроков, которые стали основой для ограничений на продолжительность наказания игрока при его отклонении от кооперации. В статье определяются стратегии игроков, где реализовано определенное нами наказание. Чаще всего, в схожих работах рассматривается наказание, которое не ограничено по времени (длиться до окончания игры) [13, 15], в нашей же работе длительность наказания зависит от шага, на котором произошло отклонение. Доказана теорема о существовании  $\varepsilon$ -равновесия в таких стратегиях и приведен численный пример.

Статья имеет следующую структуру. В разделе 2 дано описание повторяющейся игры. Определение стратегий наказания в повторяющейся игре приводится в разделе 3, и существование ситуации  $\varepsilon$ -равновесия по Нэшу в заданных стратегиях наказания доказывается в разделе 4. Теоретические результаты демонстрируются на численном примере, приведенном в разделе 5. Далее дается заключение к работе (раздел 6), где описываются основные результаты и будущие направления работы.

## 2. Постановка задачи

Рассмотрим игру «Дилемма заключенного» двух лиц, которая повторяется в периоды из множества  $\mathbf{T} = \{1, \dots, T\}$ , которые также будем называть «шагами». Введем несколько определений для описания повторяющейся игры. Для повторяющихся игр *базовой игрой* называется игра, разыгрываемая на каждом шаге. В данной работе на каждом шаге повторяющейся игры разыгрывается базовая игра в нормальной форме типа «Дилеммы заключенного», где каждый игрок имеет две стратегии:  $C$  — «довериться» и  $N$  — «предать» и матрица выигрышей имеет вид:

|     | $C$      | $N$      |
|-----|----------|----------|
| $C$ | $(a, a)$ | $(c, b)$ |
| $N$ | $(b, c)$ | $(d, d)$ |

где  $b > a > d > c$  и  $2a > c + b$ .

**Определение 2.1.** В одношаговой игре в нормальной форме  $\Gamma = \{I, \{S_i\}, \{J_i\}\}$ , где  $I$  — множество игроков,  $S_i$  — множество стратегий  $i$ -го игрока,  $J_i$  — функция выигрыша игрока  $i$ , ставящая в

соответствие каждой ситуации (набору стратегий игроков) выигрыш этого игрока, равновесием по Нэшу называется ситуация  $(s_i : i \in I)$  такая, что для  $\forall i \in I$  выполнено:

$$J_i(s_i, s_{-i}) \geq J_i(s'_i, s_{-i})$$

для  $\forall s'_i \in S_i$  [17].

Иными словами, если игрок в одиночку решит отклониться от своей стратегии в равновесии по Нэшу, то он не увеличит свой выигрыш.

**Определение 2.2.** Стратегия  $s_i \in S_i$  игрока  $i$  строго доминируется в игре  $\Gamma$ , если существует другая стратегия  $s'_i \in S_i$  такая, что для всех  $s_{-i} \in S_{-i}$  выполнено

$$J_i(s'_i, s_{-i}) > J_i(s_i, s_{-i}).$$

Для обоих игроков стратегия  $N$  строго доминирует стратегию  $C$  в одношаговой игре «Дилемма заключенного», поэтому единственным равновесием базовой игры является ситуация  $(N, N)$ .

Теперь определим конечно повторяющуюся игру, в которой выигрыши игроков на каждом шаге суммируются с учетом коэффициента дисконтирования  $\rho \in (0, 1)$ , одинакового для всех игроков.

**Определение 2.3.** [2] Конечная  $T$ -повторяющаяся игра базовой игры  $\Gamma$  с коэффициентом дисконтирования  $\rho$  — это игра, состоящая из повторения на каждом из  $T$  шагов базовой игры, когда выигрыш игрока в этой игре определяется как дисконтированная сумма выигрышей этого игрока в базовой игре.

**Определение 2.4.** [2] Ситуация равновесия по Нэшу в конечной  $T$ -повторяющейся игре — набор стратегий игроков в повторяющейся игре такой, что ни один из игроков не может увеличить свой выигрыш, индивидуально отклонившись от своей стратегии в равновесии.

**Определение 2.5.** [1, 2] Ситуация абсолютного равновесия в конечной  $T$ -повторяющейся игре — набор стратегий игроков в повторяющейся игре такой, что для любого шага  $t \in \mathbf{T}$  сужение ситуа-

*ции на подыгру, начинающуюся с шага  $t$ , является равновесием по Нэшу в этой подыгре.*

В теории повторяющихся игр [2] известен результат о том, что если базовая игра имеет единственное равновесие по Нэшу, то  $T$ -повторяющаяся игра для любого конечного  $T$  имеет единственное абсолютное равновесие, при котором на каждом шаге разыгрывается равновесие по Нэшу в базовой игре [12]. Таким образом, в  $T$ -повторяющейся игре двух лиц «Дилемма заключенного» существует равновесие по Нэшу, т.е. ситуация или набор стратегий игроков, когда оба игрока выбирают стратегию  $N$  на любом шаге  $t \in T$ .

### **3. Определение ситуации в стратегиях наказания в $T$ -повторяющейся игре**

Кооперативное поведение игроков, т. е. выбор ими стратегии  $C$  на каждом шаге, является социально привлекательным, поскольку приносит максимальный суммарный выигрыш игрокам. Но так как такое поведение игроков не является равновесным, определим стратегию поведения игроков в  $T$ -повторяющейся игре, которая позволит реализовать наказание игроков в случае их отклонения от кооперативной стратегии. Введем дополнительные правила, которые применяются в случае отклонений игроков от кооперативного поведения:

A1 Если один из игроков впервые отклонится от кооперативного поведения (стратегии  $C$ ) на шаге  $t$ , то он будет наказан на период с заданной продолжительностью  $h(t)$ , которая известна всем игрокам. Наказание заключается в разыгрывании равновесных по Нэшу стратегий (стратегия  $N$ ) на каждом шаге всего периода наказания.

A2 Если наблюдается второе отклонение от кооперативного поведения, то на следующем после этого отклонения шаге игра будет разыгрываться в некооперативном варианте на каждом шаге до своего окончания.

На основе этих правил сформулируем условия нахождения длительности наказания  $h$ :

Б1 Игрок, не отклоняющийся от кооперативного поведения, когда другой игрок от него отклоняется на шаге  $t + 1$ , получает дисконтированный выигрыш не меньше, чем выигрыш, который он может получить, сам отклонившись от кооперации сразу после окончания наказания.

Б2 Наказание должно длиться столько, чтобы выигрыш игрока при отклонении не превышал бы выигрыш, получаемый им при кооперации на всех шагах игры.

Дополнительно наложим ограничение [13, 15]:

$$\frac{b - a}{b - d} < \rho. \quad (3.1)$$

*Замечание 3.1.* Неравенство (3.1) обычно требуется в теории повторяющихся игр, и оно получается из условия, что выигрыш игрока при кооперации на каждом шаге в бесконечно повторяющейся игре «Дилемма заключенного» больше, чем выигрыш при отклонении от кооперации на данном шаге, когда начиная со следующего после отклонения шага все игроки переключаются на бесконечный выбор равновесных по Нэшу стратегий, т.е.

$$\begin{aligned} a + \rho a + \rho^2 a + \dots &> b + \rho d + \rho^2 d + \dots, \\ \frac{\rho}{1 - \rho}(a - d) &> b - a, \\ \rho(b - d) &> b - a, \end{aligned}$$

что эквивалентно неравенству (3.1).

**Лемма 3.1.** [3] Пусть выполнено условие (3.1). Продолжительность наказания  $h(t + 1)$  игрока, отклонившегося от кооперации на шаге  $t + 1$ , удовлетворяющая условиям Б1, Б2, и учитывая правила А1, А2, удовлетворяет системе неравенств:

$$\begin{cases} h(t + 1) \leq T - t - 1 - \log_{\rho} \left( \frac{a + b(\rho - 1) - d\rho}{a - d} \right), \\ h(t + 1) \geq \log_{\rho} \left( \frac{a + b(\rho - 1) - d\rho}{a - d} \right) - 1, \end{cases} \quad (3.2)$$

если выполнено:

$$T - t \geq 2 \log_{\rho} \left( \frac{a + b(\rho - 1) - d\rho}{a - d} \right). \quad (3.3)$$

Если же условие (3.3) не выполнено, то после отклонения на шаге  $t + 1$  игрока невозможно наказать, удовлетворив указанные условия и правила, и игроки больше не кооперируются и до конца игры выбирают стратегию  $N$  на каждом шаге.

*Доказательство.* Не умаляя общности, будем считать, что второй игрок решил отклониться в момент времени  $t + 1$ . Пусть  $h$  – количество раундов, которое будет длиться наказание.

Рассчитаем, сколько получит первый игрок за всю игру:

$$S_1 = \frac{a(1 - \rho^t)}{1 - \rho} + c\rho^t + \frac{d\rho^{t+1}(1 - \rho^h)}{1 - \rho} + \frac{a\rho^{t+1+h}(1 - \rho^{T-h-t-1})}{1 - \rho}.$$

Второй игрок при этом получит

$$S_2 = \frac{a(1 - \rho^t)}{1 - \rho} + b\rho^t + \frac{d\rho^{t+1}(1 - \rho^h)}{1 - \rho} + \frac{a\rho^{t+1+h}(1 - \rho^{T-h-t-1})}{1 - \rho}.$$

Рассмотрим условие B1. Напомним, что следуя правилу A2, после того, как один игрок отклонится, то любое второе отклонение в игре приведет к тому, что кооперации не будет до конца игры. Первый игрок, при этом желая максимизировать свой выигрыш, может захотеть отклониться сразу после окончания наказания (ради получения мгновенной прибыли за счет коэффициента дисконтирования будущих платежей). Предположение B1 делает этот шаг невыгодным. Условие B1 можно записать в виде:

$$\begin{aligned} \frac{a\rho^{t+1+h}(1 - \rho^{T-h-t-1})}{1 - \rho} &\geq b\rho^{t+1+h} + \frac{d\rho^{t+2+h}(1 - \rho^{T-h-t-2})}{1 - \rho}, \\ \frac{1 - \rho^{T-h-t-1}}{1 - \rho} &\geq \frac{b - d}{a - d} = k, \end{aligned} \quad (3.4)$$

где  $k > 1$ , так как  $b > a$ .

Запишем условие (3.1) через параметр  $k$ :

$$\frac{k - 1}{k} < \rho < 1.$$

Выражая  $h$  из (3.4), получаем

$$h \leq T - t - 1 - \log_\rho(k\rho - k + 1).$$

Заменяя  $k$ , получим первое неравенство из системы (3.2).

Условие Б2 можно записать следующим образом:

$$\begin{aligned} & \frac{a(1-\rho^t)}{1-\rho} + b\rho^t + \frac{d\rho^{t+1}(1-\rho^h)}{1-\rho} + \frac{a\rho^{t+1+h}(1-\rho^{T-h-t-1})}{1-\rho} \leq \\ & \leq \frac{a(1-\rho^t)}{1-\rho} + a\rho^t + \frac{a\rho^{t+1}(1-\rho^h)}{1-\rho} + \frac{a\rho^{t+1+h}(1-\rho^{T-h-t-1})}{1-\rho}, \end{aligned}$$

после преобразования получаем:

$$(b-a)\rho^t \leq \frac{(a-d)\rho^{t+1}(1-\rho^h)}{1-\rho},$$

следовательно,  $h \geq \log_\rho(g\rho + \rho - g) - 1$ , где  $g = \frac{b-a}{a-d}$ . Таким образом, получили второе неравенство из системы (3.2), что завершает доказательство.  $\square$

Введем обозначение:

$$K = 2 \log_\rho \left( \frac{a + b(\rho - 1) - d\rho}{a - d} \right). \quad (3.5)$$

Сначала определим историю шага в повторяющейся игре, далее — стратегию поведения в рассматриваемой игре.

**Определение 3.1.** *История  $\mathcal{H}(t)$  шага  $t$  в  $T$ -повторяющейся игре — это вектор размерности  $t$  такой, в котором последовательно записаны ситуации, которые разыгрывались на шагах  $1, \dots, t$  в базовой игре.*

В соответствии с правилами А1 и А2 и учитывая лемму 3.1, построим стратегию поведения, которая определяет стратегию в базовой игре на каждом шаге  $t + 1$  в зависимости от истории  $\mathcal{H}(t)$ , т.е. от последовательности векторов выбранных игроками стратегий в базовых играх, разыгранных на шагах  $1, \dots, t$ .

Введем обозначения:

- $C_{[i,k]}$  — история шагов с  $i$  по  $k$  такая, что игроки кооперируются на этих шагах, т.е. оба игрока выбирают стратегию  $C$ , начиная с шага  $i$  по шаг  $k$  включительно;

- $N_{[i,k]}$  — история шагов с  $i$  по  $k$  такая, что игроки не кооперируются, т.е. оба игрока выбирают стратегию  $N$ , начиная с шага  $i$  по шаг  $k$  включительно;
- $V_{[i]}$  (от victim — «меня предали») — один игрок кооперируется (выбирает стратегию  $C$ ), в то время как другой игрок предаёт (выбирает  $N$ ) на шаге  $i$ , т.е. это ситуация  $(C, N)$  в базовой игре на шаге  $i$ ;
- $B_{[i]}$  (от betrayer — «я предал») — один игрок не кооперируется (выбирает стратегию  $N$ ), в то время как другой игрок кооперируется (выбирает стратегию  $C$ ) на шаге  $i$ , т.е. это ситуация  $(N, C)$  в базовой игре на шаге  $i$ .

Теперь определим стратегию поведения следующим образом:

$$\phi(t+1, \mathcal{H}(t)) = \begin{cases} C, \text{ если } \mathcal{H}(t) = C_{[1,t]} \text{ или } \mathcal{H}(t) = \emptyset, \\ \\ C, \text{ если } T - m + 1 \geq K \text{ и} \\ \mathcal{H}(t) = (C_{[1,m-1]} V_{[m]} N_{[m+1, m+1+h(m)]}) \text{ или} \\ \mathcal{H}(t) = (C_{[1,m-1]} B_{[m]} N_{[m+1, m+1+h(m)]}) , \\ \text{ где } t = m + h(m) + 1, \\ \\ C, \text{ если } T - m + 1 \geq K \text{ и} \\ \mathcal{H}(t) = (C_{[1,m-1]} V_{[m]} N_{[m+1, m+1+h(m)]} C_{[m+2+h(m), t]}) \text{ или} \\ \mathcal{H}(t) = (C_{[1,m-1]} B_{[m]} N_{[m+1, m+1+h(m)]} C_{[m+2+h(m), t]}) , \\ \text{ где } t > m + h(m) + 1, \\ \\ N, \text{ в остальных случаях.} \end{cases} \quad (3.6)$$

Иначе говоря, стратегия (3.6) предписывает игроку следующие действия:

- Если на всех предыдущих шагах наблюдалась обоюдная кооперация, то нужно выбирать стратегию  $C$  (кооперироваться), также если это первый шаг игры (первый случай в правой части (3.6)),

- Если только что закончилось наказание, то следует выбирать кооперацию (второй случай в правой части (3.6)),
- Если в истории наблюдалось отклонение и последующее наказание, которое закончилось, и на предыдущем шаге была обоюдная кооперация, то нужно выбирать стратегию  $C$  (третий случай в правой части (3.6)),
- Во всех остальных случаях нужно выбрать  $N$  (не кооперироваться) (последний случай в правой части (3.6)).

*Замечание 3.2.* Согласно лемме 3.1 продолжительность наказания, удовлетворяющая неравенствам (3.2), может быть неединственной. В случае, когда можно выбрать разную продолжительность наказания, мы не уточняем правило, по которому это делается. Важным является то, что все игроки знают значение  $h(t)$  для любого  $t$ . В разделе 5, где рассмотрен численный пример,  $h(t)$  выбирается из найденного промежутка произвольно. Очевидно, что чем больше продолжительность наказания, тем меньше выигрыши игроков в случае, если они отклоняются на каком-то шаге от кооперативного поведения.

#### 4. Приближенное равновесие в конечно повторяющейся игре

Определим приближенное или  $\varepsilon$ -равновесие в  $T$ -повторяющейся игре, затем докажем его существование.

**Определение 4.1.** Набор стратегий  $\hat{\sigma}$  в конечной повторяющейся игре — (абсолютное)  $\varepsilon$ -равновесие, если для любого игрока  $i \in I$ , любой истории  $\mathcal{H}$  и стратегии поведения  $\sigma_i$  (возможной при этой истории) выполнено

$$U_i(\hat{\sigma}^{\mathcal{H}}) \geq U_i(\sigma_i^{\mathcal{H}}, \hat{\sigma}_{-i}^{\mathcal{H}}) - \varepsilon,$$

$U_i$  — выигрыш игрока  $i$  в конечно повторяющейся игре,  $\sigma^{\mathcal{H}}$  — стратегия игрока  $i$  в конечно повторяющейся игре, которая зависит от истории  $\mathcal{H}$ . [12]

**Теорема 4.1.** Набор стратегий (3.6) является  $\varepsilon$ -равновесием в  $T$ -повторяющейся игре «Дилемма заключенного» двух лиц с дополни-

тельными условиями A1, A2, B1, B2, для любого  $\varepsilon \geq \hat{\varepsilon}$ , где

$$\hat{\varepsilon} = (b - a)\rho^{T-1}. \quad (4.1)$$

*Доказательство.* Так как игроки симметричны, то, не умаляя общности, будем рассматривать случай, когда игрок 2 отклоняется от кооперации. Если оба игрока будут придерживаться стратегии (3.6), то выигрыш каждого составит

$$S_0 = \frac{a(1 - \rho^T)}{1 - \rho}.$$

Доказательство можно разбить на два случая:

1. произошло одно отклонение от стратегии (3.6) в течение игры,
2. произошло не менее двух отклонений от стратегии (3.6) (после второго отклонения игроки по правилу A2 больше не кооперируются).

Рассмотрим первый случай. Его также можно разделить на два пункта:

- а) Если игрок отклонился только один раз в момент времени  $\tau + 1$ , такой что  $T - \tau \leq K$ , тогда начиная с шага  $\tau + 2$  он получит наказание продолжительностью  $h(\tau + 1)$  и его суммарный выигрыш составит

$$S_1 = \frac{a(1 - \rho^\tau)}{1 - \rho} + b\rho^\tau + \frac{d\rho^{\tau+1}(1 - \rho^h)}{1 - \rho} + \frac{a\rho^{\tau+1+h}(1 - \rho^{T-\tau-1-h})}{1 - \rho}.$$

Вычислим прибыль, которую получит второй игрок при отклонении:

$$\begin{aligned} S_1 - S_0 &= b\rho^\tau + \frac{d\rho^{\tau+1}(1 - \rho^h)}{1 - \rho} - a\rho^\tau - \frac{a\rho^{\tau+1}(1 - \rho^h)}{1 - \rho} = \\ &= (b - a)\rho^\tau + (d - a)\frac{\rho^{\tau+1}(1 - \rho^h)}{1 - \rho}, \end{aligned}$$

которая не положительна по условию B2 и поскольку  $h$  удовлетворяет условию (3.2).

Таким образом, игроку невыгодно отклоняться от кооперации на шаге  $\tau + 1$ , таком что  $T - \tau \leq K$ .

- б) Если игрок отклонился только один раз в момент времени  $\tau + 1$ , такой что  $T - \tau > K$ , то в этом случае наказание продолжительностью

$h(\tau + 1)$  невозможно реализовать и согласно стратегии (3.6) игроки переходят на разыгрывание стратегии  $N$  до конца игры. Суммарный выигрыш за всю игру отклонившегося игрока будет равен

$$S_2 = \frac{a(1 - \rho^\tau)}{1 - \rho} + b\rho^\tau + \frac{d\rho^{\tau+1}(1 - \rho^{T-\tau-1})}{1 - \rho}.$$

Дополнительную прибыль, которую получит второй игрок в этом случае, равна

$$\varepsilon = S_2 - S_1 = b\rho^\tau + \frac{d\rho^{\tau+1}(1 - \rho^{T-\tau-1})}{1 - \rho} - a\rho^\tau - \frac{a\rho^{\tau+1}(1 - \rho^{T-\tau-1})}{1 - \rho}.$$

Найдем максимальное значение  $\varepsilon$ , которое игрок может получить, выбирая шаг  $\tau$  такой, что  $\tau \leq T - 1$  и  $\tau > T - K$ :

$$\begin{aligned} \varepsilon &= (b - a)\rho^\tau - (a - d)\frac{\rho^{\tau+1}}{1 - \rho}(1 - \rho^{T-\tau-1}) \\ &= \frac{\rho^T}{1 - \rho}(a - d) + (b - a)\rho^\tau - \frac{\rho^{\tau+1}}{1 - \rho}(a - d) \rightarrow \max_{\substack{\tau: \tau \leq T-1, \\ \tau > T-K}} \end{aligned}$$

Первое слагаемое в последнем выражении не зависит от  $\tau$  и на максимум не влияет, потому рассмотрим выражение  $(b - a)\rho^\tau - \frac{\rho^{\tau+1}}{1 - \rho}(a - d)$ , сделав следующие преобразования:

$$\begin{aligned} (b - a)\rho^\tau - \frac{\rho^{\tau+1}}{1 - \rho}(a - d) &= \frac{\rho^{\tau+1}}{1 - \rho} \left( b - a - \rho(b - d) \right) = \\ &= \frac{\rho^\tau}{1 - \rho} (b - d) \left( \frac{b - a}{b - d} - \rho \right) \rightarrow \max_{\substack{\tau: \tau \leq T-1, \\ \tau > T-K}} \end{aligned}$$

Так как  $\frac{(b-d)}{1-\rho} \left( \frac{b-a}{b-d} - \rho \right) < 0$ , поскольку выполняется неравенство (3.1), то максимум достигается при минимальном значении  $\rho^\tau$ , и так как  $\rho \in (0, 1)$ , то  $\tau$  требуется выбрать равным максимальному допустимому значению, т. е.  $\varepsilon$  достигает максимальное значение при  $\tau = T - 1$ :

$$\begin{aligned} \max_{\substack{\tau: \tau \leq T-1, \\ \tau > T-K}} \varepsilon &= \frac{\rho^T}{1 - \rho}(a - d) + (b - a)\rho^{T-1} - \frac{\rho^{T-1+1}}{1 - \rho}(a - d) \\ &= (b - a)\rho^{T-1} > 0. \end{aligned}$$

Подводя итог по первому случаю, делаем вывод, что максимальная прибыль при индивидуальном отклонении игрока равна  $(b - a)\rho^{T-1}$  и она достигается, если он отклоняется на шаге  $T$ .

Теперь рассмотрим второй случай, когда игрок отклонится за игру не менее двух раз. Напомним, что согласно дополнительному правилу А2, после второго отклонения игроки не будут больше кооперироваться до конца игры. Таким образом, если игрок захочет отклониться строго больше, чем два раза, то он только потеряет, так как отклонением будет считаться выбор стратегии  $C$ , в то время как другой игрок придерживается стратегии (3.6), отклонившийся игрок получит  $c$  вместо  $d$  на этих этапах игры, и на последующих шагах (после того, как уже произошли два отклонения). Больше, чем  $d$ , игрок не получит, если будет отклоняться в третий и последующие разы, поскольку будет разыгрываться равновесие по Нэшу на каждом шаге, отклонение от которого невыгодно по определению. Поэтому можно ограничиться рассмотрением случая с двумя отклонениями, так как только в этом варианте во втором случае игрок может получить прибыль от отклонения.

Если первое отклонение произошло в момент  $t + 1$ , такой что  $T - t > K$ , то после этого согласно (3.6) игроки реализуют некооперативное поведение, во время которого игрокам невыгодно отклоняться в индивидуальном порядке. Тогда для первого отклонения в  $t + 1$  справедливо, что  $T - t \leq K$ , а второе отклонение, рассуждая аналогично, может произойти после  $h(t + 1)$  шагов наказания, т. е. начиная с шага  $t + 1 + h + 1$ . Опишем последовательно реализацию этой ситуации:

1. Первое отклонение игрока 2 произошло в момент  $t_1 + 1$  такой, что  $T - t_1 \leq K$ , после этого наказание реализуется на протяжении шагов  $t_1 + 2, \dots, t_1 + 1 + h$ .
2. Второе отклонение игрока 2 произошло в момент  $t_2 + 1$  такой, что  $t_2 - h - t_1 - 1 > 0$  и  $t_2 \leq T - 1$ .

В этом случае игрок 2 получит выигрыш, равный

$$S_3 = \frac{a(1 - \rho^{t_1})}{1 - \rho} + b\rho^{t_1} + \frac{d\rho^{t_1+1}(1 - \rho^h)}{1 - \rho} + \frac{a\rho^{t_1+1+h}(1 - \rho^{t_2-t_1-1-h})}{1 - \rho}$$

$$+ b\rho^{t_2} + \frac{d\rho^{t_2+1}(1 - \rho^{T-t_2-1})}{1 - \rho}.$$

Дополнительная прибыль, которую получит второй игрок при отклонении от кооперативного поведения, равна:

$$\begin{aligned} \varepsilon = S_3 - S_0 &= (b - a)\rho^{t_1} + \frac{\rho^{t_1+1}(1 - \rho^h)}{1 - \rho}(d - a) \\ &+ (b - a)\rho^{t_2} + \frac{\rho^{t_2+1}(1 - \rho^{T-t_2-1})}{1 - \rho}(d - a). \end{aligned}$$

Найдем максимум этого выражения по  $t_1$  и  $t_2$  при заданных на них ограничениях.

Пусть  $\varepsilon = \mathcal{I}_1 + \mathcal{I}_2$ , где

$$\begin{aligned} \mathcal{I}_1 &= (b - a)\rho^{t_1} + \frac{\rho^{t_1+1}(1 - \rho^h)}{1 - \rho}(d - a), \\ \mathcal{I}_2 &= (b - a)\rho^{t_2} + \frac{\rho^{t_2+1}(1 - \rho^{T-t_2-1})}{1 - \rho}(d - a). \end{aligned}$$

По условию Б2 имеем  $\mathcal{I}_1 \leq 0$ , значит,  $\varepsilon = \mathcal{I}_2 - |\mathcal{I}_1| \leq \mathcal{I}_2$ .

Выражение  $\mathcal{I}_2$  было проанализировано в пункте б) первого случая, и оно принимает максимальное значение при  $t_2 = T - 1$ , и это означает, что  $\varepsilon \leq (b - a)\rho^{T-1}$ .

Таким образом, максимальную прибыль, которую второй игрок может получить во втором случае (не менее двух отклонений) не превысит его максимальную прибыль при одном отклонении. Следовательно, отклонившись от стратегии (3.6), игрок не получит при отклонении больше, чем  $(b - a)\rho^{T-1}$ , что и требовалось доказать.  $\square$

## 5. Пример

Пусть  $T = 10$ ,  $\rho = 0,9$ , и матрица выигрышей игроков на каждом шаге (базовая игра) есть

|     | $C$        | $N$        |
|-----|------------|------------|
| $C$ | $(10, 10)$ | $(-1, 15)$ |
| $N$ | $(15, -1)$ | $(1, 1)$   |

Заметим, что условие (3.1), которое записывается в виде:

$$\frac{b-a}{b-d} < \rho,$$

выполнено.

Найдем  $K$  по формуле (3.5), получаем

$$K = 2 \log_{0,9} \left( \frac{10 + 15(-0,1) - 0,9}{10 - 1} \right) \approx 3,209.$$

В таблице 1 приведены нижние и верхние границы продолжительности наказания в зависимости от шага игры, рассчитанные в соответствие с леммой 3.1, а также множества допустимых значений  $h$ :

Таблица 1. Расчеты продолжительности наказания  $h(\cdot)$

| $t + 1$ | Ниж. граница $h$ | Верх. граница $h$ | Допустимые $h$            |
|---------|------------------|-------------------|---------------------------|
| 1       | 0,6047           | 7,39              | $\{1, 2, 3, 4, 5, 6, 7\}$ |
| 2       | 0,6047           | 6,39              | $\{1, 2, 3, 4, 5, 6\}$    |
| 3       | 0,6047           | 5,39              | $\{1, 2, 3, 4, 5\}$       |
| 4       | 0,6047           | 4,39              | $\{1, 2, 3, 4\}$          |
| 5       | 0,6047           | 3,39              | $\{1, 2, 3\}$             |
| 6       | 0,6047           | 2,39              | $\{1, 2\}$                |
| 7       | 0,6047           | 1,39              | $\{1\}$                   |
| 8       | 0,6047           | 0,39              | $\emptyset$               |
| 9       | 0,6047           | -0,60             | $\emptyset$               |
| 10      | 0,6047           | -1,60             | $\emptyset$               |

Пусть выбраны следующие продолжительности наказания из допустимых множеств (см. последний столбец в таблице 1):

| Шаг $t + 1$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------------|---|---|---|---|---|---|---|
| $h(t + 1)$  | 5 | 4 | 5 | 3 | 2 | 2 | 1 |

Отметим, что допустимых наказаний, продолжительность которых удовлетворяет условию леммы 3.1, не существует, если отклонение произошло на шаге 8, 9 или 10.

Вычислим значение  $\hat{\varepsilon}$  по формуле (4.1), получим:

$$\hat{\varepsilon} = (15 - 10) \cdot 0,9^9 \approx 1,937.$$

Согласно теореме 4.1 для любого  $\varepsilon > \hat{\varepsilon} = 1,937$  существует  $\varepsilon$ -равновесие в 10-шаговой повторяющейся игре «Дилемма заключенного» в стратегиях поведения вида (3.6). Отметим, что если игроки кооперируются на каждом шаге на протяжении всей игры, то выигрыш каждого игрока в 10-шаговой повторяющейся игре равен  $S_0 = 65,1322$ .

Теперь проверим этот результат, посчитав, сколько получит игрок, отклонившись от стратегии (3.6) в каждый возможный момент времени. Обозначим выигрыш игрока через  $S_t$ , где  $t$  — шаг, на котором происходит первое и единственное отклонение игрока, когда другой игрок придерживается стратегии (3.6). Значения выигрышей  $S_t$  приведены в таблице 2.

Таблица 2. Выигрыши игроков при однократном отклонении от кооперации  $S_t$ ,  $t = 1, \dots, 10$ .

|       |         |         |         |         |         |
|-------|---------|---------|---------|---------|---------|
| $t$   | 1       | 2       | 3       | 4       | 5       |
| $S_t$ | 36,9618 | 44,5618 | 42,3142 | 52,7749 | 58,3153 |
| $t$   | 6       | 7       | 8       | 9       | 10      |
| $S_t$ | 58,9970 | 63,4850 | 60,1630 | 63,7980 | 67,0690 |

Теперь посчитаем, сколько игрок может получить, отклонившись два раза от кооперации на шагах  $t_1$  и  $t_2$  соответственно, обозначив этот выигрыш через  $S_{t_1, t_2}$ . Значения  $S_{t_1, t_2}$  для численного примера представлены в таблице 3.

Нетрудно заметить, что

$$\begin{aligned} \max_t (S_t - S_0) &= S_{10} - S_0 \approx 1,9368, \\ \max_{t_1, t_2} (S_{t_1, t_2} - S_0) &= S_{7, 10} - S_0 \approx 0,2896, \end{aligned}$$

что подтверждает результат теоремы (4.1), что для  $\forall \varepsilon > 1,9368$  существует  $\varepsilon$ -равновесие в 10-шаговой повторяющейся игре «Дилемма заключенного» в стратегиях поведения вида (3.6).

Таблица 3. Выигрыши игроков при двукратном отклонении от кооперации  $S_{t_1, t_2}$ .

|                |          |          |          |          |
|----------------|----------|----------|----------|----------|
| $(t_1, t_2)$   | (1, 7)   | (1, 8)   | (1, 9)   | (1, 10)  |
| $S_{t_1, t_2}$ | 27, 9534 | 31, 9992 | 35, 6274 | 38, 8988 |
| $(t_1, t_2)$   | (2, 7)   | (2, 8)   | (2, 9)   | (2, 10)  |
| $S_{t_1, t_2}$ | 35, 5534 | 39, 5923 | 43, 2274 | 46, 4985 |
| $(t_1, t_2)$   | (3, 9)   | (3, 10)  |          |          |
| $S_{t_1, t_2}$ | 40, 9798 | 44, 2513 |          |          |
| $(t_1, t_2)$   | (4, 8)   | (4, 9)   | (4, 10)  |          |
| $S_{t_1, t_2}$ | 47, 8054 | 51, 4404 | 54, 7120 |          |
| $(t_1, t_2)$   | (5, 8)   | (5, 9)   | (5, 10)  |          |
| $S_{t_1, t_2}$ | 53, 3458 | 56, 9808 | 60, 2524 |          |
| $(t_1, t_2)$   | (6, 9)   | (6, 10)  |          |          |
| $S_{t_1, t_2}$ | 57, 6625 | 60, 9341 |          |          |
| $(t_1, t_2)$   | (7, 9)   | (7, 10)  |          |          |
| $S_{t_1, t_2}$ | 62, 1502 | 65, 4218 |          |          |

## 6. Заключение

В статье рассматривается конечно повторяющаяся игра «Дилемма заключенного», для которой построена новая ситуация приближенного равновесия для поддержки кооперативного поведения игроков. При построении ситуации приближенного или  $\varepsilon$ -равновесия моделируется наказание отклонившегося от кооперации игрока. Особенностью стратегии наказания является то, что наказание в случае отклонения игрока не длится до окончания игры, а имеет конечную продолжительность. Таким образом, игроки могут вернуться к кооперации после того, как наказание будет осуществлено. Выигрыши при таком наказании будут больше, чем в случае, когда наказание длится до конца игры. В наших будущих исследованиях планируется построение платежной схемы, основанной на предложенном в текущей работе равновесии для конечно повторяющихся игр, подобно тому, как это было сделано в работе [15].

# СПИСОК ЛИТЕРАТУРЫ

1. Петросян Л.А., Зенкевич Н.А., Семина Е.А. *Теория игр*. М.: Высшая школа, 1998.
2. Печерский С.Л., Беляева А.А. *Теория игр для экономистов. Вводный курс*. СПб.: Издательство Европейского университета в Санкт-Петербурге, 2001.
3. Писарева А.М. *Построение стратегии наказания в повторяющихся играх «дилемма заключенного»* // Процессы управления и устойчивость. 2023. № 1. С. 472–476.
4. Axelrod R. *On six advances in cooperation theory* // Analyse & Kritik. 2000. V. 22. N. 1. P. 130–151.
5. Benoit J., Krishna V. *Finitely Repeated Games* // Econometrica. 1985. V. 53. N 4. P. 905–922.
6. Everett H. *Recursive games* // Contributions to the Theory of Games. 1957. V. 3. N 39. P. 47–78.
7. Friedman J. *A Non-cooperative Equilibrium for Supergames* // The Review of Economic Studies. 1971. V. 38. N 1. P. 1–12.
8. Fudenberg D., Maskin E. *The Folk Theorem in Repeated Games with Discounting or with Incomplete Information* // Econometrica. 1986. V. 54. N 3. P. 533–554.
9. Grauer L.V., Petrosyan L.A. *Multistage games* // Journal of Applied Mathematics and Mechanics. 2004. V. 68. N. 4. P. 597–605.
10. Grinikh A.L., Petrosyan L.A. *An effective punishment for an n-person prisoner's dilemma on a network* // Труды института математики и механики УрО РАН. 2021. Т. 27. № 3, С. 256–262.
11. Kuzyutin D., Smirnova N. *Subgame consistent cooperative behavior in an extensive form game with chance moves* // Mathematics. 2020. V. 8. N. 7. Art.no. 1061.
12. Mailath G., Samuelson L. *Repeated games and reputations: long-run relationships*. Oxford University Press, 2006.

13. Maschler M., Solan E., Zamir S. *Game Theory*. Cambridge University Press, 2013.
14. Parilina E., Zaccour G. *Approximated cooperative equilibria for games played over event trees* // Operations Research Letters. 2015. V. 43 N 5. P. 507–513.
15. Parilina E., Zaccour G. *Payment schemes for sustaining cooperation in dynamic games* // Journal of Economic Dynamics and Control. 2022. V. 139. Art. no. 104440.
16. Radner R. *Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives* // Journal of Economic Theory. 1980. V. 22. N 2. P. 136–154
17. Radner R. *Can bounded rationality resolve the prisoner's dilemma* // Essays in honor of Gerard Debreu. 1986. P. 387-399.
18. Vasin A. *The Folk theorem for dominance solutions* // International Journal of Game Theory. 1999. V. 28. N. 1. P. 15–24.
19. Vasin A. *The folk theorems in the framework of evolution and cooperation* // Annals of the International Society of Dynamic Games. 2006. V. 8. P. 197–207.

## APPROXIMATE EQUILIBRIUM IN A FINITELY REPEATED “PRISONER’S DILEMMA”

**Alena M. Pisareva**, Saint Petersburg State University, Faculty of Applied Mathematics and Control Processes, student (st055836@student.spbu.ru),

**Elena M. Parilina**, Saint Petersburg State University, Faculty of Applied Mathematics and Control Processes, Dr.Sc., professor (e.parilina@spbu.ru).

*Abstract:* The paper studies finitely repeated Prisoner’s Dilemma. To maintain cooperation in the game, a new profile of behavioral strategies is proposed, where the deviation of a player is not punished until the end of the game, but for a given number of stages depending on the stage of the game. The existence of an approximate equilibrium in these strategies is proven, and the maximum value of benefit of a player deviating from the approximate equilibrium is found.

*Keywords:* repeated games, prisoner’s dilemma, approximate equilibrium, trigger strategy.