Доклады Академии наук СССР
1970. Том 195, № 3

УДК 518-9

МАТЕМАТИКА

Л. А. ПЕТРОСЯН

ДИФФЕРЕНЦИАЛЬНЫЕ ИГРЫ С НЕПОЛНОЙ ИНФОРМАЦИЕЙ

(Представлено академиком Л. С. Понтрягиным 27 IV 1970)

1. Неформальное описание. Рассмотрим дифференциальную игру с предписанной продолжительностью T и уравнениями движения $\dot{x} = f(x, u, v)$, где $x \in R^n$, $u \in U \subset R^e$, $v \in V \subset R^s$. Управляющий параметр u находится в распоряжении игрока P (минимизирующий игрок), а управляющий параметр v находится в распоряжении игрока E (максимизирующий игрок). Движение начинается из начального состояния x_0, которое считается известным обоим игрокам. Фиксируется некоторое число $l > 0$, которое мы будем называть информационным лагом. На отрезке времени $[0, l]$ игроку P известно только начальное состояние x_0 и свои выборы $u(\tau)$, далее при $l \leqslant t \leqslant T$ в каждый момент времени игроку P известно состояние процесса в момент времени $t - l$, $x(t-l)$ и свои выборы $u(\tau)$, $\tau \leqslant t$. Игроку E в каждый момент времени t известно состояние процесса в этот момент времени $x(t)$. Пусть $x(t)$ — некоторая траектория, реализовавшаяся в процессе игры. Выигрыш игрока P определяется как некоторый функционал от $x(t)$, $F(x_0; x(t))$. Выигрыш E равен $-F$.

2. Дискретная модель. Зададимся разбиением отрезка $[0, T]$ с шагом δ. Пусть t_k — моменты разбиения, здесь $t_k - t_{k-1} = \delta$. Заменим уравнение движения для непрерывного случая разностным уравнением

$$x_{k+1} = x_k + \delta f(x_k, u_k, v_k). \qquad (1)$$

Для простоты будем предполагать, что l кратно δ. Игра протекает следующим образом. Игрок E на каждом шаге k выбирает управление v_k, зная $x_0, \ldots, x_{k-1}$. Игрок P при $0 \leqslant k \leqslant l/\delta$ выбирает на каждом шаге свое управление u_k, зная состояние x_0 и свои выборы на предыдущих шагах $i < k$, на последующих шагах P выбирает свое управление, зная состояние x_{k-l}, свои предыдущие выборы, предысторию $x_0, \ldots, x_{k-1}$. Игра прекращается в момент времени T. Игрок P получает выигрыш равный $F(x_0, x_1, \ldots, x_{T/\delta})$. E получает $-F$.

Определим стратегии в игре. Под стратегией P мы будем понимать правило, которое каждой информации игрока ставит в соответствие некоторое управление $u \in U$, т. е. стратегиями игрока P являются всевозможные измеримые функции $u(x_0, \ldots, x_{k-l-1}, u_0, \ldots, u_{k-1})$ со значениями в U. Аналогично стратегия E ставит в соответствие каждой информации игрока E некоторое управление $v \in V$, т. е. стратегиями E являются всевозможные измеримые функции $v(x_0, \ldots, x_k)$ со значениями в V. Для простоты стратегии игроков мы будем обозначать через $u(I_k^P)$ и соответственно $v(I_k^E)$, понимая под I_k^P текущую информацию, доступную игроку P и под I_k^E — таковую для игрока E.

Обычным образом показывается, что любое начальное условие x_0 и фиксированная пара стратегий $u(I)$, $v(I)$ однозначно определяет «партию» — последовательность $x_0, \ldots, x_{T/\delta}$, а следовательно, и выигрыш $K(x_0; u(I), v(I)) = F(x_0, \ldots, x_{T/\delta})$. Обозначим множество всех стратегий игрока P через A и множество всех стратегий игрока E через B.

Определение. Множество партий $x_0, \ldots, x_k$, которые получаются при зафиксированных $x_0, \ldots, x_{k-l}$ и управлений $u_0, \ldots, u_{k-1}$ для всевозможных выборов управлений игрока E на шагах $k-l \leqslant i < k$, называется информационным множеством игрока P k-го уровня. Очевидно, что информационные множества одного уровня не пересекаются. Информационное множество уровня k игрока E совпадает с некоторой последовательностью $x_0, \ldots, x_k$ (является одноэлементным).

Класс всех информационных множеств уровня k для игрока P мы обозначим через J_k^P, аналогично, класс всех информационных множеств игрока E уровня k мы обозначим через J_k^E. Все J_k^E и все J_k^P мы будем считать стандартными измеримыми пространствами. (Пространство называется стандартным, если оно конечно или счетно с дискретной структурой или оно изоморфно единичному интервалу.)

Под решением антагонистической игры естественно понимать ситуацию равновесия, т. е. такую пару стратегий, что для всех $u(I) \in A$, $v(I) \in B$ имеет место

$$K(x_0;\, u(I),\, v^*(I)) \geqslant K(x_0;\, u^*(I),\, v^*(I)) \geqslant K(x_0;\, u^*(I),\, v(I)). \tag{2}$$

Однако хорошо известно [1], что в играх с неполной информацией такая ситуация существует в исключительных случаях и для нашей задачи понятие решения (2) оказывается непригодным.

Далее мы сделаем то же, что сделал фон Нейман при решении матричных игр: мы расширим понятие стратегии, включая в него возможность случайного выбора.

3. Смешанные стратегии и стратегии поведения. Будем считать, что множества U и V являются экземплярами единичного интервала. Это означает, что они изоморфны единичному интервалу, т. е. существует взаимно однозначное изоморфное в обе стороны соответствие. Можно также считать, что множества U, V изменяются от шага к шагу, и писать U_k, V_k.

Обозначим через Ω пространство с мерой, которое получается из единичного интервала, если на нем ввести лебегову меру. Все наши пространства выборок будут изоморфны Ω. Можно считать далее, что все J_k^P, J_k^E являются экземплярами единичного интервала, так как если какое-нибудь из них конечно или счетно, то всегда можно добавить к нему континуум одинаковых элементов. Элементы декартовых произведений xJ_k^P и J_k^E будут обозначаться через $I^{P(E)} = (I_1^{P(E)}, \ldots, I_{T/\delta}^{P(E)})$.

Определение. Смешанной стратегией P называется последовательность $m = (m_1, \ldots, m_{T/\sigma})$ измеримых отображений $m_i\colon \Omega \times J_k \to U$, где Ω есть фиксированное пространство выборок. Стратегией поведения называется такая смешанная стратегия b, что для $i \neq k$, $b(\cdot, I_i^P)$ и $b(\cdot, I_k^P)$ суть независимые случайные величины ($I_i^P \in J_i^P$ и $I_k^P \in J_k^P$ произвольны). Определение для E аналогично.

Каждая тройка $(\omega;\, m,\, v(I))$, состоящая из элемента пространства выборок, смешанной стратегии и стратегии противника, однозначно определяет элемент $u(\omega;\, m,\, v(I))$ из U; последовательность $u = (u_1, \ldots, u_{T/\delta})$ определяется рекуррентно равенством $u_i = m_i(\omega, I_i^P)$, где $I_i^P \in J_i^P$. Грубо говоря, это и есть последовательность фактически выбранных в процессе игры управлений. Кроме того, каждая пара $(m, v(I))$ однозначно определяет меру μ в xU_i, для любого измеримого множества $\bar{B} \in xU_i$ она определяется равенством

$$\mu(\bar{B}) = \mu(\bar{B};\, m,\, v(I)) = \lambda(\omega;\, u(\omega;\, m,\, v(I)) \in \bar{B})$$

(здесь λ — мера на Ω). Таким образом, μ есть распределение случайной величины $u(\cdot;\, m,\, v(I))$. Две смешанные стратегии называются эквивалентными, если для каждого $v(I) \in B$ они определяют одно и то же распределение в xU_i.

Следующая теорема является аналогом теоремы Куна ([1]) для дифференциальной игры с информационным лагом.

Теорема. *В дискретной игре с информационным лагом каждая смешанная стратегия имеет эквивалентную ей стратегию поведения.*

4. Ситуация равновесия в стратегиях поведения. Предыдущая теорема позволяет искать ситуацию равновесия в классе стратегий поведения. Из нее вытекает, что любая ситуация равновесия в классе стратегий поведения является ситуацией равновесия и в смешанных стратегиях.

Теорема. *Пусть выполнены следующие условия:* а) *функция выигрыша является непрерывной функцией от партии;* б) *каждое из информационных множеств является компактным множеством;* в) *класс всех информационных множеств каждого из игроков одного уровня является компактным в метрике Хаусдорфа. При выполнении условий* а), б), в), *в дискретной игре с информационным лагом существует ситуация равновесия в стратегиях поведения.*

При доказательстве теоремы вводятся вспомогательные подыгры по аналогии с конечным случаем (см. ([2])). Индукцией по длине подыгры показывается, что каждая из подыгр обладает ситуацией равновесия в стратегиях поведения. Переход от подыгры к основной игре осуществляется так же как и в ([2]). Для вывода функционального уравнения для значений подыгр определим подыгру.

Рассмотрим некоторое информационное множество игрока P уровня k. Зафиксируем это состояние информации I_k^P. Предположим, что игроку P известно также распределение вероятностей на выборах игрока E, $v_{k-l}, \ldots, v_{k-1}$, которые мы обозначим через $p_k(\cdot)$. Подыгра протекает следующим образом. Ходы $v_{k-l}, \ldots, v_{k-1}$ рандомизируются согласно распределению $p_k(\cdot)$ и сообщаются игроку E, но не P. После этого E выбирает v_k и выбор v_{k-l} сообщается игроку P. Выбор u_k сообщается обоим игрокам после того, как он сделан, но согласно информационным условиям выбор v_k содержится в секрете от игрока E до тех пор, пока тот не совершит выбор $u_{k+(l-1)}$. После этого производятся ходы u_{k+1} и v_{k+1}, соответственно, и выбор v_{k+1} объявляется игроку P. Эта последовательность продолжается до тех пор, пока не будут объявлены все случайные ходы, после чего игра продолжается с использованием схемы состояния информации, принятой для исходной игры. Из построения $p_k(\cdot)$ видно, что $p_k(\cdot)$ есть просто распределение вероятностей на I_k^P. Функцией выигрыша будет функция выигрыша в исходной игре. Подыгру мы будем обозначать как $\Gamma_k = \Gamma_k[I_k^P;\ p_k(\cdot)]$. Пусть $V[I_k^P,\ p_k(\cdot)]$ — значение подыгры Γ_k, тогда можно показать, что функция $V[I_k^P,\ p_k(\cdot)]$ непрерывна по I_k^P, $p_k(\cdot)$ (близость между информационными множествами одного уровня понимается в метрике Хаусдорфа) и $V[I_k^P,\ p_k(\cdot)]$ и удовлетворяет следующему функциональному уравнению

$$V[I_k^P; p_k(\cdot)] = \max_{F(v_k|v_{k-l},\ldots,v_{k-1})} \min_{\hat{u}_k} \int V[I_{k+1}^P, p_{k+1}(\cdot\,|\,\hat{v}_{k-l})]\, dF_{v_{k-l}}. \qquad (3)$$

Здесь $F(v_k | v_{k-l}, \ldots, v_{k-1})$ — стратегия поведения игрока E на k-м шаге, $p_{k+1}(\cdot)\hat{v}_{k-l})$ распределение на I_{k+1}^P при фиксированном $\hat{v}_{k-l}$ (выбор $\hat{v}_{k-l}$ стал известен P после k-го шага), которое индуцировано распределением $p_k(\cdot)$ и стратегией поведения $F(v_k | v_{k-l}, \ldots, v_{k-1})$. Функциональное уравнение (3) может быть использовано для нахождения оптимальной стратегии поведения игрока E.

Заметим, что до сих пор мы рассматривали дискретные игры.

Определение. Пусть существует предел значений дискретных игр с информационным лагом при произвольном измельчении дробления отрезка времени δ к нулю, тогда предельное значение будет называться обобщенным значением дифференциальной игры с информационным лагом.

5. Пример. Рассмотрим игру с лагом (лаг l имеет P). Движения игроков независимы.

$$\dot{x} = u, \ \dot{y} = v, \ |u| \leqslant 1, \ |v| \leqslant \lambda, \ \lambda < 1,$$

P управляет параметром u, E управляет параметром v. Игра имеет предписанную продолжительность T и выигрыш равен

$$K(x_0, y_0; u(I), v(I)) = -\rho(x(T), y(T)).$$

Оптимальная стратегия n^* игрока P чистая. В каждый момент времени t стратегия u^* выбирает направление движения на точку $y(t-l)$ при $t > l$ и на точку y_0 при $0 \leqslant t \leqslant l$. Если существует такой момент $t' < T$, что $x(t') = y(t'-l)$, то на отрезке $t' \leqslant t \leqslant T$, P стремится сохранить равенство $x(t) = y(t-l)$.

Оптимальная стратегия v^* игрока E. На отрезке времени $0 \leqslant t \leqslant T - l$ стратегия v^* выбирает направление от точки $x(t)$. В момент $T - l$, E с вероятностями $(1/2, 1/2)$ выбирает одно из двух направлений перпендикулярных отрезку $[x(T-l), \ y(T-l)]$, если $x(T-l) \neq y(T-l)$ и любые два противоположных направления, если $x(T-l) = y(T-l)$. Далее на отрезке $T - l < t \leqslant T$ он придерживается направления, выбранного в момент времени $T - l$ с помощью случайного механизма $(1/2, 1/2)$.

Если в ситуации равновесия существует момент t^1, такой что $x(t^1) = y(t^1 - l)$ и $t^1 < T$, то значение игры не зависит от T и равно λl.

Ленинградский государственный университет
им. А. А. Жданова

Поступило
15 IV 1970

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] Г. У. Кун, Позиционные игры и проблема информации, Позиционные игры, «Наука», 1967. [2] Х. Э. Скарф, Л. С. Шэпли, Игры с неполной информацией, Применение теории игр в военном деле, М., 1961.