D. B. Rokhlin, G. A. Ougolnitsky, Stackelberg equilibrium in a dynamic stimulation model with complete information, *Avtomat. i Telemekh.*, 2018, Issue 4, 152–166

Интеллектуальные системы управления, анализ данных

© 2018 г. Д.Б. РОХЛИН, д-р физ-мат. наук (rokhlin@math.rsu.ru),
Г.А. УГОЛЬНИЦКИЙ, д-р. физ.-мат. наук (ougoln@mail.ru)
(Южный федеральный университет, Ростов-на-Дону)

РАВНОВЕСИЕ ШТАКЕЛЬБЕРГА В ДИНАМИЧЕСКОЙ МОДЕЛИ СТИМУЛИРОВАНИЯ С ПОЛНОЙ ИНФОРМАЦИЕЙ[1]

Рассмотрена модель стимулирования с марковской динамикой и дисконтированными критериями оптимальности в случае дискретного времени и бесконечного горизонта планирования. В указанной модели регулятор оказывает экономическое воздействие на исполнителя, выбирая стимулирующую функцию, зависящую от состояния системы и действий исполнителя, который использует позиционные стратегии управления. Динамика системы, доходы регулятора и затраты исполнителя зависят от состояния системы и действий исполнителя. Показано, что отыскание приближенного решения (обратной) игры Штакельберга сводится к решению задачи оптимального управления с критерием, равным разности между доходом регулятора и затратами исполнителя. При этом ε-оптимальная стратегия регулятора состоит в том, чтобы экономически мотивировать исполнителя следовать данной оптимальной стратегии управления.

Ключевые слова: двухуровневая модель стимулирования, обратная игра Штакельберга, дисконтированный критерий оптимальности, уравнение Беллмана.

1. Введение

Мотивация к труду (стимулирование) по очевидным причинам составляет одну из ключевых проблем менеджмента. Выбор метода стимулирования является задачей регулятора (лидера, центра, ведущего), который анализирует реакцию исполнителя (агента, ведомого) и стремится максимизировать свой критерий оптимальности. Часто рассматривается и случай нескольких ведомых. Ставший классическим подход для описания подобных иерархических игровых задач предложен Штакельбергом [1]. Общее введение в теорию и экономические приложения динамических игр можно найти в монографиях [2–4]. Недавний обзор динамических игровых моделей Штакельберга применительно к проблемам менеджмента и маркетинга содержится в [5].

Версия игры Штакельберга, приспособленная к задаче стимулирования, получила название обратной игры Штакельберга. Этот термин впервые

[1] Работа выполнена при финансовой поддержке Российского научного фонда (проект № 17-19-01038).

появился в [6]. В такой игре регулятор выбирает из некоторого класса стимулирующую функцию, зависящую от действий исполнителя, и сообщает ее исполнителю. Если указанный класс содержит только константы, получаем обычную игру Штакельберга. Дальнейшую информацию об обратных играх Штакельберга можно найти в [7–10]. Следует подчеркнуть, что в западных публикациях не отражен вклад Ю.Б. Гермейера, который рассматривал обратную игру Штакельберга (получившую в дальнейшем название "игра Гермейера Γ_2") значительно раньше [6]: см. [11, 12]. Отметим также, что в отечественных публикациях математический аппарат исследования иерархических игр для класса стратегий с обратной связью по состоянию и управлению разрабатывался А.Ф. Кононенко и его соавторами: см. [13, 14]. В основе данного подхода лежит построение взаимовыгодной программы действий, реализация которой обеспечивается механизмами поощрения–наказания, предлагаемыми ведущим игроком ведомому.

Рассматриваемую в настоящей статье модель стимулирования можно отнести к теории контрактов. Это одно из наиболее типичных приложений обратных игр Штакельберга. Другие приложения касаются сетевого ценообразования [15], взимания платы за поезд [16], назначения цен на электроэнергию [17] и управления организационными системами [18].

В идейном плане постановка задачи аналогична рассмотренной в [19, 20]. В статической (одношаговой) задаче ε-оптимальный механизм стимулирования компенсирует затраты исполнителя с некоторой надбавкой $\delta > 0$ при выборе исполнителем желаемого для регулятора действия. Указанное действие определяется из задачи максимизации разности между доходом регулятора и затратами исполнителя. Этот результат с соответствующими дополнениями сохраняет силу для многошаговых задач стимулирования и задач стимулирования с несколькими исполнителями ([20], теоремы 1–3).

В настоящей статье рассматривается динамическая модель стимулирования с марковской динамикой, бесконечным горизонтом планирования и дисконтированными критериями оптимальности, в которой регулятор располагает полной информацией о состоянии системы и действиях исполнителя, а исполнитель использует позиционные стратегии управления. Таким образом, исполнитель знает размер материального поощрения в зависимости от состояния системы и собственных действий. Динамика системы и доходы регулятора также зависят от состояния системы и действий исполнителя. Оказывается, что качественно структура оптимальной стимулирующей функции аналогична описанной выше: а именно желаемая для регулятора стратегия исполнителя является решением задачи оптимального управления с критерием, равным разности между доходом регулятора и затратами исполнителя, а ε-оптимальная стратегия регулятора состоит в том, чтобы экономически мотивировать исполнителя следовать данной стратегии.

Статья построена следующим образом. В разделе 2 приводится динамическая постановка задачи стимулирования. В разделе 3 найдена ε-оптимальная стратегия регулятора в классе полунепрерывных сверху функций (теорема 1). В разделе 4 этот результат распространен на класс универсально измеримых функций и η-оптимальных стратегий исполнителя (теорема 2) и сформули-

рована гипотеза о том, что полученные результаты сохраняют силу, если регулятор и исполнитель используют разные коэффициенты дисконтирования.

2. Игровая динамическая модель стимулирования

Рассмотрим управляемую систему с дискретным временем, пространством состояний которой является компактное метрическое пространство $\mathcal{X}$, наделенное борелевской σ-алгеброй $\mathcal{B}(\mathcal{X})$. Предположим, что множество допустимых действий $\Gamma(x)$ является непустым замкнутым подмножеством отрезка $[0,1]$. Многозначное отображение $\Gamma : \mathcal{X} \mapsto 2^{[0,1]}$ предполагается непрерывным в смысле стандартного определения [21, гл. 9; 22, определение 6.1.2(c)]. Динамика системы описывается переходным ядром $q(B|x,a)$, т.е. если система находится в состоянии $x \in \mathcal{X}$, то при использовании действия $a \in \Gamma(x)$ вероятность того, что следующее состояние окажется в множестве $B \in \mathcal{B}(\mathcal{X})$, равна $q(B|x,a)$. Для каждой пары $(x,a) \in \mathcal{X} \times [0,1]$ функция $q(\cdot|(x,a))$ является вероятностной мерой на $\mathcal{B}(\mathcal{X})$, и для каждого $B \in \mathcal{B}(\mathcal{X})$ функция $q(B|\cdot,\cdot)$ на $\mathcal{X} \times \mathcal{A}$ является борелевской.

В рассматриваемой задаче считается, что a — усилие *исполнителя*, направленное на управление системой и приводящее к затратам $g(x,a)$. Одновременно *регулятор*, который непосредственно не управляет системой, но может оказывать экономическое воздействие на исполнителя, получает выигрыш $f(x,a)$. Предполагается, что функции $f, g : \mathcal{X} \times [0,1] \mapsto \mathbb{R}_+$ непрерывны, $f(x,0) \leqslant 0$, $g(x,0) = 0$ и

$$\max_{a \in [0,1]} (f(x,a) - g(x,a)) > 0. \tag{2.1}$$

Указанное экономическое воздействие состоит в том, что регулятор сообщает исполнителю *стимулирующую функцию* c, принадлежащую семейству $\mathcal{L}$ функций $c : \mathcal{X} \times [0,1] \mapsto \mathbb{R}_+$, полунепрерывных сверху. Напомним, что функция c называется полунепрерывной сверху, если множества $\{(x,a) : c(x,a) \geqslant \lambda\}$, $\lambda \in \mathbb{R}$, замкнуты.

Исполнитель использует управления u в форме обратной связи из семейства $\mathcal{U}$ борелевских функций $u : \mathcal{X} \mapsto [0,1]$, удовлетворяющих условию $u(x) \in \Gamma(x)$. Фиксируя такую функцию и начальное состояние x, получаем вероятностную меру $\mathsf{P}_{x,u}$ на пространстве $(\mathcal{A} \times \mathcal{X})^\infty$ траекторий, которая формально может быть записана следующим образом (см., например, [23, приложение C]):

$$\mathsf{P}_{x,u}(dx_0, da_0, dx_1, da_1, \dots) =$$
$$= \delta_x(dx_0)\delta_{u(x_0)}(da_0)q(dx_1|x_0,a_0)\delta_{u(x_1)}(da_1)q(dx_2|x_1,a_1)\dots,$$

где δ_y — мера Дирака, сосредоточенная в точке y. Математическое ожидание по мере $\mathsf{P}_{x,u}$ обозначим через $\mathsf{E}_{x,u}$.

В рассматриваемой далее игре между регулятором и исполнителем считается, что если регулятор выбирает стимулирующую функцию $c \in \mathcal{L}$, а исполнитель — управление $u \in \mathcal{U}$, то дисконтированные доходы регулятора и

исполнителя определяются формулами:

$$J_1(x,c,u) = \mathsf{E}_{x,u} \sum_{t=0}^{\infty} \beta^t (f(x_t,a_t) - c(x_t,a_t)),$$

$$J_2(x,c,u) = \mathsf{E}_{x,u} \sum_{t=0}^{\infty} \beta^t (c(x_t,a_t)) - g(x_t,a_t)),$$

где $\beta \in [0,1)$ — коэффициент дисконтирования. Регулятор является лидером, а исполнитель — ведомым. Передавая исполнителю функцию c, регулятор вычисляет оптимальную стратегию u^* исполнителя, а затем максимизирует свой выигрыш по всем стимулирующим функциям c. Точная постановка задачи дана далее.

При фиксированной функции $c \in \mathcal{L}$ оптимальная стратегия исполнителя может быть найдена с помощью метода динамического программирования. Предположим, что переходное ядро слабо непрерывно по (x,a):

$$\int\limits_{\mathcal{X}} h(y)\, q(dy|x_n,a_n) \to \int\limits_{\mathcal{X}} h(y)\, q(dy|x,a) \quad \text{при} \quad (x_n,a_n) \to (x,a)$$

для любой непрерывной функции $f : \mathcal{X} \mapsto \mathbb{R}$. Тогда функция Беллмана

$$V_2(x;c) := \sup_{u \in \mathcal{U}} J_2(x,u,c) \tag{2.2}$$

исполнителя является единственным полунепрерывным сверху решением уравнения

$$V_2(x;c) = \sup_{a \in \Gamma(x)} \left\{ c(x,a) - g(x,a) + \beta \int\limits_{\mathcal{X}} V_2(y;c) q(dy|x,a) \right\},$$

а его оптимальные стратегии $u^* \in \mathcal{U}$ существуют и в точности описываются соотношением

$$u^*(x) \in \arg\max_{a \in \Gamma(x)} \left\{ c(x,a) - g(x,a) + \beta \int\limits_{\mathcal{X}} V_2(y;c) q(dy|x,a) \right\}.$$

При сделанных выше предположениях это утверждение вытекает из общих результатов [24; 25, предложение 2.1; 26, гл. 9; 27, предложение 3.1]. Если функция c непрерывна, то $V_2(\cdot;c)$ также будет непрерывной.

Обозначим множество оптимальных стратегий u^* исполнителя через $T(c)$. Рассмотрим *игру Штакельберга*, в которой регулятор оценивает свой выигрыш, исходя из пессимистического сценария, считая, что исполнитель использует одну из своих оптимальных стратегий

$$G(x,c) = \inf_{u \in T(c)} J_1(x,c,u). \tag{2.3}$$

Оптимальный выигрыш регулятора при этом равен

$$V_1(x) = \sup_{c \in \mathcal{L}} G(x, c).$$

Назовем V_1 ценой игры (лидера). Функция $\overline{c} \in \mathcal{L}$ называется решением Штакельберга, если $G(x, \overline{c}) = V_1(x)$, $x \in \mathcal{X}$ [2, определение 4.6].

Как будет показано далее, решения Штакельберга в рассматриваемой игре не существует. В связи с этим введем понятие ε-решения. Для $\varepsilon > 0$ функция $\overline{c}_\varepsilon \in \mathcal{L}$ называется ε-решением Штакельберга ([2, определение 4.7]), если

$$G(x, \overline{c}_\varepsilon) \geqslant V_1(x) - \varepsilon.$$

При этом пара $(\overline{c}_\varepsilon, u)$, $u \in T(c)$, называется ε-равновесием Штакельберга.

Отметим, что в данной постановке задачи предполагается, что регулятор обладает полной информацией о действиях исполнителя. Если же регулятору известно только состояние системы, то естественно рассматривать класс позиционных стратегий $c = c(x)$.

В прикладных задачах управляемый марковский процесс x_t часто задается рекуррентным соотношением

$$x_{t+1} = b(x_t, a_t, \xi_{t+1}), \quad x_0 = x,$$

где ξ_t — последовательность независимых одинаково распределенных случайных величин. Соответствующее переходное ядро имеет вид

$$q(B|(x, a)) = \mathsf{P}(b(x, a, \xi) \in B).$$

Пример 1 (регулирование вылова). Пусть динамика рыбной популяции описывается рекуррентным соотношением

$$x_{t+1} = b(x_t - a_t, \xi_{t+1}), \quad t \in \mathbb{Z}_+.$$

Здесь $x_t \in \mathcal{X} = [0, 1]$ — количество рыбы, $b : [0, 1] \times [0, 1] \mapsto [0, 1]$ — функция воспроизводства, ξ_t — независимые одинаково распределенные случайные величины со значениями на отрезке $[0, 1]$. Функция полезности регулятора (государства) $f = f(a)$ обычно предполагается возрастающей и вогнутой, функция затрат $g = g(a)$ исполнителя (рыбака) предполагается возрастающей и выпуклой. Множество допустимых действий $\Gamma(x) = [0, x]$ описывает допустимые объемы вылова.

Пример 2 (управление запасами). Фирма, являющаяся продавцом товара и владельцем склада (регулятор), получает товар от производителя (исполнителя). Количество товара x_t на складе определяется формулой

$$x_{t+1} = \min\{(x_t + a_t - \xi_{t+1})^+, M\}, \quad t \in \mathbb{Z}_+.$$

Здесь a_t — количество товара, произведенное исполнителем и доставленное на склад, M — емкость склада, ξ_t — независимые одинаково распределенные случайные величины, описывающие спрос на товар, $x^+ = \max\{x, 0\}$. Функция

полезности регулятора учитывает прибыль p от продажи товара, затраты q на покупку (и доставку) товара и затраты на его хранение h:

$$f(x,a) = \mathsf{E}p(\min\{\xi_{t+1}, x+a\}) - q(a) - h(x+a).$$

Функция затрат производителя $g = g(a)$ предполагается возрастающей и выпуклой, $\Gamma(x) = [0, \overline{a}]$. Будем считать, что согласно контракту владелец склада покупает у производителя весь произведенный товар в количестве a_t (излишки, $(x_t + a_t - M)^+$, которые не помещаются на склад, пропадают). Влияние регулятора на исполнителя осуществляется посредством выбора стимулирующей функции c.

3. ε-решение Штакельберга

Рассмотрим вспомогательную задачу оптимального управления с целевой функцией

$$J(x,v) = \mathsf{E}_{x,v} \sum_{t=0}^{\infty} \beta^t (f(x_t, a_t) - g(x_t, a_t))$$

и тем же множеством $\mathcal{U}$ допустимых управлений. Положим

$$V(x) = \sup_{v \in \mathcal{U}} J(x,v). \tag{3.1}$$

Как и задача исполнителя (2.2), данная задача имеет решение $\overline{v} \in \mathcal{U}$. Кроме того, поскольку функции f и g непрерывны, то функция Беллмана V непрерывна. Условие (2.1) обеспечивает положительность V:

$$V(x) > 0, \quad x \in \mathcal{X}.$$

Далее, если решение $\overline{v} \in \mathcal{U}$ задачи (3.1) единственно, то функция $\overline{v}$ непрерывна. Это вытекает из представления

$$\{\overline{v}(x)\} = \arg\max_{a \in \Gamma(x)} \left\{ f(x,a) - g(x,a) + \beta \int_{\mathcal{X}} V(y) q(dy|x,a) \right\}$$

и теоремы Бержа о максимуме ([21, теорема 9.17]).

Заметим, что для любых $c \in \mathcal{L}$ и $u^* \in T(c)$ верно неравенство

$$J_2(x,c,u^*) \geqslant J_2(x,c,0) \geqslant 0.$$

Отсюда следует, что

$$\begin{aligned} G(x,c) \leqslant J_1(x,c,u^*) \leqslant (J_1 + J_2)(x,c,u^*) = \\ = \mathsf{E}_{x,u^*} \sum_{t=0}^{\infty} \beta^t (f(x_t,a_t) - g(x_t,a_t)) \leqslant V(x), \\ V_1(x) = \sup_{c \in \mathcal{L}} G(x,c) \leqslant V(x). \end{aligned} \tag{3.2}$$

Следующий результат показывает, что при соответствующем дополнительном предположении в (3.2) имеет место в равенство.

Теорема 1. Пусть существует непрерывное оптимальное решение $\overline{v} \in \mathcal{U}$ *задачи* (3.1). *Тогда для описанной игры Штакельберга на* $(\mathcal{L}, \mathcal{U})$ *справедливы следующие утверждения*:

i. Цена игры совпадает с оптимальным значением задачи (3.1):

$$V_1(x) = V(x); \tag{3.3}$$

ii. Полунепрерывная сверху функция

$$\overline{c}_\varepsilon(x,a) = g(x,a) + \varepsilon(1-\beta) I_{\{a=\overline{v}(x)\}}, \quad \varepsilon > 0, \tag{3.4}$$

является ε*-решением. При этом* $T(c_\varepsilon) = \{\overline{v}\}$.

Доказательство. Полунепрерывность сверху функции c_ε вытекает из непрерывности $\overline{v}$. Покажем, что $\overline{v}$ — единственная оптимальная стратегия исполнителя, соответствующая $\overline{c}_\varepsilon$, т.е. $T(\overline{c}_\varepsilon) = \{\overline{v}\}$. Заметим, что $\overline{c}_\varepsilon - g \leqslant (1-\beta)\varepsilon$. Следовательно,

$$V_2(x;\overline{c}_\varepsilon) = \sup_{u\in\mathcal{U}} J_2(x,\overline{c}_\varepsilon,u) \leqslant \sum_{t=0}^{\infty} \beta^t (1-\beta)\varepsilon = \varepsilon. \tag{3.5}$$

Вычислим выигрыш исполнителя при использовании стратегии $\overline{v}$. Учитывая, что $\overline{c}_\varepsilon(x,\overline{v}(x)) = g(x,\overline{v}(x)) + \varepsilon(1-\beta)$, находим

$$J_2(x,\overline{c}_\varepsilon,\overline{v}) = \mathsf{E}_{x,\overline{v}} \sum_{t=0}^{\infty} \beta^t (\overline{c}_\varepsilon(x_t,a_t) - g(x_t,a_t)) = \sum_{t=0}^{\infty} \beta^t (1-\beta)\varepsilon = \varepsilon. \tag{3.6}$$

В силу (3.5) это означает, что $V_2(x;\overline{c}_\varepsilon) = \varepsilon$ и $\overline{v} \in T(\overline{c}_\varepsilon)$. В частности, функция Беллмана V_2 является константой и

$$\begin{aligned} \arg\max_{a\in\Gamma(x)} \left\{ \overline{c}_\varepsilon(x,a) - g(x,a) + \beta \int_{\mathcal{X}} V_2(y;c) q(dy|x,a) \right\} = \\ = \arg\max_{a\in\Gamma(x)} \{\overline{c}_\varepsilon(x,a) - g(x,a)\} = \{\overline{v}(x)\}. \end{aligned}$$

Следовательно, $T(\overline{c}_\varepsilon) = \{\overline{v}\}$.

Вычислим соответствующий выигрыш регулятора. С учетом (3.6) имеем

$$\begin{aligned} V_1(x) \geqslant J_1(x,\overline{c}_\varepsilon,\overline{v}) &= \mathsf{E}_{x,\overline{v}} \sum_{t=0}^{\infty} \beta^t (f(x_t,a_t) - \overline{c}_\varepsilon(x_t,a_t)) = \\ &= \mathsf{E}_{x,\overline{v}} \sum_{t=0}^{\infty} \beta^t (f(x_t,a_t) - g(x_t,a_t)) - \varepsilon = V(x) - \varepsilon. \end{aligned} \tag{3.7}$$

Из неравенств (3.2) и (3.7) вытекает, что $\overline{c}_\varepsilon$ является ε-оптимальным решением Штакельберга и имеет место равенство (3.3). □

Полученный результат можно описать следующим образом. Регулятор должен представить, что он сам занимается управлением системой и его затраты совпадают с затратами исполнителя. Определив оптимальное управление $\overline{v}$, регулятор должен мотивировать исполнителя к реализации данного управления, объявив стимулирующую функцию $\overline{c}_\varepsilon$. Если исполнитель будет действовать оптимально с точки зрения своих интересов, то выигрыш регулятора $J_1(x, \overline{c}_\varepsilon, \overline{v}) = V(x) - \varepsilon$ будет отличаться на ε от его оптимального выигрыша: см. (3.7), а выигрыш исполнителя будет равен ε: см. (3.5), (3.6). Второе слагаемое в (3.4) может трактоваться как мотивирующая надбавка, которая тем больше, чем сильнее близорукость исполнителя ($\beta \to 0$).

Указанное в теореме 1 ε-решение, конечно, не является единственным. Например, функции

$$\overline{c}^1_\varepsilon(x,a) = g(x,a) + \varepsilon(1-\beta)(1 - |a - \overline{v}(x)|),$$
$$\overline{c}^2_\varepsilon(x,a) = (g(x,\overline{v}(x)) + \varepsilon(1-\beta))I_{\{a=\overline{v}(x)\}}$$

также будут ε-решениями. Заметим, что если $\overline{v}$ непрерывна, то функция $\overline{c}^1_\varepsilon$ непрерывна.

В случае конечного $\mathcal{X}$ условие непрерывности $\overline{v}$ выполняется автоматически. Как уже отмечено, в общем случае непрерывность оптимального решения вытекает из его единственности. Обсуждение достаточных условий единственности можно найти в [28]. Не будем останавливаться на этом более подробно, так как в полученном далее более общем результате (теорема 2) не требуется выполнения условия непрерывности $\overline{v}$.

Замечание 1. Покажем, что решения Штакельберга не существует. В условиях теоремы 1 предположим, что $\overline{c}$ — такое решение. Пусть для некоторых $u \in T(\overline{c})$ и $x \in \mathcal{X}$ выигрыш исполнителя положителен: $J_2(x, \overline{c}, u) > 0$. Тогда поскольку

$$(J_1 + J_2)(x, \overline{c}, u) = \mathsf{E}_{x,u} \sum_{t=0}^{\infty} \beta^t (f(x_t, a_t) - g(x_t, a_t)) \leqslant V(x),$$

то получаем противоречие с определением решения Штакельберга:

$$J_1(x, \overline{c}, u) < V_1(x) = V(x).$$

Таким образом, если $\overline{c}$ — решение Штакельберга, то

$$J_2(x, \overline{c}, u) = 0, \quad x \in \mathcal{X},\ u \in T(\overline{c}).$$

Но поскольку $J_2(x, \overline{c}, 0) \geqslant 0$, то $0 \in T(\overline{c})$, что также ведет к противоречию: $J_1(x, \overline{c}, 0) \leqslant 0 < V(x)$.

Замечание 2. В рассмотренной игре регулятор рассматривает пессимистический сценарий: см. (2.3). Такая игра Штакельберга известна как "слабая" [29]. В сильной игре Штакельберга лидер рассматривает оптимистический сценарий, при котором его выигрыш равен $\sup_{u \in T(c)} J_1(x, c, u)$. В рассматриваемом случае такая игра тривиальна. Действительно, полагая $c = g$,

заключаем, что выигрыш исполнителя равен нулю при любой стратегии. Таким образом, $T(g) = \mathcal{U}$. Далее, для $u \in \mathcal{U}$ выигрыш регулятора равен

$$J_1(x, g, u) = J(x, u) = \mathsf{E}_{x,u} \sum_{t=0}^{\infty} \beta^t (f(x_t, a_t) - g(x_t, a_t)).$$

Выбирая в качестве u любое решение $\overline{v}$ задачи (3.2), исполнитель обеспечит регулятору выигрыш $V(x)$. Но из неравенства вида (3.3) вытекает, что выигрыш регулятора не может быть больше этого значения при любой стратегии $c \in \mathcal{L}$. Таким образом, g — сильное решение Штакельберга:

$$J_1(x, g, \overline{v}) = \sup_{c \in \mathcal{L}} \sup_{u \in T(c)} J_1(x, c, u) = V(x).$$

Очевидный недостаток сильного решения в данных обстоятельствах состоит в том, что наиболее естественной оптимальной стратегией исполнителя является "нулевая стратегия" $u^* = 0$, не требующая от исполнителя приложения усилий. При этом выигрыш регулятора будет неположительным.

4. (ε, η)-решение Штакельберга

Наложенное на стимулирующую функцию c условие полунепрерывности сверху связано с желанием обеспечить существование решения задачи исполнителя. Рассмотрим ситуацию, когда это требование доставляет неудобства.

Если задача (3.1) не имеет непрерывного решения $\overline{v}$, то функция $\overline{c}_\varepsilon$ (а также любая из функций $\overline{c}^1_\varepsilon$, $\overline{c}^2_\varepsilon$) может не быть полунепрерывной сверху. Попытка перейти к полунепрерывной сверху оболочке $\overline{c}^*_\varepsilon$ не приводит к успеху, так как решение задачи

$$\overline{c}^*_\varepsilon(x, a) - g(x, a) \to \max_{a \in \Gamma(x)} \tag{4.1}$$

может быть неединственным. Пусть, например, $\mathcal{X} = [0, 1]$, функция $\overline{v}$ непрерывна справа и $\overline{v}(x-) := \lim_{y \uparrow x} \overline{v}(y) < \overline{v}(x)$, $x > 0$. Ясно, что $\overline{c}^*_\varepsilon \leqslant g + (1 - \beta)\varepsilon$. С другой стороны,

$$\overline{c}^*_\varepsilon(x, \overline{v}(x-)) \geqslant \lim_{y \uparrow x} \overline{c}_\varepsilon(y, \overline{v}(y)) =$$
$$= \lim_{y \uparrow x} g(y, \overline{v}(y)) + \varepsilon(1 - \beta) = g(x, \overline{v}(x-)) + \varepsilon(1 - \beta).$$

Следовательно,

$$\overline{c}^*_\varepsilon(x, \overline{v}(x-)) = g(x, \overline{v}(x-)) + (1 - \beta)\varepsilon$$

и функция (4.1) достигает максимума не только в точке $\overline{v}(x)$, но и в точке $\overline{v}(x-)$. Это приводит к тому, что $T(\overline{c}^*_\varepsilon) \neq \{\overline{v}\}$ и рассуждения из доказательства теоремы 1 не проходят.

В то же время для функции $\overline{c}_\varepsilon$ доказательство теоремы 1 формально работает и в том случае, когда она не является полунепрерывной сверху. Чтобы

сделать $\overline{c}_\varepsilon$ допустимой стратегией в общем случае, естественно попытаться расширить класс $\mathcal{L}$, включив в него все борелевские функции. При этом решение задачи исполнителя может не существовать и естественно считать, что он будет использовать приближенно оптимальные решения. Однако, как показано в [30, пример 2], для общей борелевской функции c в задаче (2.2) даже ε-оптимальных стратегий из класса $\mathcal{U}$ может не существовать.

В связи с этим оказывается, что удобно рассматривать классы универсально измеримых функций. Расширим класс $\mathcal{L}$ до класса $\mathcal{M}$ всех универсально измеримых функций $c : \mathcal{X} \times [0,1] \mapsto \mathbb{R}_+$. Как известно: см. [26, предложение 9.19; 31, теорема 2], в этом случае ε-оптимальные стратегии также существуют в классе универсально измеримых функций $u : \mathcal{X} \mapsto [0,1]$, $u(x) \in \Gamma(x)$, который обозначим через $\mathcal{V}$. Напомним, что функция h, определенная на компактном множестве K, называется универсально измеримой, если она измерима относительно пополнения σ-алгебры $\mathcal{B}(K)$ относительно любой вероятностной меры.

Итак, рассмотрим множество η-оптимальных решений исполнителя в классе $\mathcal{V}$:

$$T_\eta(c) = \left\{ u \in \mathcal{V} : J_2(x,u,c) \geqslant \sup_{v \in \mathcal{V}} J_2(x,v,c) - \eta \right\}, \quad \eta \geqslant 0.$$

Следуя известному определению [32, определение 4.1] или [33, определение 11], элемент $\overline{c}_{\varepsilon,\eta} \in \mathcal{M}$ назовем (ε,η)-решением Штакельберга, если

$$\inf_{u \in T_\eta(\overline{c}_{\varepsilon,\eta})} J_1(x, \overline{c}_{\varepsilon,\eta}, u) \geqslant \sup_{c \in \mathcal{M}} \inf_{u \in T_\eta(c)} J_1(x,c,u) - \varepsilon.$$

Обозначим через $\mathcal{N}$ множество всех последовательностей $(\varepsilon_n, \eta_n) > 0$, удовлетворяющих условию $(\varepsilon_n, \eta_n) \to 0$. Величину

$$V_1(x) = \sup_{(\varepsilon_n,\eta_n) \in \mathcal{N}} \lim_{(\varepsilon_n,\eta_n) \to 0} \inf\{ J_1(x, \overline{c}_{\varepsilon_n,\eta_n}, u) : u \in T_{\eta_n}(\overline{c}_{\varepsilon_n,\eta_n}) \} \tag{4.2}$$

назовем ценой игры (лидера).

Заметим, что рассматриваемая игра не является корректной в смысле определений [32], так как равновесия Штакельберга не существует. Тем не менее справедлив следующий результат.

Теорема 2. Для описанной игры Штакельберга на $(\mathcal{M}, \mathcal{V})$ имеют место следующие утверждения:

i. Цена игры совпадает с оптимальным значением задачи (3.1)*:*

$$V_1(x) = V(x);$$

ii. При

$$\eta < \frac{1-\beta}{1+\beta}\varepsilon \tag{4.3}$$

функция (3.4) *является (ε,η)-решением и $T(\overline{c}_\varepsilon) = \{\overline{v}\}$.*

Доказательство. Верхняя оценка (3.2) цены игры сохраняет силу. Для доказательства теоремы 2 достаточно установить, что $T_\eta(\overline{c}_\varepsilon) = \{\overline{v}\}$ при вы-

полнении условия (4.3), так как тогда неравенство (3.7)

$$\inf_{u\in T_\eta(\overline{c}_\varepsilon)} J_1(x,\overline{c}_\varepsilon,u) = J_1(x,\overline{c}_\varepsilon,\overline{v}) \geqslant V(x)-\varepsilon$$

позволяет в определении (4.2) выбрать последовательность (ε_n,η_n), например, следующим образом: $(1/n,(1-\beta)/(2(1+\beta)n))$.

Доказательство теоремы 1 показывает, что $\overline{v}\in T(\overline{c}_\varepsilon)\subset T_\eta(\overline{c}_\varepsilon)$. Доказательство соотношения $T_\eta(\overline{c}_\varepsilon)=\overline{v}$ основано на идеях [34], которые в данном случае работают с существенными упрощениями. Пусть

$$H(x,a)=\varepsilon(1-\beta)I_{\{a=\overline{v}(x)\}}+\beta\int\limits_{\mathcal{X}} V_2(y;\overline{c}_\varepsilon)q(dy|x,a).$$

Рассмотрим произвольную η-оптимальную стратегию u исполнителя:

$$J_2(x,\overline{c}_\varepsilon,u)\geqslant V_2(x;\overline{c}_\varepsilon)-\eta.$$

Имеем

$$H(x,\overline{v}(x))=\varepsilon(1-\beta)+\beta\int\limits_{\mathcal{X}} V_2(y;\overline{c}_\varepsilon)q(dy|x,\overline{v}(x)),$$

$$H(x,u(x))=\varepsilon(1-\beta)I_{\{u(x)=\overline{v}(x)\}}+\beta\int\limits_{\mathcal{X}} V_2(y;\overline{c}_\varepsilon)q(dy|x,u(x)). \tag{4.4}$$

Но $V_2=\varepsilon$, поэтому

$$H(x,\overline{v}(x))-H(x,u(x))=\varepsilon(1-\beta)I_{\{u(x)\neq\overline{v}(x)\}}. \tag{4.5}$$

С другой стороны, с учетом оптимальности $\overline{v}$ и принципа динамического программирования имеем

$$H(x,\overline{v}(x))=V_2(x;\overline{c}_\varepsilon), \tag{4.6}$$

$$J_2(x,\overline{c}_\varepsilon,u)=\varepsilon(1-\beta)I_{\{u(x)=\overline{v}(x)\}}+\beta\int\limits_{\mathcal{X}} J_2(y,\overline{c}_\varepsilon,u)q(dy|x,u(x)). \tag{4.7}$$

Используя равенства (4.4), (4.6), (4.7), находим

$$|H(x,\overline{v}(x))-H(x,u(x))|=\Bigg|V_2(x;\overline{c}_\varepsilon)-J_2(x,\overline{c}_\varepsilon,u)+\beta\int\limits_{\mathcal{X}} J_2(y,\overline{c}_\varepsilon,u)q(dy|x,u(x))-$$

$$-\beta\int\limits_{\mathcal{X}} V_2(y;\overline{c}_\varepsilon)q(dy|x,u(x))\Bigg|\leqslant(1+\beta)\eta. \tag{4.8}$$

Последнее неравенство вытекает из свойства η-оптимальности u. Если $u(x)\neq v(x)$ при некотором x, то неравенства (4.5), (4.8) противоречат друг другу при выполнении условия (4.3). □

Полученный результат затрагивает также важную проблему, касающуюся последствий (с точки зрения регулятора) использования исполнителем приближенно оптимальных стратегий. Теорема 2 указывает на то, что при жестком регулировании, заложенном в стимулирующую стратегию (3.4), исполни-

тель может отклониться от управления $\overline{v}$, в котором заинтересован регулятор, только за счет больших убытков.

Всюду выше предполагалось, что игроки используют одинаковые коэффициенты дисконтирования. Предположим теперь, что исполнитель использует, вообще говоря, другой коэффициент дисконтирования γ:

$$J_2(x,c,u) = \mathsf{E}_{x,u} \sum_{t=0}^{\infty} \gamma^t (c(x_t,a_t) - g(x_t,a_t)) \to \max_{u\in\mathcal{U}}, \quad \gamma \in [0,1).$$

Из доказательств теорем 1 и 2 ясно, что для любого $\gamma \in [0,1)$ по-прежнему будут верны соотношения $\overline{v} \in T(\overline{c}_\varepsilon) \subset T_\eta(\overline{c}_\varepsilon)$ и $\{\overline{v}\} = T_\eta(\overline{c}_\varepsilon)$ при

$$\eta < \frac{1-\gamma}{1+\gamma}\varepsilon \tag{4.9}$$

и неравенство (3.7): $V_1 \geqslant V - \varepsilon$. Таким образом, цены V_1 обоих описанных выше игр будут не меньше V. Поскольку рассуждение, которое привело к верхней оценке (3.2) цены игры здесь неприменимо, возникает вопрос, может ли выигрыш регулятора быть больше?

Гипотеза 1. Теоремы 1 и 2 сохраняют силу (*при замене условия* (4.3) *условием* (4.9)) *при любом коэффициенте дисконтирования* $\gamma \in [0,1)$ *исполнителя.*

Покажем, что данная гипотеза верна в тривиальном случае "близорукого" исполнителя: $\gamma = 0$. Пусть $c \in \mathcal{L}$ или $c \in \mathcal{M}$. Для η-оптимального решения $u^* \in U$ задачи исполнителя

$$c(x,a) - g(x,a) \to \max_{a\in\Gamma(x)}$$

верно неравенство $c(x,u^*(x)) - g(x,u^*(x)) \geqslant c(x,0) - \eta$. Выигрыш регулятора тогда удовлетворяет оценке

$$\begin{aligned} J_1(x,c,u^*) &= \mathsf{E}_{x,u^*} \sum_{t=0}^{\infty} \beta^t (f(x_t,a_t) - c(x_t,a_t)) \leqslant \\ &\leqslant \mathsf{E}_{x,u^*} \sum_{t=0}^{\infty} \beta^t (f(x_t,a_t) - g(x_t,a_t) - c(x_t,0) + \eta) \leqslant V(x) + \eta. \end{aligned} \tag{4.10}$$

Рассмотренным в теоремах 1 и 2 постановкам задач соответствуют случаи $\eta = 0$ и $\eta > 0$. В любом из них из (4.10) получаем верхнюю оценку цены игры: $V_1 \leqslant V$.

Таким образом, регулятор не может получить выигрыш больше V в игре против близорукого исполнителя. Интуиция, подкрепляющая гипотезу 1, говорит о том, что ситуация не может стать лучше, если исполнитель не является близоруким.

5. Заключение

В статье предложена динамическая постановка базовой задачи стимулирования "регулятор — исполнитель". Желаемая для регулятора стратегия ис-

полнителя является решением задачи оптимального управления с критерием, равным разности между доходом регулятора и затратами исполнителя, а ε-оптимальная стратегия регулятора состоит в том, чтобы экономически мотивировать исполнителя следовать данной стратегии. Данное утверждение справедливо как для класса полунепрерывных сверху стимулирующих функций, так и для класса универсально измеримых функций и η-оптимальных стратегий исполнителя. Полученные результаты можно рассматривать как обобщение теорем 1 и 2 из [20] на модель стимулирования с марковской динамикой, бесконечным горизонтом планирования и дисконтированными критериями оптимальности, в которой регулятор располагает полной информацией о состоянии системы и действиях исполнителя, а исполнитель использует позиционные стратегии управления. Результат теоремы 2 имеет прямое отношение к сформулированной в [10] открытой проблеме робастности ε-оптимального решения в рамках обратной игры Штакельберга.

Представляется целесообразным продолжение исследований в следующих направлениях:

— анализ рассмотренной модели в непрерывном времени;
— изучение моделей с дискретным временем и несколькими (слабо и сильно связанными) агентами;
— исследование моделей с несколькими агентами в непрерывном времени;
— анализ моделей с неполной информацией, которые могут более адекватно описывать ряд ситуаций, например, в теории контрактов.

СПИСОК ЛИТЕРАТУРЫ

1. *von Stackelberg H.* Marktform und Gleichgewicht. Vienna: Springer, 1934.
2. *Basar T., Olsder G.J.* Dynamic noncooperative game theory. Philadelphia: SIAM, 1999.
3. *Dockner E., Jørgensen S., Van Long N., Sorger G.* Differential games in economics and management science. Cambridge: Cambridge University Press, 2000.
4. *Van Long, N.* A survey of dynamic games in economics. Singapore: World Scientific, 2010.
5. *Li T., Sethi S.P.* A Review of Dynamic Stackelberg Game Models // Discrete Cont. Dyn.–B. 2017. V. 22. No. 1. P. 125–159.
6. *Ho Y.-C., Luh P., Muralidharan R.* Information Structure, Stackelberg Games, and Incentive Controllability // IEEE Trans. Automat. Control. 1981. V. 26. No. 2. P. 454–460.
7. *Olsder G.J.* Phenomena in Inverse Stackelberg Games. Part 1: Static Problems // J. Optim. Theory Appl. 2009. V. 143. No. 3. P. 589–600.
8. *Olsder G.J.* Phenomena in Inverse Stackelberg Games. Part 2: Dynamic Problems // J. Optim. Theory Appl. 2009. V. 143. No. 3. P. 601–618.
9. *Groot N., De Schutter B., Hellendoorn H.* Reverse Stackelberg Games. Part I: Basic Framework / Control Applications (CCA), 2012 IEEE Int. Conf. on Control Applications. 2012. P. 421–426.
10. *Groot N., De Schutter B., Hellendoorn H.* Reverse Stackelberg Games. Part II: Results and Open Issues / Control Applications (CCA), 2012 IEEE Int. Conf. on Control Applications. 2012. P. 427–432.

11. *Гермейер Ю.Б.* Об играх двух лиц с фиксированной последовательностью ходов // Докл. АН СССР. 1971. Т. 198. № 5. С. 1001–1004.

12. *Гермейер Ю.Б.* Игры с непротивоположными интересами. М.: Наука, 1976.

13. *Кононенко А.Ф.* Теоретико-игровой анализ двухуровневой иерархической системы управления // Ж. вычисл. матем. и матем. физ. 1974. Т. 14. № 5. С. 1161–1170.

    Kononenko A.F. Game-theory Analysis of a Two-level Hierarchical Control System // USSR Comput. Math. Mathemat. Physics. 1974. V. 14. No. 5. P. 72–81.

14. *Горелов М.А., Кононенко А.Ф.* Динамические модели конфликтов. III. Иерархические игры // АиТ. 2015. № 2. С. 89–106.

    Gorelov M.A., Kononenko A.F. Dynamic Models of Conflicts. III. Hierarchical Games // Autom. Remote Control. 2015. V. 76. No. 2. P. 264–277.

15. *Shen H., Başar T.* Incentive-Based Pricing for Network Games with Complete and Incomplete Information / Jørgensen S., Quincampoix M., Vincent Th.L. (eds). Advances in Dynamic Game Theory: Numerical Methods, Algorithms, and Applications to Ecology and Economics. Boston: Birkhäuser, 2007. P. 431–458.

16. *Staňková K., Olsder G.J., Bliemer M.C.J.* Comparison of Different Toll Policies in the Dynamic Second-best Optimal Toll Design Problem. Case study on a three-link network // Eur. J. Transp. Infrast. Res. 2009. V. 4. No. 9. P. 331–346.

17. *Luh P., Ho Y., Muralidharan R.* Load Adaptive Pricing: An Emerging Tool for Electric Utilities // IEEE Trans. Autom. Control. 1982. V. 27. No. 2. P. 320–329.

18. *Burkov V.N., Goubko M., Korgin N., Novikov D.* Introduction to theory of control in organizations. Boca Raton: CRC Press, 2015.

19. *Новиков Д.А.* Стимулирование в социально-экономических системах (базовые математические модели). М.: ИПУ РАН, 1998.

20. *Новиков Д.А., Шохина Т.Е.* Механизмы стимулирования в динамических активных системах // АиТ. 2003. № 12. С. 94–104.

    Novikov D.A., Shokhina T.E. Incentive Mechanisms in Dynamic Active Systems // Autom. Remote Control. 2003. V. 64. No. 12. P. 1912–1921.

21. *Sundaram R.K.* A first course in optimization theory. Cambridge: Cambridge University Press, 1996.

22. *Papageorgiou N.S., Kyritsi-Yiallourou S.Th.* Handbook of applied analysis. Dordrecht: Springer, 2009.

23. *Hernández-Lerma O., Lasserre J.B.* Discrete-time Markov control processes: basic optimality criteria. N.Y.: Springer, 1996.

24. *Maitra A.* Discounted dynamic programming on compact metric spaces // Sankhyā: Indian J. Statist. Ser. A. 1968. V. 30. No. 2. P. 211–216.

25. *Schäl M.* Average Optimality in Dynamic Programming with General State Space // Math. Oper. Res. 1993. V. 18. No. 1. P. 163–172.

26. *Bertsekas D., Shreve S.* Stochastic optimal control: the discrete time case. Belmont: Athena Sci., 1996.

27. *Feinberg E.A., Lewis M.E.* Optimality Inequalities for Average Cost Markov Decision Processes and the Stochastic Cash Balance Problem // Math. Oper. Res. 2007. V. 32. No. 4. P. 769–783.

28. *Cruz-Suárez D., Montes-de-Oca R., Salem-Silva F.* Conditions for the Uniqueness of Optimal Policies of Discounted Markov Decision Processes // Math. Oper. Res. 2004. V. 60. No. 3. P. 415–436.

29. *Breton M., Alj A., Haurie A.* Sequential Stackelberg Equilibria in Two-person Games // J. Optim. Theory Appl. 1998. V. 59. No. 1. P. 71–97.

30. *Blackwell D.* Discounted Dynamic Programming // Ann. Math. Statist. 1965. V. 36. No. 1. P. 226–235.

31. *Shreve S.E., Bertsekas D.P.* Universally Measurable Policies in Dynamic Programming // Math. Oper. Res. 1979. V. 4. No. 1. P. 15–30.

32. *Morgan J.* Constrained well-posed two-level optimization problems / Clarke F.H., Dem'yanov V.F., Giannessi F. (eds). Nonsmooth optimization and related topics. Boston: Springer, 1989. P. 307–325.

33. *Patrone F.* Well-posedness for Nash equilibria and related topics / Lucchetti R., Revalski J. (eds). Recent developments in well-posed variational problems. Dordrecht: Springer, 1995. P. 211–227.

34. *Montes-De-Oca R., Lemus-Rodríguez E.* When are the Value Iteration Maximizers Close to an Optimal Stationary Policy of a Discounted Markov Decision Process? Closing the Gap between the Borel Space Theory and Actual Computations // WSEAS Trans. Math. 2010. V. 9. No. 3. P. 151–160.

Статья представлена к публикации членом редколлегии Е.Я. Рубиновичем.

Поступила в редакцию 09.08.2017