Доклады Академии наук СССР
1971. Том 198, № 4

УДК 62-50

КИБЕРНЕТИКА И ТЕОРИЯ РЕГУЛИРОВАНИЯ

Г. Д. ПЕНЕВ, В. А. ЯКУБОВИЧ

О НЕКОТОРЫХ ЗАДАЧАХ АДАПТИВНОГО УПРАВЛЕНИЯ

(Представлено академиком В. И. Смирновым 25 XII 1970)

1°. Рассмотрим дискретную систему управления

$$x_{t+1} = A(\xi)x_t + b(\xi)u_t + f(t, x_t, \xi), \quad \zeta_t = c(\xi)^* x_t, \tag{1}$$

где переменная t принимает значения $t = 0,1,\ldots$; ξ — неизвестный вектор варьируемых параметров, принимающий значение в заданном множестве M; $A(\xi)$ — квадратная матрица порядка n; u_t — вещественная переменная, называемая управлением; $f(t, x, \xi)$ — вещественная векторная функция («малое возмущение») со значением в $\mathbf{R}^n$; $b(\xi)$, $c(\xi) \in \mathbf{R}^n$, $x_t \in \mathbf{R}^n$; ζ_t — вещественная переменная.

Постановку задачи поясним на примере задачи построения «робота-велосипедиста», который должен сам обучаться кататься на велосипеде, двигающемся с постоянной скоростью. При этом параметры велосипеда и его скорость неизвестны конструктору робота; они могут принимать любые значения из заданного множества M. Задача построения адаптивной модели велосипедиста рассматривалась ранее в ([1], [2]). Здесь рассматривается более общий случай. Уравнения движения велосипеда в предположении, что скорость его постоянна, угол поворота руля является кусочно-постоянной функцией времени и при других естественных предположениях, сводятся к уравнениям вида (1). При этом ζ_t — угол между плоскостью рамы велосипеда и вертикальной плоскостью в момент времени t, u_t — угол поворота руля. Вначале велосипед устанавливается вертикально (задается значение x_0 так, что $|\zeta_0| < \varepsilon$, где $\varepsilon > 0$ — заданное число), затем движется, управляемый роботом в соответствии с уравнением (1). Это движение имеет место до тех пор, пока велосипед не упал, т. е. пока $|\zeta_t| \leqslant d$, где $d > \varepsilon$ — заданное число, и пока время этого движения (называемое временем первой игры) не превосходит заданного числа θ. Затем велосипед снова устанавливается вертикально (в указанном выше смысле) и движется до тех пор, пока велосипед не упал и пока время второй игры не превосходит θ, и т. д. Робот должен обучиться вырабатывать управления u_t так, чтобы $|u_t| \leqslant q$ для всех t, где $q > 0$ — заданное число, и так, чтобы для всех игр, начиная с некоторой, выполнялось требование $|\zeta_t| < \varepsilon$. Это должно иметь место для любого $\xi \in M$. Совершенно аналогично ставится задача в общем случае.

Уточним постановку задачи для общего случая. Определим сигнал игры μ_t следующим образом: $\mu_t = 1$, если $|\zeta_t| \leqslant d$ и если продолжительность текущей игры не превысила числа θ; в противном случае $\mu_t = 0$. Обозначим через $[t_k', t_k'']$ интервал k-й игры. (Считаем, что $t_1' = t'_{k+1} = t_k'' + 2$.) Обозначим через $x_k^0 = x_{t'_k}$, задаваемые начальные значения для k-й игры. Изменяя класс M, будем считать ε варьируемым параметром. Будем считать сенсором ([1], [2]) вектор $\sigma_t = \{w_t, \mu_t, \operatorname{sign}(\varepsilon - |\zeta_t|)\}$, где $w_t = \{\zeta_{t-n+1}, \ldots, \zeta_t, u_{t-n+1}, \ldots, u_{t-1}\}$. Подлежит определению множество $\{\tau\}$ (τ-«тактика») «уравнения мозга»

$$u_t = u(\sigma_t, \tau_t, t), \quad \tau_{t+1} = T(\tau_t, \sigma_t, \sigma_{t+1}, t). \tag{2}$$

Эти уравнения должны быть выбраны так, чтобы $|u_t| \leqslant q$, где q — заданное число, чтобы правые части (2) не зависели от варьируемых параметров, и так, чтобы для всех игр, начиная с некоторой, было выполнено «целевое условие» (ц.у.): если $\mu_t = 1$, то $|\zeta_{t+1}| < \varepsilon$. Если это выполняется для любого $\xi \in M$, то будем говорить, что робот разумен в классе задач M. Отметим, что в теореме 3 будет рассмотрено и другое целевое условие.

Введем обозначения: $\det(\lambda I_n - A) = \delta_n\lambda^n + \ldots + \delta_0$ $(\delta_n = 1)$, $\pi_k = c^*A^kb$, $k = 0, 1, \ldots, n-1$, $\operatorname{sign}\pi_0 = s$, $\alpha_k = -\delta_{k-1}$ $(k = 1, \ldots, n)$, $\alpha_{n+k} = \pi_0\delta_k + \ldots + \pi_{n-k}\delta_n$ $(k = 1, \ldots, n-1)$, $a = \|\alpha_k\|$, $k = 1, \ldots, 2n-1$, $\tau_* = -\pi_0^{-1}a$, $g_{t+k-1} = A^{k-1}f_t + A^{k-2}f_{t+1} + \ldots + f_{t+k-1}$, $\psi_t = \delta_1 c^*g_{t-n+1} + \ldots$ $\ldots + \delta_n c^*g_t$. Через r обозначим число исправлений тактики.

Теорема 1. *Пусть выполнены предположения:*

I) *для каждой игры* $x_j^0 \in E = \{x \in \mathbf{R}^n \colon |c^*A^kx| < \varepsilon' < \varepsilon,\ k = 0, 1, \ldots$ $\ldots, n-1\}$;

II) *пусть* $0 < \rho < 1$. *Возмущение* f_t *настолько мало, что для любого* t *и для любых* x_{t+k}, $k = 0, 1, \ldots, n-1$, *для которых* $|\zeta_{t+k}| \leqslant d$, *выполнено неравенство* $|\psi_t| \leqslant \rho\varepsilon$, *а если* $x_t \in E$, *то и* $|g_{t+k-1}| \leqslant \varepsilon - \varepsilon'$;

III) *выполнено* $0 < \varepsilon \leqslant d$, $0 < |\pi_0| \leqslant \pi'$ *и величины* $s = \operatorname{sign}\pi_0$, d, π' *известны (не варьируемые параметры)*;

IV) *для любого* $\xi \in M$ *и для некоторого заданного числа* $q_* < q$ *выполнено неравенство* $|a|D \leqslant q_*|\pi_0|$, *где* $D = (nd^2 + (n-1)q^2)^{1/2}$. *Положим* $\varkappa = 4\pi'D^2$, $\beta = \varkappa^{-1}(1-\rho^2)$.

Тогда при любом $\tau_0 \in \mathbf{R}^{2n-1}$ *и* $j = 1, 2, \ldots$ *следующие уравнения являются «уравнениями мозга», разумного в классе задач* M *робота*:

$$u_t = \begin{cases} 0, & \text{если } t \in [t_j', t_j' + n - 2], \\ w_t^*\tau_t, & \text{если } t \in [t_j' + n - 1, t_j'']; \end{cases} \tag{3}$$

$\tau_{t+1} = \tau_t$, *если либо* $\mu_t = 0$, *либо* $t \in [t_j', t_j' + n - 3]$, *либо* $|\zeta_{t+1}| < \varepsilon$ *и* $|w_{t+1}^*\tau_t| \leqslant q$;

$\tau_{t+1} = |\tau_t|^{-1}D^{-1}q_*\tau_t$, *если* $t \in [t_j' + n - 2, t_j'']$, $|\zeta_{t+1}| < \varepsilon$ *и* $w_{t+1}^*\tau_t| > q$; $\tau_{t+1} = \tau_t - 2\beta s\zeta_{t+1}w_t$, *если либо* $t = t_j''$ *и* $|\zeta_{t+1}| \geqslant \varepsilon$, *либо* $t \in$ $\in [t_j' + n - 2, t_j'' - 1)$, $|\zeta_{t+1}| \geqslant \varepsilon$ *и* $|w_{t+1}^*(\tau_t - 2\beta s\zeta_{t+1}w_t)| \leqslant q$;

$\tau_{t+1} = |\tau_t - 2\beta s\zeta_{t+1}w_t|^{-1}D^{-1}q_*(\tau_t - 2\beta s\zeta_{t+1}w_t)$, *если* $t \in [t_j' + n - 2, t_j'' - 1]$, $|\zeta_{t+1}| \geqslant \varepsilon$ *и* $|w_{t+1}^*(\tau_t - 2\beta s\zeta_{t+1}w_t)| > q$. *При этом для любого* $t \in [t_j', t_j'']$, $j = 1, 2, \ldots$, *выполнено неравенство* $|u_t| \leqslant q$. *Для числа* r *и, тем более, для числа моментов времени, при которых не выполнено ц.у. справедлива оценка*

$r \leqslant |\tau_0 - \tau_*|^2\delta^{-1}$, где $\delta = \min\{D^{-2}(q - q_*)^2,\ |\pi_0|^{-1}\varkappa^{-1}(1-\rho^2)^2\varepsilon^2\}$.

Теорема 1 дает решение поставленной выше задачи в предположении, что для любого ζ_t, для которого $|\zeta_t| \leqslant d$, существует «идеальное» управление $u_t^{\text{ид}}$ такое, что $|\zeta_{t+1}| < \varepsilon$. Это жесткое требование далеко не всегда выполняется. Следующая теорема дает решение той же задачи в более слабом предположении существования $u_t^{\text{ид}}$ такого, что если $\varepsilon\gamma^{-1} \leqslant |\zeta_t| \leqslant d$, то $|\zeta_{t+1}| < \gamma|\zeta_t|$, а если $|\zeta_t| < \varepsilon\gamma^{-1}$, то $|\zeta_{t+1}| < \varepsilon$. Здесь $0 < \gamma < 1$. Обозначим $\varepsilon_t = \gamma|\zeta_t|$, если $\varepsilon\gamma^{-1} \leqslant |\zeta_t| \leqslant d$; $\varepsilon_t = \varepsilon$, если $|\zeta_t| < \varepsilon\gamma^{-1}$.

Теорема 2. *Пусть выполнены предположения* I, III *теоремы* 1 *и следующие предположения:*

V) *возмущение* f_t *настолько мало, что для любого* t *и для любых* x_{t+k}, $k = 0, 1, \ldots, n-1$, *для которых* $|\zeta_{t+k}| \leqslant d$ *выполнено неравенство* $|\psi_t| \leqslant \psi = \text{const}$, *а если* $x_t \in E$, *то и* $|g_{t+k-1}| \leqslant \varepsilon - \varepsilon'$; *здесь* ε' — *число в условии* I, *а значение* ψ *используется ниже*;

VI) *пусть* $0 < \gamma_* < \gamma < 1$. *Для любого* $\xi \in M$ *существует* $K > 1$ *такое, что выполнены неравенства*

$$D|a| \leqslant K|\pi_0|q_*, \quad |\alpha_n|\varepsilon + d\pi(|\alpha_1| + \ldots + |\alpha_{n-1}|) +$$
$$+ q\gamma(|\alpha_{n+1}| + \ldots + |\alpha_{2n-1}|) + K\gamma\psi(K-1)^{-1} \leqslant K\gamma_*\varepsilon(K-1)^{-1}.$$

Положим $\varkappa = 4\pi' D^2$, $\beta' = \varkappa^{-1}(1 - \gamma_*^2\gamma^{-2})$. *Определим первое «уравнение мозга» соотношением* (3), *второе «уравнение мозга» — соотношениями*

$\tau_{t+1} = \tau_t$, *если либо* $\mu_t = 0$, *либо* $t \in [t_j', t_j' + n - 3]$, *либо* $|\zeta_{t+1}| < \varepsilon_t$ *и* $|w_{t+1}^* \tau_t| \leqslant q$;

$\tau_{t+1} = |\tau_t|^{-1} D^{-1} q_* \tau_t$, *если* $|\zeta_{t+1}| < \varepsilon_t$ *и* $|w_{t+1}^* \tau_t| > q$;

$\tau_{t+1} = \tau_t - 2\beta' s \zeta_{t+1} w_t$, *если либо* $|\zeta_{t+1}| > d$, *либо* $\varepsilon_t \leqslant |\zeta_{t+1}| \leqslant d$ *и* $|w_{t+1}^* (\tau_t - 2\beta' s \zeta_{t+1} w_t)| \leqslant q$;

$\tau_{t+1} = |\tau_t - 2\beta' s \zeta_{t+1} w_t|^{-1} D^{-1} q_* (\tau_t - 2\beta' s \zeta_{t+1} w_t)$, *если* $\varepsilon_t \leqslant |\zeta_{t+1}| \leqslant d$ *и* $|w_{t+1}^* (\tau_t - 2\beta' s \zeta_{t+1} w_t)| > q$. *Здесь* τ_0 — *любой* $(2n - 1)$-*мерный вектор.*

Тогда робот разумен в классе задач М. При этом выполнено неравенство $|u_t| \leqslant q$ при $t \in [t_j', t_j'']$, $j = 1, 2, \ldots$ Для числа r *справедлива оценка, указанная в теореме* 1 *при* $\rho = \gamma_* \gamma^{-1}$ *и* $\tau_* = -\pi_0^{-1} K^{-1} a$.

В теоремах 1 и 2 предполагалось, что в начале каждой игры «велосипед устанавливается вертикально», именно, предполагалось, что векторы начального состояния настолько малы, что в течение первых n моментов каждой игры выполнено неравенство $|\zeta_t| < \varepsilon$. Следующая теорема относится к случаю, когда в начале каждой игры велосипед устанавливается, вообще, не вертикально, а робот должен обучиться «выравнивать» велосипед в течение каждой игры. Именно, в общем случае вектор начального состояния может быть большим, но «не слишком»: он должен быть таким, чтобы в течение n первых моментов при отсутствии управления ($u_t = 0$) «велосипед бы не упал», т. е. в общем случае, чтобы было выполнено $|\zeta_t| \leqslant d$ для n первых моментов игры. При этом, естественно, меняется и ц.у.: для заданного целого m (заданная длина переходного режима) должно быть выполнено $|\zeta_t| < \varepsilon$ спустя $n + m$ моментов после начала игры. Робот называется **разумным** в классе M, если для любого $\xi \in M$ целевое условие выполняется для всех игр, начиная с некоторой.

Теорема 3. *Пусть выполнены* III, VI *и предположения* I, V *с заменой* ε *на* d. *Пусть целое* $m > 0$ *такое, что* $m + n < \theta$ *и* $d\gamma^{m+1} < \varepsilon$ *при любом* $\xi \in M$.

Тогда уравнения, указанные в теореме 2, *дополненные соотношениями* $\tau_{t+1} = \tau_t$, *если* $t = t_j' + n - 2$ *и* $|w_{t+1}^* \tau_t| \leqslant q$; $\tau_{t+1} = |\tau_t|^{-1} D^{-1} q_* \tau_t$, *если* $t = t_j' + n - 2$ *и* $|w_{t+1}^* \tau_t| > q$, *является «уравнениями мозга» робота, разумного в классе задач М. Для числа исправлений тактики и, тем более, для числа игр, для которых не выполнено ц.у., справедлива оценка, указанная в теореме* 2.

2°. Ниже пояснена лишь идея доказательства этих теорем. В силу сделанных предположений $t_j'' > t_j' + n - 1$ при любом $j = 1, 2, \ldots$ Для $t \in [t_j' + n - 1, t_j'']$, пользуясь теоремой Гамильтониана — Кэли, получим равенство $\zeta_{t+1} = \pi_0 u_t + \eta_t$, где $\eta_t = w_t^* a + \psi_t$. Ищем уравнение в виде $u_t = w_t^* \tau_t$ (первое «уравнение мозга»). Пусть $t \in [t_j' + n - 2, t_j'']$. В случае теоремы 1 ц.у. $|\zeta_{t+1}| < \varepsilon$ примет вид $|\pi_0 w_t^* \tau_t + \eta_t| < \varepsilon$. В случае теоремы 2, 3 имеем $|\pi_0 w_t^* \tau_t + \eta_t| < \alpha_t$, где $\alpha_t = \gamma|\zeta_t|$, если $|\zeta_t| \geqslant \varepsilon\gamma^{-1}$, и $\alpha_t = \varepsilon$, если $|\zeta_t| < \varepsilon\gamma^{-1}$. Мы пришли к ситуации, разобранной в [3].

Алгоритм определения τ_t (второе «уравнение мозга») должен быть выбран так, чтобы $\tau_t = \mathrm{const}$ при $t \geqslant t_0$ и чтобы все неравенства $|\pi_0 w_t^* \tau_t + \eta_t| < \alpha_t$ были выполнены при $t \geqslant t_0$. (При этом w_t, η_t, α_t зависят от выбранного алгорифма.) Такие «конечно сходящиеся» алгорифмы приведены в [3, 4]. Однако в данном случае имеются дополнительные требования. Алгорифм должен быть таким, чтобы: 1) значение τ_{t+1} не зависело от варьируемых параметров, 2) было обеспечено условие $|u_t| \leqslant q$. В теоремах 1—3 используется алгорифм теоремы 4 [3], «исправленный» следующим образом. Если доставляемое им значение $\tau_{t+1} = \tau_{t+1}'$ таково, что $|u_{t+1}| > q$, то в качестве τ_{t+1} берем точку сферы $|\tau| = D^{-1} q_*$, ближайшую к точке τ_{t+1}'. Поскольку $|\tau_*| \leqslant D^{-1} q_*$, то $|\tau_{t+1} - \tau_*|^2 \leqslant |\tau_t - \tau_*|^2 - \delta$, где $\delta > 0$. Это же неравенство выполнено для значений τ_{t+1}, доставляемых алгорифмом теоремы 4 ([3]). Из этого неравенства следует, что число исправлений тактик и, тем более, число значений t, для которых не выполнено неравенство $|\pi_0 w_t^* \tau_t + \eta_t| < \alpha_t$ (а значит, и ц.у.), конечно. Оценка для числа r следует из неравенства $|\tau_0 - \tau_*|^2 \geqslant r\delta$.

Заметим, что в случае, когда π_0 не зависит от варьируемых параметров, для построения второго «уравнения мозга» целесообразнее использовать теорему 4 ([4]).

3°. Были проведены эксперименты на ЭВМ по моделированию процесса самообучения робота-велосипедиста, «уравнения мозга» которого были взяты из теоремы 1. При этом были взяты параметры обычного велосипеда. Велосипедист «упал» шесть раз и, начиная с седьмой игры, обучился кататься. В других экспериментах было введено запаздывание для управления $\Delta t = 0{,}01 - 0{,}1$ сек. (Уравнения движения имели снова вид (1).) Велосипедист обучался кататься 10—20 игр.

Поступило
17 XII 1970

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] В. А. Якубович, ДАН, 182, № 3 (1968). [2] В. А. Якубович, ДАН, 183, № 2 (1968). [3] В. А. Якубович, ДАН, 189, № 3 (1969). [4] В. А. Якубович, ДАН, 166, № 6 (1966).