

МОДЕЛІ РЕЗОНАНСНОГО КЕРУВАННЯ НЕЛІНІЙНИМИ ГАМІЛЬТОНОВИМИ СИСТЕМАМИ

Н. М. Моніна^{1, а}, С. А. Смирнов^{1, б}, П. О. Наказной^{1, в}

¹ Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»,
НН Фізико-технічний інститут

Анотація

Метою роботи є побудова алгоритмів оптимального керування цілком інтегровними нелінійними гамільтоновими системами. На прикладі модельної задачі проаналізовано можливість застосування методу швидкісного градієнта та алгоритму навчання з підкріпленням. Для представлення динамічних систем було використано змінні дія-кут. В роботі виконано теоретичний аналіз збурюючого впливу резонансного керування на обрану систему, створено програмні реалізації зазначених методів та проведено комп'ютерне моделювання еволюції досліджуваної системи з побудовою оптимального керування в режимі реального часу.

Ключові слова: гамільтонові системи, теорія збурень, оптимальне нелінійне керування, метод швидкісного градієнта, навчання з підкріпленням, нейронні мережі, змінні дія-кут, комп'ютерне моделювання

Вступ

Моделювання явищ навколишнього світу часто призводить до нелінійних диференціальних рівнянь [1]. Оскільки аналітичні методи інтегрування нелінійних моделей розроблені недостатньо, аналіз отриманих моделей є достатньо складним. Ця проблема, зокрема, виникає при керуванні нелінійними процесами, що виникають на виробництві, в економіці, екології, робототехніці тощо [2]. Насамперед, актуальною проблемою є аналіз відгуків системи на керуюче збудження та побудова алгоритмів керування на основі оберненого зв'язку для досягнення цілей керування.

У даній роботі розглядатимуться деякі методи резонансного керування в режимі реального часу нелінійними гамільтоновими системами, що є цілком інтегровними.

1. Метод змінних «дія-кут» для представлення гамільтонових систем

Будемо розглядати цілком інтегровну гамільтонову систему, тоді її рух задається рівняннями Гамільтона [3]:

$$\begin{cases} \dot{q}_j = \frac{\partial H_0}{\partial p_j}, \\ \dot{p}_j = -\frac{\partial H_0}{\partial q_j}, \end{cases} \quad (j = \overline{1, \dots, N}) \quad (1)$$

тут q_i, p_j – узагальнені координати та імпульси відповідно, N – кількість ступенів свободи системи, $H_0(q_1, \dots, q_n, p_1, \dots, p_N)$ – функція Гамільтона

Тоді змінні «дія-кут» (I_i, θ_i) вводяться наступними співвідношеннями [4]:

$$\begin{cases} I_i = \frac{1}{2\pi} \oint p_i dq_i = I_i(p, q), \\ \theta_i = \frac{\partial S(q, I)}{\partial I_i} = \theta_i(p, q), \end{cases} \quad (2)$$

де $S(q, I) = \int^q p(q, H_0) dq$, — скорочена «дія». Для системи (2) виконується, що H_0 не залежить від часу t явно, та система здійснює фінітний рух.

Рівняння Гамільтона для незбуреної системи (1) в змінних (I, θ) мають вигляд:

$$\begin{cases} \dot{I}_i = -\frac{\partial H_0(I)}{\partial \theta_i} = 0, \\ \dot{\theta}_i = \frac{dH_0(I)}{dI_i} \equiv \omega_i(I), \end{cases} \quad (3)$$

звідки:

$$I_i = \text{Const}; \quad \theta_i = \omega_i(I)t + \theta_{i0}. \quad (4)$$

2. Резонансне керування гамільтоновими системами

В даній роботі буде розглядатись модель, в якій цілком інтегровна система з $H_0 = H_0(I_1, \dots, I_N) \equiv H_0(I)$ збурюється внаслідок резонансного керування:

$$H = H_0(I) + \varepsilon(t)V(I, \theta; t), \quad (5)$$

де $\varepsilon(t)$ – змінний малий параметр керування ($|\varepsilon| < 1$), а $V(I, \theta; t)$ – вид потенціалу збурення [4].

В якості моделі для прикладу розглянемо осцилятор з кубічною нелінійністю, що збурюється внаслідок керування, яке містить резонансний член:

$$H = I_1 a_1 + I_2 a_2 + \frac{1}{30} \left(I_1^3 + \frac{1}{2} I_2^3 \right) + \varepsilon \cos(\theta_2 - \theta_1). \quad (6)$$

^аnatmon-ipt22@iit.kpi.ua

^бs.smirnov@kpi.ua

^вp.nakaznoy@kpi.ua

Рівняння руху для неї набувають вигляду:

$$\begin{aligned}\dot{I}_1 &= -\varepsilon \sin(\theta_2 - \theta_1), \\ \dot{I}_2 &= \varepsilon \sin(\theta_2 - \theta_1), \\ \dot{\theta}_1 &= \frac{1}{10} I_1^2 + a_1, \\ \dot{\theta}_2 &= \frac{1}{20} I_2^2 + a_2.\end{aligned}\quad (7)$$

За допомогою $\varepsilon \cos(\theta_2 - \theta_1)$ здійснюється резонансне керування системи, тому необхідно аналітично дослідити його вплив.

З рівнянь 7 одразу випливає, що величина

$$I_1 + I_2 = \text{const} \quad (8)$$

зберігається.

Нехай загальна ціль керування, відповідно до виконання умови (8), задається граничним співвідношенням досягнення величини I_1 заданого значення I_1^* :

$$Q = \lim_{t \rightarrow \infty} \frac{1}{2} [I_1 - I_1^*]^2 = 0. \quad (9)$$

2.1. Теоретичний аналіз миттєвого впливу керування

Для побудови чисельної моделі оберемо проміжок часу, на якому буде проводитись моделювання, та розіб'ємо його на частини, що відповідатимуть моментам часу, в які будуть проводитись виміри. Будемо вважати, що всі стани вимірюються.

Побудуємо аналітичну модель впливу миттєвого збудження моделі (6) із застосуванням теорії збурень за параметром ε [4]. Оберемо момент часу t_k , та будемо оцінювати еволюцію системи в інтервалі до t_{k+1} . Обмежимося лише першими двома членами розкладу. Тоді маємо в нульовому наближенні:

$$\begin{aligned}\dot{I}_1^{(0)}(t) &= 0, \\ \dot{I}_2^{(0)}(t) &= 0, \\ \dot{\theta}_1^{(0)}(t) &= \frac{1}{10} I_1^2 + a_1 \equiv \omega_1(I_1), \\ \dot{\theta}_2^{(0)}(t) &= \frac{1}{20} I_2^2 + a_2 \equiv \omega_2(I_2).\end{aligned}\quad (10)$$

Проводячи процедуру інтегрування:

$$\begin{aligned}I_1^{(0)}(t) &= I_1(t_k) = \text{const}, \\ I_2^{(0)}(t) &= I_2(t_k) = \text{const}, \\ \theta_1^{(0)}(t) &= \theta_1(t_k) + \left[\frac{1}{10} I_1^2(t_k) + a_1 \right] (t - t_k), \\ \theta_2^{(0)}(t) &= \theta_2(t_k) + \left[\frac{1}{20} I_2^2(t_k) + a_2 \right] (t - t_k), \\ t &\in (t_k, t_{k+1}).\end{aligned}\quad (11)$$

Для знаходження першого наближення будемо підставляти в систему рівнянь нульові наближення. Отримуємо вирази для дій та кутів:

$$\begin{aligned}\dot{I}_1^{(1)}(t) &= -\varepsilon \sin([\omega_2(I_2^{(0)}) - \omega_1(I_1^{(0)})]t), \\ \dot{I}_2^{(0)}(t) &= \varepsilon \sin([\omega_2(I_2^{(0)}) - \omega_1(I_1^{(0)})]t), \\ \dot{\theta}_1^{(1)}(t) &= \frac{d\omega_1(I_1^{(0)})}{dI_1} I_1^{(1)} = \frac{1}{5} I_1^{(0)}(t) I_1^{(1)}(t), \\ \dot{\theta}_2^{(1)}(t) &= \frac{d\omega_1(I_2^{(0)})}{dI_2} I_2^{(1)} = \frac{1}{10} I_2^{(0)}(t) I_2^{(1)}(t).\end{aligned}\quad (12)$$

Інтегруючи дані вирази та враховуючи, що $I^{(0)} = \text{const} = I(t_k)$, обраховуються перші наближення I :

$$\begin{aligned}I_1^{(1)}(t) &= A \cos(\Delta\omega t) - B, \\ I_2^{(1)}(t) &= -A \cos(\Delta\omega t) + B, \\ \Delta\omega &= \omega_2(I_2(t_k)) - \omega_1(I_1(t_k)), \\ A &= \varepsilon / \Delta\omega, \\ B &= \varepsilon \cos(\Delta\omega t_k) / \Delta\omega,\end{aligned}\quad (13)$$

та θ :

$$\begin{aligned}\theta_1^{(1)}(t) &= I_1(t_k) \frac{A}{5\Delta\omega} [\sin(\Delta\omega t) - C] - B(t - t_k), \\ \theta_2^{(1)}(t) &= -I_2(t_k) \frac{A}{10\Delta\omega} [\sin(\Delta\omega t) - C] + B(t - t_k), \\ C &= \sin(\Delta\omega t_k).\end{aligned}\quad (14)$$

Звідси отримуємо наближені аналітичні вирази для впливу керування на майбутню еволюцію:

$$\begin{aligned}I_1(t) &\approx I_1(t_k) + \varepsilon (A \cos(\Delta\omega t) - B), \\ I_2(t) &\approx I_2(t_k) - \varepsilon (A \cos(\Delta\omega t) - B), \\ \theta_1(t) &\approx \theta_1(t_k) + \left[\frac{1}{10} I_1^2(t_k) + a_1 - \varepsilon B \right] (t - t_k) + \\ &\quad + \varepsilon I_1(t_k) \frac{A}{5\Delta\omega} [\sin(\Delta\omega t) - C], \\ \theta_2(t) &\approx \theta_2(t_k) + \left[\frac{1}{20} I_2^2(t_k) + a_2 + \varepsilon B \right] (t - t_k) - \\ &\quad - \varepsilon I_2(t_k) \frac{A}{10\Delta\omega} [\sin(\Delta\omega t) - C].\end{aligned}\quad (15)$$

З отриманих апроксимацій, миттєві значення для дій набувають гармонічний характер, при цьому в самому наближенні зберігається інваріант (8).

2.2. Метод швидкісного градієнта

Для побудови алгоритмів оптимізації для систем афінних за керуванням зручно подавати їх у наступній формі з заміною $x = (I_1, I_2, \theta_1, \theta_2)$:

$$\dot{x} = f(x) + g(x)u = \begin{bmatrix} 0 \\ 0 \\ a_1 + \frac{x_1^2}{10} \\ a_2 + \frac{x_2^2}{20} \end{bmatrix} + \begin{bmatrix} -\sin(x_4 - x_3) \\ \sin(x_4 - x_3) \\ 0 \\ 0 \end{bmatrix} u. \quad (16)$$

Застосуємо метод швидкісного градієнта [5, стор. 35] для оптимізації задачі (16). Як критерій керування обрано саму ціль регулювання 9. Тоді швидкість зміни критерію:

$$\begin{aligned}\alpha &= \dot{Q} = \nabla_x Q \cdot \dot{x}, \\ \alpha &= (x_1 - x_1^*) (-\sin(x_4 - x_3)) u.\end{aligned}\quad (17)$$

Відповідно до цього миттєве покращення $u = u(t)$ набуває вигляду:

$$\frac{du}{dt} = -\gamma \nabla_u \alpha = \gamma (x_1 - x_1^*) \sin(x_4 - x_3). \quad (18)$$

У дискретному вигляді функція керування являє собою східчасту структуру та будується наступним чином:

$$u(t_{k+1}) = u_{k+1} = u_k + \frac{du}{dt}(t_k) \cdot \Delta, \quad (19)$$

де Δ — інтервал між вимірами t_k та t_{k+1} .

2.3. Оптимізація керування за допомогою навчання з підкріпленням

Для розв'язку задачі (16) за допомогою model-based навчання з підкріпленням [6] для задоволення цілі (9) оберемо в якості критерію керування інтегральне співвідношення вигляду:

$$V(x) = \inf_u \int_{t_0}^{t_f} r(x(\tau; t_0, x_0, u), u(\tau)) d\tau, \quad (20)$$

$$r(x) = \frac{1}{2} [x_1 - x_1^*]^2 + Ru^2.$$

При цьому вважається, що задачу можна розбивати на частини за припущенням Беллмана [7]:

$$V(x(t_k)) = \inf_u \int_{t_k}^{t_{k+1}} r(x, u) d\tau + V(x(t_{k+1})). \quad (21)$$

Основна ідея розв'язання задачі полягає у використанні апроксимацій розв'язків $V(x)$ та $u(x)$ за допомогою нейронних мереж з нелінійними активаційними функціями.

$$V(x) \rightarrow \hat{V}(x, W_c) \equiv W_c \cdot \sigma(x), \quad (22)$$

$$u(x) \rightarrow \hat{u}(x, W_a) \equiv -\frac{1}{2} R^{-1} g^T(x) \nabla_x \sigma^T(x) W_a,$$

де $\sigma(x) \in \mathbf{R}^L$ — L -вимірний набір активаційних функцій, а W_c та W_a — оцінки ваг для W , що рівномірно обмежено апроксимує розв'язок $V(x) = W \cdot \sigma(x) + \varepsilon(x)$, а $\varepsilon(x)$ — похибка наближення.

Для розв'язання задачі оптимальності намагаються мінімізувати миттєву похибку Беллмана:

$$\delta(t) = \nabla_x \hat{V}(x, W_c(t)) (f(x) + g(x) \hat{u}(x, W_a(t))) + r(x, \hat{u}(x, W_a(t))), \quad (23)$$

причому кожна оцінка даної похибки виступає як отриманий «досвід».

Для апроксимації задачі (16) обрано ядро $\sigma : \mathbf{R}^N \rightarrow \mathbf{R}^L$, що складається функцій $x_1^2, x_1 x_2, x_2^2, \sin(x_3), \sin(x_4), \cos(x_3), \cos(x_4), \sin(x_4 - x_3), \cos(x_4 - x_3)$ та обчислюватиметься в точках $x = (x_1 - x_1^*, x_2 + x_1^*, x_3, x_4) = (I_1 - I_1^*, I_2 + I_1^*, \theta_1, \theta_2)$.

Тепер перейдемо до безпосереднього опису алгоритму. Оскільки навчання з підкріпленням передбачає оптимізацію мереж на основі зроблених спостережень, тому для кращої оцінки параметрів W_c та W_a можна сформулювати задачу мінімізації похибки Беллмана (23) не тільки для поточного стану $x(t)$, але й екстраполювати її на деяку множину близьких станів $x_i, i = (1, \dots, N)$.

Відповідно до аналізу стаціонарності вагових розв'язків [6, стор. 106] ваги W_c для оцінки критерію втрат оновлюються як:

$$\dot{W}_c(t) = -\eta_{c1} \Gamma(t) \frac{w(t)}{\rho(t)} \delta(t) - \frac{\eta_{c2}}{N} \Gamma(t) \sum_{i=1}^N \frac{w_i(t)}{\rho_i(t)} \delta_i(t), \quad (24)$$

$$\dot{\Gamma}(t) = \left(\beta \Gamma(t) - \eta_{c1} \frac{\Gamma(t) w(t) w(t)^T \Gamma(t)}{\rho^2(t)} \right) \mathbf{1}_{\|\Gamma\| \leq \bar{\Gamma}}, \quad (25)$$

де $\Gamma : \mathbf{R}_{t>t_0} \rightarrow \mathbf{R}^{L \times L}$ — матриця приросту найменших квадратів; $\bar{\Gamma} > 0$ — граничне значення її норми;

$$w, w_i \in \mathbf{R}^L \text{ — вектори регресії, що визначають як}$$

$$w(t) = \nabla_x \sigma(x(t)) (f(x(t)) + g(x(t)) \cdot \hat{u}(x(t), W_a(t))),$$

$$w_i(t) = \nabla_x \sigma(x_i) (f(x_i) + g(x_i) \cdot \hat{u}(x_i, W_a(t))),$$

$$\rho(t) = 1 + \nu w^T(t) \Gamma(t) w(t),$$

$$\rho_i(t) = 1 + \nu w_i^T(t) \Gamma(t) w_i(t),$$

$$i = (1, \dots, N) \quad (26)$$

ν — константа нормалізації приросту; $\beta > 0$ — фактор «забуття»; та $\eta_{c1}, \eta_{c2} > 0$ — константи адаптаційних приростів.

Закон оновлення ваг W_a , що використовуються для побудови керування як функції $u(x(t))$ задається наступним чином:

$$\dot{W}_a(t) = -\eta_{a1} (W_a(t) - W_c(t)) - \eta_{a2} W_a(t) +$$

$$+ \eta_{c1} \frac{G_\sigma^T(t) W_a(t) w^T(t)}{4\rho(t)} W_c(t) +$$

$$+ \eta_{c2} \sum_{i=1}^N \frac{G_{\sigma i}^T W_a(t) w_i^T(t)}{4\rho_i(t)} W_c(t), \quad (27)$$

де $\eta_{a1}, \eta_{a2} > 0$ — адаптаційні прирости; $G_\sigma \in \mathbf{R}^{L \times L}$

$$G_\sigma(t) = \nabla_x \sigma(x(t)) g(x(t)) R^{-1} g^T(x(t)) \nabla_x \sigma^T(x(t)),$$

$$G_{\sigma i} = \nabla_x \sigma(x_i) g(x_i) R^{-1} g^T(x_i) \nabla_x \sigma^T(x_i). \quad (28)$$

Таким чином, аналогічно до алгоритму швидкісного градієнта, в дискретні моменти часу $t_k = \Delta k$ вимірюється вектор станів системи, оновлюються вектори вагових коефіцієнтів W_c, W_a та матриця Γ за методом Ейлера:

$$W_c(t_{k+1}) = W_{ck+1} = W_{ck} + \frac{dW_c}{dt}(t_k) \cdot \Delta,$$

$$W_a(t_{k+1}) = W_{ak+1} = W_{ak} + \frac{dW_a}{dt}(t_k) \cdot \Delta, \quad (29)$$

$$\Gamma(t_{k+1}) = \Gamma_{k+1} = \Gamma_k + \frac{d\Gamma}{dt}(t_k) \cdot \Delta.$$

3. Результати моделювання

Для всіх ситуацій моделювання було обрано проміжок часу рівним 100 секунд. Початкова система ініціалізувалась відповідно до (табл. 1).

Табл. 1. Параметри моделювання системи (7)

a_1	a_2	$I_1(0)$	$I_2(0)$	$\theta_1(0)$	$\theta_2(0)$	I_1^*
10	20	6	4	0	0	8

Було проведено серію випробувань градієнтного методу із використанням різних показників γ . Результати роботи методу з обраним коефіцієнтом $\gamma = 0.3$ та первинним значенням $u = 0.1$ зображено на (рис. 1).

Швидкісний градієнт намагається локально мінімізувати значення свого критерію, проте не досягає цілі керування глобально. Це можна пояснити тим, що даний метод за побудовою не може забезпечити швидку осциляцію малого параметра керування.

Під час тренування нейронної мережі обиралися різні коефіцієнти адаптації $\eta_{c1}, \eta_{c2}, \eta_{a1}, \eta_{a2}$ та кількість точок екстраполювання N , та проводились перевірки швидкості обрахування нового коефіцієнтів

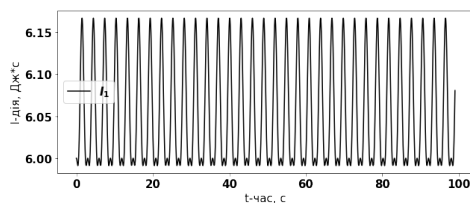
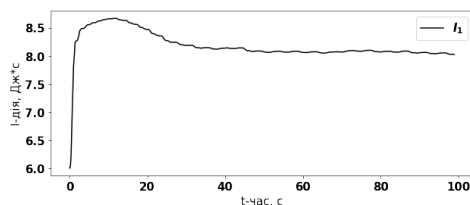
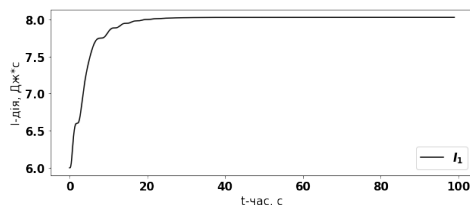


Рис. 1. Швидкісний градієнт

протягом відрізка часу між замірами. Порівнюючи результативність відносно ініціалізації, в якості базових параметрів мережі було обрано множину $N = 10$, $W_c = (1, \dots, 1) = W_a$, $\Gamma = 10\mathbf{I}$, $\bar{\Gamma} = 100$, $\eta_{a1} = 10$, $\eta_{c1} = \eta_{c2} = \eta_{a2} = \nu = 1$, а відповідна їй еволюція зображена на (рис. 2(а)-2(б)).



(а) Навчання нейронної мережі



(б) Випробування натренованої мережі

Рис. 2. Навчання з підкріпленням

В процесі еволюції системи нейронна мережа адаптується та зводить розв'язки до бажаної цілі в режимі реального часу. Окрім того, під час навчання спостерігається збіжність ваг оптимальних оцінок W_c та W_a , які апроксимують функцію критерію $V(x)$, тобто базисні функції ядра було обрано вдало:

$$W_c = (0.98, 1.21, 0.61, 0, 0, 0, 0, 0.59, 0.57)^T, \quad (30)$$

$$W_a = (0.97, 1.19, 0.60, 0, 0, 0, 0, 0.59, 0.57)^T. \quad (31)$$

Також навчену мережу було успішно випробувано на системах з різними початковими умовами й цілями керування (табл. 2).

Табл. 2. Тестування навченої мережі на різних вхідних даних та цілях

$I_1(0)$	$I_2(0)$	I_1^*	Успіх
7	3	8	досягнуто
10	7	8	досягнуто
12	7	10	досягнуто
15	30	20	досягнуто

Результати моделювань підтверджують гіпотезу про те, що виконується твердження (21), та розв'язана задача динамічного програмування є за-

стосовною не тільки для конкретної початкової умови та цілі. За підсумками навчання було побудовано апроксимації для функцій $V(x)$ та $u(x)$:

$$\begin{aligned} \hat{V}(x) = & 0.98(x_1 - x_1^*)^2 + 1.21(x_1 - x_1^*)(x_2 + x_1^*) + \\ & + 0.61(x_2 + x_1^*)^2 + 0.59 \sin(x_4 - x_3) + \\ & + 0.57 \cos(x_4 - x_3), \end{aligned} \quad (32)$$

$$\hat{u}(x) = \frac{1}{2} [0.75(x_1 - x_1^*) + 0.0(x_2 + x_1^*)] \sin(x_4 - x_3).$$

Висновки

Аналіз отриманих результатів показує, що вплив мінливого амплітудного керування при швидкоосцилюючому резонансному збуренні визначає також осцилюючий характер поведінки змінних дії, що відповідає теоретичним очікуванням, але і накладає обмеження при малих амплітудах. Можна зробити висновок, що класичний метод швидкісного градієнта є обмежено застосовним для такої задачі та потребує удосконалень. Застосування нейронних мереж з підкріпленням виявилось більш ефективним: в процесі еволюції алгоритм забезпечує досягнення бажаної цілі в режимі реального часу. При цьому побудована апроксимація оптимального керування як функції поточних станів системи, що робить модель застосовною для керування динамічними системами, які описуються заданим гамільтоніаном, проте можуть мати різні початкові умови та цілі керування. Також виділяються функції, що не корелюють із розв'язком, тому при подальшому вдосконаленні нейронної мережі можна зменшити розмір її базисного ядра.

Перелік використаних джерел

1. S. Strogatz. [Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering](#). — 2 ed. — Westview Press, 2015. — 532 p.
2. K.M. Hangos, J. Bokor, G. Szederkényi. [Analysis and Control of Nonlinear Process Systems](#). — 6330 Cham, Switzerland : Springer-Verlag London Limited 2004, 2004. — 335 p.
3. В. Арнольд. Математические методы классической механики. — Москва : Наука, 1974. — 432 с.
4. Г. Заславский, Р. Сагдеев. Введение в нелинейную физику. От маятника до турбулентности и хаоса. — Наука, 1988. — 338 с.
5. A. Fradkov. [Cybernetical Physics: From Control of Chaos to Quantum Control](#). — Springer-Verlag Berlin Heidelberg, 2007. — 248 p.
6. [Reinforcement Learning for Optimal Feedback Control](#) / Kamalapurkar Rushikesh, Walters Patrick, Rosenfeld Joel, and Dixon Warren. — Springer International Publishing AG, 2018. — 305 p.
7. Zhang Hao, Li Shaoyuan, Zheng Yi. [Q-Learning-Based Model Predictive Control for Nonlinear Continuous-Time Systems](#) // [Industrial & Engineering Chemistry Research](#). — 2020. — Vol. 59. — P. 17987–17999.