

BỘ ĐIỀU KHIỂN TỐI ƯU BỀN VỮNG DỰA TRÊN BIẾN TRƯỢT CHO CÁC HỆ CƠ KHÍ MÔ HÌNH PHI TUYẾN BẤT ĐỊNH

Vũ Văn Tú, Nguyễn Thị Thu Hiền
Khoa Điện - Cơ, Trường Đại học Hải Phòng
Email: tuv@dhhp.edu.vn

Ngày nhận bài: 01/4/2023

Ngày PB đánh giá: 28/5/2023

Ngày duyệt đăng: 14/7/2023

TÓM TẮT: Trong bài báo này, nhiệm vụ chính của chúng ta là thiết kế một bộ điều khiển tối ưu bền vững được áp dụng cho Rô-bốt Planar sử dụng một mạng nơ-ron kết hợp với biến trượt. Một thuật toán học tăng cường thích nghi (ARL) được phát triển để giải nghiệm phương trình Hamilton-Jacobi-Bellman (HJB) trong điều kiện của điều khiển tối ưu. Sự hội tụ của kỹ thuật được đề xuất với phân tích nghiệm của phương trình HJB được đảm bảo. Ngoài ra hiệu quả của bám quỹ đạo cũng được đề cập. Các nghiên cứu mô phỏng được đưa ra để đánh giá chất lượng của phương pháp đề xuất.

Từ khóa: Học tăng cường thích nghi, Điều khiển tối ưu, Quy hoạch động thích nghi, Điều khiển trượt, Rô-bốt Planar.

SLIDING VARIABLE-BASED ON OPTIMAL ROBUST CONTROLLER FOR UNCERTAIN NONLINEAR MECHANICAL SYSTEMS

ABSTRACT: In this paper, our main mission is to design a sustainable robust optimal controller applied for Planar Robot using a neural network combined a sliding variable. An adaptive reinforcement learning (ARL) is then developed to solve Hamilton-Jacobi-Bellman (HJB) equation in optimal control term. The convergence of the proposed technique with solution analysis of HJB equation is guaranteed. In addition, the efficiency of trajectory tracking is also mentioned. Simulation studies are given to evaluate the quality of the proposed method.

Keywords: Adaptive Reinforcement Learning, Optimal control, Adaptive dynamics programming (ADP), Sliding mode control (SMC), Robot Planar.

1. ĐẶT VẤN ĐỀ

Các hệ thống tay máy Rô-bốt nói chung và Rô-bốt Planar nói riêng được sử dụng rộng rãi trong công nghiệp. Chính vì lý do này, để nâng cao năng suất sản xuất và đem lại hiệu quả kinh tế, nhiệm vụ quan trọng của kỹ thuật điều khiển cánh tay máy là kiểm soát và điều chỉnh hoạt động của Rô-bốt sao cho chúng thực hiện các tác vụ một cách hiệu quả và tối ưu [1]. Vấn đề điều khiển Rô-bốt nhìn chung vẫn là bài toán khó bởi lẽ mô hình toán học của chúng thường có dạng Euler-Lagrange phi tuyến [2].

Để giải quyết bài toán điều khiển cánh tay máy Rô-bốt, các phương pháp tiếp cận dựa trên lý thuyết ổn định của Lyapunov cho thấy ưu thế khi có thể đảm bảo sai lệch điều khiển cho các trạng thái của hệ thống luôn nằm trong một lân cận của góc tọa độ. Thuật toán tiêu biểu cho cách tiếp cận này có thể kể tới bộ điều khiển trượt [3], [4]. Bằng cách tổng hợp mặt trượt từ các biến trạng thái của hệ thống, bộ điều khiển có thể được thiết kế sao cho không chỉ đảm bảo lái các trạng thái của Rô-bốt đến vị trí yêu cầu mà còn

kiểm soát được sự thay đổi về mặt tốc độ. Tuy nhiên, sự phụ thuộc vào mô hình hệ thống để xác định tín hiệu điều khiển là một điểm hạn chế của bộ điều khiển trượt bởi lẽ trong thực tế, yêu cầu về việc biết chính xác mô hình toán học của Rô-bốt là không dễ thực hiện. Chính vì vậy, để giảm thiểu sự ảnh hưởng của thành phần bất định hệ thống, các bộ quan sát [5], [6], [7] có thể được sử dụng thông qua việc ước lượng và loại bỏ tác động của chúng lên hệ thống. Ngoài ra, các phương pháp điều khiển thích nghi nhằm đảm bảo chất lượng bám của các quỹ đạo trạng thái trong nhiều tình huống làm việc khác nhau của Rô-bốt mà ở đó mạng nơ-ron được sử dụng như một công cụ hữu hiệu để xây dựng tín hiệu điều khiển cũng đã được nghiên cứu [8], [9] và cho kết quả tốt. Điều này đã khẳng định khả năng ứng dụng của mạng nơ-ron trong việc xử lý bài toán điều khiển Rô-bốt. Đối với yêu cầu điều khiển hệ thống Rô-bốt hiện này, bài toán đặt ra không chỉ là lái các trạng thái theo đúng quỹ đạo mong muốn mà còn phải thực hiện quá trình bám một cách tối ưu. Vấn đề này có thể được đưa về bài toán tối thiểu hóa

hàm chi phí và được giải quyết bằng các phương pháp học tăng cường thích nghi hoặc phương pháp tối ưu [10] thông qua việc tìm nghiệm của phương trình HJB [11]. Tuy nhiên, việc tìm nghiệm giải tích trực tiếp của phương trình HJB nhìn chung là rất khó khăn, trong nhiều trường hợp hầu như không thể tìm được [12]. Do đó, các thuật toán nhằm xấp xỉ nghiệm của phương trình HJB được ưa dùng hơn trong đó quá trình tối thiểu hóa hàm chi phí và tín hiệu điều khiển được thực hiện nhờ mạng nơ-ron [13]. Cấu trúc điều khiển học tăng cường thích nghi được thiết kế cho cánh tay Rô-bốt thường được thiết kế theo dạng Actor-Critic [14], [15], [16] với hai mạng nơ-ron, một mạng để xấp xỉ hàm đánh giá điều khiển tối ưu và một mạng để xấp xỉ luật học tăng cường thích nghi tối ưu [16]. Việc sử dụng cấu trúc hai mạng nơ-ron sẽ đem lại sự tiện lợi cho việc quản lý và phát triển bộ điều khiển nhưng điều này sẽ gia tăng độ phức tạp tính toán và là yếu tố chính cản trở khả năng áp dụng của phương pháp này trong thực tế. Vì vậy, phương pháp điều khiển học tăng cường thích nghi chỉ sử dụng một mạng nơ-ron

trong toàn bộ quá trình thiết kế [17] cho thấy sự hiệu quả hơn trong việc giảm thiểu khối lượng tính toán.

1. Từ những đánh giá thu được, trong bài báo này phương pháp điều khiển học tăng cường thích nghi chỉ sử dụng một mạng nơ-ron thông qua việc tổng hợp trạng thái của hệ thống thành biến trượt và điều khiển mặt trượt về gốc tọa độ sẽ được trình bày.

2. Qua đó quá trình mô phỏng sẽ được thực hiện nhằm khẳng định tính đúng đắn và hiệu quả của thuật toán.

2. TỔNG QUAN NGHIÊN CỨU

Mô hình động học của Rô-bốt Planar có thể được biểu diễn dưới dạng Euler-Lagrange [5]:

$$M(\underline{q})\ddot{\underline{q}} + C(\underline{q}, \dot{\underline{q}})\dot{\underline{q}} + G(\underline{q}) = \underline{\tau} + \underline{\tau}_d \quad (1)$$

trong đó $\underline{q} \in \mathbb{R}^2$ là véc-tơ trạng thái của các biến khớp, $\dot{\underline{q}}$ và $\ddot{\underline{q}}$ lần lượt là đạo hàm cấp một và đạo hàm cấp hai của \underline{q} theo thời gian. Ma trận $M \in \mathbb{R}^{2 \times 2}$ là ma trận quán tính của hệ thống thỏa mãn tính chất đối xứng và xác định dương theo [11]. $C \in \mathbb{R}^{2 \times 2}$ là ma trận hướng tâm và Criolit, $G \in \mathbb{R}^2$ là véc-tơ trọng trường. Tín hiệu điều khiển đặt vào hệ thống $M(\underline{q})\ddot{\underline{q}} + C(\underline{q}, \dot{\underline{q}})\dot{\underline{q}} + G(\underline{q}) = \underline{\tau} + \underline{\tau}_d$ được biểu diễn qua véc-tơ $\underline{\tau}$ và nhiều

tác động lên rô-bốt trong quá trình làm việc được mô tả bởi $\underline{\tau}_d$. Một số giả thiết được đặt ra nhằm phục vụ thiết kế bộ điều khiển như sau:

Giả thiết 1: Các trạng thái của hệ thống cùng đạo hàm của chúng là tồn tại, liên tục theo thời gian và bị chặn, tức là tồn tại hằng số dương sao cho trong đó kí hiệu là chuẩn của véc-tơ $\underline{\tau}_d$ (kí hiệu này cũng được sử dụng với chuẩn của ma trận).

Giả thiết 2: Quỹ đạo đặt $\underline{r}(t)$ thiết kế sao cho $\dot{\underline{r}}(t)$ cùng các đạo hàm của nó tồn tại và bị chặn, tức tồn tại hằng số dương b_r sao cho $\|\underline{r}(t)\| \leq b_r$.

Giả thiết 3: Véc-tơ nhiễu ngoài $\underline{\tau}_d$ tồn tại đạo hàm bậc nhất, bậc hai và bậc ba và luôn bị chặn bởi một hằng số $\|\underline{\tau}_d\| \leq a$.

Mục tiêu của bài toán điều khiển học tăng cường thích nghi đó là xác định tín hiệu điều khiển $\underline{\tau}$ sao cho đảm bảo được các trạng thái bám theo một quỹ đạo đặt cho trước đồng thời tối ưu hóa hàm chi phí đánh giá. Để xây dựng hàm chi phí này, dựa trên tài liệu [13] trước tiên sai lệch bám của hệ thống được định nghĩa bởi: $e_1 = \underline{r} - \underline{q}$ (2) và mặt trượt

được xác định theo công thức:

$$\underline{s} = \dot{e}_1 + \lambda_1 e_1$$
 (3)

Từ đây, ta thu được mô hình động học của biến trượt:

$$M\dot{\underline{s}} = -C\underline{s} - \underline{\tau} + \underline{\xi}$$
 (4) với $\underline{\xi}$ là một hàm bất định phi tuyến:

$$\underline{\xi} = M(\ddot{\underline{r}} + \lambda_1 \dot{e}_1) + C(\dot{\underline{r}} + \lambda_1 e_1) + G - \underline{\tau}_e$$
 (5)

Với việc sự ảnh hưởng của thành phần nhiễu tổng $\underline{\xi}$ có thể được loại bỏ thông qua bộ quan sát [5], [6] hoặc phương pháp tích phân dấu của sai lệch (RISE) [13], bằng cách sử dụng biến trượt ta sẽ đưa hệ thống từ dạng bậc hai về dạng bậc nhất từ đó thuận lợi hơn trong việc áp dụng thuật toán quy hoạch động trực tuyến để giải bài toán học tăng cường thích nghi. Khi đó, tín hiệu điều khiển có thể được thiết kế theo công thức: $\underline{\tau} = \hat{\underline{\xi}} - \underline{u}$ (6)

với \underline{u} là tín hiệu điều khiển tối ưu được xác định qua thuật toán quy hoạch động thích nghi trực tuyến và là thành phần xấp xỉ của nhiễu tổng $\underline{\xi}$. Giả sử có thể hoàn toàn loại bỏ ảnh hưởng của $\underline{\xi}$ thông qua $\hat{\underline{\xi}}$, từ cách thiết kế trên hệ thống (4) trở thành:

$$M\dot{\underline{s}} = -C\underline{s} + \underline{u}$$
 (7) và nếu đặt biến

trạng thái mới là $\underline{x} = [e_1^T, s^T]^T$, xây dựng được mô hình:

$$\dot{\underline{x}} = A\underline{x} + B\underline{u} \quad (8) \quad \text{trong đó}$$

$$A = \begin{bmatrix} -\lambda_1 & I_2 \\ \Theta_2 & -M^{-1}C \end{bmatrix}, B = \begin{bmatrix} \Theta_2 \\ M^{-1} \end{bmatrix} \text{ với } \Theta_2$$

và I_2 lần lượt là ma trận không và ma trận được vị kích thước 2×2 . Véc-tơ điều khiển \underline{u} thuộc tập tín hiệu điều khiển U . Hàm chi phí lúc này được xác định bởi:

$$J = \int_0^{\infty} (\underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u}) dt \quad (9) \quad \text{với}$$

$Q \in \mathbb{R}^{4 \times 4}$ và $R \in \mathbb{R}^{2 \times 2}$ là các ma trận tham số đối xứng xác định dương.

Như vậy, bài toán điều khiển học tăng cường thích nghi cho Robot Planar sẽ được thực hiện thông qua việc thiết kế bộ điều khiển tối ưu u nhằm tối ưu hóa hàm chi phí J . Chi tiết thuật toán để giải quyết bài toán trên sẽ được trình bày ở phần tiếp theo.

Dựa trên [15], tín hiệu điều khiển tối ưu cho hệ (8) có thể được xây dựng thông qua việc tìm nghiệm của phương trình Hamilton-Jacobi-Bellman (HJB).

$$H(\underline{x}, \underline{u}, V_x) = V_x^T (A\underline{x} + B\underline{u}) + \underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u} \quad (10)$$

trong đó:

$$V(\underline{x}(t)) = \int_0^{\infty} (\underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u}) dt \quad (11)$$

là hàm đánh giá luật điều khiển và $V_x = \partial V / \partial \underline{x}$. Hàm đánh giá chất lượng điều khiển tối ưu được định nghĩa qua [15]:

$$V^*(\underline{x}(t)) = \min_{\underline{u} \in U} \int_0^{\infty} (\underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u}) dt \quad (12)$$

Khi đó ta thu được hàm Hamilton tối ưu:

$$H(\underline{x}, \underline{u}, V_x^*) = (V_x^*)^T (A\underline{x} + B\underline{u}) + \underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u} \quad (13)$$

và phương trình HJB trở thành:

$$\min_{\underline{u} \in U} H(\underline{x}, \underline{u}, V_x^*) = 0 \quad (14)$$

Tín hiệu điều khiển tối ưu \underline{u}^* trong trường hợp này sẽ được xác định dựa theo điều kiện dừng [15]:

$$\frac{\partial H(\underline{x}, \underline{u}, V_x^*)}{\partial \underline{u}} = 0 \quad (15)$$

và ta thu được:

$$\underline{u}^* = -\frac{1}{2} R^{-1} B^T V_x^* \quad (16)$$

Như vậy có thể thấy rằng, nếu hàm đánh giá tối ưu V^* được xác định, ta hoàn toàn có thể tính toán được tín hiệu điều khiển tối ưu \underline{u}^* . Tuy nhiên, việc giải phương trình vi phân HJB để tìm nghiệm giải tích là rất phức tạp, trong nhiều trường hợp ta gần như không thể giải được [12]. Chính vì vậy các phương pháp học tăng cường thích nghi để xấp xỉ trực tuyến nghiệm tối ưu V^* là cách tiếp cận phù hợp và mạng nơ-ron là một công cụ hữu dụng để giải quyết bài toán này.

Đầu tiên, hàm đánh giá lý tưởng $V^*(x)$ có thể được xấp xỉ bởi một mạng nơ-ron [17] như sau:

$V^*(\underline{x}) = W^T \phi(\underline{x}) + \varepsilon(\underline{x})$ (17) trong đó $W \in \mathbb{R}^m$ là trọng số của mạng, $\phi(\underline{x}) : \mathbb{R}^n \rightarrow \mathbb{R}^m$ là véc-tơ hàm kích hoạt với m là số nơ-ron lớp ẩn và $\varepsilon(\underline{x})$ là sai số xấp xỉ của mạng. Khi đó, phương trình HJB được viết lại thành:

$$\begin{aligned} H(\underline{x}, \underline{u}, W) &= W^T \phi_x (A\underline{x} + B\underline{u}) \\ &+ \underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u} \\ &+ \varepsilon_x^T (A\underline{x} + B\underline{u}) = 0 \end{aligned} \quad (18)$$

$$\text{với } \phi_x = \partial \phi / \partial \underline{x}, \quad \varepsilon_x = \partial \varepsilon / \partial \underline{x}$$

Tiếp theo, thay tín hiệu điều khiển tối ưu được tính theo công thức

(16) vào phương trình (18) và biểu diễn hàm đánh giá lý tưởng $V^*(\underline{x})$ qua mạng nơ-ron (17) thu được:

$$\begin{aligned} H(\underline{x}, W) &= W^T \phi_x A \underline{x} + \underline{x}^T Q \underline{x} \\ &- \frac{1}{4} W^T \phi_x B_c \phi_x^T W + \varepsilon_{hjb} = 0 \end{aligned} \quad (19)$$

với $B_c = BR^{-1}B^T$ và ε_{hjb} sai số của phương trình HJB được tính:

$$\varepsilon_{hjb} = \varepsilon_x^T A \underline{x} - \frac{1}{2} W^T \phi_x B_c \varepsilon_x - \frac{1}{4} \varepsilon_x^T B_c \varepsilon_x \quad (20)$$

Nhiệm vụ của bài toán điều khiển lúc này chính là xác định trọng số lý tưởng W sao cho thỏa mãn phương trình (19). Đối với phương pháp quy hoạch động, giá trị lý tưởng của W sẽ được ước lượng dần dần thông qua việc cập nhật trọng số xấp xỉ \hat{W} trong quá trình hoạt động của hệ thống bằng cách tối ưu hóa hàm H . Lúc này, hàm đánh giá $V(\underline{x})$ sẽ được thực tính theo công thức:

$$\hat{V}(\underline{x}) = W^T \phi(\underline{x}) \quad (21)$$

Khi sử dụng $\hat{V}(\underline{x})$ để thiết kế bộ điều khiển, từ (16) ta thu được tín hiệu điều khiển tối ưu xấp xỉ:

$$\hat{u}(\underline{x}) = -\frac{1}{2} R^{-1} B^T \phi_x^T \hat{W} \quad (22)$$

Phương trình HJB lúc này trở thành:

$$\begin{aligned} H(\underline{x}, \underline{u}, \widehat{W}) &= \widehat{W}^T \phi_x A \underline{x} \\ &+ \underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u} = \widehat{\varepsilon}_{hjb} \end{aligned} \quad (23)$$

với $\widehat{\varepsilon}_{hjb}$ là sai số Hamilton sinh ra bởi mạng nơ-ron xấp xỉ và tín hiệu điều khiển tối ưu xấp xỉ. Việc tối thiểu hóa hàm H tương đương với việc tối thiểu hóa $\widehat{\varepsilon}_{hjb}$. Để thực hiện điều này, luật cập nhật trọng số cho \widehat{W} sẽ được xây dựng nhằm tối thiểu hóa $\Phi = \frac{1}{2} \widehat{\varepsilon}_{hjb}^T \widehat{\varepsilon}_{hjb}$ đồng thời vẫn đảm bảo tính ổn định hệ thống. Dựa trên [10], luật cập nhật trọng số mạng được đề xuất như sau:

$$\dot{\widehat{W}} = \begin{cases} \widehat{W}_1, & \underline{x}^T (A \underline{x} + B \underline{u}) \leq 0 \\ \widehat{W}_1 + \widehat{W}_c, & \underline{x}^T (A \underline{x} + B \underline{u}) > 0 \end{cases} \quad (24)$$

trong đó \widehat{W}_1 được xây dựng dựa trên thuật toán Levenberg-Marquardt cải tiến [10] và xác định bởi:

$$\begin{aligned} \dot{\widehat{W}} &= -\alpha \frac{\sigma}{(\sigma^T \sigma + 1)^2} \\ &(\sigma^T \widehat{W} + \underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u}) \end{aligned} \quad (25)$$

với $\sigma(\underline{x}) = \phi_x(A \underline{x} + B \underline{u})$ và W_c là luật hiệu chỉnh nhằm đảm bảo tính ổn định của hệ thống cho bởi

$$W_c = \frac{1}{2} \beta \phi_x B_c \underline{x} \quad (26)$$

Hệ số α và β được chọn là hai hằng số dương. Sự ổn định của hệ kín khi sử dụng bộ điều khiển (22) với luật cập nhật (25) đã được chứng minh trong [17] qua đó đảm bảo rằng trọng

số xấp xỉ \widehat{W} tiến tới trọng số lý tưởng W đồng thời đảm bảo sai lệch bám của trạng thái tiến về một lân cận của không. Tổng kết lại, với luật cập nhật trọng số mạng nơ-ron xấp xỉ (25), hàm đánh giá tối ưu xấp xỉ $\widehat{V}(\underline{x})$ sẽ được tính theo (21) từ đó tín hiệu điều khiển tối ưu xấp xỉ được xác định bởi (22). Tín hiệu này sẽ được đưa vào điều khiển hệ (8) trong suốt quá trình hoạt động. Có thể thấy trong toàn bộ quá trình thiết kế, chỉ duy nhất một mạng nơ-ron được sử dụng do đó khối lượng tính toán đã được giảm đi đáng kể.

Theo định nghĩa ổn định UUB được nghiên cứu của công trình [18], bộ điều khiển tối ưu thiết kế cho hệ thống trên đảm bảo ổn định.

3. KẾT QUẢ MÔ PHỎNG THUẬT TOÁN TRÊN MATLAB-SIMULINK

3.1. Tham số mô phỏng

Cấu trúc chuyển động của Rô-bốt lanar 2 bậc tự do có thể được cho bởi Hình. 1 với biến trạng thái q_1, q_2 lần lượt là góc quay khớp 1 và khớp 2. Với mô hình toán học dạng Euler-Lagrange (1), các phân tử của ma trận M , ma trận C và véc-tơ G dựa trên [16] được xác định lần lượt như sau:

$$M = \begin{bmatrix} \gamma_1 + 2\gamma_2 \cos q_2 & \gamma_3 + \gamma_2 \cos q_2 \\ \gamma_3 + \gamma_2 \cos q_2 & \gamma_3 \end{bmatrix}$$

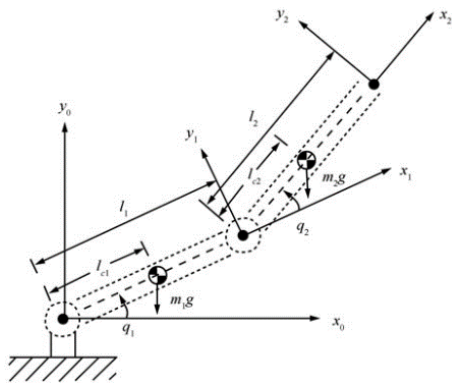
$$C = \begin{bmatrix} -\gamma_2 \sin q_2 \dot{q}_2 & -\gamma_2 \sin q_2 (\dot{q}_1 + \dot{q}_2) \\ \gamma_2 \sin q_2 \dot{q}_1 & 0 \end{bmatrix}$$

$$G = \begin{bmatrix} \gamma_4 \cos q_1 + \gamma_5 \cos q_1 + q_2 \\ \gamma_5 \cos q_1 + q_2 \end{bmatrix}$$

Các thông số mô hình sử dụng trong mô phỏng được cho bởi Bảng 1:

Bảng 1. Tham số mô phỏng của Rô-bốt Planar

Tham số	γ_1	γ_2	γ_3
Giá trị	5	1	1
Tham số	γ_4	γ_5	
Giá trị	11.77	9.81	



Hình 1: Cấu trúc Rô-bốt Planar

Các ma trận hàm chi phí được chọn theo [13]:

$$Q = \begin{bmatrix} 40 & 2 & -4 & 4 \\ 2 & 40 & 4 & -6 \\ -4 & 4 & 4 & 0 \\ 4 & -6 & 0 & 4 \end{bmatrix}; R = \begin{bmatrix} 0.25 & 0 \\ 0 & 0.25 \end{bmatrix}$$

Tham số thiết kế bộ điều khiển lần lượt là

$$\alpha = 5, \beta = 1, \lambda_1 = \begin{bmatrix} 15.6 & 10.6 \\ 10.6 & 10.4 \end{bmatrix},$$

hàm kích hoạt $\phi(\underline{x})$ dựa theo [13] được chọn bằng:

$$\phi(\underline{x}) = [\phi_1; \phi_2]$$

$$\phi_1(\underline{x}) = [x_1^2, x_1 x_2, x_2^2]^T$$

$$\phi_2(\underline{x}) = [x_3^2, x_3^2 \cos q_2, x_3 x_4 \cos q_2, x_4^2]^T$$

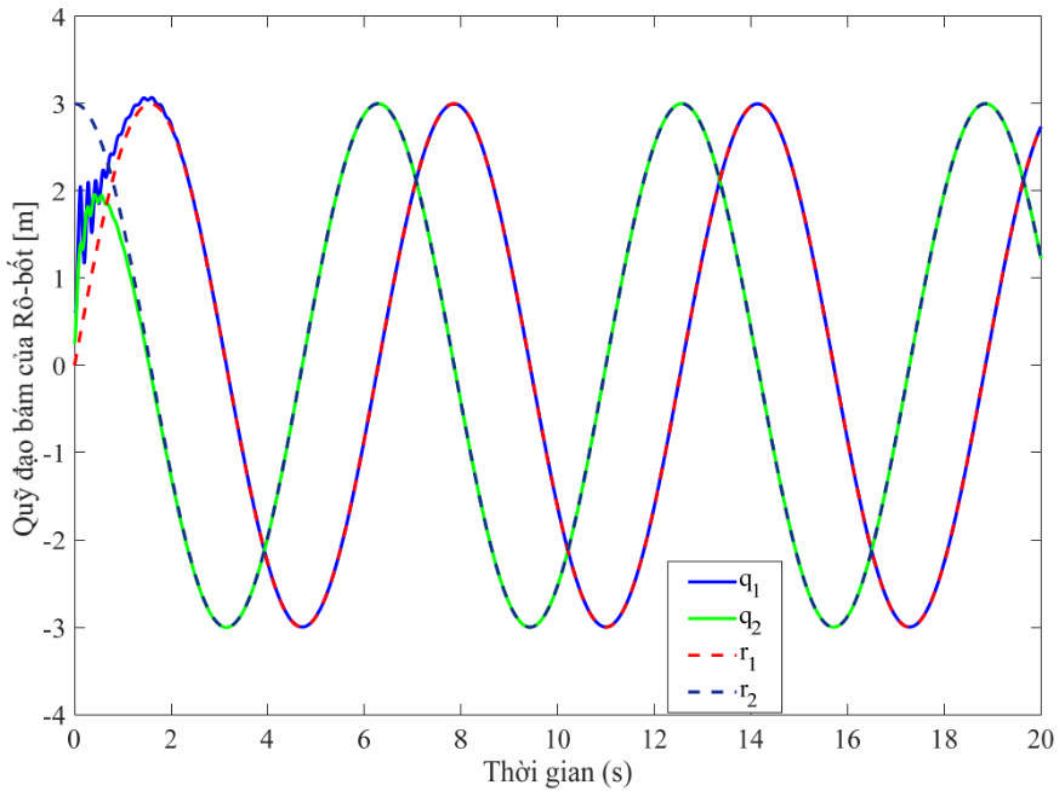
Giá trị của \hat{W} được tính chính xác theo lý thuyết:

$$\hat{W} = [2; 1; 3; 2.5; 1; 1; 1; 0.5]^T \quad (27)$$

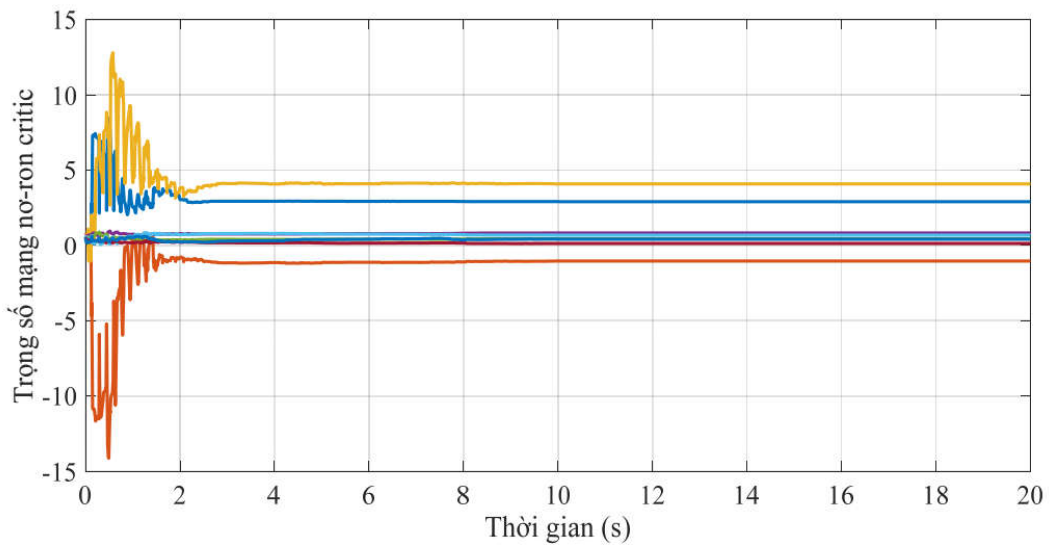
Bộ điều khiển (22) sẽ được sử dụng để điều khiển rô-bốt với tín hiệu đặt có dạng thay đổi tuần hoàn theo thời gian:

$$\underline{r} = [3 \sin(t) \quad 3 \cos(t)]^T \quad (28)$$

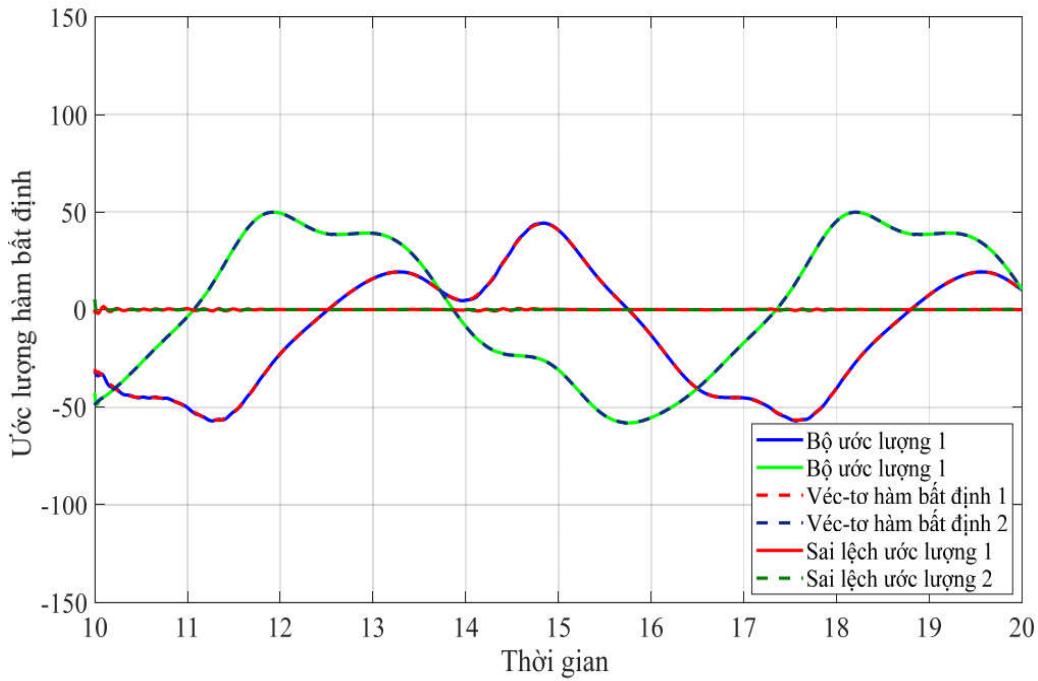
3.2. Kết quả mô phỏng
Trường hợp 1: Véc-tơ nhiễu ngoài được sử dụng là: $\underline{\tau}_d(t) = [50 \sin(t) \quad 50 \cos(t)]^T$



Hình 2. Đáp ứng quỹ đạo đặt của Rô-bốt Planar



Hình 3. Sự hội tụ của trọng số mạng nơ-ron critic



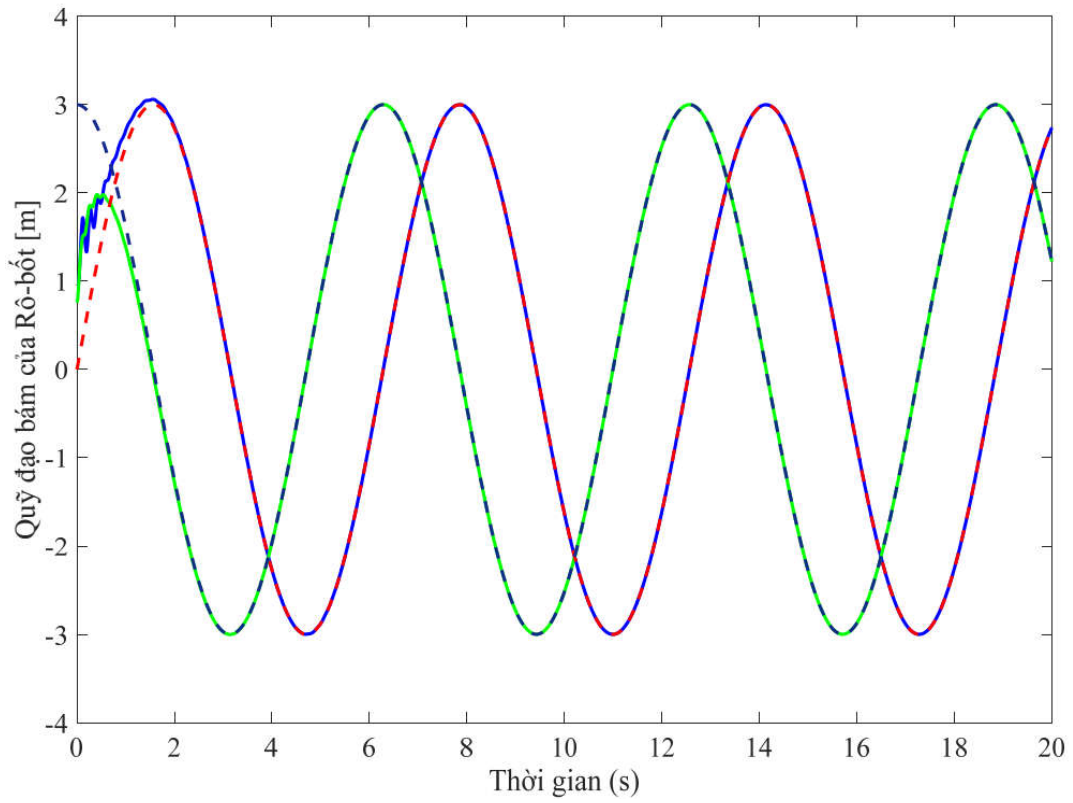
Hình 4. Ước lượng hàm bất định và sai số ước lượng

Bảng 2. So sánh giữa trọng số mạng nơ-ron thuật toán đề xuất và theo lý thuyết

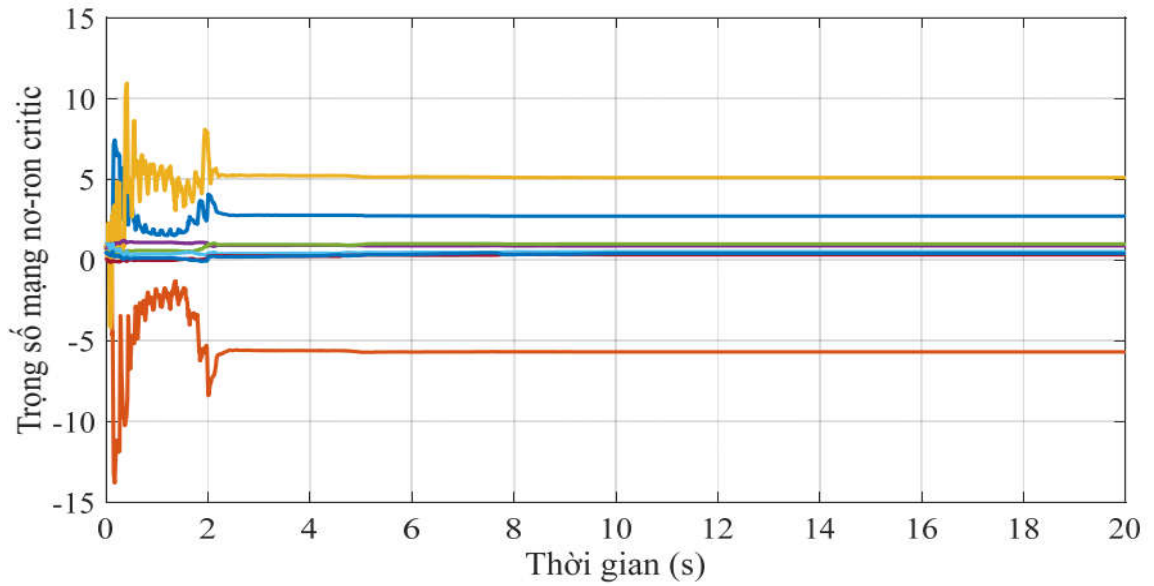
W	Trọng số của thuật toán đề xuất	Trọng số tính theo lý thuyết theo (27)
W_1	2.02	2.00
W_2	-3.95	-4.00
W_3	2.98	3.00
W_4	2.50	2.50
W_5	1.00	1.00
W_6	1.00	1.00
W_7	1.00	1.00
W_8	0.50	0.50

Trường hợp 2: Véc-tơ nhiễu ngoài được sử dụng là:

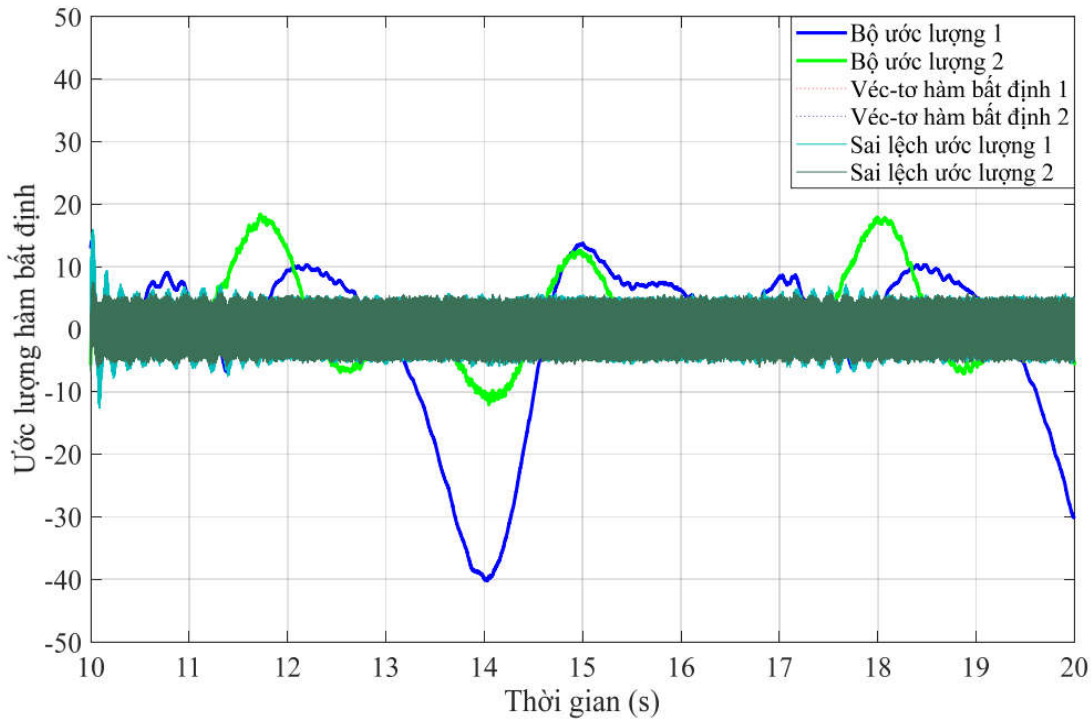
$$\underline{r}_a(t) = [10\text{rand}(1) \quad 10\text{rand}(1)]^T$$



Hình 4. Đáp ứng quỹ đạo đặt của Rô-bốt Planar



Hình 5. Sự hội tụ của trọng số mạng nơ-ron critic



Hình 6. Ước lượng hàm bất định và sai số ước lượng

Nhận xét:

Kết quả mô phỏng thu được qua Hình 2 và Hình 4 cho thấy các trạng thái hệ thống nhanh chóng bám tín hiệu đặt với các nhiễu tác động khác nhau, từ đó khẳng định sự hiệu quả của bộ điều khiển. Sự hội tụ của các trọng số mạng nơ-ron xấp xỉ được thể hiện qua Hình 3 và Hình 5 với các nhiễu khác nhau, kết quả hội tụ so sánh với giá trị tính toán lý thuyết cho sai số rất nhỏ, như vậy cấu trúc điều khiển được

thiết kế hoàn toàn đáp ứng được yêu cầu bài toán.

Kết quả ước lượng véc-tơ hàm bất định có chất lượng tốt, sai lệch ước lượng đảm bảo nằm trong miền ổn định của hệ thống thể hiện trên Hình 4 và Hình 6.

4. KẾT LUẬN

Bài báo đã trình bày phương pháp điều khiển học tăng cường thích nghi cho Rô-bốt Planar sử dụng mạng nơ-ron thông qua biến trượt. Với việc chỉ sử dụng một mạng nơ-ron duy nhất trong quá trình thiết kế, độ phức

tạp tính toán của bộ điều khiển được giảm thiểu qua đó nâng cao khả năng áp dụng của phương pháp trong thực tế. Đồng thời, bằng cách sử dụng biến trượt làm biến trạng thái để điều khiển, ta có thể kiểm soát tốc độ chuyển động của các khớp Rô-bốt thông qua ổn định mặt trượt. Các kết quả mô phỏng cho thấy sự hiệu quả của mạng nơ-ron trong việc xấp xỉ nghiệm tối ưu của phương trình HJB từ đó đảm bảo cũng như nâng cao chất lượng điều khiển hệ thống. Trong tương lai, để tiếp tục đơn giản hóa cấu trúc điều khiển, thành phần nhiều tổng có thể được loại bỏ song song với quá trình tìm tín hiệu điều khiển tối ưu thông qua một mạng nơ-ron duy nhất thay vì sử dụng bộ quan sát. Khi đó, mạng nơ-ron sẽ tiếp tục được khẳng định như một công cụ hữu hiệu cho nhiệm vụ giải quyết các bài toán điều khiển.

TÀI LIỆU THAM KHẢO

1. M. Blatnický, J. Dizohtps, J. Gerlici1, M. Sága, T. Lack and E. Kuba, “*Design of a robotic manipulator for handling products of automotive industry*”, International Journal of Advanced Robotic Systems. 17(1), (2020).
2. H. Li, C. L. Liu, Y. Zhang and Y. Y. Chen, “*Adaptive neural networks-based fixed-time faulttolerant consensus tracking for uncertain multiple Euler-Lagrange systems*”, ISA Transactions, (2021).
3. M. Rahmani, H. Komijani, M. H. Rahman, “*New Sliding Mode Control of 2-DOF Robot Manipulator Based on Extended Grey Wolf Optimizer*”, International Journal of Control, Automation and Systems, 18, 1572-1580, (2020).
4. K. Shao, R. Tang, F. Xu, X. Wang, J. Zheng, “*Adaptive sliding mode control for uncertain Euler-Lagrange systems with input saturation*”, Journal of the Franklin Institute, 358(16), 8356-8376, (2021).
5. M. Ramírez-Neria, R. Madonski, A. Luviano-Juárez, Z. Gao, H. Sira-Ramírez, “*Design of ADRC for Second-Order Mechanical Systems without TimeDerivatives in the Tracking Controller*”, 2020 American Control Conference (ACC), (2020).

6. W. Ha, J. Back, “*A Disturbance Observer-based Robust Tracking Controller for Uncertain Robot Manipulators*”, International Journal of Control, Automation and Systems, 16 417-425, (2018).
7. Z. Zhang, M. Leibold and D. Wollherr, “*Integral Sliding-Mode Observer-Based Disturbance Estimation for Euler-Lagrangian Systems*”, IEEE Transactions on Control Systems Technology, 28(6), 2377-2389, (2019).
8. Q. Zhou, S. Zha,; H. Li, R. Lu, C. Wu, “*Adaptive Neural Network Tracking Control for Robotic Manipulators With Dead Zone*”, IEEE Transactions on Neural Networks and Learning Systems, 30(12), 3611 - 3620, (2018).
9. A. Jouila, K. Nouri, “*An adaptive robust nonsingular fast terminal sliding mode controller based on wavelet neural network for a 2-DOF robotic arm*”, Journal of the Franklin Institute, 357(18), 13259-13282, (2020).
10. P. Ioannou and B. Fidan, “*Advances in design and control, Adaptive control tutorial*”. PA: SIAM, (2006).
11. L. Tang, Y. J. Liu, S. Tong, “*Adaptive neural control using reinforcement learning for a class of robot manipulator*”, Neural Computing and Applications, 25, 135-141, (2014).
12. F.L. Lewis and D. Vrabie, “*Reinforcement learning and adaptive dynamic programming for feedback control*”, IEEE Circuits and Systems Magazine, 9(3), 32-50, (2009).
13. K. Dupree, P. M. Patre, Z. D. Wilcox and W.E. Dixon, “*Asymptotic optimal control of uncertain nonlinear Euler-Lagrange systems*”, Automatica, 1, 99-107, (2011).
14. Y. Hu and B. Si, “*A Reinforcement Learning Neural Network for Robotic Manipulator Control*”, Neural Computation, 30(7), 1983 - 2004, (2018).
15. K.G. Vamvoudakis and F.L., Lewis, “*Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem*”, Automatica, 46, 878-888, (2010).
16. V. T. Vu, P. N. Dao, P. T. Loc, T. Q. Huy, “*Sliding Variable-based*

- Online Adaptive Reinforcement Learning of Uncertain/Disturbed Nonlinear Mechanical Systems*”, Journal of Control, Automation and Electrical Systems, 32, 281-290, (2021).
17. Nguyễn Tân Lũy, “*Nghiên cứu giải thuật học củng cố trong điều khiển thích nghi bền vững hệ phi tuyến*”, Luận án tiến sĩ, Mã số chuyên ngành: 62.52.60.01, Trường Đại học Bách khoa, Đại học Quốc gia Tp.Hồ Chí Minh.
18. H. K. Khalil (2002), "*Nonlinear Systems, 3rd Edition*".