

ĐIỀU KHIỂN BẮM TỐI ƯU CHO TÀU MẶT NƯỚC ĐỦ CƠ CẤU CHẤP HÀNH DỰA TRÊN THUẬT TOÁN HỌC TĂNG CƯỜNG TÍCH PHÂN TRỰC TUYẾN

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

Ngày nhận bài: 27/11/2023

Ngày PB đánh giá: 08/12/2023

Ngày duyệt đăng: 15/12/2023

TÓM TẮT: Trong bài báo này, nhiệm vụ chính của chúng ta là thiết kế một bộ điều khiển bám tối ưu được áp dụng cho các hệ thống tàu mặt nước đủ cơ cấu chấp hành với một phần động học chưa biết. Một thành phần feed-forward trong bộ điều khiển đề xuất được giới thiệu để có được mô hình sai lệch bám tự trị. Một thuật toán học tăng cường tích phân trực tuyến (OIRL) 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: Tàu mặt nước (SVs); học tăng cường tích phân trực tuyến (OIRL); quy hoạch động thích nghi (ADP); lý thuyết ổn định Lyapunov.

OPTIMAL TRACKING CONTROL FOR FULLY-ACTUATED SURFACE VESSELS BASED ONLINE ON-POLICY INTEGRAL REINFORCEMENT LEARNING

ABSTRACT: In this paper, our main mission is to design an optimal tracking controller applied to fully-actuated Surface Vessel Systems with partially unknown dynamics. A feed-forward component in the proposed controller is introduced to obtain a corresponding autonomous tracking error model. An

Online On-Policy Integral Reinforcement Learning algorithm (OIRL) is then developed to solve Hamilton-Jacobi-Bellman (HJB) equation under the optimal control term. The convergence of the proposed technique to the solution analysis of the HJB equation is guaranteed. Additionally, the trajectory tracking effectiveness is also mentioned. Simulation studies are given to evaluate the quality of the proposed method.

Keywords: Surface Vessels (SVs); Online On-Policy Integral Reinforcement Learning (OIRL); Adaptive Dynamic Programming (ADP); Lyapunov Stability Theory.

1. Đặt vấn đề

Trong những năm gần đây điều khiển tối ưu được ứng dụng một cách rộng rãi để phát triển, cải thiện hiệu suất của hệ thống điều khiển SVs. Việc ứng dụng lý thuyết điều khiển tối ưu chủ yếu tập trung giải nghiệm xấp xỉ của phương trình HJB sử dụng kỹ thuật học tăng cường (Reinforcement Learning (RL)) bởi vì phương trình HJB là phương trình dạng vi phân phi tuyến không thể giải được nghiệm chính xác. Các cấu trúc điều khiển cho SVs thường là các cấu trúc điều khiển tầng [1-4]. Trong công trình [1], vấn đề điều khiển toàn bộ trạng thái của hệ thống SVs được thực hiện bằng cách chia thành hai nhiệm vụ phụ đó là cấu trúc điều khiển chuyển động tịnh tiến và bộ điều khiển chuyển động quay được thiết kế với bộ quan sát chính xác. Ngoài ra, thiết kế một bộ điều khiển trượt thời gian hữu hạn được phát triển cho SVs với biến trượt có thể thu được

từ bộ quan sát bao gồm cả thành phần bất định, nhiễu ngoài và thành phần sai lệch vận tốc [2]. Trong công trình [3], mặc dù hệ thống điều khiển tầng được xử lý nhưng rõ ràng nó khác với các phương pháp có trong [1, 2], kỹ thuật Tan-Barrier Lyapunov đã giải quyết vấn đề ràng buộc sai lệch và điều khiển thời gian hữu hạn. Ngoài ra các tác giả trong [3] đã giải quyết tình trạng bão hòa của cơ cấu chấp hành bằng cách bổ xung một thành phần vào cấu trúc điều khiển và sau đó thu được luật cập nhật của nó. Trong công trình [4], các tác giả coi mô hình của SVs giống với mô hình của WMR nhưng phần thiết kế hệ thống điều khiển thì hoàn toàn khác với thiết kế hệ thống điều khiển của WMR, hệ thống con vòng ngoài là đủ cơ cấu chấp hành và hệ thống con vòng trong là hệt cơ cấu chấp hành. Ngoài ra, độ bão hòa của cơ cấu chấp hành và bộ quan sát cũng được thực hiện trong [2, 3]. Ứng dụng của bộ điều khiển trượt

(SMC) cho SVs được mở rộng sang kỹ thuật trượt tích phân (ISMC) trong công trình [5]. Trong những năm gần đây thuật toán RL đã được đề cập rất nhiều với nhiều phương pháp tiếp cận như Actor/Critic, kỹ thuật học tăng cường tích phân trực tuyến, không trực tuyến, Q-learning,...[6-11]. Trong công trình [6], luật cập nhật cho các mạng nơ-ron Actor và Critic được thực hiện đồng bộ dựa trên việc xem xét hàm Hamilton. Phương pháp này cũng được sử dụng trong công trình [7] cho tất cả các vòng điều khiển của bộ điều khiển tầng với mạng nơ-ron Critic được cải tiến cho phù hợp. Tuy nhiên, do quỹ đạo bám mong muốn thay theo thời gian, mô hình sai lệch bám còn được hiểu là hệ thống tự trị, do đó nó dẫn tới hiệu suất của bộ điều khiển được thiết kế. Các tác giả trong công trình [9] đã đề xuất bổ xung thêm thành phần của hàm Nussbaum để nghiên cứu hướng điều khiển chưa biết với chỉ số hiệu suất cải tiến.

Trong bài báo này, thuật toán OIRL được sử dụng để tìm ra giải pháp cho vấn đề điều khiển tối ưu bám của các hệ thống SVs bất định mô hình. Một phương pháp biến đổi mô hình của SVs thành mô hình sai lệch bám tự trị cũng được giới thiệu để phát triển thuật toán học tăng cường. Do đó phương pháp điều khiển được

xây dựng trong bài báo này khác hoàn toàn phương pháp hiện có đã được giới thiệu trong tài liệu [7], thực thi cấu trúc Actor/Critic của thuật toán học tăng cường cho các hệ thống không tự trị. Cũng trong tài liệu [7], các tác giả sử dụng nhiều mạng nơ-ron dẫn đến tốn tài nguyên, tốc độ hội tụ giảm làm cho bộ điều khiển được thiết kế phức tạp, tốc độ xử lý của các máy tính số không xử lý kịp các tính toán của các bộ điều khiển.

Phương pháp đề xuất trong bài báo có thể thực hiện với các hệ thống tự trị với số biến trạng thái nhỏ hơn so với các công trình [9, 10] mặc dù thuộc tính không tự trị của hệ kín theo quỹ đạo tham chiếu khác nhau.

Từ những đánh giá thu được, trong bài báo này một thuật toán học tăng cường tích phân trực tuyến (OIRL) được phát triển để giải nghiệm phương trình Hamilton-Jacobi-Bellman ứng dụng cho điều khiển bám tối ưu cho mô hình tàu thủy mà không cần biết một phần động học của tàu mà các công trình [1-5] chưa giải quyết được.

Đề xuất cấu trúc điều khiển tối ưu bám quỹ đạo cho hệ lái tàu thủy dựa trên giải thuật ADP với cấu trúc điều khiển là OIRL, critic NN dùng để xấp xỉ hàm chi phí tối ưu và từ đó tính toán luật điều khiển tối ưu.

2. Tổng quan nghiên cứu

Phương trình động lực học ba bậc tự do mô tả chuyển động tàu mặt nước trong mặt phẳng ngang như sau:

$$\begin{cases} \dot{\eta}(t) = J(\eta)v \\ M\dot{v} + C(v)v + D(v)v + g(\eta) = \tau + \Delta(\eta, v) \end{cases} \quad (1)$$

Trong đó: $\eta = [x, y, \psi]^T$ biểu thị vec-tơ vị trí và hướng trong hệ trục tọa độ Trái đất (*n-frame*), $v = [u, \nu, r]^T$ biểu thị vec-tơ vận tốc dài và vận tốc góc trong hệ tọa độ gắn thân tàu (*b-frame*) và $J(\eta)$ ma trận chuyển đổi, $M = M^T$ là ma trận quán tính hệ thống, $C(v)$ ma trận Coriolis và lực hướng tâm hệ thống, $D(v)$ ma trận suy giảm thủy động lực học, $g(\eta) \in \mathbb{R}^3$ là vec-tơ lực đẩy và lực trọng trường, $\Delta(\eta, v)$ vec-tơ của nhiễu ngoài biến đổi theo thời gian (gió, sóng, dòng chảy đại dương,...) τ là vec-tơ của đầu vào điều khiển. Các ma trận $C(v)$ và ma trận $D(v)$ như sau:

$$M = \begin{bmatrix} m_{11} & 0 & 0 \\ 0 & m_{22} & m_{23} \\ 0 & m_{23} & m_{33} \end{bmatrix}; C(v) = \begin{bmatrix} 0 & 0 & -c_{13} \\ 0 & 0 & c_{23} \\ c_{13} & -c_{23} & 0 \end{bmatrix}; D(v) = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & d_{23} \\ 0 & d_{32} & d_{33} \end{bmatrix}$$

2.1. Xây dựng mô hình động học tàu mặt nước cho bài toán điều khiển bám tối ưu

Định nghĩa vec-tơ sai lệch bám vị trí, vận tốc $z_\eta, z_v \in \mathbb{R}^3$

$$z_\eta = \eta - \eta_d \quad (2)$$

$$z_v = v - v_d \quad (3)$$

Trong đó $\eta_d = [x_d, y_d, \psi_d]^T$ là vec-tơ của các giá trị vị trí và góc Yaw cài đặt, $\alpha_1 \in \mathbb{R}^3$ là tín hiệu điều khiển ảo cho hệ con thứ nhất. Đạo hàm $\dot{z}_\eta$ theo thời gian ta có:

$$\dot{z}_\eta = J(\eta)v - \dot{\eta}_d \quad (4)$$

Chọn tín hiệu điều khiển ảo:

$$\alpha_1 = J^{-1}(\eta)(-K_1 z_\eta + \dot{\eta}_d) \quad (5)$$

Do ma trận $J(\eta) \in \mathbb{R}^{3 \times 3}$ là một ma trận vuông. Coi v_d là bộ điều khiển ảo của hệ con thứ nhất, dễ dàng thiết kế $v_d = \alpha_1$ làm hệ (1) ổn định tiệm cận:

$$v_d(z_\eta, z_v) = \alpha_1 = J^{-1}(\eta)(-K_1 z_\eta + \dot{\eta}_d) \quad (6)$$

Đạo hàm công thức (6) ta được:

$$\dot{z}_v = \dot{v} - \dot{v}_d \quad (7)$$

Kết hợp (7) với hệ phương trình mô tả động học của tàu mặt nước không xét tới nhiễu ngoài (1) ta thu được:

$$\dot{z}_v = M^{-1}[-C(v) - D(v) + g(\eta) + \tau] - \dot{v}_d \quad (8)$$

Thay $v = v_d$ vào phương trình (9) ta được:

$$M\dot{v}_d = -C(v_d)v_d - D(v_d)v_d - g(\eta) + \tau_d \quad (9)$$

Luật điều khiển feedforward để duy trì cho hệ (9) khi hệ đạt tới trạng thái xác lập $z_v = 0$ và (τ_d, v_d) cũng thỏa mãn hệ (9). Bộ điều khiển τ_d tương ứng với quỹ đạo đặt v_d mong muốn:

$$\tau_d = M\dot{v}_d + C(v_d)v_d - D(v_d)v_d + g(\eta) \quad (10)$$

Định nghĩa biến trạng thái mới $X = [z_\eta^T, z_v^T, \eta_d^T]^T$ và các quan hệ thỏa mãn $\dot{\eta}_d = h_1(\eta_d)$; $\dot{\eta}_d = h_2(\eta_d)$ cùng với các công thức (4), (6) và (8) ta được hệ phương trình trạng thái mới của tàu mặt nước:

$$\frac{d}{dt} \begin{bmatrix} z_v \\ z_\eta \\ \eta_d \end{bmatrix} = \begin{bmatrix} A(z_v + v_d) - h_d(v_d) + M^{-1}\tau_d \\ J(z_\eta + v_d)z_v - K_1 z_\eta \\ h_1(\eta_d) \end{bmatrix} + \begin{bmatrix} M^{-1} \\ 0_{3 \times 3} \\ 0_{3 \times 3} \end{bmatrix} u \quad (11)$$

Ta có thể viết lại như sau:

$$\dot{X} = F(X) + G(X)u \quad (12)$$

Trong đó:

$$F(X) = \begin{bmatrix} A(z_v + v_d) - h_d(v_d) + M^{-1}\tau_d \\ J(z_\eta + v_d)z_v - K_1 z_\eta \\ h_1(\eta_d) \end{bmatrix}; G(X) = \begin{bmatrix} M^{-1} \\ 0_{3 \times 3} \\ 0_{3 \times 3} \end{bmatrix}; \tau = u + \tau_d$$

$$A(z_v + v_d) = M^{-1}[-C(v)v - D(v)v + g(\eta)]$$

2.2. Cơ sở của điều khiển tối ưu

Định nghĩa hàm Bellman:

$$V^*(X) = \min_{u(X) \in \psi(\Omega)} \int_t^\infty r(X, u) dt \quad (13)$$

Trong đó $r(X, u) = X^T Q X + u^T R u$. Với $Q = Q^T \geq 0$ là hàm xác định dương của $R = R^T$ là ma trận đối xứng xác định dương. Hàm chi phí tối ưu thỏa mãn phương trình HJB

$$0 = \min_{u(X) \in \psi(\Omega)} [H(X, u, \nabla V_X^*)] \quad (14)$$

Giả sử mức tối thiểu của vế phải phương trình (14) tồn tại và là duy nhất thì luật điều khiển tối ưu như sau:

$$u^*(X) = \frac{-1}{2} R^{-1} G^T(X) \nabla V_X^*(X) \quad (15)$$

Sử dụng phương trình (14) vào phương trình (15) và $r(X, u) = X^T Q X + u^T R u$ ta được phương trình HJB theo tham số $\nabla V^*(X)$ như sau:

$$\begin{cases} (\nabla V_X^*)^T F(X) - \frac{1}{4} (\nabla V_X^*)^T G(X) R^{-1} G^T(X) \nabla V_X^* + X^T Q X = 0 \\ V^*(0) = 0 \end{cases} \quad (16)$$

2.3. Thuật toán lặp PI để giải phương trình HJB

Thay vì chuyển phương trình HJB về dạng phương trình vi phân (16), để phương trình HJB ở dạng tích phân như sau:

$$V^*(X) = \int_t^{t+T} r(X, u) dt + V^*(X(t+T)) \quad (17)$$

Thuật toán lặp PI xây dựng dựa trên phương trình HJB tích phân (17).

Thuật toán 1: Thuật toán OIRL

Bước 1: $\forall X \in \Omega$ khởi tạo luật điều khiển chấp nhận được $u^0(X)$ và giá trị $V^0(X) = 0, i = 0$

Bước 2: Xác định $V^{(i)}(X)$ từ hệ phương trình:

$$\begin{cases} V^{(i)}(X) = \int_t^{t+T} r(X, u^{(i)}) dt + V^{(i)}(X(t+T)) \\ V^{(i)}(0) = 0 \end{cases} \quad (18)$$

Bước 3: i) Cập nhật luật điều khiển cho vòng lặp kế tiếp theo.

$$u^{(i+1)}(X) = \frac{-1}{2} R^{-1} G^T(X) \nabla V_X^{(i)}(X) \quad (19)$$

ii) Đến khi $u^{(i+1)}(X)$ hội tụ đến luật điều khiển tối ưu $u^*(X) \in \psi(\Omega)$ và hàm chi phí $V^{(i)}(X)$ hội tụ đến hàm chi phí tối ưu $V^*(X)$.

Định lý 1: Giải hệ phương trình (18) để tìm nghiệm $V^{(i)}(X)$ tương đương với việc tìm nghiệm của phương trình:

$$\begin{cases} (\nabla V_X^{(i)})^T (F(X) + G(X)u^{(i)}(X)) + r(X, u^{(i)}(X)) = 0 \\ V^{(i)}(0) = 0 \end{cases} \quad (20)$$

Chứng minh: Định lý được chứng minh trong [13].

Hệ phương trình (18) và hệ phương trình (20) có chung nghiệm nhưng hệ phương trình (18) không yêu cầu về động học nội của hệ thống (1) là $F(X)$. Do đó, thuật toán OIRL chỉ yêu cầu một phần động học nội của hệ thống (1) là $G(X)$ để tính toán luật điều khiển (19).

Định lý 2 (sự hội tụ): Thuật toán OIRL sẽ đưa $V^{(i)}(X)$ và $u^{(i)}(X)$ hội tụ đều về nghiệm tối ưu, tức là:

$$\forall \varepsilon > 0, \exists i \geq i_0$$

$$\sup_{X \in \Omega} |V^{(i)}(X) - V^*(X)| < \varepsilon; \sup_{X \in \Omega} |V^{(i)}(X) - u^*(X)| < \varepsilon; \quad (21)$$

Chứng minh:

Trong công trình [14] và [15] đã chỉ ra rằng việc lặp đi lặp lại trên hệ phương trình (20) và phương trình (19) được điều hòa bởi một luật điều khiển ban đầu chấp nhận được $u^{(0)}(X)$, tất cả các luật điều khiển tiếp theo sẽ được

chấp nhận và lặp đi lặp lại $V^{(i)}(X)$ và $u^{(i)}(X)$ sẽ hội tụ tới nghiệm của phương trình HJB, do đó biểu thức (21) sẽ thỏa mãn.

2.4. Ứng dụng mạng nơ-ron cho thuật toán OIRL

Thuật toán OIRL này sử dụng một mạng nơ-ron nhằm xấp xỉ hàm chi phí tối ưu $V^{(i)}(X)$ với $X \in \Omega$ như sau:

$$\hat{V}_L^{(i)}(X) = (\hat{W}_L^{(i)})^T \phi_L(X) \quad (22)$$

Trong đó $\hat{W}_L^{(i)} \in \mathbb{R}^L$ là ma trận trọng số lý tưởng chưa biết, L là số nơ-ron, $\phi_L(X) = [\phi_1(X), \phi_2(X), \phi_3(X) \dots \phi_L(X)]^T \in \mathbb{R}^L$ là véc-tơ các hàm cơ bản phù hợp. Sử dụng mạng nơ-ron xấp xỉ cho hàm chi phí tối ưu thay công thức (22) vào công thức (18) ta được:

$$(\hat{W}_L^{(i)})^T \phi_L(X) = \int_t^{t+T} r(X(t), u^{(i)}(t)) dt + (\hat{W}_L^{(i)})^T \phi_L(X(t+T)) \quad (23)$$

Xuất hiện $e(X(t), T)$ là sai số xấp xỉ của hàm Bellman

$$e(X(t), T) = (\hat{W}_L^{(i)})^T (\phi_L(X(t+T)) - \phi_L(X(t))) + \int_t^{t+T} r(X(t), u^{(i)}(t)) dt \quad (24)$$

$$\text{Đặt } h(t) = \phi_L(X(t+T)) - \phi_L(X(t)); y(t) = \int_t^{t+T} r(X(t), u^{(i)}(t)) dt$$

Phương trình (24) được viết lại:

$$e(X(t), T) = (\hat{W}_L^{(i)})^T h(t) + y(t) \quad (25)$$

Ta chỉnh định tham số $\hat{W}_L^{(i)}$ để tối thiểu hóa bình phương sai lệch $e^2(X(t), T)$. Vì $e(X(t), T)$ là hàm tuyến tính của tham số $\hat{W}_L^{(i)}$. Dựa trên việc thu thập dữ liệu trong n trích mẫu

$$H = [h(t_1), h(t_2), h(t_3) \dots h(t_n)]; Y = [y(t_1), y(t_2), y(t_3) \dots y(t_n)]^T \text{ do đó ta có thể áp}$$

dụng thuật toán LSE để tính $\hat{W}_L^{(i)}$ tối ưu:

$$\hat{W}_L^{(i)} = -(H^T H)^{-1} H Y \quad (26)$$

Thuật toán 2: Thuật toán OIRL sử dụng mạng nơ-ron

Bước 1: Khởi tạo $\forall X \in \Omega$ chọn luật điều khiển được

$$u^{(0)}(X) \in \psi(\Omega), i = 1, \varepsilon_w.$$

Bước 2: Vòng lặp

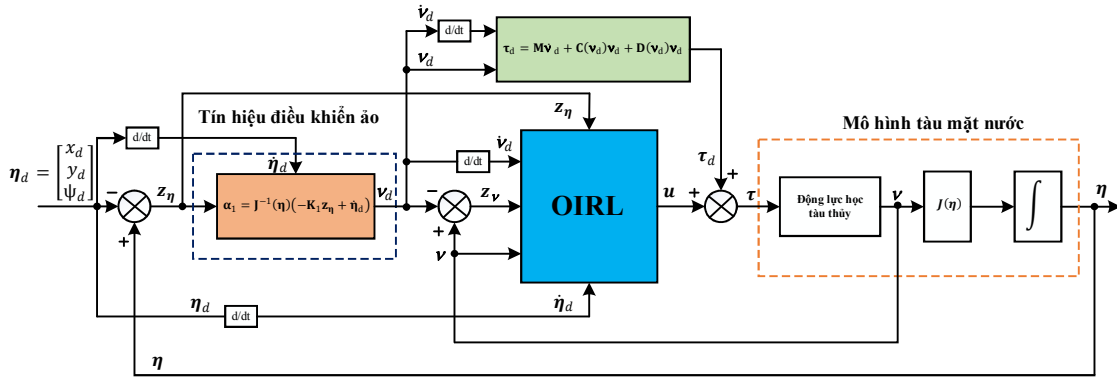
i) Sử dụng các thông tin đã thu thập về hệ thống để tính H và Y . Xác định $\hat{W}_L^{(i)}$ từ phương trình (26). Cập nhật luật điều khiển cho vòng lặp kế tiếp theo.

$$u^{(i+1)}(X) = -\frac{1}{2} R^{-1} G^T(X) \left(\frac{\partial \phi(X)}{\partial X} \right)^T \hat{W}_L^{(i)} \quad (27)$$

Bước 3: Đến khi nếu thỏa mãn tiêu chuẩn hội tụ sao cho $\|\hat{W}_L^{(i+1)}(X) - \hat{W}_L^{(i)}(X)\| < \varepsilon$, kết thúc giải thuật. Nếu không thỏa mãn, gán $i \leftarrow i + 1$, cho tín hiệu $u^{(i)}(X)$ vào hệ thống và thu thập thông tin cần thiết của hệ thống về trạng thái, tín hiệu điều khiển tại n trích mẫu khác nhau trong khoảng thời gian T rồi quay lại bước 2.

2.5. Cấu trúc điều khiển cho mô hình tàu mặt nước

Mô hình của tàu thủy truyền ngược, do đó hệ thống điều khiển được thiết lập với hệ thống điều khiển như hình 1. Thứ nhất bộ điều khiển ảo được thiết kế dựa trên hàm Lyapunov, sẽ tính toán vec-tơ vận tốc đặt mong muốn v_d cho bộ điều khiển tối ưu OIRL. Thứ hai, bộ điều khiển động lực học là cấu trúc của bộ điều khiển tối ưu và luật điều khiển Feed Forward



Hình 1. Cấu trúc điều khiển của mô hình tàu mặt nước sử dụng thuật toán OIRL

3. MÔ PHỎNG VÀ ĐÁNH GIÁ KẾT QUẢ

3.1. Tham số mô phỏng

Để kiểm chứng tính đúng đắn của thuật toán của bộ điều khiển tối ưu OIRL, bài báo thực hiện mô phỏng số trên phần mềm Matlab với mô hình tàu

được thu nhỏ với tỷ lệ so với tàu thật 1:75. Khối lượng của tàu là $m = 21kg$, chiều dài và chiều rộng tương ứng là 1.2m và 0.3m, với các tham số được tham khảo trong tài liệu [7] như sau:

$$M = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 19.2 & 0.72 \\ 0 & 0.72 & 2.7 \end{bmatrix}; C(v) = \begin{bmatrix} 0 & 0 & -19\nu + 0.72r \\ 0 & 0 & 20u \\ 19\nu + 0.72r & -20u & 0 \end{bmatrix}$$

$$D(v) = \begin{bmatrix} 0.72 + 1.3|u| + 5.8|u|^2 & 0 & 0 \\ 0 & 0.86 + 36|u| + 3|r| & -0.1 - 2|\nu| + 2|r| \\ 0 & -0.1 - 5|\nu| + 3|r| & 6 + 4|\nu| + 4|r| \end{bmatrix}$$

$$g_1(\eta) = g_2(\eta) = g_3(\eta) = 0$$

Trong bài báo sử dụng mạng nơ-ron một lớp truyền thẳng với số nút của mạng là $L = 12$, quỹ đạo bám mong muốn của tàu mặt nước là $\eta_d(t) = [12 \sin(0.2t), -12 \sin(0.2t), 0.2t]^T$ tham số của bộ điều khiển

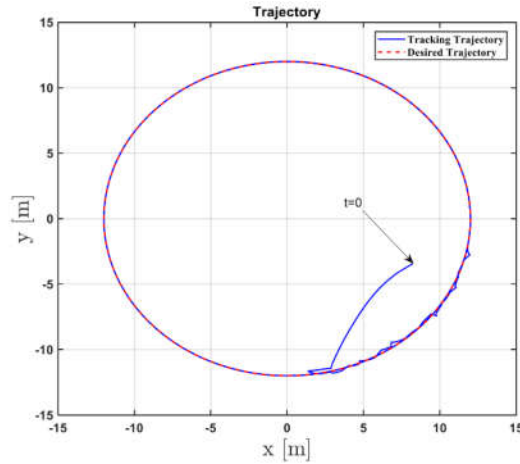
$$K_1 = 0.3I_{3 \times 3}, Q_T = \begin{bmatrix} Q & 0_{3 \times 3} \\ 0_{6 \times 3} & 0_{6 \times 3} \end{bmatrix}; Q = \text{diag}([1;1;1]), R = Q = \text{diag}([1;1;1])$$

Véc-tơ hàm tác động $\phi(X) \in \mathbb{R}^{12}$ và vec-tơ trọng số mạng nơ-ron được chọn như sau:

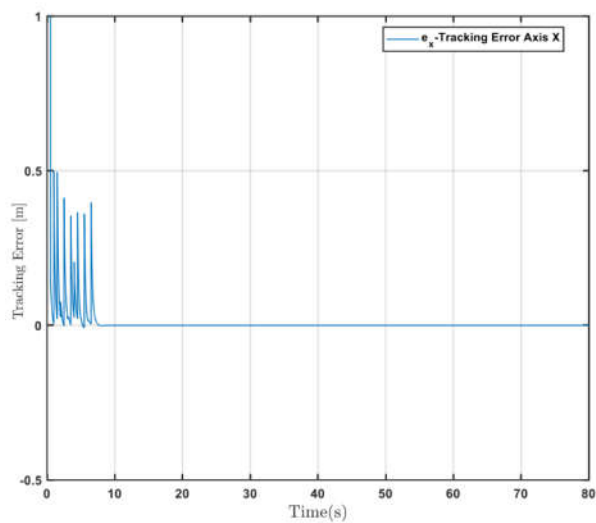
$$\phi(X) = [X_1^2, X_1X_2, X_1X_3, X_2^2, X_2X_3, X_3^2, X_1^2X_7, X_2^2X_8, X_3^2X_9, X_1^2X_4, X_2^2X_5, X_3^2X_6]^T \quad (28)$$

Luật cập nhật cho mạng nơ-ron được thực hiện dựa trên (26).

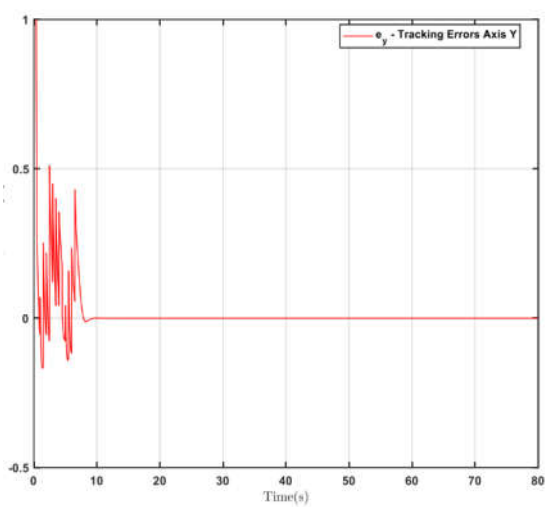
4.2. Kết quả mô phỏng



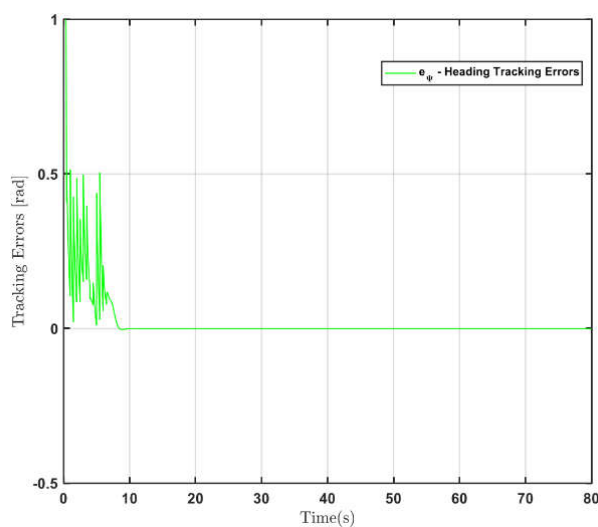
Hình 2. Quỹ đạo bám của tàu mặt nước theo quỹ đạo tròn sử dụng thuật toán OIRL



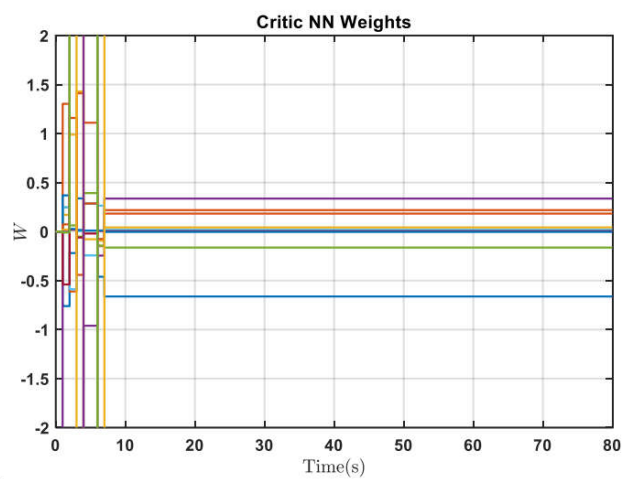
Hình 3. Sai số bám quỹ đạo theo trục x



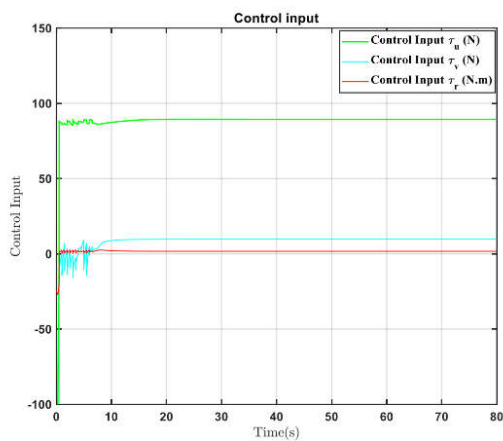
Hình 4. Sai số bám quỹ đạo theo trục y



Hình 5. Sai số bám quỹ đạo theo trục ψ



Hình 6. Sự hội tụ của trọng số W của mạng nơ-ron Critic với thuật toán OIRL



Hình 7. Tín hiệu điều khiển đầu vào của tàu mặt nước

Trọng số mạng nơ-ron W hội tụ về giá trị:

$$W = \begin{bmatrix} -0,6620 & 0,2190 & 0,0425 & 0,3367 & -0,1631 & 0,0173 \\ 0,0045 & -0,0031 & 0,1835 & 8,1800 & 6,8034 & -4,6509 \end{bmatrix}^T \quad (27)$$

3.3. Nhận xét kết quả

Trong kết quả mô phỏng của thuật toán, đáp ứng đầu ra bám quỹ đạo của tàu mặt ở hình 2 có chất lượng rất tốt. Điểm ban đầu $t=0$ được chọn ngẫu nhiên, do thuật toán dựa trên học lặp thu thập động học của hệ thống một cách trực tuyến, một giá trị nhiễu ngẫu nhiên được thêm vào thỏa mãn điều kiện PE. Sau khi thu thập đủ động học của hệ thống để tính toán ma trận trọng số của mạng nơ-ron, sau đó tính toán luật điều khiển tối ưu cho tàu mặt nước, sai lệch bám theo x, y, ψ về xấp xỉ về 0 sau 7s thể hiện trên các hình 3, hình 4, hình 5, ma trận trọng số của mạng nơ-ron hội tụ như hình 6. Hình 7 cho thấy tín hiệu điều khiển lúc ban đầu dao động, quá trình này rất nhanh, ngay sau đó thì tín hiệu điều khiển chỉ dao động trong khoảng $5[N][N.m] \div 80[N][N.m]$. Như vậy, bộ điều khiển bám tối ưu dựa trên thuật toán OIRL hoàn toàn đáp ứng yêu

cầu điều khiển bám quỹ đạo của hệ thống tàu mặt nước.

4. Kết luận

Bài báo đã trình bày phương pháp thiết kế bộ điều khiển bám tối ưu dựa trên thuật toán OIRL được thiết kế cho tàu mặt nước có bất định trong mô hình. Trong cấu trúc điều khiển một bộ điều khiển ảo được thiết kế dựa trên hàm điều khiển Lyapunov và bộ điều khiển Feed forward được thiết kế cho hệ thống không dừng với số lượng biến trạng thái nhỏ. Nó cho phép phát triển thuật toán OIRL cho quỹ đạo phụ thuộc thời gian. Kết quả mô phỏng chứng minh tính khả thi của thuật toán.

TÀI LIỆU THAM KHẢO

1. N. Wang, G. Xie, X. Pan, and S. F. Su, 2019. *Full-State Regulation Control of Asymmetric Underactuated Surface Vehicles*. IEEE Trans. Ind. Electron., vol. 66, no. 11, pp. 8741–8750.
2. N. Wang, H. R. Karimi, H. Li, and S. F. Su, 2019. *Accurate*

- Trajectory Tracking of Disturbed Surface Vehicles: A Finite-Time Control Approach.* IEEE/ASME Transactions on Mechatronics, vol. 24, no. 3. pp. 1064–1074.
3. H. Qin, C. Li, Y. Sun, X. Li, Y. Du, and Z. Deng, 2020. *Finite-time trajectory tracking control of unmanned surface vessel with error constraints and input saturations.* Journal of the Franklin Institute, vol. 357, no. 16. pp. 11472–11495.
 4. B. S. Park, J. W. Kwon, and H. Kim, 2017. *Neural network-based output feedback control for reference tracking of underactuated surface vessels.* Automatica, vol. 77. pp. 353–359.
 5. M. Van, 2019. *Adaptive neural integral sliding-mode control for tracking control of fully actuated uncertain surface vessels.* International Journal of Robust and Nonlinear Control, vol. 29, no. 5. pp. 1537–1557.
 6. Z. Zheng, L. Ruan, M. Zhu, and X. Guo, 2020. *Reinforcement learning control for underactuated surface vessel with output error constraints and uncertainties.* Neurocomputing, vol. 399. pp. 479–490.
 7. G. Wen, S. S. Ge, C. L. P. Chen, F. Tu, and S. Wang, 2019. *Adaptive tracking control of surface vessel using optimized backstepping technique.* IEEE Trans. Cybern., vol. 49, no. 9, pp. 3420–3431.
 8. G. Wen, C. L. P. Chen, and S. S. Ge, 2020. *Simplified Optimized Backstepping Control for a Class of Nonlinear Strict-Feedback Systems With Unknown Dynamic Functions* IEEE Trans. Cybern., pp. 1–14.
 9. R. C. Xinxin Guo, Weisheng Yan, 2019. *Integral Reinforcement Learning-Based Adaptive Systems With Unknown Control Directions.* IEEE Trans. Syst. Man, Cybern. Syst., vol. PP, pp. 1–10.
 10. Z. Yin, W. He, C. Yang, and C. Sun, 2018. *Control Design of a Marine Vessel System Using Reinforcement Learning.* Neurocomputing, vol. 311, pp. 353–362.
 11. R. Cui, C. Yang, Y. Li, and S. Sharma, 2017. *Adaptive Neural Network Control of AUVs with Control Input Nonlinearities Using Reinforcement Learning.* IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, no. 6. pp. 1019–1029.
 12. R. Kamalapurkar, H. Dinh, S. Bhasin, and W. E. Dixon, 2015. *Approximate optimal trajectory tracking for continuous-time*

- nonlinear systems*. Automatica, vol. 51. pp. 40-48.
13. F. L. Draguna Vrabie, 2009. *Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems*. Neural Networks, vol. 22, no. 3, pp. 237–246.
14. R. W. Beard, G. N. Saridis, and J. T. Wen, 1997. *Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation*,” Automatica, vol. 33, no. 12. pp. 2159–2177.
15. Murad Abu-Khalaf and Frank L Lewis, 2005. *Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network hjb approach*. Automatica, vol. 41(5), pp. 779-791.