

## ĐIỀU KHIỂN BÁM TÀU MẶT NƯỚC BẤT ĐỊNH MÔ HÌNH VÀ NHIỀU NGOÀI THÔNG QUA BỘ ĐIỀU KHIỂN HỌC TĂNG CƯỜNG THÍCH NGHI TRỰC TUYẾN VÀ RISE

Vũ Văn Tú<sup>1, 2\*</sup>, Đào Phương Nam<sup>1</sup>, Phan Xuân Minh<sup>1</sup>

**Tóm tắt:** Bài viết đề cập tới cách tiếp cận một bộ điều khiển bám quỹ đạo áp dụng cho tàu bề mặt có mô hình bất định và chịu ảnh hưởng nhiễu sử dụng cấu trúc mới dựa trên thuật toán học tăng cường thích nghi (ARL) và tích phân bền vững theo hàm dấu sai lệch bám RISE. Để đảm bảo hiệu suất bám tiệm cận, RISE được sử dụng trong thiết kế điều khiển phản hồi để bù các thành phần bất định trong mô hình và nhiễu loạn bên ngoài. Kỹ thuật học tăng cường thích nghi động được áp dụng để hiệu chỉnh đồng bộ mạng nơ-ron Actor-Critic dùng để xấp xỉ điều khiển tối ưu và hàm chi phí tương ứng. Sự hội tụ của trọng số cũng như vấn đề điều khiển bám được phân tích dựa trên lý thuyết tối ưu và hàm Lyapunov. Cuối cùng, mô phỏng kỹ thuật số được thực hiện để chứng minh sự hiệu quả của thuật toán đề xuất.

**Từ khóa:** Điều khiển tối ưu; Quy hoạch động thích nghi; Tích phân bền vững hàm dấu sai lệch bám RISE; Hệ thống tàu bề mặt; Điều khiển bám quỹ đạo.

### 1. ĐẶT VẤN ĐỀ

Trong những thập kỷ qua, điều khiển bám quỹ đạo của phương tiện hàng hải đã nhận được sự chú ý đáng kể trong công nghệ hàng hải và kỹ thuật điều khiển bởi các ứng dụng rộng rãi của nó như giao thông vận tải, khảo sát môi trường, giám sát và nhiều ứng dụng quân sự và thương mại [1, 2]. Vì các tàu làm việc trong điều kiện khắc nghiệt của môi trường, hệ thống điều khiển bám quỹ đạo phải đối mặt với những thách thức không nhỏ do sự xuất hiện của bất định mô hình và nhiễu loạn bên ngoài [3, 4].

Tất cả các nghiên cứu trên là cơ sở để thiết kế thuật toán điều khiển bám quỹ đạo để giải quyết các vấn đề như cơ cấu chấp hành bão hòa, dự báo hiệu suất, nhiễu loạn môi trường và bất định mô hình sử dụng mạng nơ-ron và bộ quan sát. Tuy nhiên, việc tối ưu hóa hiệu suất chưa được nghiên cứu đầy đủ. Bộ điều khiển tối ưu không chỉ giải quyết vấn đề điều khiển bám quỹ đạo của hệ thống phi tuyến mà còn đảm bảo hiệu suất bám của hệ thống được tối ưu hóa.

Trong những năm gần đây, giải thuật học tăng cường (Reinforcement Learning) đã được phát triển để thiết kế bộ điều khiển tối ưu cho mô hình tàu phi tuyến dựa trên việc giải phương trình Hamilton – Jacobi – Bellman (HJB) với nhiều cách tiếp cận như cấu trúc Actor/Critic, kỹ thuật học tăng cường tích phân trực tuyến (On Policy Integral Reinforcement Learning),... [5-11].

Các công trình [5, 6] đã đề xuất thuật toán điều khiển tối ưu bằng cách sử dụng giải thuật học tăng cường (RL) cho các hệ thống tàu mặt nước có động học đã biết. Thuật toán điều khiển tối ưu dựa trên giải thuật lặp PI (Policy Iteration) và nó có thể thu được các giá trị xấp xỉ thích hợp của hàm chi phí và luật điều khiển được tối ưu hóa. Giải thuật RL dựa trên cấu trúc mạng nơ-ron actor/critic (AC-NNs), trong đó, mạng nơ-ron critic (critic NN) nhằm ước tính hàm chi phí và mạng nơ-ron actor (actor NN) được sử dụng để thiết kế đầu vào phù hợp và giảm thiểu sai lệch bám. AC-NN đã được mở rộng sang lĩnh vực điều khiển bám quỹ đạo cho các phương tiện hàng hải bao gồm tàu thủy thiếu cơ cấu chấp hành (USVs) và phương tiện tự hành dưới nước [10, 11].

Trong công trình [7], kỹ thuật điều khiển mới có tên là tối ưu hóa Backstepping dựa trên kỹ thuật Backstepping kết hợp với cấu trúc AC-NNs để thực hiện các giải pháp tối ưu hóa cho các hệ thống con tương ứng. Bởi vì thực hiện các giải pháp tối ưu hóa cho các hệ con, trong công trình nghiên cứu này các tác giả sử dụng nhiều mạng nơ-ron dẫn đến tốn tài nguyên, tốc độ hội tụ giảm, 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. Tuy nhiên, trong các công trình trên vấn đề ảnh hưởng của các yếu tố như bất định mô hình và nhiễu loạn bên ngoài tới bộ điều khiển đã thiết kế không được xem xét.

Công trình [9] đề xuất một phương pháp điều khiển bám thích nghi dựa trên cấu trúc AC-NNs cho USVs có thành phần bất định mô hình, nhiễu loạn môi trường và ràng buộc tín hiệu đầu ra bằng cách sử dụng kỹ thuật chuyển đổi sai lệch để xử lý vấn đề ràng buộc sai lệch đảm bảo rằng USV có thể bám chính xác quỹ đạo bám. Đối tượng áp dụng của công trình nghiên cứu này là tàu thủy thiếu cơ cấu chấp hành không phải tàu thủy đủ cơ cấu chấp hành.

Trong bài báo này, nghiên cứu sẽ tập trung vào việc giải quyết bài toán điều khiển tối ưu bám quỹ đạo cho tàu mặt nước đủ cơ cấu chấp hành có mô hình phi tuyến bất định và hoạt động trong môi trường có nhiễu loạn. Bộ điều khiển bám phản hồi được đề xuất trên cơ sở kết hợp học tăng cường thích nghi động kết hợp với RISE đảm bảo hệ thống kín ổn định, bám quỹ đạo và bù ảnh hưởng của nhiễu cũng như các thành phần bất định của tàu.

Những đóng góp chính của bài báo này tập trung vào những nội dung sau:

1. Xây dựng được cấu trúc AC-NNs để ước lượng hàm chi phí và luật điều khiển tối ưu, luật cập trọng số cả hai mạng nơ-ron đồng bộ trong thời gian thực.
2. Bộ điều khiển tối ưu được xây dựng dựa trên giải thuật RL, bổ xung bộ ước lượng bất định mô hình và nhiễu ngoài RISE. Bằng cách này, bộ điều khiển đề xuất đảm bảo điều khiển tối ưu bằng thuật toán học tăng cường thích nghi động cho phương trình HJB cho hệ có mô hình bất định và chịu ảnh hưởng của nhiễu.
3. Bộ điều khiển được đề xuất cải thiện tốc độ hội tụ của sai số bám và độ quá điều chỉnh nhỏ. Sự kết hợp của RISE và học tăng cường thích nghi động (Adaptive Reinforcement Learning) trong thiết kế bộ điều khiển tối ưu bám giúp giải quyết được vấn đề bất định mô hình và nhiễu, những vấn đề mà chỉ sử dụng hoặc tăng cường thích nghi động chưa giải quyết được [5-7].

## **2. MÔ HÌNH TÀU ĐỘNG HỌC TÀU MẶT NƯỚC**

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

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

Trong đó:  $\eta = [x, y, \psi]^T$  biểu thị véc-tơ vị trí và hướng trong hệ trục tọa độ Trái đất (*n-frame*),  $v = [u, v, r]^T$  biểu thị véc-tơ vận tốc dài và vận tốc góc trong hệ tọa độ gắn thân tàu (*b-frame*) và  $J(\eta)$  ma trận chuyển đổi và là ma trận trực giao  $J^{-1}(\eta) = J^T(\eta)$ ,  $M = M^T$  là ma trận quán tính hệ thống,  $C(v) = -C^T(v)$  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 véc-tơ lực đẩy và lực trọng trường,  $d(t) = [d_1(t), d_2(t), d_3(t)]^T$  véc-tơ của nhiễu ngoài biến đổi theo thời gian (gió, sóng, dòng chảy đại dương, ...)  $\tau = [\tau_u, \tau_v, \tau_r]^T$  là véc-tơ của đầu vào điều khiển. Các ma trận  $M, C(v), D(v)$  và ma trận  $J(\eta)$  như sau:

$$J(\eta) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}; \quad C(v) = \begin{bmatrix} 0 & 0 & -c_{13} \\ 0 & 0 & c_{23} \\ c_{13} & -c_{23} & 0 \end{bmatrix};$$

$$D(v) = \begin{bmatrix} d_{11} & 0 & 0 \\ 0 & d_{22} & d_{23} \\ 0 & d_{32} & d_{33} \end{bmatrix}; \quad M = \begin{bmatrix} m_{11} & 0 & 0 \\ 0 & m_{22} & m_{23} \\ 0 & m_{23} & m_{33} \end{bmatrix}.$$

Để thuận tiện trong việc thiết kế bộ điều khiển, chuyển đổi động học của hệ (1) thành phương trình sau:

$$\bar{M}(\eta)\ddot{\eta} + \bar{C}(\eta, \dot{\eta}, v)\dot{\eta} + \bar{D}(\eta, v)\dot{\eta} + \bar{g}(\eta) + \tau_n(t) = \tau_d \quad (2)$$

Trong đó:  $\bar{M} \triangleq J^T M J^{-1}$ ,  $\bar{C} \triangleq J^T (C - M J^{-1} \dot{J}) J^{-1}$ ,  $\bar{D} \triangleq J^T D J^{-1}$ ,  $\bar{g} \triangleq J^T g$ ,  $\tau_n \triangleq J^T \tau$ ,  $\tau_d \triangleq J^T \tau$ . Sự phát triển của mô hình (2) dựa trên các giả thiết là  $\eta$  và  $v$  là đo được và  $\bar{M}$ ,  $\bar{C}$ ,  $\bar{D}$ ,  $\bar{g}$  và  $\tau_n$  là những hàm chưa biết.

**Giả thiết 1:** Véc-tơ vị trí và hướng  $\eta$  và đạo hàm  $\dot{\eta}$  của nó bị chặn bởi hằng số dương  $\bar{\eta}_1, \bar{\eta}_2 \in \mathbb{R}^+$  sao cho  $\|\eta\| \leq \bar{\eta}_1, \|\dot{\eta}\| \leq \bar{\eta}_2$ .

**Giả thiết 2:** Nếu  $\eta, \dot{\eta}$  bị chặn thì tất cả những hàm này  $\bar{C}(\eta, \dot{\eta}, v), \bar{D}(\eta, v), \bar{g}(\eta)$  bị chặn. Hơn thế, nếu  $\eta, \dot{\eta}$  bị chặn thì đạo hàm riêng phần bậc nhất và bậc hai của các thành phần  $\bar{C}(\eta, \dot{\eta}, v), \bar{D}(\eta, v), \bar{g}(\eta)$  theo  $\eta$  là tồn tại và bị chặn.

**Giả thiết 3:** Quỹ đạo mong muốn  $\eta_d$  được thiết kế sao cho đạo hàm theo thời gian thứ  $i$  của  $\eta_d \forall i = 0, 1, \dots, 4$  là tồn tại và bị chặn.

**Giả thiết 4:** Thành phần nhiễu phi tuyến và đạo hàm bậc nhất, bậc hai của nó  $\tau_n, \dot{\tau}_n, \ddot{\tau}_n$  bị chặn bởi hằng số đã biết.

**Giả thiết 5:** Với  $\eta_d$  là một quỹ đạo đặt bị chặn và giả thiết rằng tồn tại một hàm Lipschitz toàn cục  $h_d(\cdot)$  thỏa mãn  $\frac{d}{dt}\eta_d = h_d(\eta_d)$ ,  $\ddot{\eta}_d = h_d(\dot{\eta}_d)$ .

Sai lệch bám quỹ đạo của tàu mặt nước được tính như sau:

$$e_1 = \eta_d - \eta \quad (3)$$

Để tạo điều kiện cho việc phân tích tiếp theo, các sai số bám đã được lọc, ký hiệu là  $e_2 \in \mathbb{R}^3$  được định nghĩa là:

$$e_2 = \dot{e}_1 + \lambda_1 e_1 \quad (4)$$

Thiết lập sai số bám quỹ đạo đã lọc được định nghĩa:

$$r = \dot{e}_2 + \lambda_2 e_2 \quad (5)$$

Trong đó,  $\lambda_1 \in \mathbb{R}^{3 \times 3} > 0$ ,  $\lambda_2 \in \mathbb{R}^+$  là một số dương.

Để phát triển một mô hình không gian trạng thái cho sai lệch bám trong (3) và (4), đạo hàm theo thời gian (4) sau đó nhân hai vế với  $\bar{M}$  và thay thế (2), (3) vào ta được:

$$\bar{M}e_2 = -(\bar{C} + \bar{D})e_2 + f + \tau_n - \tau_d \quad (6)$$

Trong đó, hàm phi tuyến  $f_d(\eta, \dot{\eta}, e_1, \eta_d, \dot{\eta}_d, \ddot{\eta}_d) \in \mathbb{R}^3$  được định nghĩa như sau:

$$f \triangleq \bar{M}(\ddot{\eta}_d + \lambda_1 \dot{e}_1) + (\bar{C} + \bar{D})(\dot{\eta}_d + \lambda_1 e_1) + \bar{g} \quad (7)$$

Để thuận tiện triệt tiêu các đại lượng trong công thức đầu vào điều khiển của hệ được thiết kế như sau:

$$\tau_d \triangleq \mathbf{f} + \tau_n - \mathbf{u} \quad (8)$$

Kết hợp (4), (6), (8) và dựa trên giả thiết 5 chuyển đổi từ mô hình phụ thuộc theo thời gian thành mô hình không phụ thuộc thời gian, với biến trạng thái mới  $\mathbf{X} = [\mathbf{e}_1^T, \mathbf{e}_2^T, \boldsymbol{\eta}_d^T, \dot{\boldsymbol{\eta}}_d^T]^T$  được hệ mới như sau:

$$\frac{d}{dt} \begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \boldsymbol{\eta}_d \\ \dot{\boldsymbol{\eta}}_d \end{bmatrix} = \begin{bmatrix} -\lambda_1 \mathbf{e}_1 + \mathbf{e}_2 \\ -\bar{\mathbf{M}}^{-1}(\bar{\mathbf{C}} + \bar{\mathbf{D}})\mathbf{e}_2 \\ \mathbf{h}_d(\boldsymbol{\eta}_d) \\ \mathbf{h}_d(\dot{\boldsymbol{\eta}}_d) \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{3 \times 3} \\ \bar{\mathbf{M}}^{-1} \\ \mathbf{0}_{3 \times 3} \\ \mathbf{0}_{3 \times 3} \end{bmatrix} \mathbf{u} \quad (9)$$

Có thể viết lại hệ mới như sau:

$$\frac{d}{dt} \mathbf{X} = \mathbf{F}(\mathbf{X}) + \mathbf{G}(\mathbf{X})\mathbf{u} \quad (10)$$

Trong đó, tín hiệu  $\mathbf{u}$  được thiết kế bằng cách sử dụng thuật toán điều khiển tối ưu cấu trúc ARL để tối thiểu hóa hàm chi phí.

### 3. THIẾT KẾ BỘ ĐIỀU KHIỂN CHO TÀU MẶT NƯỚC

#### 3.1. Học tăng cường thích ứng thiết kế dựa trên điều khiển tối ưu

Bộ điều khiển  $\mathbf{u}$  được thiết kế để tối thiểu hàm chi phí:

$$V(\mathbf{X}, \mathbf{u}) = \int_t^\infty (\mathbf{Q}(\mathbf{X}) + \mathbf{u}^T \mathbf{R} \mathbf{u}) ds \quad (11)$$

Trong đó:  $\mathbf{Q}(\mathbf{X}) \in \mathbb{R}^+ > 0$ ;  $\mathbf{Q}(\mathbf{X}) = \mathbf{X}^T \mathbf{Q}_T \mathbf{X}$ ;  $\mathbf{Q}_T \in \mathbb{R}^{12 \times 12}$ ;  $\mathbf{R} \in \mathbb{R}^{3 \times 3}$ .

Theo lý thuyết về phương trình Hamilton – Jacobi – Bellman (HJB) kinh điển bộ điều khiển tối ưu  $\mathbf{u}^*(\mathbf{X})$  và hàm chi phí tối ưu tương ứng  $V^*(\mathbf{X})$  được suy ra như sau:

$$\mathbf{u}^*(\mathbf{X}) = -\frac{1}{2} \mathbf{R}^{-1} \mathbf{G}(\mathbf{X}) \frac{\partial V^*(\mathbf{X})}{\partial \mathbf{X}} \quad (12)$$

$$V^*(\mathbf{X}, \mathbf{u}) = \min_{\mathbf{u}(\mathbf{X}) \in \mathcal{U}(\Omega)} \int_t^\infty (\mathbf{X}^T \mathbf{Q}_T \mathbf{X} + \mathbf{u}^T \mathbf{R} \mathbf{u}) ds \quad (13)$$

Do đó, việc học trực tuyến dựa trên giải pháp trực tuyến được xem xét bằng cách sử dụng mạng nơ-ron để xấp xỉ hàm chi phí tối ưu và bộ điều khiển tối ưu tương đương [12]:

$$V^*(\mathbf{X}) = \mathbf{W}^T \boldsymbol{\phi}(\mathbf{X}) + \varepsilon_v(\mathbf{X}); \mathbf{u}^*(\mathbf{X}) = -\frac{1}{2} \mathbf{R}^{-1} \mathbf{G}^T(\mathbf{X}) \left( \left( \frac{\partial \boldsymbol{\phi}}{\partial \mathbf{X}} \right)^T \mathbf{W} + \left( \frac{\partial \varepsilon_v}{\partial \mathbf{X}} \right)^T \right) \quad (14)$$

**Giả thiết 6:** Ma trận  $\mathbf{G}(\mathbf{X})$  trong biểu thức là đã biết và bị chặn, tồn tại một hằng số dương đã biết  $\bar{G} \in \mathbb{R}^+$ , thỏa mãn  $0 < \|\mathbf{G}(\mathbf{X})\| \leq \bar{G}$ .

**Giả thiết 7:** Các trọng số lý tưởng  $\mathbf{W}$  là bị chặn bởi một số dương chưa biết  $\bar{W} \in \mathbb{R}^+$ ,  $\|\mathbf{W}\| \leq \bar{W}$ .

**Giả thiết 8:** Véc-tơ hàm tác động  $\boldsymbol{\phi}(\mathbf{X})$  và đạo hàm riêng theo  $\mathbf{X}$  bị chặn.

Xét số nơ-ron cố định  $N$ , mạng nơ-ron Critic  $\hat{V}(\mathbf{X})$  và mạng nơ-ron Actor  $\hat{\mathbf{u}}(\mathbf{X})$  được sử dụng để xấp xỉ hàm chi phí và bộ điều khiển tối ưu như:

$$\hat{V}(\mathbf{X}) = \hat{\mathbf{W}}_c^T \phi(\mathbf{X}); \hat{u}(\mathbf{X}) = -\frac{1}{2} \mathbf{R}^{-1} \mathbf{G}^T(\mathbf{X}) \left( \frac{\partial \phi}{\partial \mathbf{X}} \right)^T \hat{\mathbf{W}}_a \quad (15)$$

Luật thích nghi của các trọng số mạng nơ-ron Critic  $\hat{\mathbf{W}}_c$  và mạng nơ-ron Actor  $\hat{\mathbf{W}}_a$  được thực hiện đồng thời để giảm tích phân sai số bình phương Bellman  $\delta_{hjb}$  tương ứng:

$$\begin{aligned} \delta_{hjb} = & \hat{H} \left( \mathbf{X}, \hat{u}, \frac{\partial \hat{V}}{\partial \mathbf{X}} \right) - H^* \left( \mathbf{X}, u^*, \frac{\partial V^*}{\partial \mathbf{X}} \right) = -\tilde{\mathbf{W}}_c^T \boldsymbol{\omega} + \frac{1}{4} \tilde{\mathbf{W}}_a^T \left( \frac{\partial \phi}{\partial \mathbf{X}} \right) \mathbf{G} \mathbf{R}^{-1} \mathbf{G}^T \left( \frac{\partial \phi}{\partial \mathbf{X}} \right)^T \\ & - \frac{1}{4} \left( \frac{\partial \varepsilon_v}{\partial \mathbf{X}} \right) \mathbf{G} \mathbf{R}^{-1} \mathbf{G}^T \left( \frac{\partial \varepsilon_v}{\partial \mathbf{X}} \right)^T - \left( \frac{\partial \varepsilon_v}{\partial \mathbf{X}} \right) (\mathbf{F} + \mathbf{G} u^*) \end{aligned} \quad (16)$$

Trong đó:  $\boldsymbol{\omega}(\mathbf{X}, \hat{u}) = \frac{\partial \phi}{\partial \mathbf{X}} (\mathbf{F} + \mathbf{G} \hat{u})$ ;  $\mathbf{F}(\mathbf{X}) = \mathbf{F}$ ;  $\mathbf{G}(\mathbf{X}) = \mathbf{G}$ .

Trong tài liệu [12], luật cập nhật cho trọng số  $\hat{\mathbf{W}}_c$  của mạng nơ-ron Critic được đưa ra như sau:

$$\frac{d}{dt} \tilde{\mathbf{W}}_c = \boldsymbol{\Omega}_{nom} + \Delta_{per} \quad (17)$$

Trong đó,  $\boldsymbol{\Omega}_{nom}(\tilde{\mathbf{W}}_c, t) \triangleq -k_c \boldsymbol{\Gamma} \boldsymbol{\psi} \boldsymbol{\psi}^T \tilde{\mathbf{W}}_c$

$$\begin{aligned} \Delta_{per} \triangleq & k_c \boldsymbol{\Gamma} \frac{\boldsymbol{\omega}}{1 + \gamma \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}} \left[ \frac{1}{4} \tilde{\mathbf{W}}_a^T \left( \frac{\partial \phi}{\partial \mathbf{X}} \right) \mathbf{G}(\mathbf{X}) \mathbf{R}^{-1} \mathbf{G}^T(\mathbf{X}) \left( \frac{\partial \phi}{\partial \mathbf{X}} \right)^T \tilde{\mathbf{W}}_a \right. \\ & \left. - \frac{1}{4} \left( \frac{\partial \varepsilon_v}{\partial \mathbf{X}} \right) \mathbf{G}(\mathbf{X}) \mathbf{R}^{-1} \mathbf{G}^T(\mathbf{X}) \left( \frac{\partial \varepsilon_v}{\partial \mathbf{X}} \right)^T - \left( \frac{\partial \varepsilon_v}{\partial \mathbf{X}} \right) (\mathbf{F}(\mathbf{X}) + \mathbf{G}(\mathbf{X}) u^*) \right] \end{aligned}$$

Với  $k_c, \gamma \in \mathbb{R}$  là các số dương,  $\boldsymbol{\Gamma} \in \mathbb{R}^{N \times N}$  là ma trận hệ số ước lượng đối xứng được tính và có thể bị chặn như sau [12]:

$$\frac{d}{dt} \boldsymbol{\Gamma} = -k_c \boldsymbol{\Gamma} \frac{\boldsymbol{\omega}}{1 + \gamma \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}} \boldsymbol{\Gamma}; \boldsymbol{\Gamma}(t_r^+) = \boldsymbol{\Gamma}(0) = \varphi_0 \mathbf{I} \quad (18)$$

Luật cập nhật thích nghi cho mạng nơ-ron Actor được mô tả như sau:

$$\frac{d}{dt} \tilde{\mathbf{W}}_a = -k_{a1} \frac{1}{\sqrt{1 + \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}}} \left( \frac{\partial \phi}{\partial \mathbf{X}} \right) \mathbf{G} \mathbf{R}^{-1} \mathbf{G}^T \left( \frac{\partial \phi}{\partial \mathbf{X}} \right)^T (\tilde{\mathbf{W}}_a - \tilde{\mathbf{W}}_c) \delta_{hjb} - k_{a2} (\tilde{\mathbf{W}}_a - \tilde{\mathbf{W}}_c) \quad (19)$$

Trong đó:  $\tilde{\mathbf{W}}_c = \mathbf{W} - \hat{\mathbf{W}}_c$ ;  $\tilde{\mathbf{W}}_a = \mathbf{W} - \hat{\mathbf{W}}_a$  và  $\boldsymbol{\psi}(t) = \frac{\boldsymbol{\omega}}{\sqrt{1 + \gamma \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}}}$  là véc-tơ ước tính hồi quy của mạng Critic tiêu chuẩn hóa bị chặn bởi:

$$\|\boldsymbol{\psi}\| \leq \frac{1}{\sqrt{\gamma \varphi_1}} \quad (20)$$

Sự hội tụ của các trọng số ước lượng bởi mạng nơ-ron Actor/Critic  $\hat{\mathbf{W}}_a$  và  $\hat{\mathbf{W}}_c$  phụ thuộc vào điều kiện PE bởi  $\frac{\boldsymbol{\omega}}{\sqrt{1 + \gamma \boldsymbol{\omega}^T \boldsymbol{\Gamma} \boldsymbol{\omega}}} \in \mathbb{R}^N$  trong [13].

$$\alpha_2 \mathbf{I} \geq \int_{t_0}^{t_0 + \delta} \boldsymbol{\psi}(s) \boldsymbol{\psi}^T(s) ds \geq \alpha_1 \mathbf{I}, \forall t_0 \geq 0, \alpha_1 > 0, \alpha_2 > 0 \quad (21)$$

### 3.2. Thiết kế bộ điều khiển phản hồi RISE

Nhân hai vế của (5) với  $\bar{\mathbf{M}}$  và kết hợp với công thức (7), ta được:

$$\bar{M}\dot{r} = -(\bar{C} + \bar{D})e_2 + f + \tau_n - \tau_d + \lambda_2 \bar{M}e_2 \quad (22)$$

Thiết kế điều khiển (8) được hoàn thiện bằng cách ước lượng  $f + \tau_n$  được thiết kế dựa trên khung RISE [14].

$$\mu \approx \Delta \triangleq f + \tau_n \text{ với sai lệch } \|\mu - \Delta\| \leq \sigma_c \text{ vô cùng nhỏ} \quad (23)$$

Để tạo điều kiện cho phân tích ổn định của hệ ở phần tiếp theo, ta định nghĩa một hàm phụ  $f_d(\eta_d, \dot{\eta}_d, \ddot{\eta}_d) \in \mathbb{R}^3$  như sau:

$$f_d = \bar{M}(\eta_d)\ddot{\eta}_d + \bar{C}(\eta_d, \dot{\eta}_d, v_d)\dot{\eta}_d + \bar{D}(\eta_d, v_d)\dot{\eta}_d + \bar{g}(\eta_d) \quad (24)$$

Hàm  $\bar{f}(\eta, \dot{\eta}, \eta_d, \dot{\eta}_d, \ddot{\eta}_d) \in \mathbb{R}^3$  cho bởi:

$$\bar{f} = f - f_d \quad (25)$$

Thay (8), (24) và (25) vào vế phải (22) và đạo hàm hai vế thời gian có thể viết lại như sau:

$$\bar{M}\dot{r} = -\frac{1}{2}\dot{\bar{M}}r + \tilde{N} + N_D - e_2 - \mu \quad (26)$$

Trong công thức (26) có  $\tilde{N}(\hat{W}_a, \eta_d, \dot{\eta}_d, \ddot{\eta}_d, e_1, \dot{e}_1, e_2, \dot{e}_2, r, t), N_D(\hat{W}_a, \eta_d, \dot{\eta}_d, e_1, \dot{e}_1, t) \in \mathbb{R}^3$  là các thành phần phụ được định nghĩa như sau:

$$\tilde{N} = -(\bar{C} + \bar{D})e_2 - (\bar{C} + \bar{D})\dot{e}_2 - \frac{1}{2}\dot{\bar{M}}r + \bar{f} + \lambda_2 \dot{\bar{M}}e_2 + \lambda_2 \bar{M}\dot{e}_2 + e_2 + \frac{\partial H(\hat{W}_a, X)}{\partial X} \dot{X} - \dot{\mu} \quad (27)$$

$$N_D = \dot{f}_d + \dot{\tau}_n + \frac{\partial H(\hat{W}_a, X)}{\partial \hat{W}_a} \dot{\hat{W}}_a \quad (28)$$

$$\text{Trong công thức (27): } u = \hat{u}(X) = -\frac{1}{2}R^{-1}G^T(X)\left(\frac{\partial \phi}{\partial X}\right)^T \hat{W}_a = H(\hat{W}_a, X).$$

Theo định lý giá trị trung bình (Mean Value Theorem) và các giả thiết 1-8 có thể sử dụng cho giới hạn trên cho nhóm các thành phần phụ:

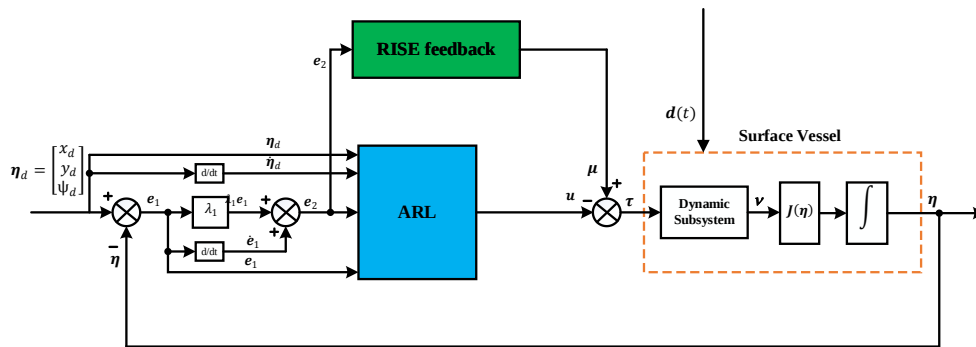
$$\|\tilde{N}\| \leq \rho(\|y\|)\|y\|; \|N_D\| \leq \xi_1, \|\dot{N}_D\| \leq \xi_2 \quad (29)$$

Trong đó,  $y = [e_1^T \ e_1^T \ r]^T \in \mathbb{R}^9$ . Hàm giới hạn  $\rho(\|y\|) \in \mathbb{R}$  là một hàm không giảm, khả nghịch toàn cục dương. Dựa trên công thức (26), thành phần điều khiển  $\mu$  được thiết kế dựa trên RISE [14] như sau:

$$\mu = \tau_n + f = (k_s + 1)e_2 - (k_s + 1)e_2(0) - \delta(t) \quad (30)$$

Với  $k_s \in \mathbb{R}^+$ ,  $\beta_1 \in \mathbb{R}^+$ ,  $\delta(t) \in \mathbb{R}^3$  được tính bởi phương trình toán học sau:

$$\delta(t) = (k_s + 1)\lambda_2 e_2 + \beta_1 \text{sgn}(e_2) \quad (31)$$



Hình 1. Cấu trúc điều khiển ARL-RISE cho mô hình tàu thủy.

### 3.3. Phát biểu định lý và chứng minh tính ổn định của hệ kín

**Định lý 1.1.** Hệ thống điều khiển kín bao gồm mô hình tàu thủy (1) và (2) thỏa mãn các giả thiết 1-8 bộ điều khiển ARL-RISE (8) với luật điều khiển RISE (30), luật điều khiển tối ưu (15), luật cập nhật trọng số (17), (19), tín hiệu véc-tơ  $\psi(t)$  thỏa mãn điều kiện PE (21), các tham số thiết kế được chọn như sau:

$$\lambda_{\min}(\lambda_1) > \frac{1}{2}; \lambda_2 > 1; \beta_1 > \xi_1 + \frac{1}{\lambda_2} \xi_2 \quad (32)$$

Đảm bảo:

1. Sai số của các trọng số các mạng nơ-ron Actor-Critic  $\tilde{W}_c$  và  $\tilde{W}_a$  sẽ UUB.
2. Sai số bám  $y \in \mathbb{R}^9$  của hệ thống tàu mặt nước cũng UUB.

#### Chứng minh Định lý 1.1:

Chọn hàm ứng viên Lyapunov cho toàn bộ hệ thống (2):

$$V_L = \frac{1}{2} \mathbf{r}^T \bar{\mathbf{M}} \mathbf{r} + P + \frac{1}{2} \mathbf{e}_1^T \mathbf{e}_1 + \frac{1}{2} \mathbf{e}_2^T \mathbf{e}_2 + V^* + V_c(\tilde{W}_c, t) + \frac{1}{2} \tilde{W}_a^T \tilde{W}_a \quad (33)$$

Với  $V^*$  là hàm chi phí tối ưu (14), hàm phụ  $P$  được định nghĩa [14]:

$$P = \beta_1 \sum_{i=1}^n |e_{2i}(0)| - \mathbf{e}_2^T(0) N_D(0) - L \quad (34)$$

Trong công thức (34),  $\dot{L} = \mathbf{r}^T (N_D - \beta_1 \text{sgn}(\mathbf{e}_2))$ . Đạo hàm  $V_L$  theo thời gian, thay thế (4), (26) và (34) vào thu được:

$$\begin{aligned} \dot{V}_L = \dot{V}_1 + \dot{V}_2 = & -\mathbf{e}_1^T \lambda_1 \mathbf{e}_1 + \mathbf{e}_2^T \mathbf{e}_1 - \lambda_2 \mathbf{e}_2^T \mathbf{e}_2 + \mathbf{r}^T \tilde{N} - (k_s + 1) \mathbf{r}^T \mathbf{r} \\ & + \frac{\partial V^*}{\partial \mathbf{X}} \mathbf{F} + \frac{\partial V^*}{\partial \mathbf{X}} \mathbf{G} \hat{\mathbf{u}} + \frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_c} \Omega_{nom} + \frac{\partial V_c}{\partial \tilde{W}_c} \Delta_{per} - \tilde{W}_a^T \dot{\tilde{W}}_a \end{aligned} \quad (35)$$

Lại có:

$$\frac{\partial V^*}{\partial \mathbf{X}} \mathbf{F} = \frac{\partial V^*}{\partial \mathbf{X}} \mathbf{G} \mathbf{u}^* - \mathbf{Q}(\mathbf{X}) - \mathbf{u}^{*T} \mathbf{R} \mathbf{u}^* \quad (36)$$

Trong đó:  $\frac{\partial V^*}{\partial \mathbf{X}} \mathbf{G} = -2\mathbf{u}^{*T} \mathbf{R}$ ;  $\mathbf{e}_2^T \mathbf{e}_1 \leq \frac{1}{2} \|\mathbf{e}_1\|^2 + \frac{1}{2} \|\mathbf{e}_2\|^2$ ;  $\mathbf{e}_1^T \lambda_1 \mathbf{e}_1 \leq \lambda_{\min}(\lambda_1) \|\mathbf{e}_1\|^2$ ;  $\lambda_2 \mathbf{e}_2^T \mathbf{e}_2 \leq \lambda_2 \|\mathbf{e}_2\|^2$ ;  $(k_s + 1) \mathbf{r}^T \mathbf{r} \leq (k_s + 1) \|\mathbf{r}\|^2$ ;  $\mathbf{r}^T \tilde{N} \leq \rho(\|\mathbf{y}\|) \|\mathbf{r}\| \|\mathbf{y}\|$ .

Sử dụng các biểu thức (19), (26), (29), (30), (31), (36) và thay thế  $\mathbf{u}^*, \hat{\mathbf{u}}, \delta_{hjb}, \Delta_{per}$  bằng cách sử dụng các công thức (14), (15), (16), (17) tương ứng, có  $\lambda_3 \triangleq \min \left\{ \left( \lambda_{\min}(\lambda_1) - \frac{1}{2} \right), \left( \lambda_2 - \frac{1}{2} \right), 1 \right\}$ .

Từ các giả thiết 6-8 có phát triển các giới hạn, chọn  $c_3 > k_{a1} k_1 k_2$ , sử dụng biến đổi  $ab \leq \gamma a^2 + \frac{1}{4\gamma} b^2$

và lựa chọn  $0 < \theta < 1$ ;  $\lambda_3 - \frac{\rho^2(\|\mathbf{y}\|)}{4k_s} > 0 \Rightarrow \|\mathbf{y}\| < \rho^{-1}(2\sqrt{\lambda_3 k_s})$  vào biểu thức (35),  $\dot{V}_L$  bị chặn như sau:

$$\begin{aligned} \dot{V}_L \leq & -\lambda_4 \|\mathbf{y}\|^2 - \mathbf{Q}(\mathbf{X}) - (1-\theta)(c_3 - k_{a1} k_1 k_2) \|\tilde{W}_c\|^2 - k_{a2} \|\tilde{W}_a\|^2 \\ & + \frac{1}{4\theta(c_3 - k_{a1} k_1 k_2)} \left[ c_4 \frac{k_c \varphi_0}{2\sqrt{\gamma \varphi_1}} k_3 + k_{a1} k_1 k_2 k_3 + k_{a1} k_1^2 k_2 + k_{a1} k_1 \right]^2 + k_{a1} k_1^2 k_2 k_3 + k_4 \end{aligned} \quad (37)$$

Trong đó:  $V_c(\tilde{W}_c, t)$  thỏa mãn các bất đẳng thức được xây dựng trong [12], biến đổi  $\lambda_3 - \frac{\rho^2(\|y\|)}{4k_s} > 0 \Rightarrow \lambda_4 = \min \left( \lambda_3 - \frac{\rho^2(\|y\|)}{4k_s} \right) > 0 \Rightarrow \lambda_4 \leq \left( \lambda_3 - \frac{\rho^2(\|y\|)}{4k_s} \right)$ . Ta định nghĩa véc – tơ  $z = [y^T \quad \tilde{W}_c^T \quad \tilde{W}_a^T]^T$  để phân tích vấn đề sai lệch bám của hệ kín. Có thể thấy rằng, tồn tại 2 hàm lớp K là  $\alpha_5$  và  $\alpha_6$  thỏa mãn:

$$\alpha_5(\|z\|) \leq \lambda_4 \|y\|^2 + Q(X) + (1-\theta)(c_3 - k_{a1}k_1k_2) \|\tilde{W}_c\|^2 + k_{a2} \|\tilde{W}_a\|^2 \leq \alpha_6(\|z\|) \quad (38)$$

Dựa trên biểu thức (38), bất đẳng thức (37) được viết lại thành:

$$\dot{V}_L \leq -\alpha_5(\|z\|) + \frac{1}{4\theta(c_3 - k_{a1}k_1k_2)} \left[ c_4 \frac{k_c \varphi_0}{2\sqrt{\gamma \varphi_1}} k_3 + k_{a1}k_1k_2k_3 + k_{a1}k_1^2k_2 + k_{a1}k_1 \right]^2 + k_{a1}k_1^2k_2k_3 + k_4 \quad (39)$$

Có thể thấy rõ ràng,  $\dot{V}_L(\cdot)$  là âm nếu  $z(t)$  nằm ngoài miền hấp dẫn:

$$\Omega_z \triangleq \left\{ z : \|z\| \leq \alpha_5^{-1} \left( \frac{1}{4\theta(c_3 - k_{a1}k_1k_2)} \left[ c_4 \frac{k_c \varphi_0}{2\sqrt{\gamma \varphi_1}} k_3 + k_{a1}k_1k_2k_3 + k_{a1}k_1^2k_2 + k_{a1}k_1 \right]^2 + k_{a1}k_1^2k_2k_3 + k_4 \right) \right\} \quad (40)$$

Từ (40), ta có  $\|z\|$  tiến từ  $z(t)$  tiến đến  $\Omega_z$  là miền hấp dẫn. Gọi  $T$  là thời gian  $z(t)$  tiến vào  $\Omega_z$ . Ta phải chứng minh  $T$  hữu hạn, khi  $t_0 \leq t \leq T$  có  $V_L$  ở công thức (33) đơn điệu giảm dần từ  $V_L(t_0)$  đến  $V_L(T)$ . Khi đó, ta có thể ký hiệu  $V_L(t_0) = c, V_L(T) = \varepsilon, 0 < \varepsilon < c$  định nghĩa các tập compact sau:  $\Omega_c = \{V_L(z) \leq c\}, \Omega_\varepsilon = \{V_L(z) \leq \varepsilon\}, A = \{\varepsilon \leq V_L(z) \leq c\}$ .

Sử dụng bất phương trình sau:

$$\dot{V}_L(z, t) \leq -W_3(z) \quad (41)$$

$W_3(z)$  là liên tục và xác định dương, ta có:

$$k = \min_{z \in A} W_3(z) > 0 \quad (42)$$

Từ (41) và (42) thu được bất đẳng thức sau:

$$\dot{V}(z, t) \leq -k, \forall z \in A, \forall t \geq t_0 \geq 0 \quad (43)$$

Tích phân hai vế (43) ta được:

$$V(z(t)) \leq V(z(t_0)) - k(t - t_0) \leq c - k(t - t_0) \quad (44)$$

Như vậy,  $z(t) \in A$  trong khoảng thời gian  $[t_0, t_0 + T]$  với  $T$  được xác định từ:

$$V(z(T)) = \varepsilon = c - k(T - t_0) \Rightarrow T = (c - \varepsilon) / k + t_0 \quad (45)$$

và hiển nhiên nó là hữu hạn.

#### 4. MÔ PHỎNG, TÍNH TOÁN, THẢO LUẬN

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  $m = 21$  (kg) chiều dài và chiều rộng tương ứng là 1.2

m và 0.3 m, với các tham số được tham khảo trong tài liệu [7]:

$$c_{13}(v) = -19v - 0.72r; c_{23}(v) = 20u; c_{31}(v) = 19v + 0.72r; c_{32}(v) = -20u$$

$$d_{11}(v) = 0.72 + 1.3|u| + 5.8u^2; d_{22}(v) = 0.86 + 36|v| + 3|r|$$

$$d_{23}(v) = -0.1 - 2|v| + 2|r|; d_{32}(v) = -0.1 - 5|v| + 3|r|; d_{33}(v) = 6 + 4|v| + 4|r|$$

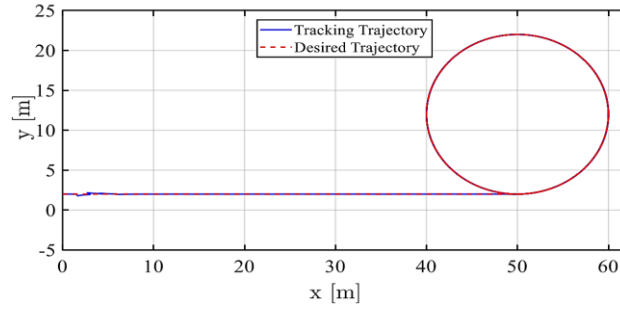
$$M = \begin{bmatrix} 20 & 0 & 0; & 0 & 19.2 & 0.72; & 0 & 0.72 & 2.7 \end{bmatrix},, g_1(\eta) = g_2(\eta) = g_3(\eta) = 0$$

Các tham số điều khiển được chọn:

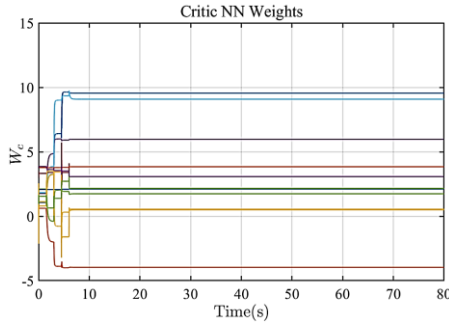
$$k_{a1} = 0.01, k_c = 2, k_{a2} = 50, \gamma = 0.01, \lambda_2 = 60, \beta_1 = 5, k_s = 100$$

Mạng nơ-ron Critic và Actor có số nơ-ron là  $N = 12$ , nhiều ngoài thỏa mãn giả thiết 4 được lấy như sau:

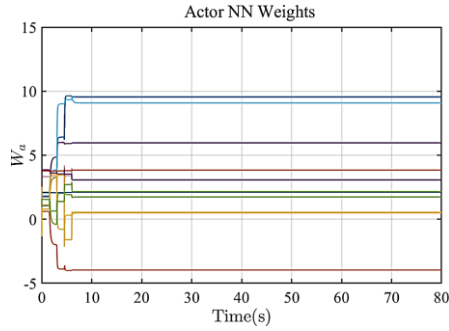
$$d(t) = \left[ 4 + 1.5\sin(0.8t) + 1.5\cos\left(0.8t + \frac{\pi}{4}\right); 4 + \sin\left(t + \frac{\pi}{6}\right) + 1.2\cos(0.6t); 4 + \sin\left(0.8t + \frac{\pi}{6}\right) \right]$$



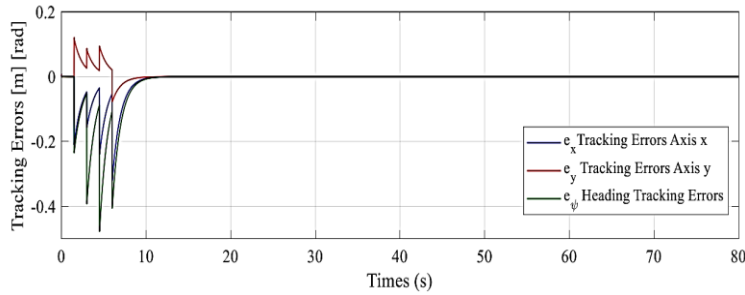
**Hình 2.** Quỹ đạo bám với mô hình tàu thủy có bất định và nhiễu ngoài sử dụng bộ điều khiển ARL-RISE.



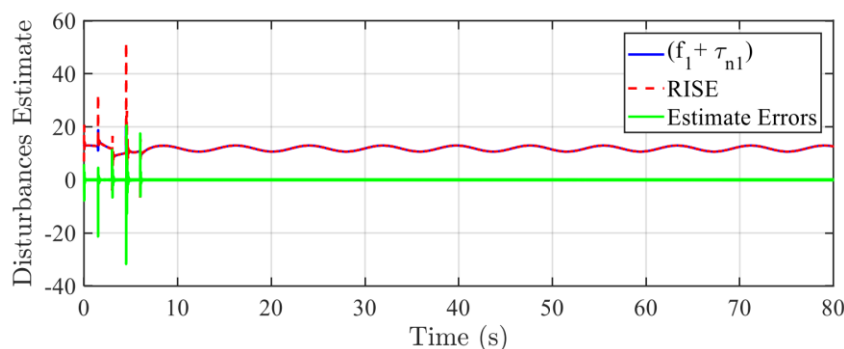
**Hình 3.** Sự hội tụ của ma trận trọng số mạng nơ-ron Critic.



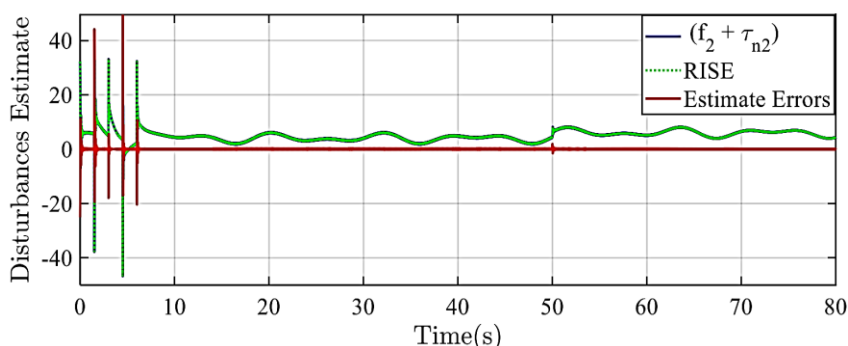
**Hình 4.** Sự hội tụ của ma trận trọng số mạng nơ-ron Actor.



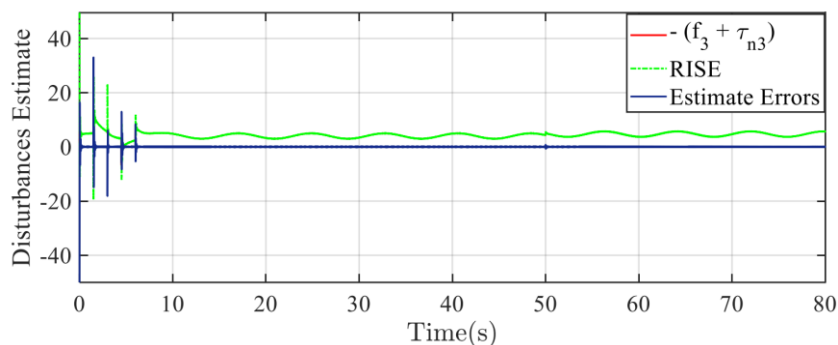
**Hình 5.** Sai lệch bám quỹ đạo theo trục  $x, y$  và theo góc  $\psi$ .



**Hình 6.** Ước lượng nhiễu ngoài và thành phần bất định  $f_1 + \tau_{n1}$  bởi bộ RISE và sai lệch ước lượng.



**Hình 7.** Ước lượng nhiễu ngoài và thành phần bất định  $f_2 + \tau_{n2}$  bởi bộ RISE và sai lệch ước lượng.



**Hình 8.** Ước lượng nhiễu ngoài và thành phần bất định  $f_3 + \tau_{n3}$  bởi bộ RISE và sai lệch ước lượng.

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, tàu bám quỹ đạo đặt. Đáp ứng quỹ đạo ban đầu của tàu còn có sai lệch bám theo  $x, y, \psi$  là  $0[m][rad] \div 0.5[m][rad]$  trong thời gian 7s thể hiện trên hình 5. 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  $x, y, \psi$  về xấp xỉ về 0 thể hiện trên các hình 5, ma trận trọng số của mạng nơ-ron hội tụ như hình 3, hình 4. Các hình 6, hình 7, hình 8 ước lượng nhiễu ngoài và phần bất định mô hình của bộ RISE có sai số nhỏ. Như vậy, bộ điều khiển bám tối ưu dựa trên thuật toán ARL – RISE hoàn toàn đáp ứng yêu cầu điều khiển bám quỹ đạo của hệ thống tàu mặt nước.

## 5. KẾT LUẬN

Bài báo đã đề xuất một bộ điều khiển bám thích nghi mới cho tàu bề mặt có mô hình bất định và chịu ảnh hưởng của nhiễu dựa trên kỹ thuật học tăng cường thích nghi động và RISE. Kỹ thuật học tăng cường thích nghi động nhằm xác định điều khiển tối ưu bằng phương pháp xấp xỉ nghiệm của phương trình HJB, RISE để bù thành phần bất định và nhiễu. Với cách thiết kế như vậy, bộ điều khiển tối ưu được đề xuất trong bài báo này có thể áp dụng cho lớp mô hình phi tuyến bất định, có nhiễu tác động. Các kết quả đạt được của bài báo đã được chứng minh bằng lý thuyết và mô phỏng số. Các kết quả này cho thấy hiệu quả của bộ điều khiển đề xuất và khả năng ứng dụng trong thực tế của bộ điều khiển này.

## TÀI LIỆU THAM KHẢO

- [1]. B. Xiao, X. Yang, and X. Huo, "A Novel Disturbance Estimation Scheme for Formation Control of Ocean Surface Vessels," IEEE Transactions on Industrial Electronics, **vol. 64**, no. 6. (2017) pp. 4994–5003,
- [2]. S. L. Dai, M. Wang, and C. Wang, "Neural Learning Control of Marine Surface Vessels with Guaranteed Transient Tracking Performance," IEEE Transactions on Industrial Electronics, **vol. 63**, no. 3. (2016) pp. 1717–1727.
- [3]. Z. Zhao, W. He, and S. S. Ge, "Adaptive neural network control of a fully actuated marine surface vessel with multiple output constraints," IEEE Transactions on Control Systems Technology, **vol. 22**, no. 4. (2014) pp. 1536–1543.
- [4]. C. S. W. He, Z. Yin, "Adaptive neural network control of a marine vessel with constraints using the asymmetric barrier Lyapunov function," IEEE Trans. Cybern., **vol. 47(7)**, (2017) pp. 1641–1651.
- [5]. Z. Yin, W. He, C. Yang, and C. Sun, "Control Design of a Marine Vessel System Using Reinforcement Learning," Neurocomputing, **vol. 311** (2018) pp. 353–362.
- [6]. Z. Yin, W. He, C. Sun, G. Li, and C. Yang, "Adaptive control of a marine vessel based on reinforcement learning," Chinese Control Conference, CCC, **vol. 2018-July**. (2018) pp. 2735–2740.
- [7]. G. Wen, S. S. Ge, C. L. P. Chen, F. Tu, and S. Wang, "Adaptive tracking control of surface vessel using optimized backstepping technique," IEEE Trans. Cybern., **vol. 49**, no. 9, (2019) pp. 3420–3431.
- [8]. R. C. Xinxin Guo, Weisheng Yan, "Integral Reinforcement Learning-Based Adaptive Systems With Unknown Control Directions," IEEE Trans. Syst. Man, Cybern. Syst., **vol. PP**, (2019) pp. 1–10.
- [9]. Z. Zheng, L. Ruan, M. Zhu, and X. Guo, "Reinforcement learning control for underactuated surface vessel with output error constraints and uncertainties," Neurocomputing, **vol. 399**. (2020) pp. 479–490.
- [10]. P. Walters, R. Kamalapurkar, F. Voight, E. M. Schwartz, and W. E. Dixon, "Online Approximate Optimal Station Keeping of a Marine Craft in the Presence of an Irrotational Current," IEEE Trans. Robot., **vol. 34**, no. 2, (2018) pp. 486–496.
- [11]. A. B. Martinsen, A. M. Lekkas, S. Gros, J. A. Glomsrud, and T. A. Pedersen, "Reinforcement Learning-Based Tracking Control of USVs in Varying Operational Conditions," Frontiers in Robotics and AI, **vol. 7** (2020).
- [12]. S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, "A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems," Automatica, **vol. 49**, no. 1 (2013) pp. 82–92.
- [13]. K. G. Vamvoudakis and F. L. Lewis, "Online actor critic algorithm to solve the continuous-time infinite horizon optimal control problem," Proceedings of the International Joint Conference on Neural Networks (2009) pp. 3180–3187.
- [14]. B. Xian, D. M. Dawson, M. S. De Queiroz, and J. Chen, "A Continuous Asymptotic Tracking Control Strategy for Uncertain Nonlinear Systems," IEEE Trans. Automat. Contr., **vol. 49**, no. 7, (2004) pp. 1206–1211.

**ABSTRACT**

**TRACKING CONTROL FOR UNCERTAIN SURFACE VESSEL  
WITH EXTERNAL DISTURBANCE VIA ON-POLICY  
ADAPTIVE REINFORCEMENT LEARNING – RISE CONTROLLER**

*This article addresses a trajectory tracking control approach for uncertain/disturbed surface vessels using the new structure of adaptive reinforcement learning (ARL) algorithm and Robust Integral of the Sign of the Error (RISE). To obtain an asymptotic tracking performance, a robust integral of the sign of the error (RISE) feedback term is introduced in feedback control design to compensate for the uncertain components in the model and external disturbances. The adaptive reinforcement learning technique is proposed by tuning simultaneously the actor-critic network to approximate the control policy and the cost function, respectively. The convergence of weight as well as tracking control problem was determined by theoretical analysis. Finally, the numerical example is investigated to validate the effectiveness of the proposed control scheme.*

**Keywords:** Optimal Control; Adaptive Dynamic Programming (ADP); Surface Vessel (SV) systems; Robust Integral of the Sign of the Error (RISE); Trajectory Tracking Control.

*Nhận bài ngày 23 tháng 01 năm 2021*

*Hoàn thiện ngày 14 tháng 6 năm 2021*

*Chấp nhận đăng ngày 29 tháng 7 năm 2021*

*Địa chỉ:* <sup>1</sup>Bộ môn Điều khiển tự động, Trường Đại học Bách Khoa Hà Nội;

<sup>2</sup>Khoa Điện Cơ, Trường Đại học Hải Phòng;

\*Email : tuvv@dhhp.edu.vn.