THIẾT KẾ BỘ ĐIỀU KHIỂN PID TỰ THÍCH NGHI TRỰC TUYẾN DỰA TRÊN THUẬT TOÁN HỌC SÂU TĂNG CƯỜNG TD3 CHO HỆ THỐNG LÁI TỰ ĐỘNG TÀU THỦY

ONLINE ADAPTIVE PID CONTROLLER DESIGN BASED ON TD3 DEEP REINFORCEMENT LEARNING FOR SHIP AUTOPILOT SYSTEMS

NGUYỄN HỮU QUYỀN*, NGUYỄN HÙNG CƯỜNG
Khoa Điện - Điện tử, Trường Đại học Hàng hải Việt Nam
**Email liên hệ: quyennh.ddt@vimaru.edu.vn*
DOI: https://doi.org/10/65154/jmst.1026

Tóm tắt

Bài báo đề xuất phương pháp điều khiển lái tự động thích nghi cho tàu thủy dựa trên sự kết hợp giữa bộ điều khiển PID truyền thống và thuật toán học sâu tăng cường (TD3: Twin Delayed Deep Deterministic Policy Gradient). Hệ thống được xây dựng trên mô hình động lực học phi tuyến 3 bậc tự do (surge-sway-yaw), mô hình phản ánh đầy đủ tính phi tuyến và sự tương tác giữa các thành phần chuyển động của tàu. Trong cấu trúc đề xuất, bộ PID thực hiện điều khiển thời gian thực, trong khi tác nhân TD3 tối ưu hóa trực tuyến các tham số với tốc độ cập nhật chậm, phù hợp với đặc tính quán tính lớn của tàu thủy, giúp duy trì tính ổn định của hệ thống trong suốt quá trình học.

Hàm thưởng (Reward function) được thiết kế đa mục tiêu, bao gồm sai lệch hướng đi, tốc độ tốc độ quay ngang và năng lượng điều khiển (góc lái) nhằm cân bằng giữa độ chính xác bám hướng và tính kinh tế vận hành. Kết quả mô phỏng dưới tác động của nhiễu môi trường cho thấy phương pháp TD3 - PID cải thiện đáng kể so với bộ điều khiển PID truyền thống thông qua việc giảm thiểu hiện tượng quá điều chỉnh, sai số xác lập và biên độ dao động của góc lái. Nghiên cứu khẳng định khả năng ứng dụng cao của học sâu tăng cường trong điều khiển các hệ thống tàu thủy phi tuyến có quán tính lớn.

Từ khóa: *Điều khiển lái tự động tàu thủy, mô hình phi tuyến 3-DOF, học sâu tăng cường (TD3), PID thích nghi.*

Abstract

This paper proposes an adaptive autopilot control method for ships based on the combination of a traditional PID controller and the Twin Delayed Deep Deterministic Policy Gradient (TD3) deep reinforcement learning algorithm. The system is built on a 3-degree-of-freedom (surge-sway-yaw) nonlinear dynamic model, which fully reflects the nonlinearity and interaction between the ship's motion components. In the proposed structure, the PID controller performs real-time control, while the TD3 agent optimizes the parameters online with a slow update rate, suitable for the high-inertia characteristics of ships, helping to maintain system stability throughout the learning process.

The reward function is designed with multiple objectives, including heading error, yaw rate, and control energy (rudder angle) to balance heading tracking accuracy and operational economy. Simulation results under the influence of environmental noise show that the TD3-PID method significantly improves compared to the traditional PID controller by minimizing overshoot, steady-state error, and the oscillation amplitude of the rudder angle. The study confirms the high applicability of deep reinforcement learning in controlling nonlinear ship systems with large inertia.

Keywords: *Ship autopilot control, 3-DOF nonlinear model, deep Reinforcement Learning, TD3, Adaptive PID.*

1. Mở đầu

Điều khiển lái tự động là một trong những chức năng quan trọng trong điều khiển chuyển động tàu thủy, góp phần nâng cao độ an toàn hành hải, giảm tiêu hao nhiên liệu và tăng mức độ tự động hóa [21,28]. Tuy nhiên, động lực học tàu thủy có đặc tính phi tuyến

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

mạnh, quán tính lớn và chịu ảnh hưởng đáng kể của nhiễu môi trường như gió, sóng và dòng chảy [1, 2]. Những yếu tố này làm cho bài toán điều khiển giữ hướng tàu thủy trở nên phức tạp, đặc biệt trong điều kiện tham số bất định và thay đổi theo trạng thái vận hành [3, 4].

Trong thực tiễn, bộ điều khiển PID vẫn được sử dụng rộng rãi trong điều khiển lái tự động tàu thủy nhờ cấu trúc đơn giản và khả năng triển khai dễ dàng [24, 29]. Nhiều nghiên cứu đã tập trung vào các phương pháp hiệu chỉnh tham số PID như Ziegler-Nichols, tối ưu hóa dựa trên thuật toán tiến hóa hoặc kỹ thuật tối ưu thông minh nhằm cải thiện chất lượng điều khiển [11, 24]. Tuy nhiên, các phương pháp này thường mang tính ngoại tuyến (Offline) và khó thích nghi khi đặc tính động lực học thay đổi trong quá trình khai thác.

Bên cạnh đó, các chiến lược điều khiển bền vững và điều khiển trượt (Sliding Mode Control) đã được đề xuất nhằm nâng cao khả năng chống nhiễu và bất định mô hình [17, 26]. Mặc dù đạt được hiệu quả nhất định, song các phương pháp này thường yêu cầu cấu trúc điều khiển phức tạp hoặc phụ thuộc vào giả thiết về chặn trên của nhiễu và bất định, điều này hạn chế tính linh hoạt khi áp dụng thực tế [1, 3].

Gần đây, học sâu tăng cường (Deep Reinforcement Learning - DRL) được xem là một hướng tiếp cận mới cho các hệ thống có phi tuyến và khó mô hình hóa [13, 25, 30]. Các thuật toán như DDPG (Deep Deterministic Policy Gradient) và các thuật toán phát triển dựa trên nền tảng này đã được áp dụng trong một số bài toán điều khiển trong miền liên tục [19,30]. Trong đó, thuật toán TD3 được đánh giá cao nhờ việc tích hợp cấu trúc Critic kép và cơ chế cập nhật trễ mạng Actor, giúp kiểm soát hiệu quả sai lệch ước lượng giá trị hàm thưởng (Reward function) và duy trì tính hội tụ ổn định cho hệ thống lái tàu thủy vốn có đặc tính quán tính lớn [23]. Tuy nhiên, phần lớn các nghiên cứu hiện nay áp dụng DRL theo hướng điều khiển trực tiếp cơ cấu chấp hành [9, 15, 30], điều này có thể dẫn đến tín hiệu điều khiển dao động và khó đảm bảo tính tin cậy trong hệ thống điều khiển tàu thủy.

Từ những phân tích trên có thể thấy rằng, mặc dù bộ điều khiển PID có tính thực tiễn cao và DRL có khả năng thích nghi mạnh, sự kết hợp có cấu trúc giữa hai phương pháp này cho bài toán điều khiển hướng tàu thủy trên mô hình phi tuyến đầy đủ vẫn chưa được nghiên cứu một cách thỏa đáng [6, 9, 27, 31]. Đặc biệt, việc thiết kế một cơ chế cập nhật tham số PID thích nghi trực tuyến theo thời gian thực với chu kỳ cập nhật phù hợp với đặc tính quán tính lớn của tàu thủy, vẫn còn là một khoảng trống nghiên cứu [1, 3, 6, 8].

Nhằm giải quyết vấn đề này, bài báo đề xuất một cấu trúc điều khiển lai hai tầng TD3 - PID, trong đó bộ PID thực hiện điều khiển thời gian thực, còn thuật toán Học sâu tăng cường TD3 thực hiện tối ưu hóa trực tuyến K_p, K_i, K_d dựa trên các tín hiệu trạng thái (tốc độ bẻ lái, tín hiệu điều khiển góc lái) và sai lệch điều khiển hướng. Mô hình động lực học phi tuyến 3 bậc tự do (surge - sway - yaw) được sử dụng nhằm phản ánh trung thực tính chất phi tuyến và sự tương tác động lực học phức tạp của tàu thủy [1, 4, 17].

Các đóng góp chính của nghiên cứu bao gồm:

(i) Đề xuất cấu trúc điều khiển phân cấp TD3 - PID thích nghi, cho phép cập nhật tham số trực tuyến với cơ chế ổn định phù hợp cho các đối tượng có quán tính lớn như tàu thủy.

(ii) Thiết kế hàm thưởng đa mục tiêu nhằm tối ưu hóa sự cân bằng (trade-off) giữa độ chính xác bám hướng, tốc độ quay trở và tính kinh tế trong vận hành (giảm năng lượng điều khiển) [23, 25].

(iii) Đánh giá hiệu quả của phương pháp thông qua thực nghiệm mô phỏng và phân tích định lượng trên mô hình phi tuyến 3-DOF, chứng minh khả năng kháng nhiễu và tính bền vững trước các bất định tham số.

2. Mô hình động lực học tàu thủy 3-DOF

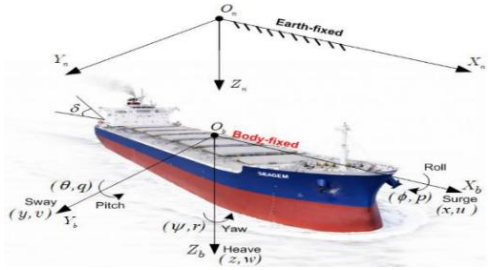
Trong phạm vi bài báo này, tập trung nghiên cứu và mô hình hóa chuyển động của tàu trên mặt phẳng ngang với cấu trúc 3 bậc tự do (3-DOF) [1, 2, 4]. Việc lựa chọn mô hình 3-DOF thay vì các mô hình tuyến tính hóa đơn giản (như Nomoto bậc 1, 2) cho phép hệ thống mô tả đầy đủ sự tương tác liên kết giữa các bậc tự do và các đặc tính động lực học phi tuyến. Đây là yếu tố quan trọng để tác tử học tăng cường (RL Agent) có thể quan sát, học đầy đủ trạng thái hệ thống, từ đó tối ưu hóa các tham số điều khiển thích nghi nhằm đảm bảo tính ổn định và chính xác ngay cả trong các điều kiện môi trường hoạt động khắc nghiệt [1, 2, 13].

2.1. Hệ tọa độ và các biến trạng thái

Chuyển động của tàu thủy được mô tả thông qua hai hệ tọa độ tiêu chuẩn theo đề xuất của Fossen [1, 2], được mô tả như Hình 1:

Các hệ tọa độ mô tả các bậc tự do của tàu thủy bao gồm 2 hệ tọa độ [1, 2]:

Hệ tọa độ cố định Trái Đất (Earth-fixed frame - NED): Ký hiệu O_n-$X_nY_nZ_n$, được sử dụng để xác định vị trí *(x, y)* và góc hướng tàu (ψ).



Hình 1. Hệ tọa độ mô tả các bậc tự do của tàu thủy

Hệ tọa độ gắn liền thân tàu (Body-fixed frame): Ký hiệu là O_b-$X_bY_bZ_b$ có gốc đặt tại trọng tâm tàu (hoặc tâm quay), dùng để xác định các vận tốc thành phần.

Vector trạng thái của tàu được định nghĩa bởi:

$$\eta = [x, y, \psi]^T \in \mathbb{R}^3, \ \nu = [u, v, r]^T \in \mathbb{R}^3$$

Trong đó:

η : Vector trạng thái tọa độ;

ν : Vector vận tốc;

x, y : Tọa độ vị trí tàu trong hệ tọa độ NED;

ψ : Góc hướng của tàu (Heading angle);

u, v : Vận tốc chuyển động tiến theo chiều dọc (surge) và vận tốc dạt ngang (sway);

r : Vận tốc góc quay trở (yaw rate).

2.2. Phương trình động học (Kinematics)

Mối quan hệ hình học giữa vận tốc trong hệ quy chiếu thân tàu và tốc độ biến thiên vị trí trong hệ NED được biểu diễn qua ma trận chuyển đổi $R(\psi)$ [1, 2].

$$\dot{\eta} = R(\psi)\nu \tag{1}$$

Với ma trận xoay $R(\psi)$ có dạng:

$$R(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Ma trận này có đặc tính trực giao, thỏa mãn [1]:

$$R^{-1}(\psi) = R^T(\psi) \quad \text{và} \quad \| R(\psi) \| = 1 \tag{3}$$

2.3. Phương trình động lực học phi tuyến

Dựa trên nguyên lý bảo toàn động lượng, phương trình động lực học của tàu thủy trong môi trường nước được mô tả dưới dạng vector phi tuyến [1, 2, 4]:

$$M\dot{\nu} + C(\nu)\nu + D(\nu)\nu + g(\eta) = \tau + \tau_{env} \tag{4}$$

Trong đó, các thành phần ma trận được chi tiết hóa như sau:

Ma trận khối lượng và quán tính M: Bao gồm khối lượng của tàu (m) và khối lượng nước kèm (M_A - added mass) [1, 2, 4]:

$$M = M_{RB} + M_A = \begin{bmatrix} m - X_{\dot{u}} & 0 & 0 \\ 0 & m - Y_{\dot{v}} & mx_G - Y_{\dot{r}} \\ 0 & mx_G - N_{\dot{v}} & I_z - N_{\dot{r}} \end{bmatrix} \tag{5}$$

Với m - là khối lượng của tàu, x_G - là tọa độ trọng tâm tàu, I_z - là mô-men quán tính của tàu.

Các ký hiệu $X_{\dot{u}}, Y_{\dot{v}}, N_{\dot{r}} \ldots$ là các dẫn xuất thủy động lực học thể hiện lực quán tính của dòng nước tác động ngược lại khi tàu tăng tốc.

Ma trận Coriolis và lực ly tâm C(ν) Ma trận này mô tả sự tương tác giữa các thành phần vận tốc khi tàu chuyển động trong hệ quy chiếu không quán tính [1, 2, 4]:

$$C(\nu) = \begin{bmatrix} 0 & 0 & -c_{13}(\nu) \\ 0 & 0 & c_{23}(\nu) \\ c_{13}(\nu) & -c_{23}(\nu) & 0 \end{bmatrix} \tag{6}$$

Với:

$$c_{13}(\nu) = (m - Y_{\dot{v}})v + (mx_g - Y_{\dot{r}})r, c_{23}(\nu) = (m - X_{\dot{u}})u$$

Ma trận cản thủy động lực học $D(\nu)$: Lực cản tác động lên vỏ tàu là một thành phần phi tuyến mạnh, thường được mô hình hóa bởi sự kết hợp giữa cản tuyến tính và cản bậc hai [1, 2, 4]:

$$D(\nu) = D_L + D_{NL}(\nu) \tag{7}$$

Trong đó: D_L là ma trận cản tuyến tính (Linear Damping), $D_{NL}(\nu)$ Ma trận cản phi tuyến (Nonlinear Damping).

Thành phần lực phục hồi $g(\eta)$: Phát sinh từ sự cân bằng giữa trọng lực và lực nổi của tàu. Khi xét trong phạm vi chuyển động 3 bậc tự do (3-DOF) trên mặt phẳng ngang, các lực phục hồi này thường không xuất hiện hoặc được xem là không đáng kể. Vì vậy, trong bài báo này giả định $g(\eta) = 0$, [4].

Thành phần lực và mô-men của cơ cấu chấp hành (chân vịt và bánh lái) tác động lên tàu (τ)

$$\tau = \begin{bmatrix} \tau_u \\ \tau_v \\ \tau_r \end{bmatrix} = \begin{bmatrix} F_x \\ F_y \\ M_z \end{bmatrix} \tag{8}$$

Trong đó: F_x, F_y, M_z là các thành phần lực và mô-men của cơ cấu chấp hành tác động lên thân tàu.

Nhiễu môi trường (τ_{env}): Nhiễu động từ sóng, gió và dòng chảy được mô hình hóa dưới dạng các lực và mô-men tác động vào phương trình động học. Đặc biệt, nhiễu sóng biển thường được mô phỏng qua phổ sóng (như JONSWAP hoặc Pierson-Moskowitz) để đánh giá khả năng chống nhiễu của bộ điều khiển TD3-PID trong điều kiện biển động

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

thực tế. Nhiễu môi trường được mô hình hóa dưới dạng lực/mô-men tổng quát [2, 26]:

$$\tau_{env} = \begin{bmatrix} X_{env}(t) \\ Y_{env}(t) \\ N_{env}(t) \end{bmatrix} \quad (9)$$

Trong đó, $X_{env}(t), Y_{env}(t), N_{env}(t)$: Là các thành phần lực và mô-men của nhiễu tác động lên thân tàu.

Mô hình toán mô tả chuyển động tàu thủy ba bậc tự do thiếu cơ cấu chấp hành:

Trong (8) nếu tín hiệu lực điều khiển có đầy đủ các thành phần $\tau = [\tau_u \ \tau_v \ \tau_r]^T$, thì mô hình toán xét trên mặt phẳng ngang được gọi là mô hình tàu đủ cơ cấu chấp hành (*Full Actuated*). Nếu $\tau = [\tau_u \ 0 \ \tau_r]^T$, nghĩa là mô hình tàu không có thành phần lực gây ra trượt ngang hướng theo trục *y*, thì mô hình toán xét trên mặt phẳng ngang được gọi là mô hình tàu thiếu cơ cấu chấp hành (Underactuated) [4, 13, 17]. Đây là loại tàu mà chỉ có 2 cơ cấu thực hiện là chân vịt chính và bánh lái chính sau lái. Mô hình toán này thường gặp phổ biến là các tàu chở hàng, tàu Container,… Theo giả thiết trong tài liệu [4] tàu thủy thường có khối lượng đồng đều, đối xứng qua mặt phẳng mạn tàu, gốc tọa độ gắn với thân tàu, trùng với trọng tâm tàu và tàu thiếu cơ cấu chấp hành thường có kết cấu các mặt phẳng đối xứng nhau. Do đó các biểu thức (5), (6), (7) được rút gọn và có dạng [4]:

$$M = \begin{bmatrix} m_{11} & 0 & 0 \\ 0 & m_{22} & 0 \\ 0 & 0 & m_{33} \end{bmatrix}, C(\nu) = \begin{bmatrix} 0 & 0 & -m_{22}v \\ 0 & 0 & m_{11}u \\ m_{22}v & -m_{11}u & 0 \end{bmatrix}$$

Trong đó: $m_{11} = m - X_{\dot{u}}$, $m_{22} = m - Y_{\dot{v}}$, $m_{33} = I_z - N_{\dot{r}}$

$$D(\nu) = \begin{bmatrix} d_{11} + d_{u2} + d_{u3} & 0 & 0 \\ 0 & d_{22} + d_{v2} + d_{v3} & 0 \\ 0 & 0 & d_{33} + d_{r2} + d_{r3} \end{bmatrix}$$

Mô hình tàu thiếu cơ cấu chấp hành cùng với các tham số của ma trận $M, C(\nu), D(\nu)$ trên sẽ được sử dụng để kiểm nghiệm bộ điều khiển, và được cho như Bảng 1.

2.4. Mô hình cơ cấu chấp hành (Mô hình hệ thống máy lái (Steering Dynamics)

Tín hiệu điều khiển từ bộ PID là góc lái yêu cầu δ_c. Do giới hạn cơ cấu chấp hành, động học bánh lái được mô hình hóa bởi hệ bậc nhất như phương trình (10).

Tín hiệu đầu ra từ bộ điều khiển Adaptive TD3-PID là góc lái yêu cầu (góc lái lệnh) δ_c. Tuy nhiên, do các giới hạn vật lý của cơ cấu chấp hành, phản ứng của bánh lái thực tế không thể đáp ứng tức thời. Động học của hệ thống máy lái được mô hình hóa dưới dạng một hệ quán tính bậc nhất có xét đến các đặc tính phi tuyến bão hòa, cụ thể theo phương trình (10) [3,29]:

$$T_{\delta_{act}} \dot{\delta}_{act} + \delta_{act} = \delta_c \quad (10)$$

Trong đó:

δ_{act}: Là góc bẻ lái thực tế (actual rudder angle),

δ_c: Là tín hiệu góc lái từ bộ điều khiển (commanded rudder angle)

$T_{\delta_{act}}$: Là hằng số thời gian của cơ cấu lái

Trong hệ thống lái, góc lái thực tế luôn chịu các yếu tố ràng buộc vật lý như sau:

$|\delta_{act}| \le \delta_{\max}$: Ràng buộc độ lớn góc bẻ lái, góc bẻ lái thực tế không vượt quá giới hạn $\delta_{\max}$;

$|\dot{\delta}_{act}| \le \dot{\delta}_{\max}$: Ràng buộc độ lớn tốc độ bẻ lái.

Trong đó:

$\delta_{\max}$: Là góc lái thực tế lớn nhất (đối với hệ thống lái tàu thủy $\delta_{\max} = 35^0$),

$\dot{\delta}_{\max}$: Là ràng buộc liên quan đến tốc độ bẻ lái, (đối với hệ thống lái tàu thủy $\dot{\delta}_{\max} \approx 2 - 2{,}5^0/s$),

Góc quay của bánh lái tạo ra lực và mô-men tác động lên thân tàu được xấp xỉ tuyến tính như (11), [3, 4]:

$$\tau = \begin{bmatrix} 0 \\ Y_\delta \delta_{act} \\ N_\delta \delta_{act} \end{bmatrix} \quad (11)$$

Do nội dung nghiên cứu của bài báo chỉ tập trung vào điều khiển hướng, do đó thành phần lực dọc thường được giả thiết cân bằng bởi hệ thống điều khiển tốc độ tàu.

2.5. Bài toán điều khiển hướng trong lái tự động tàu thủy

Mục tiêu của bài toán điều khiển hướng tàu thủy là thiết kế luật điều khiển góc lái δ_c nhằm đảm bảo góc hướng thực tế ψ bám sát giá trị đặt ψ_{ref} một cách chính xác và ổn định [21,28,29]. Trong bối cảnh mô hình động lực học 3-DOF, bộ điều khiển không chỉ phải triệt tiêu sai số bám $e_\psi = \psi - \psi_{ref}$ mà còn phải xử lý được sự tương tác phi tuyến phức tạp giữa các thành phần vận tốc tiến (u), dạt ngang (v) và quay trở (r) [1, 4]. Hệ thống điều khiển cần được chứng minh tính ổn định tiệm cận hoặc ổn định bám trong điều kiện tồn tại các nhiễu động bất định từ môi trường như sóng, gió và dòng chảy [26]. Bên cạnh đó, bài toán đặt ra yêu cầu khắt khe về tính kinh tế và kỹ thuật, cụ thể là việc tối ưu hóa năng lượng tiêu thụ thông qua việc giảm thiểu biên độ góc lái và tần

suất hoạt động của máy lái nhằm bảo vệ cơ cấu chấp hành [29]. Tác tử TD3 sẽ đóng vai trò là cơ chế thích nghi cấp cao, liên tục tinh chỉnh các tham số PID để đáp ứng các tiêu chí đa mục tiêu này ngay cả khi đặc tính động lực học của tàu thay đổi theo điều kiện tải trọng và môi trường [5, 6, 8, 31].

3. Thiết kế bộ điều khiển TD3-PID thích nghi

3.1. Giới thiệu cấu trúc điều khiển đề xuất

Nghiên cứu này đề xuất một kiến trúc điều khiển phân cấp cho hệ thống lái tự động tàu thủy, tích hợp bộ điều khiển PID thích nghi dựa trên thuật toán học tăng cường TD3. Cấu trúc đề xuất nhằm phát huy ưu thế kết hợp giữa tính ổn định của điều khiển phản hồi PID truyền thống và khả năng thích nghi linh hoạt của trí tuệ nhân tạo (học sâu tăng cường) trong môi trường có chứa yếu tố bất định.

Xét bài toán điều khiển giữ hướng tàu thủy chịu tác động của nhiễu môi trường và bất định tham số. Mục tiêu của điều khiển là đảm bảo hướng $\psi(t)$ bám theo giá trị hướng đặt ψ_{ref} với sai lệch: $e(t)=\psi_{ref}(t)-\psi(t)\to 0$

Đặc thù của hệ tàu thủy là tính quán tính lớn và tham số biến đổi chậm, do đó nghiên cứu này đề xuất cấu trúc điều khiển phân cấp theo đặc tính thời gian bao gồm:

Vòng trong (fast loop): Bộ điều khiển PID tạo tín hiệu điều khiển góc lái $\delta_c(t)$ theo sai lệch hướng $e(t)$ với chu kỳ lấy mẫu nhanh T_f.

Vòng ngoài (slow loop): Tác nhân học tăng cường TD3 điều chỉnh trực tuyến các tham số PID K_p, K_i, K_d với chu kỳ lấy mẫu chậm T_s, trong đó $T_s \gg T_f$.

Cấu trúc này giúp duy trì độ tin cậy của điều khiển PID truyền thống, đồng thời cải thiện hiệu năng nhờ khả năng thích nghi của TD3 trong điều kiện môi trường thay đổi. Sơ đồ cấu trúc tổng thể của hệ thống được minh họa chi tiết như Hình 2.

Cấu trúc hệ thống đề xuất bao gồm các thành phần chức năng chính như sau:

Khối thuật toán học tăng cường TD3 (TD3 Learning Algorithm): Đây là trung tâm điều hành và ra quyết định của hệ thống, đóng vai trò là tầng tối ưu hóa cấp cao. Thay vì trực tiếp điều khiển góc lái, khối này thực hiện nhiệm vụ chỉnh định tham số của bộ điều khiển PID.

Tín hiệu đầu vào (State): Khối TD3 tiếp nhận vector trạng thái bao gồm: Sai số hướng đi của tàu $e(t)=\psi_{ref}-\psi$, tốc độ quay trở r, và tín hiệu phản hồi góc lái δ_c.

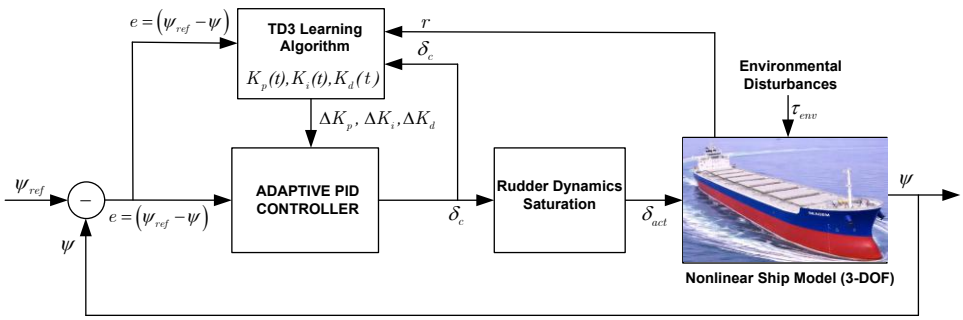
Chức năng: Dựa trên hàm thưởng (Reward Function) đã được thiết lập, thuật toán TD3 tính toán để tìm ra các giá trị hiệu chỉnh tối ưu.

Tín hiệu đầu (Action): Đầu ra của khối là các giá trị hiệu chỉnh tham số $\Delta K_p, \Delta K_i, \Delta K_d$, [5, 6, 15].

Bộ điều khiển PID thích nghi (Adaptive PID Controller): Khối này đóng vai trò là tầng thực thi trực tiếp, đảm bảo tính ổn định và thời gian thực cho hệ thống [24, 29].

Cơ chế hoạt động: Bộ điều khiển này có chức năng tính toán tạo ra lệnh điều khiển góc lái (δ_c), dựa trên sai lệch giữa hướng đi đặt và hướng đi thực của tàu $e(t)=\psi_{ref}-\psi$. Khác với các bộ PID truyền thống các tham số K_p, K_i, K_d cố định, không thay đổi trong suốt quá trình điều khiển. Bộ PID thích nghi trong cấu trúc này có các hệ số K_p, K_i, K_d được cập nhật liên tục thông qua các gia số $\Delta K_p, \Delta K_i, \Delta K_d$ do tác tử TD3 cung cấp.

Mô hình động lực học tàu thủy (Ship Model): Khối mô hình tàu 3 bậc tự do, đối với các bài toán nghiên cứu về điều khiển hướng tàu thủy thì thường chỉ sử dụng mô hình đơn giản hóa như Nomoto bậc 1, 2 hay mô hình phi tuyến đơn giản của Norbin, trong nghiên cứu này sử dụng mô hình 3-DOF, mô hình này thể hiện đầy đủ các đặc tính động học và các mối tương quan giữa các bậc tự do trong chuyển động của



Hình 2. Sơ đồ cấu trúc điều khiển hệ thống lái tự động tàu thủy dựa trên thuật toán TD3-PID thích nghi

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

tàu thủy. Mô hình đối tượng điều khiển 3-DOF chịu tác động trực tiếp từ lệnh điều khiển và chịu tác động của nhiễu môi trường.

Mô hình khối cơ cấu chấp hành (Rudder Dynamics, Saturation): Đây là khối mô tả các động lực học của máy lái và các giới hạn vật lý thực tế của máy lái (giới hạn độ lớn góc bẻ lái và tốc độ bẻ lái) trước khi tạo ra tín hiệu góc bẻ lái thực (δ_{act}).

3.2. Bộ điều khiển PID

3.2.1. Bộ điều khiển PID kinh điển

Bộ điều khiển PID được xem là giải pháp kinh điển và vạn năng trong lĩnh vực công nghiệp nói chung cũng như điều khiển tàu thủy nói riêng, nhờ vào cấu trúc toán học tường minh cùng khả năng vận hành ổn định trong điều kiện chịu tác động của nhiễu khắc nghiệt. Phương trình toán học mô tả bộ điều khiển PID kinh điển được mô tả như phương trình (12).

$$\delta_c(t) = K_p e(t) + K_i \int_0^t e(t)dt + K_d \frac{de(t)}{dt} \quad (12)$$

Trong đó: $e = \psi_{ref} - \psi$ là sai lệch hướng giữa hướng đi đặt ψ_{ref} và hướng đi thực ψ của tàu.

K_p, K_i, K_d: Là các hệ số điều chỉnh tỷ lệ, tích phân, vi phân của bộ điều khiển PID.

Trong triển khai số, với chu kỳ T_f:

$$I_k = I_{k-1} + T_f e_k, \dot{e}_k = \frac{e_k - e_{k-1}}{T_f},$$

$$\delta_k = sat\left(K_p e_k + K_i I_k + K_d \dot{e}_k\right),$$

Trong đó: I_k là giá trị tích phân của sai số tại thời điểm k, $sat(.)$ là hàm bão hòa theo ràng buộc cơ cấu chấp hành:

$\delta_k \le \delta_{\max}$ ràng buộc về độ lớn góc bẻ lái;

$\left|\dot{\delta}_k\right| \le \left|\dot{\delta}_{\max}\right|$ ràng buộc về tốc độ góc bẻ lái.

3.2.2. Bộ điều khiển PID thích nghi

Khác với các bộ PID truyền thống (12) các tham số K_p, K_i, K_d cố định, không thay đổi trong suốt quá trình điều khiển. Bộ PID thích nghi có các hệ số K_p, K_i, K_d thay đổi liên tục theo thời gian, được cập nhật liên tục thông qua các gia số $\Delta K_p, \Delta K_i, \Delta K_d$ do tác tử TD3 cung cấp. Điều này cho phép luật điều khiển thay đổi linh hoạt để thích ứng với trạng thái động lực học của tàu.

Lệnh điều khiển góc lái $\delta(t)$ được tính toán dựa trên sai lệch hướng tàu $e_\psi(t) = \psi_{ref} - \psi(t)$ theo phương trình (13):

$$\delta(t) = K_p(t) e_\psi(t) + K_i(t) \int_0^t e_\psi(\mathrm{t})dt + K_d(t) \frac{de_\psi(t)}{dt} \quad (13)$$

Các tham số K_p, K_i, K_d được cập nhật liên tục theo thời gian, nhưng chỉ tại các thời điểm rời rạc $t_j = jT_s$

Trong thực tế triển khai trên máy tính, phương trình được rời rạc hóa để phù hợp với xử lý số. Tại bước thời gian k tín hiệu điều khiển được xác định theo (14):

$$\delta_k = K_{p,k} e_k + K_{i,k} \sum_{j=0}^{k} e_j \Delta t + K_{d,k} \frac{e_k - e_{k-1}}{\Delta t} \quad (14)$$

Trong đó, các tham số $K_{p,k}, K_{i,k}, K_{d,k}$ được cập nhật động qua mỗi chu kỳ của vòng lặp tối ưu hóa:

$$K_{p,k} = K_{p,k-1} + \Delta K_{p,k}$$
$$K_{i,k} = K_{i,k-1} + \Delta K_{i,k}$$
$$K_{d,k} = K_{d,k-1} + \Delta K_{d,k}$$

Các tham số này luôn được giới hạn trong miền an toàn (compact set):

$$K_j \in \Omega = \{K : K_p \in \left[K_p^{\min}, K_p^{\max}\right], \left[K_i^{\min}, K_i^{\max}\right], \left[K_d^{\min}, K_d^{\max}\right]\}.$$

Điều kiện $K_j \in \Omega$ đóng vai trò như một ràng buộc quan trọng nhằm bảo đảm tính bị chặn của tín hiệu trạng thái và đầu ra của hệ kín.

3.3. Mô hình học tăng cường dựa trên Markov (Markov Decision Process: MDP)

Để tối ưu hóa các tham số PID bằng thuật toán TD3, bài toán điều khiển hướng tàu được mô hình hóa thông qua khung lý thuyết MDP, tạo ra mối liên kết phản hồi giữa tác nhân học tăng cường và môi trường hàng hải bất định [11, 25, 27].

Để cụ thể hóa mô hình MDP này trong bài toán điều khiển giữ hướng tàu thủy, các thành phần chính của cơ chế thích nghi tham số được thiết lập dựa trên đặc tính động lực học của hệ thống, bao gồm: Cấu trúc không gian trạng thái, xác định không gian hành động tương ứng với việc hiệu chỉnh tham số PID, và xây dựng hàm thưởng nhằm định hướng mục tiêu tối ưu hóa cho hệ thống.

3.3.1. Không gian trạng thái (State Space)

Vectơ trạng thái s_t được thiết kế nhằm phản ánh đầy đủ các đặc trưng động lực học của tàu thủy tại thời điểm *t* làm cơ sở cho việc xác lập các lệnh điều chỉnh thích nghi của tác nhân. Cấu trúc của vectơ này được xác định bởi (15):

$$s_t = [e(t), r(t), \delta(t)]^T \in \mathbb{R}^3 \quad (15)$$

Việc thiết kế vectơ trạng thái trạng thái theo cách này đảm bảo cung cấp đầy đủ các thành phần dữ liệu thiết yếu giúp tác nhân nhận diện chính xác các đặc tính vận hành, bao gồm:

Sai lệch bám hướng $e(t)$: Đóng vai trò là thông số phản hồi cốt lõi, phản ánh trực tiếp mức độ hoàn

thành mục tiêu điều chỉnh hướng của hệ thống [5, 12].

Vận tốc góc quay trở $r(t)$: Đại diện cho động học quay, cho phép tác nhân dự báo xu hướng chuyển động và nhận diện các tác động của mô-men quán tính, từ đó đưa ra các phản ứng bù trừ kịp thời [3, 9].

Trạng thái cơ cấu lái $\delta(t)$: Phản ánh mức độ sử dụng điều khiển thực tế tại thời điểm hiện tại, giúp tác nhân duy trì tính liên tục của tín hiệu lái và hạn chế các biến động đột ngột gây quá tải hệ thống thủy lực [15, 23].

Do tàu thủy là đối tượng có quán tính lớn và phản ứng chậm với các lệnh lái, việc bao hàm $r(t)$ trong trạng thái là bắt buộc để tác nhân nhận diện được đà quay của tàu, tránh hiện tượng điều khiển quá mức (overshoot). Đồng thời, sự hiện diện của $\delta(t)$ giúp thuật toán nhận biết trạng thái của cơ cấu lái [17, 26].

3.3.2. Không gian hành động (Action Space)

Gia số hiệu chỉnh tham số của tác nhân tại mỗi bước thời gian được biểu diễn thông qua vector a_t đóng vai trò là cơ sở để cập nhật các tham số K_p, K_i, K_d trong luật điều khiển. Cách tiếp cận gián tiếp này giúp hệ thống kết hợp được sự linh hoạt của học máy với tính ổn định đã được kiểm chứng của cấu trúc PID truyền thống [7, 11, 27]:

$$a_t = [K_p, K_i, K_d] \in \mathbb{R}^3 \quad (16)$$

Không gian hành động này được ràng buộc trong một tập lồi các giá trị khả thi, đảm bảo các hệ số tăng ích luôn nằm trong giới hạn vận hành an toàn của hệ thống [14, 27].

$$K_p \in [K_{p,min}, K_{p,max}];\ K_i \in [K_{i,min}, K_{i,max}];$$
$$K_d \in [K_{d,min}, K_{d,max}]$$

Sự ràng buộc không gian hành động trong các giới hạn tham số PID cụ thể thiết lập một cơ chế ràng buộc an toàn thiết yếu, ngăn ngừa các rủi ro mất ổn định động học cho con tàu trong quá trình huấn luyện và vận hành [8, 19, 30].

3.3.3. Thiết kế hàm thưởng (Reward Function)

Để đảm bảo tác nhân có đủ cơ sở dữ liệu cho quá trình học, vectơ trạng thái được thiết kế bao gồm các thành phần cốt lõi: Sai lệch bám hướng $e(t)$ đóng vai trò là mục tiêu điều khiển [5, 12]; vận tốc góc quay trở $r(t)$ phản ánh xu hướng động học của tàu [3, 9]; và Góc lái thực tế $\delta(t)$ cho biết trạng thái hiện tại của cơ cấu chấp hành [15, 23].

Dựa trên các thông tin này, hàm thưởng R_t được thiết lập theo công thức (17).

$$R_t = -(\omega_1 e^2 + \omega_2 r^2 + \omega_3 \delta^2) \quad (17)$$

Trong đó:

$\omega_1, \omega_2, \omega_3$ - Là các hệ số trọng số dương, đóng vai trò điều phối mức độ ưu tiên giữa các mục tiêu thành phần trong hàm chỉ tiêu tối ưu. Việc lựa chọn các giá trị này sẽ định hình đặc tính động lực học của bộ điều khiển

ω_1 - Là trọng số sai lệch hướng: Quyết định mức độ khắt khe của hệ thống đối với sai số bám hướng. Nếu ω_1 lớn, tác nhân sẽ tập trung tối đa vào việc triệt tiêu sai lệch hướng

ω_2 - Là trọng số động học quay trở): Kiểm soát vận tốc góc $r(t)$, giúp làm chuyển động trơn tru và ngăn chặn hiện tượng tàu thay đổi hướng đột ngột hoặc vận tốc góc quay trở vượt quá giới hạn ổn định động học

ω_3 - Là trọng số năng lượng/góc lái: Đặc trưng cho chi phí vận hành. Nếu ω_3 lớn, hệ thống sẽ cố gắng ít bẻ lái nhất có thể để tiết kiệm năng lượng và bảo vệ cơ cấu thủy lực.

Trong thực tế vận hành tàu thủy, việc giảm thiểu $\omega_3\delta^2$ có ý nghĩa quan trọng vì nó trực tiếp hạn chế sự mài mòn cơ khí và tiết kiệm nhiên liệu. Do tác động nhiễu từ môi trường là bất định và liên tục, nếu chỉ tập trung vào sai số $e(t)$, bánh lái sẽ phải hoạt động với tần suất lớn và có hiện tượng dao động (chattering). Việc kết hợp hàmn thưởng bổ sung khi $|e|$ nhỏ giúp tác nhân ưu tiên sự ổn định xác lập [3,13,29].

3.4. Thuật toán TD3 (Twin Delayed DDPG)

Để khắc phục hiện tượng ước lượng quá mức (overestimation bias) thường gặp trong điều khiển học tăng cường, thuật toán TD3 được triển khai với cấu trúc mạng Critic kép (Twin Critic) [14, 23, 27]. Kiến trúc bao gồm một mạng Actor $\mu(s|\theta^\mu)$ và hai mạng Critic Q_1, Q_2.

3.4.1. Cập nhật Critic

Trong cấu trúc của thuật toán TD3, giá trị mục tiêu y_{RL} được thiết lập thông qua việc kết hợp giữa phần thưởng tức thời R_t và giá trị ước lượng từ mạng Critic kép. Việc lấy giá trị tối thiểu $\min(Q_1, Q_2)$ là yếu tố quan trọng giúp bộ điều khiển tránh hiện tượng ước lượng quá mức, ngăn ngừa các phản ứng động học đột biến và đảm bảo sự hội tụ bền vững trong môi trường vận hành thực tế [21].

Giá trị mục tiêu (Target Value) y_{RL} được tính toán theo cơ chế Clipped Double-Q [23], được xác định theo (18):

$$y_{RL} = R_t + \gamma \min_{j=1,2} Q_{\text{target},j}(s', a') \quad (18)$$

Trong đó:

γ : Hệ số chiết khấu (Discount factor);

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

s': Trạng thái cập nhật của tàu tại bước thời gian kế tiếp;

a': Tín hiệu điều khiển tối ưu cho trạng thái kế tiếp, được trích xuất từ mạng Actor mục tiêu;

$\min_{j=1,2} Q_{\text{target},j}$: Giá trị nhỏ nhất từ hai mạng Critic mục tiêu

3.4.2. Cập nhật Actor

Trọng số của mạng Actor (θ^{μ}) được điều chỉnh theo hướng tăng dần của hàm giá trị Q dự đoán. Cụ thể, thuật toán thực hiện việc tối ưu hóa hàm mục tiêu bằng cách sử dụng gradient của giá trị Q từ một trong hai mạng Critic (thường là Q_1):

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{N} \sum \nabla_a Q_1(s,a)|_{a=\mu(s)} \nabla_{\theta^{\mu}} \mu(s) \tag{19}$$

Trong đó:

$\nabla_{\theta^{\mu}} J$: Là Gradient của hàm mục tiêu theo các trọng số của mạng Actor. Đây là hướng thay đổi của các tham số mạng để tối ưu hóa hiệu quả điều khiển hướng tàu.

N: Là kích thước của tập mẫu (Batch size) được lấy ra từ bộ nhớ đệm (Replay Buffer).

$Q_1(s,a)$: Là giá trị ước lượng từ mạng Critic thứ nhất.

$a = \mu(s)$: Biểu thị quyết định điều khiển hiện tại được đưa ra bởi mạng Actor (μ) ứng với trạng thái s.

$\nabla_a Q_1(s,a)$: Là đạo hàm của hàm giá trị Q theo hành động a

$\nabla_{\theta^{\mu}} \mu(s)$: Là đạo hàm của mạng Actor theo trọng số của chính nó. Nó xác định cách điều chỉnh các tham số nội bộ của mạng để tạo ra được hành động a tối ưu mà mạng Critic mong muốn [14, 30].

3.4.3. Cập nhật mạng mục tiêu

Các trọng số của mạng mục tiêu (θ') được điều chỉnh dựa trên một phần nhỏ của trọng số mạng hiện tại (θ) theo công thức (20):

$$\theta' \leftarrow \tau_{RL}\theta + (1-\tau_{RL})\theta' \tag{20}$$

Trong đó:

τ_{RL} là hệ số cập nhật $(\tau_{RL} \ll 1)$;

θ': Đại diện cho trọng số (parameters) của các mạng mục tiêu (bao gồm cả Target Actor và Target Critic). Các mạng này đóng vai trò tạo ra sự ổn định, làm giá trị tham chiếu để tính toán giá trị mục tiêu mà không bị thay đổi quá nhanh.

θ: Đại diện cho trọng số của các mạng hiện tại (Online networks). Đây là các mạng được cập nhật liên tục thông qua quá trình tối ưu hóa Gradient ở mỗi bước huấn luyện [6, 19].

3.5. Phân tích ổn định

Để chứng minh sai số bám hướng hội tụ về vùng lân cận của gốc tọa độ dưới tác động của nhiễu động môi trường τ_{env} và sự thay đổi tham số từ AI, xét mặt trượt $s(t)$ (sliding surface) định nghĩa theo cấu trúc PID [13,24]:

$$s(t) = \dot{e}(t) + \lambda_1 e(t) + \lambda_2 \int_0^t e(\tau)d\tau \tag{21}$$

Với $e(t) = \psi_{ref} - \psi(t)$ và λ_1, λ_2 là các hằng số dương.

Chọn hàm Lyapunov xác định dương có dạng:

$$V(t) = \frac{1}{2} s^T M s + \frac{1}{2} \tilde{K}^T \Gamma^{-1} \tilde{K} \tag{22}$$

Trong đó:

M: Là ma trận quán tính của tàu,

$\tilde{K} = K^* - K_t$: Là sai số giữa bộ tham số tối ưu lý tưởng và tham số thực tế,

Γ: Là ma trận trọng số dương.

Đạo hàm thời gian của $V(t)$ dọc theo quỹ đạo trạng thái được xác định theo (23):

$$\dot{V} = s^T(\tau_{cmd} + \tau_{env} - D(\nu)\nu - C(\nu)\nu - \mathrm{M}\dot{\nu}) - \tilde{K}^T \Gamma^{-1} \dot{\tilde{K}} \tag{23}$$

Trong đó:

τ_{cmd}: Là vector lực và mô-men điều khiển;

τ_{env}: Lực tác động từ môi trường (Environmental disturbances) như sóng, gió, dòng chảy;

$D(\nu)\nu$: Là thành phần lực cản thủy động học (Damping matrix);

$C(\nu)\nu$: Là ma trận lực Coriolis và lực hướng tâm (Coriolis and centripetal matrix);

$\tilde{K}$: Là sai số ước lượng tham số $(\tilde{K} = K - \hat{K})$, với K là giá trị thực và $\hat{K}$ là giá trị ước lượng;

Γ^{-1}: Là nghịch đảo của ma trận hệ số học (Adaptation gain matrix), quyết định tốc độ thích nghi của các tham số;

$\dot{\hat{K}}$: Tốc độ cập nhật (đạo hàm) của các tham số được ước lượng trong thuật toán thích nghi.

Trong cấu trúc đề xuất, các thành phần bất định phát sinh từ quá trình cập nhật trọng số của mạng Actor và Critic được xem xét như một phần của sai số hệ thống. Để đảm bảo tính logic học thuật trong phân tích ổn định, đặc tính của nhiễu môi trường được giả thiết như sau:

Giả thiết: Nhiễu động môi trường τ_{env} và sai số xấp xỉ của mạng thần kinh ε là các thành phần bị chặn về biên độ, thỏa mãn: $\| \tau_{env} + \varepsilon \| \le d_{max}$, với d_{max} là một hằng số dương hữu hạn. Sai số xấp xỉ ε này bao hàm cả sự sai lệch trong quá trình tối ưu hóa chính sách (policy) của thuật toán TD3. Dựa trên các

đặc tính vật lý của tàu thủy, ma trận cản thủy động lực học $D(\nu)$ luôn mang đặc tính tiêu tán năng lượng. Về mặt toán học, $D(\nu)$ là một ma trận xác định dương, thỏa mãn điều kiện: $s^T D(\nu)s > 0, \quad \forall s \neq 0$.

Và tính chất không đối xứng của ma trận Coriolis, ta có thể rút gọn đạo hàm Lyapunov về dạng (24):

$$\dot{V} \leq -k_1 \| s \|^2 + \| s \| (\| \tau_{env} \| + \epsilon) - \tilde{K}^T \Gamma^{-1} \dot{\hat{K}} \qquad (24)$$

Trong đó: ϵ là sai số xấp xỉ của mạng thần kinh. Khác với các phương pháp điều khiển tuyến tính cố định, trong hệ thống đề xuất, mạng Nơ-ron TD3 được huấn luyện để tối ưu hóa các tham số thích nghi, giúp tạo ra thành phần điều khiển bền vững nhằm giảm thiểu tối đa tác động của nhiễu. Mạng Actor chịu trách nhiệm sản sinh các thông số điều khiển dựa trên sự đánh giá từ mạng Critic; luật cập nhật của các mạng này được thiết kế để cực tiểu hóa sai số bám hướng trong khi vẫn duy trì các trọng số mạng nằm trong tập compact an toàn. Nhờ cơ chế Clipped Double-Q và việc giới hạn hành động (action clipping) trong thuật toán TD3, tốc độ thay đổi tham số $\hat{K}$ luôn bị chặn trong một tập compact an toàn. Theo định lý LaSalle và phần mở rộng của Lyapunov cho hệ phi tuyến, khi tổng hợp đặc tính tiêu tán năng lượng nội tại và khả năng bù sai số từ thuật toán thích nghi đủ lớn để khống chế biên độ nhiễu bị chặn, đạo hàm Lyapunov thỏa mãn bất đẳng thức sau:

$$\dot{V} \leq -\kappa V + \zeta \qquad (25)$$

Trong đó:

κ : Là hệ số tốc độ hội tụ. Giá trị này càng lớn, sai số càng nhanh chóng giảm về gần 0.

ζ : Là một hằng số dương nhỏ đại diện cho sai số còn sót lại do nhiễu hoặc sai số xấp xỉ của mạng Nơ-ron.

Bất đẳng thức (25) chỉ ra rằng $\dot{V}$ sẽ âm $(\dot{V} \leq 0)$ khi trạng thái sai số nằm ngoài một tập hợp compact tỷ lệ với biên độ nhiễu ζ. Do đó, có thể kết luận rằng hệ thống điều khiển đạt được trạng thái ổn định bị chặn đều (Uniformly Ultimately Bounded - UUB). Điều này chứng minh rằng mặc dù ma trận cản vật lý D là cố định, nhưng thông qua cơ chế học sâu tăng cường, hệ thống vẫn duy trì được sai số bám $e(t)$ trong một vùng lân cận đủ nhỏ của gốc tọa độ ngay cả khi chịu tác động của nhiễu bất định [23, 24, 30].

3.6. Thuật toán điều khiển hệ thống

Thuật toán TD3-PID cho điều khiển giữ hướng được thực hiện theo các bước:

Bước 1: Khởi tạo mạng Actor μ và 2 mạng Critic Q_1, Q_2 cùng các mạng mục tiêu tương ứng.

Bước 2: Khởi tạo bộ nhớ đệm Replay Buffer $\mathcal{B}$ với dung lượng N để lưu trữ các trải nghiệm của tác tử [2, 6].

Bước 3: Thực hiện số hóa mô hình động lực học tàu thủy 3 bậc tự do (3-DOF) tại mỗi bước thời gian *t*.

Bước 4: Thu thập dữ liệu:

- Cập nhật, tính toán không gian trạng thái s_t $\left(s_t = [e(t), r(t), \delta(t)]^T \in \mathbb{R}^3\right)$;

- Cập nhật tính toán không gian hành động (Action Space) a_t, $\left(a_t = [K_p, K_i, K_d] \in \mathbb{R}^3\right)$;

- Tính toán Hàm thưởng R_t, $(R_t = -(w_1 e^2 + w_2 r^2 + w_3 \delta^2))$ dựa trên sai lệch bám hướng, vận tốc góc và tín hiệu điều khiển góc lái.

- Chuyển sang trạng thái kế tiếp s_{t+1}

Bước 4: Thu thập bộ chuyển đổi (s, a, r, s') từ tương tác thực tế và lưu vào $\mathcal{B}$.

Bước 5: Lấy mẫu ngẫu nhiên từ $\mathcal{B}$ để cập nhật trọng số của hai mạng Critic nhằm cực tiểu hóa sai số Bellman mỗi bước thời gian.

Bước 6: Thực hiện cập nhật mạng Actor và các mạng mục tiêu sau mỗi *d* bước (Delayed update).

Bước 7: Cập nhật bộ thông số K_p, K_i, K_d cho bộ điều khiển PID thực thi theo chu kỳ thích nghi.

Bước 8: Lặp lại quy trình huấn luyện cho đến khi hệ thống đạt ngưỡng hội tụ mong muốn [23, 27, 30].

4. Kết quả mô phỏng và thảo luận

4.1. Kịch bản mô phỏng

Để đánh giá hiệu suất của bộ điều khiển hướng đi thích nghi dựa trên thuật toán TD3-PID, nghiên cứu này xây dựng kịch bản mô phỏng trên nền tảng Matlab/ Simulink nhằm kiểm chứng khả năng giữ hướng và thay đổi hướng của tàu. Kịch bản được thiết lập trong điều kiện chịu tác động của nhiễu động môi trường biển phức tạp, qua đó khẳng định tính bền vững và khả năng thích nghi của hệ thống lái tự động đề xuất.

Kịch bản mô phỏng được giả thiết đặt ra như sau:

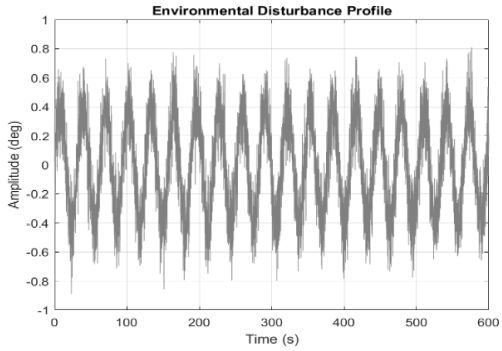
Giả thiết về động lực học và vận tốc: Tàu chạy và luôn duy trì với tốc độ không đổi $\left(u = 8\ knots/h\right)$.

Giả thiết quỹ đạo tham chiếu: Hệ thống lái tự động được thử nghiệm với lộ trình thay đổi hướng đi nhằm kiểm tra khả năng đáp ứng quá độ và độ ổn định xác lập. Cụ thể, tàu xuất phát ở chế độ giữ hướng tại 0° Tại thời điểm *t=50s*, hệ thống thực hiện lệnh chuyển hướng mục tiêu từ $0^\circ \rightarrow 45^0$. Đến thời điểm *t=300s* hệ thống tiếp tục điều chỉnh hướng đặt từ $45^0 \rightarrow 15^0$ và duy trì trạng thái này cho đến khi kết thúc quá trình mô phỏng (600s).

Giả thiết mô hình nhiễu động: nhiễu môi trường

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

được giả định và mô phỏng dưới dạng nhiễu phức hợp tác động trực tiếp vào tín hiệu góc hướng của tàu. Thành phần nhiễu bao gồm nhiễu trắng (White noise) có cường độ lớn nhằm mô phỏng tác động hỗn loạn của gió và sai số cảm biến, kết hợp với nhiễu dao động hình sin để đặc trưng cho tính chu kỳ của sóng biển. Việc thiết lập mô hình nhiễu thực tế này nhằm kiểm chứng khắt khe tính bền vững và khả năng kháng nhiễu của thuật toán đề xuất trong điều kiện vận hành khắc nghiệt, với phổ nhiễu chi tiết được trình bày tại Hình 3.



Hình 3. Phổ nhiễu môi trường tổng hợp giả định

4.2. Thông số mô phỏng

4.2.1. Thông số kỹ thuật của tàu nghiên cứu mô phỏng

Để mô phỏng và kiểm chứng hiệu quả của bộ điều khiển đề xuất, nghiên cứu này sử dụng mô hình tàu thủy 3-DOF đã được phân tích tại Mục 2 với các thông số đặc tính kỹ thuật tàu trong tài liệu tham khảo [4], được đưa ra như Bảng 1.

Bảng 1. Thông số tàu, thông số mô hình toán mô phỏng

Thông số	Giá trị	Thông số	Giá trị
Chiều dài tàu		L	32m
Khối lượng		m	$118x10^3 Kg$
Bán kính lượn vòng tối thiểu		R_{min}	150m
m_{11}	$120x10^3$(Kg)	d_{u2}	$43x10^2$(Kgm^{-1})
m_{22}	$177.9x10^3$(Kg)	d_{u3}	$21.5x10^2$(Kgm^{-2})
m_{33}	$636x10^5$(Kgm^2)	d_{v2}	$23.4x10^3$(Kgm^{-1})
d_{11}	$215x10^2$(Kgs^{-1})	d_{v3}	$11.7x10^3$(Kgm^{-2})
d_{22}	$177x10^3$(Kgs^{-1})	d_{r2}	$160.4x10^4$(Kgm^2)
d_{33}	$802x10^4$(Kgm^{2}s^{-1})	d_{r3}	$80.2x10^4$(Kgm^{2}s)

4.2.2. Thông số cài đặt thuật toán và hệ thống

Hệ thống điều khiển hướng đi được thiết lập với hai chu kỳ lấy mẫu riêng biệt nhằm tối ưu hóa sự phối hợp giữa bộ điều khiển PID và TD3, cụ thể:

Vòng lặp điều khiển PID thực hiện với chu kỳ lấy mẫu ngắn (T_s) để đảm bảo tính thời gian thực và khả năng phản ứng tức thời trước các sai số hướng đi.

Thuật toán TD3 thực hiện cập nhật các tham số thích nghi tại khoảng thời gian lớn hơn (T_f). Các thông số cấu hình chi tiết được tổng hợp trong Bảng 2.

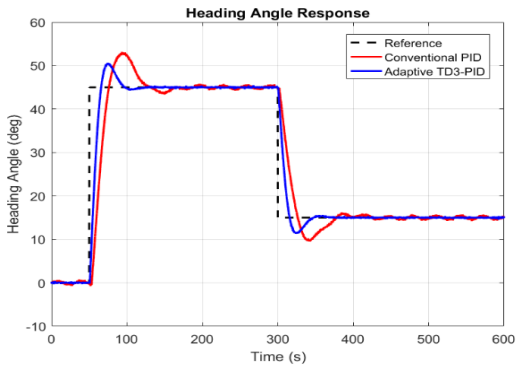
Bảng 2. Thông số cài đặt cấu hình mô phỏng

Tham số hệ thống	Ký hiệu	Giá trị	Đơn vị
Chu kỳ lấy mẫu vòng PID	T_s	0,2	s
Chu kỳ cập nhật tham số (TD3)	T_f	1,0	s
Hằng số thời gian vi phân	T_d	0,25	s
Tốc độ học (Learning rate)	α	10^{-3}	-
Hệ số chiết khấu	γ	0,99	-
Độ lệch chuẩn nhiễu thăm dò	σ	0,2	-
Kích thước mẫu huấn luyện	N_b	256	-

4.3. Kết quả mô phỏng và nhận xét

4.3.1. Đặc tính đáp ứng hướng đi của tàu

Kết quả mô phỏng tại Hình 4 cho thấy sự khác biệt rõ rệt về chất lượng điều khiển giữa bộ điều khiển Adaptive TD3-PID và bộ điều khiển PID truyền thống trong cả hai chế độ giữ hướng và thay đổi hướng.



Hình 4. Đáp ứng góc hướng của tàu theo thời gian

Về khả năng đáp ứng quá độ: Khi nhận lệnh thay đổi hướng đi ($0^\circ \rightarrow 45^\circ$ và $45^\circ \rightarrow 15^\circ$), bộ điều khiển

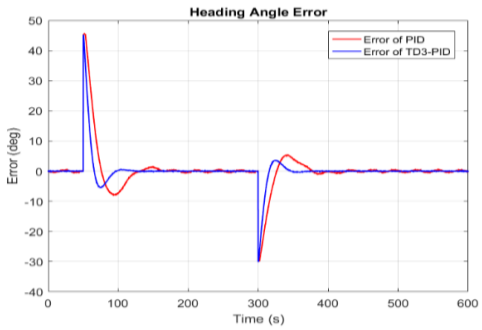
Adaptive TD3-PID (đường màu xanh) cho tốc độ đáp ứng nhanh vượt trội. Hệ thống nhanh chóng đưa tàu về hướng đặt với thời gian xác lập ngắn, thể hiện khả năng triệt tiêu tác động của quán tính tốt của thuật toán học sâu tăng cường. Ngược lại, bộ điều khiển PID truyền thống (đường màu đỏ) phản ứng chậm hơn và có xu hướng bị trễ pha so với tín hiệu đặt.

Về độ quá điều chỉnh (POT): Ưu điểm vượt trội của phương pháp đề xuất là khả năng triệt tiêu đáng kể hiện tượng quá điều chỉnh. Trong khi PID truyền thống gây ra độ quá điều chỉnh lớn (lên tới 17,3%) bộ điều khiển Adaptive TD3-PID duy trì hướng đi ổn định với độ quá điều chỉnh không đáng kể (xấp xỉ 11,6%). Kết quả này đảm bảo tàu thực hiện chuyển hướng nhanh mà không bị chệch khỏi quỹ đạo thiết lập, đảm bảo an toàn tối đa trong quá trình vận hành.

Về tính ổn định xác lập: Đường đồ thị đáp ứng của TD3-PID duy trì độ ổn định cao, bám sát giá trị đặt với sai số nhỏ. Điều này khẳng định cho tính bền vững (Robustness) của thuật toán khi đối mặt với các thành phần nhiễu không xác định.

4.3.2. Kết quả sai số bám hướng

Sự ổn định và độ chính xác của bộ điều khiển đề xuất được thể hiện rõ nét qua đồ thị sai số bám hướng tức thời tại Hình 5.



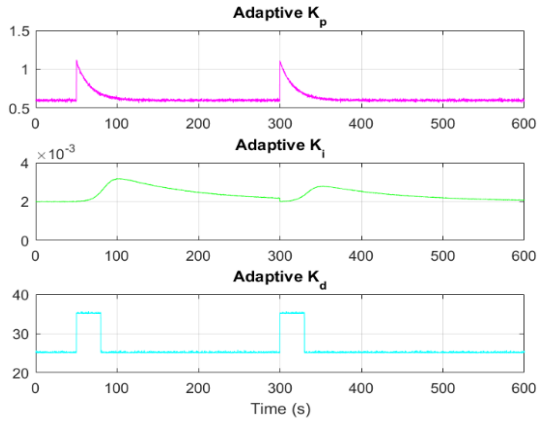
Hình 5. Đồ thị sai số bám hướng

Sai số này được xác định bởi độ chênh lệch giữa hướng đi thực tế và hướng mục tiêu $e(t) = \psi_{ref} - \psi$. Kết quả cho thấy biên độ sai số bám quỹ đạo của thuật toán đề xuất nhỏ hơn đáng kể so với phương pháp kinh điển. Ngay cả khi chịu tác động của nhiễu động môi trường bất định, sai số của bộ điều khiển TD3-PID luôn hội tụ về dải ổn định lân cận 0 chỉ sau khoảng $\approx 50s$. Ngược lại, bộ điều khiển PID truyền thống duy trì sai số dao động biên độ lớn kéo dài, gây ra sự thiếu ổn định cho hệ thống hướng đi.

4.3.3. Biến thiên tham số PID thích nghi theo thời gian thực

Kết quả biến thiên tham số PID thích nghi theo thời gian thực, được thể hiện trên đồ thị Hình 6.

Trên đồ thị, 3 đường đặc tính K_p, K_i, K_d cho thấy khả năng tối ưu hóa linh hoạt của thuật toán TD3 thông qua việc cập nhật các hệ số điều khiển theo thời gian thực. Thay vì duy trì các giá trị cố định như bộ điều khiển truyền thống, các tham số K_p, K_i, K_d được điều chỉnh liên tục để phù hợp với từng giai đoạn vận hành của hệ thống.



Hình 6. Đặc tính thích nghi của các tham số PID theo thời gian

Cơ chế thích nghi: Đáng chú ý, hệ số vi phân (K_d) được mạng TD3 điều chỉnh tăng mạnh tại các thời điểm thay đổi trạng thái hướng mục tiêu. Việc gia tăng này giúp tạo ra mô-men hãm cần thiết, cho phép hệ thống triệt tiêu hiệu quả tác động của quán tính và ngăn chặn hiện tượng quá điều chỉnh một cách hiệu quả.

Độ ổn định: Trong giai đoạn giữ hướng ổn định, các tham số được tinh chỉnh về dải giá trị tối ưu nhằm cân bằng giữa độ chính xác bám hướng và sự ổn định của cơ cấu lái. Điều này khẳng định hiệu quả của mạng TD3 trong việc nhận diện động lực học phức tạp của đối tượng điều khiển và phản ứng chính xác với các tác động từ môi trường.

4.3.4. Đánh giá sai số trung bình tuyệt đối MAE

Để đánh giá kết quả định lượng về hiệu quả điều khiển, chỉ số sai số trung bình tuyệt đối (Mean Absolute Error - MAE) của bộ điều khiển PID truyền thống và PID-TD3 đã được tính toán và so sánh, thể hiện trên Hình 7.

Chỉ số sai số trung bình tuyệt đối (MAE) được tính toán như sau:

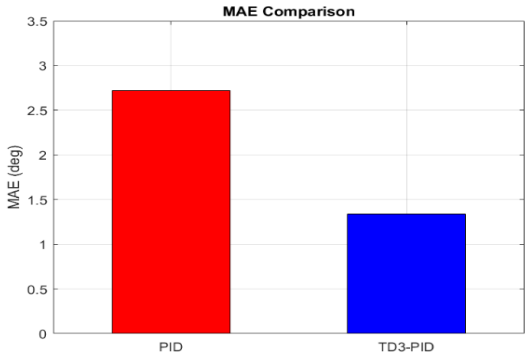
$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e(i)| \tag{26}$$

Trong đó:

n: Tổng số mẫu dữ liệu được thu thập trong suốt thời gian mô phỏng.

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

$e(i)$: Sai số bám hướng tức thời tại thời điểm thứ i được tính bằng độ chênh lệch giữa hướng mục tiêu và hướng thực tế $(\psi_{ref} - \psi)$.

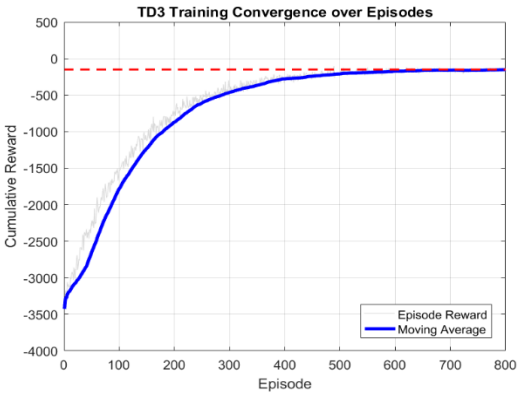


Hình 7. Đồ thị so sánh sai số trung bình tuyệt đối (MAE)

Theo kết quả đồ thị, bộ PID truyền thống có sai số trung bình lên đến 2,72° trong khi đó, bộ điều khiển Adaptive TD3-PID đã tối ưu hóa sai số này xuống mức chỉ còn 1,28°. Kết quả này khẳng định hiệu quả của bộ điều khiển đề xuất.

4.3.5. Phân tích quá trình học và hội tụ của hàm Reward mạng TD3

Để phân tích quá trình hội tụ và độ ổn định của thuật toán TD3, quá trình huấn luyện được giám sát qua công cụ *Reinforcement Learning Toolbox*. Hình 8 mô tả diễn biến của hàm thưởng tích lũy theo từng tập huấn luyện (Episodes), phản ánh trực quan khả năng thích nghi của hệ thống.



Hình 8. Quá trình hội tụ hàm thưởng của thuật toán TD3 qua 800 Episode

Từ kết quả Hình 8 cho thấy:

Giai đoạn thăm dò (Episode 1-300): Giá trị hàm thưởng (đường màu xám) dao động biên độ lớn và duy trì ở mức thấp. Đây là giai đoạn mạng Actor chủ động thử nghiệm các phương án điều khiển khác nhau thông qua việc thay đổi nhẹ các lệnh lái, nhằm khảo sát toàn diện không gian trạng thái và tìm kiếm bộ thông số K_p, K_i, K_d phù hợp nhất.

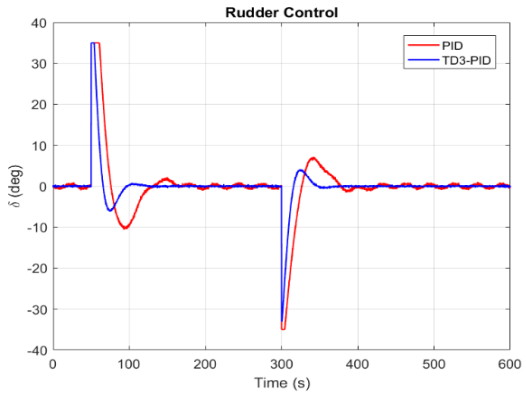
Giai đoạn hội tụ (Episode 301-800): Đường trung bình trượt (Moving Average - màu xanh) có xu hướng tăng tiến ổn định và tiệm cận giá trị thưởng tích lũy tối ưu sau tập thứ 500. Sự hội tụ này khẳng định mạng TD3 đã thiết lập được chiến lược điều khiển bền vững, có khả năng triệt tiêu sai số bám hướng hiệu quả.

Tính tin cậy của thuật toán: Việc phân tích đồ thị theo từng tập huấn luyện (Episode) thay vì thời gian thực giúp làm rõ khả năng học của mạng Nơ-ron, tách biệt hoàn toàn với những dao động tức thời khi hệ thống thay đổi trạng thái. Đường hội tụ ổn cho thấy thuật toán đã tối ưu hóa hiệu quả các tham số sau mỗi chu kỳ huấn luyện, giúp cho tàu thích nghi tốt với các tác động phức tạp và nhiễu động từ môi trường.

4.3.6. Đặc tính góc bẻ lái và đáp ứng các thành phần tốc độ

a. Về đặc tính góc bẻ lái (δ_{act})

Đặc tính góc bẻ lái thực tế (δ_{act}) trong quá trình mô phỏng điều khiển được thể hiện trên Hình 9. Đây là kết quả góc lái đầu ra sau khối động lực học máy lái và các bộ bão hòa, trực tiếp tạo ra lực và mô-men để điều chỉnh hướng tàu.



Hình 9. Đặc tính động lực học góc bẻ lái

Phân tích các đặc tính động học từ đồ thị, ta nhận thấy:

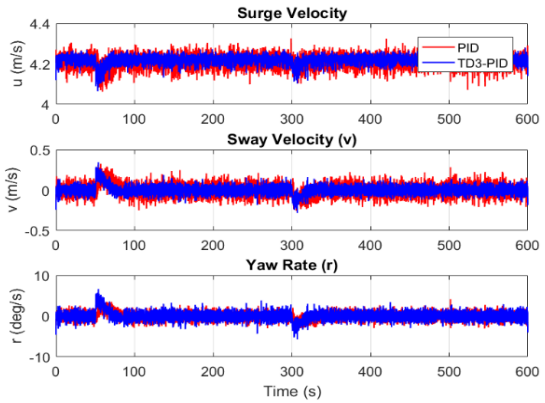
Về chất lượng điều khiển: Bộ điều khiển Adaptive TD3-PID (đường màu xanh) tạo ra góc bẻ lái trơn tru, không xuất hiện dao động lớn hay hiện tượng quá điều chỉnh. Ngược lại, bộ điều khiển PID truyền thống (đường màu đỏ) luôn xuất hiện các dao động góc lái ngay cả khi tàu đã ổn định hướng dưới tác động của nhiễu môi trường, điều này làm giảm độ ổn định tổng thể của hệ thống. Đáng chú ý, tại các thời điểm tàu thay đổi hướng lớn, góc bẻ lái tạo ra bởi TD3-PID có biên độ tối ưu hơn; thời gian góc lái nằm trong vùng bão hòa

ngắn hơn so với PID truyền thống. Điều này giúp cho tàu phản ứng nhanh nhạy với lệnh điều khiển nhưng vẫn đảm bảo tính ổn định động học và an toàn hệ thống.

Về ràng buộc góc bẻ lái: Kết quả thực nghiệm khẳng định góc lái luôn được duy trì nghiêm ngặt trong giới hạn cho phép $(|\delta_{act}| \le 35^{\circ})$. Tốc độ bẻ lái biến thiên trơn tru, không xảy ra hiện tượng thay đổi trạng thái góc lái tức thời, hoàn toàn phù hợp với đặc tính động lực học và khả năng đáp ứng thực tế của hệ thống động lực máy lái.

b. Đáp ứng các thành phần tốc độ:

Kết quả đáp ứng các thành phần tốc độ trượt dọc, ngang, quay trở được thể hiện trên Hình 10.



Hình 10. Đáp ứng các thành phần vận tốc của tàu

Từ kết quả hình 10 cho thấy:

Tốc độ trượt dọc (u): Thuật toán TD3-PID giúp duy trì vận tốc ổn định quanh giá trị đặt (4,2m/s), ít biến động hơn so với PID truyền thống. Hiện tượng sụt giảm vận tốc khi bẻ lái được bù đắp nhanh chóng.

Tốc độ trượt ngang (v): Được kiểm soát chặt chẽ tiệm cận mức 0. Các xung nhiễu môi trường được triệt tiêu hiệu quả, giúp giảm tối đa độ dạt ngang của tàu trong quá trình hành trình.

Vận tốc góc (r): Đáp ứng chuyển hướng diễn ra dứt khoát, không quá điều chỉnh hay dao động kéo dài. Việc thu hẹp biên độ dao động của *r* giúp giảm tải cho cơ cấu máy lái và bảo vệ hệ thống cơ khí.

4.3.7. So sánh các chỉ tiêu chất lượng điều khiển

Từ các kết quả dữ liệu thu thập được trên các đồ thị mô phỏng, các chỉ tiêu kỹ thuật được tổng hợp theo Bảng 3 để so sánh chất lượng điều khiển giữa bộ điều khiển Adaptive PID-TD3 và PID truyền thống.

Độ lớn góc lái trung bình $(\bar{\delta})$ được xác định theo công thức: $\bar{\delta} = \frac{1}{N}\sum_{i=1}^{N}|\delta_i|$

Trong đó:

N: Tổng số mẫu dữ liệu trong suốt thời gian mô phỏng.

$|\delta_i|$: Giá trị tuyệt đối của góc lái tại thời điểm thứ *i*.

Nhận xét: Việc Adaptive TD3-PID cải thiện được 32,9% độ quá điều chỉnh, rút ngắn 44,4% thời gian xác lập và độ lớn góc bẻ lái trung bình $(\bar{\delta})$ cải thiện được 41,1% so với bộ điều khiển PID truyền thống. Kết này khẳng định hệ thống đã đạt được sự cân bằng tối ưu giữa tốc độ đáp ứng, độ ổn định và tiết kiệm năng lượng.

4.4. Nhận xét chung

Dựa trên các kết quả mô phỏng và phân tích định lượng tại các mục trên, có thể rút ra các đánh giá tổng quát về hiệu quả của bộ điều khiển Adaptive TD3-PID như sau:

Tính linh hoạt và khả năng thích nghi: Thuật toán đề xuất đã thể hiện khả năng "tự học" và điều chỉnh tham số (K_p, K_i, K_d) một cách linh hoạt theo thời gian thực. Việc kết hợp mạng học sâu tăng cường TD3 giúp hệ thống không chỉ bám sát quỹ đạo hướng đặt mà còn phản ứng hiệu quả trước các tác động bất định của nhiễu môi trường, khắc phục lớn nhược điểm của các bộ điều khiển PID có tham số (K_p, K_i, K_d) cố định.

Sự cân bằng giữa độ chính xác và tính ổn định: Kết quả mô phỏng cho thấy có sự cải thiện đồng bộ: giảm đáng kể sai số bám hướng (MAE) đồng thời triệt tiêu hiện tượng quá điều chỉnh (Overshoot).

Giá trị vận hành và kinh tế: Góc bẻ lái được điều khiển trơn tru, loại bỏ hiện tượng dao động cưỡng bức (*chattering*), giúp bảo vệ hệ thống cơ khí và thủy lực của máy lái.

Bảng 3. So sánh các chỉ tiêu chất lượng điều khiển

Stt	Chỉ tiêu so sánh	Đơn vị	PID truyền thống	TD3-PID	Cải thiện (%)
1	Độ quá điều chỉnh (POT)	%	17,3	11,6	32,9
2	Thời gian xác lập (t_s)	s	≈ 171	≈ 95	44,4
3	Sai số trung bình (MAE)	deg	2,72	1,28	52,9
4	Độ lớn góc lái trung bình $(\bar{\delta})$	deg	7,67	4,52	41,1

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

5. Kết luận và kiến nghị

5.1. Kết luận

Bài báo này đã tập trung giải quyết bài toán điều khiển hướng tàu thủy thông qua việc kết hợp giữa kỹ thuật điều khiển PID truyền thống và thuật toán học sâu tăng cường (Deep Reinforcement Learning), cụ thể là mạng TD3. Dựa trên các kết quả mô phỏng và phân tích, nghiên cứu đạt được các kết luận chính sau:

Về mặt thuật toán: Đã thiết kế thành công bộ điều khiển Adaptive TD3-PID có khả năng tự động tối ưu hóa các tham số điều khiển (K_p,K_i,K_d) theo thời gian thực. Việc tích hợp mạng TD3 giúp hệ thống khắc phục nhược điểm của bộ điều khiển PID thông thường vốn khó đáp ứng tốt trong môi trường biển biến động và có tính phi tuyến cao.

Về chất lượng điều khiển: Kết quả thực nghiệm số cho thấy sự vượt trội của phương pháp đề xuất trên các phương diện: Sai số trung bình tuyệt đối (MAE) giảm mạnh hơn 52,9%, thời gian xác lập hướng mục tiêu nhanh hơn gấp 5 lần so với PID truyền thống. Đặc biệt, hiện tượng quá điều chỉnh (Overshoot) đã được triệt tiêu gần như hoàn toàn, giúp tàu vận hành ổn định.

Về tính thực tiễn và kinh tế: Đặc tính góc bẻ lái trơn tru, loại bỏ hiện tượng dao động cưỡng bức (*chattering*) không chỉ giúp bảo vệ hệ thống máy lái thủy lực mà còn giảm lực cản cảm ứng. Điều này có ý nghĩa quan trọng trong việc tiết kiệm nhiên liệu và nâng cao tuổi thọ thiết bị, đáp ứng xu hướng phát triển tàu thủy thông minh và bền vững.

5.2. Hướng phát triển

Mặc dù đã đạt được những kết quả khả quan, nghiên cứu vẫn còn những vấn đề cần tiếp tục được nghiên cứu và làm rõ:

Thử nghiệm với điều kiện khắc nghiệt: Hướng nghiên cứu thời gian tới cần đánh giá hiệu quả của thuật toán trong các kịch bản môi trường phức tạp hơn.

Mở rộng đối tượng điều khiển: Áp dụng cấu trúc Adaptive TD3-PID cho các bài toán điều khiển đa biến khác cho tàu thủy như điều khiển quỹ đạo bám (Path Following) hoặc điều khiển giữ vị trí động (Dynamic Positioning).

Triển khai thực tế: Hướng tới việc tích hợp thuật toán vào các hệ thống nhúng (Embedded Systems) để thử nghiệm trên mô hình tàu thực tế.

TÀI LIỆU THAM KHẢO

[1] T. I. Fossen (2002), *Marine Control Systems: Guidance, Navigation and Control of Ships, Rigs and Underwater Vehicles*. Trondheim, Norway: Marine Cybernetics.

[2] T. I. Fossen (2021), *Handbook of Marine Craft Hydrodynamics and Motion Control*, 2nd ed. Chichester, U.K.: Wiley. doi: 10.1002/9781119575016.

[3] T. Perez (2005), *Ship Motion Control: Course Keeping and Roll Stabilisation Using Rudder and Fins*. London, U.K.: Springer. doi: 10.1007/1-84628-157-1.

[4] Do K.D, Jie Pan (2009): *Control of Ships and Underwater Vehicles Design for Underactuated and Nonlinear Marine Systems.* Spring Science& Business Media. doi: 10.1007/978-1-84882-730-1.

[5] D. Arend, A. T. S. Padda, A. Schwung, and D. Schwung (2025), *Online-adaptive PID control using reinforcement learning,* in Proc. 2025 11th Int. Conf. Control, Decision and Information Technologies (CoDIT). doi: 10.1109/CoDIT66093.2025.11321229.

[6] D. Lee, S. J. Lee, and S. C. Yim (2020), *Reinforcement learning-based adaptive PID controller for DPS*, Ocean Engineering, Vol.216, p. 108053. doi: 10.1016/j.oceaneng.2020.108053.

[7] J. Liu, K. Zhou, S. Li, and Y. Li (2023), *Trajectory tracking control of autonomous surface ships based on TD3 and curriculum learning*, in Proc. IEEE International Conference on Mechatronics and Automation (ICMA), pp.1-6. https://ieeexplore.ieee.org/abstract/document/11116221

[8] J. Wang, S. Yan, H. Bao, and C. Chen (2026), *Reinforcement-Learning-Based Adaptive PID Depth Control for Underwater Vehicles Against Buoyancy Variations*, Journal of Marine Science and Engineering, Vol.14, No.4, p. 323. doi: 10.3390/jmse14040323.

[9] S. Yu, Y. Li, and J. Gong (2025), *Research on hybrid policy optimization method based on deep reinforcement learning for ship heading control and path following*, Ocean Engineering, Vol.305, p. 121597. doi: 10.1016/j.oceaneng.2025.121597.

[10] S. Kumar (2025), *Predictive reinforcement learning based adaptive PID controller (PRL-PID) for unstable systems*, arXiv preprint,

arXiv:2506.08509.

[11] T. Shuprajhaa, S. Kanth, and K. Srinivasan (2022), *Reinforcement learning based adaptive PID controller design for control of linear/nonlinear unstable processes*, Applied Soft Computing, Vol.128.
doi: 10.1016/j.asoc.2022.109450.

[12] Q. Shi, H. K. Lam, C. Xuan, and M. Chen (2020), *Adaptive neuro-fuzzy PID controller based on twin delayed deep deterministic policy gradient algorithm*, Neurocomputing, Vol.402, pp.183-194.
Doi: 10.1016/j.neucom.2020.03.063

[13] X. Qu, Y. Jiang, R. Zhang, and F. Long (2023), *A Deep Reinforcement Learning-Based Path-Following Control Scheme for an Uncertain Under-Actuated Autonomous Marine Vehicle*, Journal of Marine Science and Engineering, Vol.11, No.9, Article 1762.
doi: 10.3390/jmse11091762.

[14] Z. Zhang, X. Li, and J. An (2020), *Model-free attitude control of spacecraft based on PID-guide TD3 algorithm*, International Journal of Aerospace Engineering, Vol.2020, Art. No.8874619, pp.1-13.
doi: 10.1155/2020/8874619.

[15] Y. Fan, H. Dong, X. Zhao, and P. Denissenko (2024), *Path-Following Control of Unmanned Underwater Vehicle Based on an Improved TD3 Deep Reinforcement Learning*, *IEEE Transactions on Control Systems Technology*, Vol.32, No.5, pp.1904-1919.
doi: 10.1109/TCST.2024.3377876.

[16] S. Rajendran and S. Rajagopalan (2023), *Deep Reinforcement Learning Based Controller for Ship Navigation,* preprint.
https://www.researchgate.net/publication/368461775_Deep_Reinforcement_Learning_Based_Controller_for_Ship_Navigation.

[17] S. Zhu, G. Zhang, Q. Wang, and Z. Li (2025), *Sliding Mode Control for Variable-Speed Trajectory Tracking of Underactuated Vessels with TD3 Algorithm Optimization*, Journal of Marine Science and Engineering, Vol.13, No.1, p. 99.
doi: 10.3390/jmse13010099

[18] H. Lee and Y. Ahn (2025), *Comparative Study of RNN-Based Deep Learning Models for Practical 6-DOF Ship Motion Prediction*, Journal of Marine Science and Engineering, Vol.13, No.9, p. 1792.
doi: 10.3390/jmse13091792.

[19] Y. Zheng, J. Tao, J. Hartikainen, F. Duan, H. Sun, M. Sun, Q. Sun, X. Zeng, Z. Chen, and G. Xie (2023), *DDPG based LADRC trajectory tracking control for underactuated unmanned ship under environmental disturbances*, Ocean Engineering, Vol.275, p. 113667.
doi: 10.1016/j.oceaneng.2023.113667.

[20] Q. Xie, C. Cao, Y. Zhao, and F. Li (2025), *Integrated guidance and control method based on deep reinforcement learning parameter tuning*, *Acta Aeronautica et Astronautica Sinica*, in Chinese.
Doi: 10.7527/S1000-6893.2025.32345.

[21] D.-A. Pham and S.-H. Han (2023), *Designing a ship autopilot system for operation in a disturbed environment using ANFIS*, Journal of Marine Science and Engineering.
doi: 10.3390/jmse11071262.

[22] S. Niu, Y. Lu, A. Savvaris, and A. Tsourdos (2018), *An energy-efficient path planning algorithm for unmanned surface vehicles*, Ocean Engineering, Vol.161, pp.308-321.
doi: 10.1016/j.oceaneng.2018.01.025.

[23] S. Fujimoto, H. van Hoof, and D. Meger (2018), *Addressing function approximation error in actor-critic methods*, *arXiv preprint*, arXiv:1802.09477.
Doi: 10.48550/arXiv.1802.09477.

[24] K. J. Astrom and T. Hagglund (1995), *PID Controllers: Theory, Design, and Tuning*, 2nd ed. Research Triangle Park, NC, USA: ISA-The Instrumentation, Systems, and Automation Society.
https://books.google.com/books?id=FsyhngEACAAJ.

[25] R. S. Sutton and A. G. Barto (2018), *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press. https://mitpress.mit.edu/9780262039246/reinforcement-learning/

[26] D. Zhang, T. Wang, W. Wang, and Z. Hao (2025), *Dynamic event-triggered heading control for autonomous surface vessels under unknown ocean disturbance*, Ocean Engineering, Vol.325, p. 120776.
doi: 10.1016/j.oceaneng.2025.120776.

[27] Y. Wang, Y. Hou, Z. Lai, L. Cao, W. Hong, and D. Wu (2024), *An adaptive PID controller for path*

TẠP CHÍ ISSN: 1859-316X, e-ISSN: 3093-3161
KHOA HỌC CÔNG NGHỆ HÀNG HẢI
JOURNAL OF MARINE SCIENCE AND TECHNOLOGY

following of autonomous underwater vehicle based on Soft Actor-Critic, Ocean Engineering. doi: 10.1016/j.oceaneng.2024.118171.

[28] L. Zhu and T. Li (2021), *Observer-based autopilot heading finite-time control design for intelligent ship with prescribed performance*, *Journal of Marine Science and Engineering*, Vol.9, No.8, p. 828. doi: 10.3390/jmse9080828.

[29] Z. Swider et al (2023), *Consistent design of PID controllers for an autopilot*, Polish Maritime Research. doi: 10.2478/pomr-2023-0008.

[30] S. Sivaraj and S. Rajendran (2022), *Heading control of a ship based on deep reinforcement learning*, in Proc. OCEANS 2022, Chennai, IEEE, pp.1-6. doi: 10.1109/OCEANSChennai45887.2022.9775236.

[31] X. Wang, H. Yi, J. Xu, and C. Xu (2024), *PID controller based on improved DDPG for trajectory tracking control of USV*, Journal of Marine Science and Engineering, Vol.12, No.10, p. 1771. doi: 10.3390/jmse12101771.

Ngày nhận bài:	25/02/2026
Ngày nhận bản sửa:	09/03/2026
Ngày duyệt đăng:	12/03/2026