

BỘ GIÁO DỤC VÀ ĐÀO TẠO  
ĐẠI HỌC BÁCH KHOA HÀ NỘI  
TRƯỜNG ĐIỆN - ĐIỆN TỬ



**ĐỒ ÁN TỐT NGHIỆP**

**THIẾT KẾ HỆ THỐNG ĐIỀU KHIỂN HỌC  
TĂNG CƯỜNG CHO XE TỰ HÀNH BA  
BÁNH MECANUM**

NGUYỄN THÀNH ĐẠT  
dat.nt181385@sis.hust.edu.vn  
HOÀNG VĂN LÂM  
lam.hv181566@sis.hust.edu.vn

Ngành KT Điều khiển Tự động hóa

Giảng viên hướng dẫn: PGS TS. Đào Phương Nam

Chữ kí của GVHD

Bộ môn:

Điều khiển tự động

Hà Nội, 11/2022

**BỘ GIÁO DỤC VÀ ĐÀO TẠO      CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM**  
**ĐẠI HỌC BÁCH KHOA HÀ NỘI                      Độc lập - Tự do - Hạnh phúc**

**NHIỆM VỤ  
ĐỒ ÁN TỐT NGHIỆP**

Họ và tên sinh viên: Nguyễn Thành Đạt

Khóa: K63

Trường: Điện - Điện tử

Ngành: KT ĐK và TĐH

*1. Tên đề tài:*

Thiết kế hệ thống điều khiển học tăng cường cho xe tự hành ba bánh mecanum

*2. Nội dung đề tài:*

Thiết kế hệ thống điều khiển học tăng cường cho xe tự hành ba bánh mecanum. Sử dụng thuật toán điều khiển tối ưu cho hệ phi tuyến thông qua cấu trúc Actor-Critic. Sử dụng kiến thức đã học để chứng minh tính ổn định của hệ thống. Mô phỏng hệ thống và kiểm tra đánh giá chất lượng của phương án điều khiển bằng công cụ Matlab.

Nguyễn Thành Đạt: Tính toán mô hình xe tự hành, thiết kế bộ điều khiển, chứng minh tính ổn định hệ thống, tham gia mô phỏng hệ thống trên Matlab.

Hoàng Văn Lâm: Tính toán mô hình xe tự hành, thiết kế bộ điều khiển, chứng minh tính ổn định hệ thống, tham gia mô phỏng hệ thống trên Matlab.

*3. Thời gian giao đề tài:*

*4. Thời gian hoàn thành đề tài:*

*Ngày 05 tháng 03 năm 2023*

**CÁN BỘ HƯỚNG DẪN**

# Lời cảm ơn

Lời đầu tiên chúng em xin gửi lời cảm ơn trân thành nhất tới tất cả những người đã ủng hộ và giúp đỡ chúng em trong quá trình thực hiện đồ án.

Trước hết chúng em xin trân thành cảm ơn thầy PGS.TS. Đào Phương Nam - khoa Tự động hóa, Trường Điện - Điện tử, người đã dìu dắt chúng em không chỉ trong quá trình làm đồ án mà cả quá trình học tập tại trường. Thầy luôn đưa ra những lời nhận xét, chỉ bảo tận tình, để chúng em có thể hoàn thành đồ án một cách tốt nhất.


Xin trân thành cảm ơn tới gia đình cũng như bạn bè đã luôn ở bên tiếp sức và tạo điều kiện thuận lợi để chúng em có thể học tập và nghiên cứu đồ án.

Dù đã cố gắng để thực hiện đồ án, nhưng chúng em biết mình không thể tránh khỏi những thiếu sót cho nên chúng em mong nhận được những lời nhận xét và góp ý từ các thầy cô và các bạn.

Chúng em xin trân thành cảm ơn!

Hà Nội, ngày 05 tháng 03 năm 2023

Sinh viên thực hiện



Nguyễn Thành Đạt

# Tóm tắt nội dung đồ án


Đồ án tốt nghiệp này trình bày thuật toán điều khiển bám tối ưu cho hệ robot xe tự hành ba bánh mecanum (Mecanum Wheeled Mobile Robots (MWMRs)) với sự tác động của nhiễu từ bên ngoài tới hệ thống bằng cách sử dụng cấu trúc online actor-critic. Mô hình hệ thống MWMRs được thiết lập dựa trên việc phân tích tính chất và cấu trúc của bánh mecanum.

Bài toán điều khiển tối ưu cho hệ phi tuyến này bị ràng buộc trực tiếp bởi nghiệm của phương trình Hamilton-Jacobi-Bellman (HJB). Phương trình HJB là phương trình vi phân phi tuyến khó có thể giải ra nghiệm giải tích. Vì vậy xuất hiện bài toán xấp xỉ nghiệm phương trình HJB. Để xấp xỉ được nghiệm của phương trình HJB thì phương pháp hữu dụng nhất là sử dụng phương pháp quy hoạch động thích nghi (Adaptive Dynamic Programming (ADP)) - đây là phương pháp được phát triển từ học tăng cường (Reinforcement Learning (RL)) dựa trên quy hoạch động (Dynamic Programming (DP)) của Bellman.

Đồ án chứng minh khả năng bám quỹ đạo và sự ổn định của hệ thống dựa trên việc sử dụng và phân tích hàm Lyapunov. Nhằm cho thấy khả năng ứng dụng của thuật toán hệ thống đã được tiến hành mô phỏng trên phần mềm MATLAB.

Hà Nội, ngày 05 tháng 03 năm 2023

Sinh viên thực hiện



Nguyễn Thành Đạt

# Một số kí hiệu viết tắt

$\ \cdot\ $	Chuẩn trong không gian Euclid
MWMRs	Mecanum Wheeled Mobile Robots
RL	Reinforcement Learning
MDP	Markov Decision Process
DP	Dynamic Program
ADP	Approximate/Adaptive Dynamic Programming
NN	Neural Network
AC	Actor-Critic
UB	Ultimately Bounded
PE	Persistence of Excitation Condition

# Danh sách hình vẽ

1.1	Sơ đồ cập nhật trọng số AC . . . . .	3
1.2	Cấu trúc online AC . . . . .	4
1.3	Cấu trúc offline AC . . . . .	4
2.1	Xe tự hành . . . . .	6
2.2	Xe tự hành MWMRs . . . . .	7
3.1	Đồ thị hàm ứng viên Lyapunov và họ các đường cong khép kín . . . . .	16
4.1	Hiệu suất bám của $x, y, \theta$ với quỹ đạo đặt $q_{r1}(t)$ và $\rho = 0.1$ . . . . .	21
4.2	Sai lệch bám quỹ đạo $q_1(t) - q_{r1}(t)$ khi $\rho = 0.1$ . . . . .	21
4.3	Trọng số NN cấu trúc AC với quỹ đạo đặt $q_{r1}(t)$ và $\rho = 0.1$ . . . . .	22
4.4	Tín hiệu điều khiển với quỹ đạo đặt $q_{r1}(t)$ và $\rho = 0.1$ . . . . .	22
4.5	Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt $q_{r1}(t)$ và $\rho = 0.1$ . . . . .	22
4.6	Hiệu suất bám của $x, y, \theta$ với quỹ đạo đặt $q_{r1}(t)$ và $\rho = -0.1$ . . . . .	23
4.7	Sai lệch bám quỹ đạo $q_1(t) - q_{r1}(t)$ khi $\rho = -0.1$ . . . . .	23
4.8	Trọng số NN cấu trúc AC với quỹ đạo đặt $q_{r1}(t)$ và $\rho = -0.1$ . . . . .	23
4.9	Tín hiệu điều khiển với quỹ đạo đặt $q_{r1}(t)$ và $\rho = -0.1$ . . . . .	24
4.10	Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt $q_{r1}(t)$ và $\rho = -0.1$ . . . . .	24
4.11	Hiệu suất bám của $x, y, \theta$ với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	25
4.12	Sai lệch bám quỹ đạo $q_2(t) - q_{r2}(t)$ khi $\rho = 0.1$ . . . . .	25
4.13	Trọng số NN cấu trúc AC với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	25
4.14	Tín hiệu điều khiển với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	26
4.15	Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	26
4.16	Hiệu suất bám của $x, y, \theta$ với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$ . . . . .	27
4.17	Sai lệch bám quỹ đạo $q_2(t) - q_{r2}(t)$ khi $\rho = -0.1$ . . . . .	27
4.18	Trọng số NN cấu trúc AC với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$ . . . . .	28
4.19	Tín hiệu điều khiển với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$ . . . . .	28
4.20	Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$ . . . . .	29
4.21	Hiệu suất bám của $x, y, \theta$ khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	30
4.22	Sai lệch bám quỹ đạo $q(t) - q_r(t)$ khi không có nhiễu thăm dò với $\rho = 0.1$ . . . . .	30
4.23	Trọng số NN cấu trúc AC khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	31
4.24	Tín hiệu điều khiển khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	31
4.25	Quỹ đạo chuyển động của hệ thống MWMRs khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = 0.1$ . . . . .	32

4.26	Hiệu suất bám của $x, y, \theta$ khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$ . . . . .	32
4.27	Sai lệch bám quỹ đạo $q_2(t) - q_{r2}(t)$ khi không có nhiễu thăm dò với $\rho = -0.1$ . .	33
4.28	Trọng số NN cấu trúc AC khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$ . . . . .	33
4.29	Tín hiệu điều khiển khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$	34
4.30	Quỹ đạo chuyển động của hệ thống MWMRs khi không có nhiễu thăm dò với quỹ đạo đặt $q_{r2}(t)$ và $\rho = -0.1$ . . . . .	34

# Danh sách bảng

1.1	Một số activation function thường gặp . . . . .	5
4.1	Tham số của MWMRs . . . . .	20



# Mục lục

<b>1</b>	<b>Tổng quan về học tăng cường và thuật toán ADP</b>	<b>1</b>
1.1	Giới thiệu chung về thuật toán . . . . .	1
1.2	Quy hoạch động Bellman . . . . .	1
1.2.1	Nguyên lý tối ưu Bellman . . . . .	1
1.2.2	Bài toán . . . . .	2
1.3	Thuật toán quy hoạch động thích nghi (ADP) . . . . .	3
1.4	Nguyên lý xấp xỉ hàm mạng NN . . . . .	4
1.5	Điều kiện hội tụ Persistence of Excitation (PE) . . . . .	5
<b>2</b>	<b>Mô hình xe tự hành ba bánh mecanum (MWMR)</b>	<b>6</b>
2.1	Giới thiệu chung về xe tự hành . . . . .	6
2.2	Mô hình xe tự hành ba bánh mecanum . . . . .	7
2.2.1	Mô tả mô hình . . . . .	7
2.2.2	Mô hình động học và động lực học của MWMR . . . . .	8
2.2.3	Xây dựng lại mô hình . . . . .	10
<b>3</b>	<b>Thiết kế bộ điều khiển học tăng cường dựa trên mô hình actor-critic cho MWMR</b>	<b>12</b>
3.1	Bộ điều khiển tối ưu và phương trình HJB . . . . .	12
3.2	Tính toán bộ điều khiển dựa trên thuật toán actor-critic . . . . .	13
3.3	Phân tích tính ổn định và khả năng bám quỹ đạo . . . . .	15
3.3.1	Tiêu chuẩn ổn định Lyapunov . . . . .	15
3.3.2	Phân tích ổn định hệ thống . . . . .	17
<b>4</b>	<b>Kết quả mô phỏng</b>	<b>20</b>
4.1	Kịch bản mô phỏng . . . . .	20
4.2	Kết quả mô phỏng với quỹ đạo nửa cung tròn . . . . .	21
4.3	Kết quả mô phỏng với quỹ đạo đường tròn . . . . .	24
4.4	Kết quả mô phỏng với việc không thêm nhiều thăm dò . . . . .	29
<b>5</b>	<b>Kết luận</b>	<b>36</b>

# Lời nói đầu

Ngày nay, với sự phát triển của công nghệ sự xuất hiện của robot trong đời sống và sản xuất đã trở nên rất phổ biến. Trong đó, robot di động có bánh xe (Wheeled Mobile Robots - WMRs) đã được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau như công nghiệp, y tế, khoa học,... Nhờ những ưu điểm về tính linh hoạt chuyển động tốt, độ phức tạp cơ học thấp. Là một lớp quan trọng của WMRs, Mecanum WMRs (MWMRs) đã và vẫn là chủ đề được nghiên cứu chuyên sâu trong thời gian qua, vì khả năng di động đa hướng giúp chúng có thể chuyển động thuận lợi trong môi trường tắc nghẽn.

MWMRs rất phù hợp cho những nhiệm vụ như cứu trợ tai nạn, khi hoàn cảnh xung quanh là hạn chế về mặt di chuyển. Tuy nhiên, để hoạt động trong những môi trường như vậy cần sự chuyển động với độ chính xác cao trong khi thông số về môi trường là không rõ ràng. Từ đó, yêu cầu về bộ điều khiển thích hợp được đặt ra. Với lý do trên, chúng em đề xuất đề tài "**Thiết kế hệ thống điều khiển học tăng cường cho xe tự hành ba bánh Mecanum**" với mong muốn đây sẽ là một hướng phát triển trong việc thiết kế bộ điều khiển cho phân lớp MWMRs.

Mặc dù đã cố gắng tìm hiểu và thực hiện đề tài, nhưng không thể tránh khỏi những sai sót, kính mong thầy cô và bạn bè nhận xét, cũng như góp ý giúp đề bài hoàn thiện và phát triển hơn.

Em xin trân thành cảm ơn!

# Chương 1

## Tổng quan về học tăng cường và thuật toán ADP

### 1.1 Giới thiệu chung về thuật toán

Học tăng cường (Reinforcement Learning (RL))<sup>[9]</sup> là một trong ba thuật toán học máy chính bên cạnh học giám sát (Supervised Learning) và học không giám sát (Unsupervised Learning). Học tăng cường là một nhánh quan trọng của học máy<sup>[12]</sup>, là phương pháp tập trung vào việc làm thế nào để cho một tác tử trong môi trường hành động sao cho có thể lấy được nhiều phần thưởng nhất có thể. Thuật toán giống như cách con người tự học và lớn lên, thực hiện các hành động ngẫu nhiên và quan sát phản hồi từ môi trường, từ đó rút ra kinh nghiệm cho hành động tiếp theo. Ban đầu, thuật toán học tăng cường<sup>[19]</sup> dựa trên quá trình quyết định Markov (Markov Decision Process (MDP)) để đưa ra phương án tốt nhất. Trong lĩnh vực điều khiển, thuật toán học tăng cường được phát triển dựa trên phương pháp quy hoạch động của Bellman<sup>[18]</sup>.

### 1.2 Quy hoạch động Bellman

#### 1.2.1 Nguyên lý tối ưu Bellman

**Nguyên lý tối ưu của Bellman:** "Mọi khúc cuối của quỹ đạo trạng thái tối ưu cũng sẽ là một quỹ đạo trạng thái tối ưu"<sup>[18]</sup>.

Kiểm tra tính đúng đắn của nguyên lý thông qua hình minh họa. Giả sử quỹ đạo liên nét đi từ  $x_0$  qua  $x_1$  đến  $x_2$  là quỹ đạo tối ưu, gồm hai đoạn  $\boxed{1}$  và  $\boxed{2}$ , tương ứng với  $V^* = V_1^* + V_2^*$ . Trong đó, phần quỹ đạo  $\boxed{2}$  đi từ  $x_1$  đến  $x_2$  có  $V_2^*$  không phải là tối ưu. Vậy suy ra, tồn tại đoạn cuối tối ưu đi từ  $x_1$  đến  $x_2$  là đoạn  $2'$  với  $V_2 < V_2^*$ . Suy ra, dọc theo dọc theo đoạn  $\boxed{1} - \boxed{2'}$ , hàm  $V = V_1^* + V_2^*$  sẽ có giá trị nhỏ hơn  $V^* = V_1^* + V_2^*$  tính theo đoạn 1 – 2. Điều này trái với giả thuyết ban đầu rằng đoạn  $\boxed{1} - \boxed{2}$  là tối ưu.

Phát biểu trên của nguyên lý tối ưu đúng với một đoạn bất kỳ của quỹ đạo trạng thái tối ưu chứ không chỉ riêng đoạn cuối, tuy nhiên trong phương pháp quy hoạch động ta chỉ cần xét đến đoạn cuối.

### 1.2.2 Bài toán

Cho hệ liên tục không dừng, bậc  $n$ :

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \quad (1.1)$$

với  $\mathbf{x} \in R^n$ ,  $\mathbf{u} \in R^m$  và các điều kiện ràng buộc sau:

- Tập  $U \subseteq R^m$  là một tập con (hở hoặc đóng) trong không gian điều khiển  $R^m$ .
- Khoảng thời gian  $T$  xảy ra quá trình tối ưu là cố định cho trước.
- Điểm đầu  $\mathbf{x}(0) = \mathbf{x}_0$  là tùy ý, nhưng phải cho trước.
- Điểm cuối  $\mathbf{x}(T) = \mathbf{x}_T$  là bất kỳ.

Xác định bộ điều khiển phản hồi trạng thái tối ưu  $\mathbf{u}^* = \mathbf{u}(\mathbf{x}, t) \in U$  đưa hệ đi từ  $\mathbf{x}(0)$  đến  $\mathbf{x}(T)$  trong khoảng thời gian  $T$  sao cho hàm chi phí  $V$  cho bởi

$$V = \int_0^T g(\mathbf{x}, \mathbf{u}, t) dt \rightarrow \min \quad (1.2)$$

đạt giá trị nhỏ nhất.

#### Nội dung phương pháp

Từ nội dung lý thuyết tối ưu<sup>[2]</sup>, ta định nghĩa hàm Bellman:

$$V(\mathbf{x}, t) = \min_{\mathbf{u} \in U} \int_t^T g(\mathbf{x}, \mathbf{u}, t) d\tau \quad (1.3)$$

hay hàm Bellman chính là phần giá trị tối ưu của (1.2) tính từ thời điểm  $t$  tới thời điểm cuối  $T$  của quá trình tối ưu, tức là giá trị hàm mục tiêu tính dọc theo đoạn cuối quỹ đạo tối ưu từ  $\mathbf{x}(t) = \mathbf{x}'$  tới  $\mathbf{x}_T$  bất kỳ.

Khi đó, theo nguyên lý tối ưu Bellman thì:

$$\begin{aligned} V(\mathbf{x}, t) &= \min_{\mathbf{u} \in U} \int_t^T g(\mathbf{x}, \mathbf{u}, t) d\tau \\ &= \min_{\mathbf{u} \in U} \left\{ \int_t^{t+\delta} g(\mathbf{x}, \mathbf{u}, t) d\tau + V(\mathbf{x}(t+\delta), t+\delta) \right\} \end{aligned} \quad (1.4)$$

**Định lý 1** Nếu  $\mathbf{u}^* = \mathbf{u}(\mathbf{x}, t) \in U$  là nghiệm của bài toán tối ưu (1.1) thì hàm Bellman (1.3) phải thỏa mãn:

$$\begin{cases} \frac{\partial V(\mathbf{x}, t)}{\partial t} + \frac{\partial V(\mathbf{x}, t)}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}, \mathbf{u}^*, t) + g(\mathbf{x}, \mathbf{u}^*, t) = 0 \\ V(\mathbf{x}_T, T) = 0 \end{cases} \quad (1.5)$$

$$\mathbf{u}^* = \arg \min_{\mathbf{u} \in U} \left\{ \frac{\partial V(\mathbf{x}, t)}{\partial \mathbf{x}} \mathbf{f}(\mathbf{x}, \mathbf{u}, t) + g(\mathbf{x}, \mathbf{u}, t) \right\} \quad (1.6)$$

Phương trình vi phân đạo hàm riêng (1.5) có tên gọi là phương trình Hamilton-Jacobi-Bellman, viết tắt là HJB.

### Thực hiện thuật toán:

**Bước 1:** Khởi tạo luật điều khiển chấp nhận được  $\underline{u}^{(0)}(\underline{x})$  và giá trị  $V^{(0)}(\underline{x}) = 0$ .

**Bước 2:** Xấp xỉ  $V^{(i+1)}(\underline{x})$  ở bước  $(i + 1)$  với tín hiệu điều khiển  $\underline{u}^{(i)}$ .

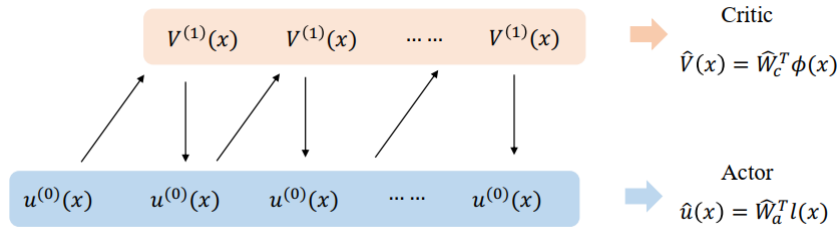
**Bước 3:** Cập nhật luật điều khiển cho vòng lặp tiếp theo

- Từ  $V^{(i+1)} \rightarrow \underline{u}^{(i+1)}$
- Nếu thỏa mãn tiêu chuẩn hội tụ  $\|V^{(i+1)} - V^{(i)}\| \leq \sigma$  với  $\sigma$  là số dương đủ nhỏ thì gán  $\underline{u}^* = \underline{u}^{(i+1)}$  và  $V^* = V^{(i+1)} \rightarrow$  Kết thúc thuật toán.
- Nếu không thỏa mãn, gán  $i \leftarrow i + 1$  và quay lại bước 2.

Nhưng việc tìm nghiệm của phương trình vi phân phi tuyến HJB là rất khó, vì vậy xuất hiện bài toán xấp xỉ nghiệm phương trình HJB và thuật toán quy hoạch động thích nghi (ADP) ra đời.

## 1.3 Thuật toán quy hoạch động thích nghi (ADP)

Thuật toán ADP ra đời nhằm giải xấp xỉ phương trình vi phân HJB, thông qua mạng nơ ron (Neural Network) xây dựng bộ điều khiển theo cấu trúc Actor - Critic (AC) [24] như trong hình 1.1.

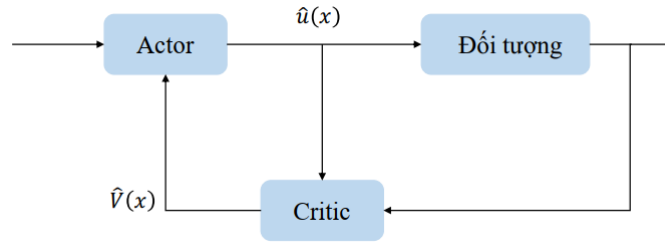


Hình 1.1: Sơ đồ cập nhật trọng số AC

Critic: đối với hệ phi tuyến  $V(\underline{x})$  không phải dạng toàn phương  $(\underline{x}^T P^{(i)} \underline{x})$  nhưng có thể xấp xỉ được  $\hat{V}(\underline{x}) = \hat{W}_c^T \phi(\underline{x})$  với  $\phi(\underline{x})$  là hàm kích hoạt (Activation Function).

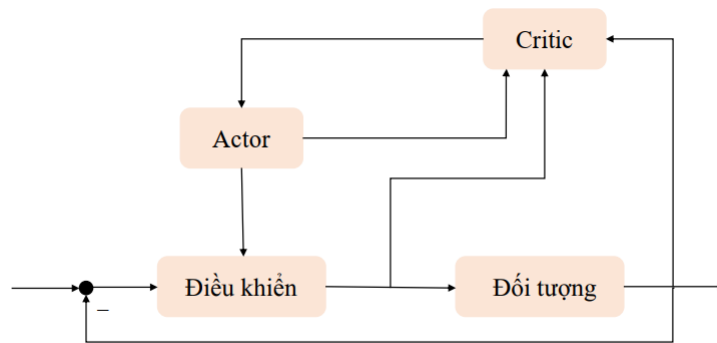
Actor: cũng có thể xấp xỉ được thành  $\hat{u} = \hat{W}_a^T l(\underline{x})$ .

Cấu trúc AC có thể thực hiện cập nhật tham số một cách tuần tự hoặc song song. Bộ điều khiển có thể xây dựng theo mô hình online AC [6] hoặc offline AC [5].



Hình 1.2: Cấu trúc online AC

- Online AC<sup>[27]</sup>: tín hiệu điều khiển từ Actor sẽ liên tục được cập nhật và đưa vào kích thích hệ thống cho đến khi tín hiệu điều khiển là tối ưu như ở hình 1.2.



Hình 1.3: Cấu trúc offline AC

- Offline AC: có một tín hiệu điều khiển khác kích thích hệ thống, song song với đó tín hiệu điều khiển từ Actor sẽ liên tục được cập nhật cho đến khi đạt tối ưu mới đưa vào kích thích hệ thống như trong hình 1.3.

Đồ án sẽ sử dụng cấu trúc online AC và cập nhật tham số song song.

## 1.4 Nguyên lý xấp xỉ hàm mạng NN

**Định lý 2 (Định lý xấp xỉ Weierstrass)** Mọi hàm liên tục định nghĩa trên khoảng đóng  $[a, b]$  có thể được xấp xỉ với độ chính xác mong muốn với hàm đa thức<sup>[8]</sup>.

Bởi vì hàm đa thức là một trong các hàm đơn giản nhất, và máy tính có thể đánh giá trực tiếp các hàm đa thức, định lý này có ý nghĩa trong cả thực tiễn và lý thuyết, đặc biệt trong nội suy đa thức.

**Định lý 3** Mạng NN có thể xấp xỉ giá trị của các hàm liên tục trên tập con của  $\mathbf{R}^n$  với độ chính xác tùy ý.

**Định lý 4** Gọi  $\varphi : \mathbf{R} \rightarrow \mathbf{R}$  là một hàm liên tục bất kỳ (còn gọi là activation function). Gọi  $K \subseteq \mathbf{R}^n$  là tập compact. Không gian các hàm liên tục trên tập  $K$  kí hiệu là  $C(K)$ . Gọi  $\mathbf{M}$  là

không gian các hàm có dạng:

$$F(X) = \sum_{i=1}^N v_i \varphi(\omega_i^T x + b_i) \quad (1.7)$$

cho tất cả số nguyên  $N \in \mathbf{N}$ , hằng số  $v_i, b_i \in \mathbf{R}$ , vector  $\omega_i \in \mathbf{R}^m$  với  $i = 1, \dots, N$  thì khi và chỉ khi  $\varphi$  không phải là đa thức, điều sau đây là đúng với bất kì  $\epsilon > 0$  và bất kì  $f \in C(K)$ , thì tồn tại  $F \in \mathbf{M}$  để:

$$|F(x) - f(x)| < \epsilon \quad (1.8)$$

với mọi  $x \in K$

Bảng 1.1: Một số activation function thường gặp

STT	Tên	Định nghĩa
1	Sigmoid	$f(x) = \frac{1}{1 + e^{-x}}$
2	Tanh	$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
3	ReLU	$f(x) = \begin{cases} 0 & (x \leq 0) \\ x & (x > 0) \end{cases}$
4	Gaussian	$f(x) = e^{-x^2}$
5	Sinc	$f(x) = \begin{cases} 1 & x = 0 \\ \frac{\sin(x)}{x} & x \neq 0 \end{cases}$

## 1.5 Điều kiện hội tụ Persistence of Excitation (PE)

Để đảm bảo sự hội tụ của  $\hat{W}$  đến  $W$  cần đảm bảo điều kiện Persistence of Excitation (PE).

Một hàm  $S$  liên tục từng phần, bị chặn toàn cục, đi từ  $[0, \infty) \rightarrow \mathbf{R}^m$  được cho là thoạt mãn điều kiện PE nếu tồn tại các hằng số dương  $\beta_1, \beta_2$  và  $T_0$  sao cho:

$$\beta_1 I \leq \int_{t_0}^{t_0+T_0} S(\tau) S(\tau)^T d\tau \leq \beta_2 I, \forall t_0 \geq 0 \quad (1.11)$$

Với  $I \in \mathbf{R}^{m \times m}$  là ma trận đơn vị. Theo định nghĩa trên, điều kiện PE yêu cầu rằng tích phân của ma trận bán xác định  $S(\tau) S(\tau)^T$  là xác định dương đều trên một khoảng thời gian  $T_0$ .

Nếu  $S$  thỏa mãn điều kiện PE trong khoảng thời gian  $[t_0, t_0 + T_0]$ , thì nó sẽ thỏa mãn điều kiện PE với mỗi khoảng có độ lớn bất kỳ  $T_1 > T_0$ .

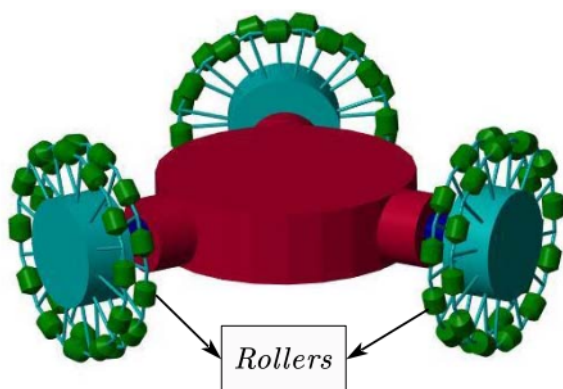
## Chương 2

# Mô hình xe tự hành ba bánh mecanum (MWMR)

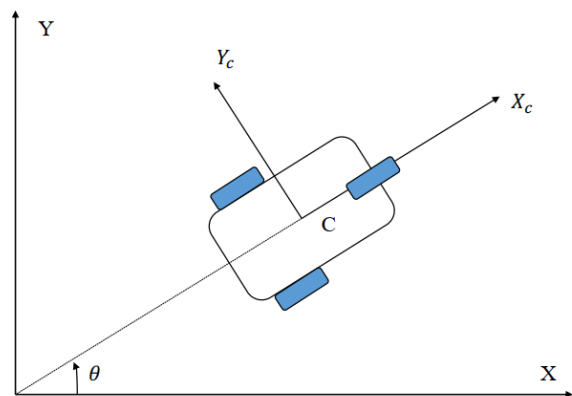
### 2.1 Giới thiệu chung về xe tự hành

Mobile robot là những thiết bị có khả năng tự di chuyển trong không gian, không bị cố định vào một vị trí thực nào. Để có thể di chuyển chính xác trong, robot cần phải định vị được vị trí của mình trong không gian và có đầy đủ thông tin về môi trường xung quanh. Do đó, trên robot tự hành thường sẽ được gắn thêm các hệ thống cảm biến nhằm thu thập những thông tin cần thiết từ đó thực hiện những phản ứng thích hợp. Về khả năng di chuyển của robot, có nhiều loại cơ cấu chấp hành như bánh xe, cánh quạt, chân, tay máy... tạo thành nhiều kiểu di động khác nhau<sup>[4]</sup>. Trong nhóm phân lớp xe tự hành (Wheeled Mobile Robots - WMRs), tức những mobile robot chuyển động nhờ bánh xe, phân loại theo khả năng điều khiển có hai loại xe tự hành holonomic mobile robot và non-holonomic mobile robot.

Holonomic là mối quan hệ giữa tổng số bậc tự do của robot và số bậc có thể điều khiển được. Nếu số bậc có thể điều khiển bằng với tổng số bậc tự do của robot thì ta gọi đó là holonomic robot. Nếu số bậc có thể điều khiển nhỏ hơn tổng số bậc tự do của robot vậy ta gọi đó là non-holonomic robot<sup>[3]</sup>.



(a) Holonomic mobile robot



(b) Non-holonomic mobile robot

Hình 2.1: Xe tự hành

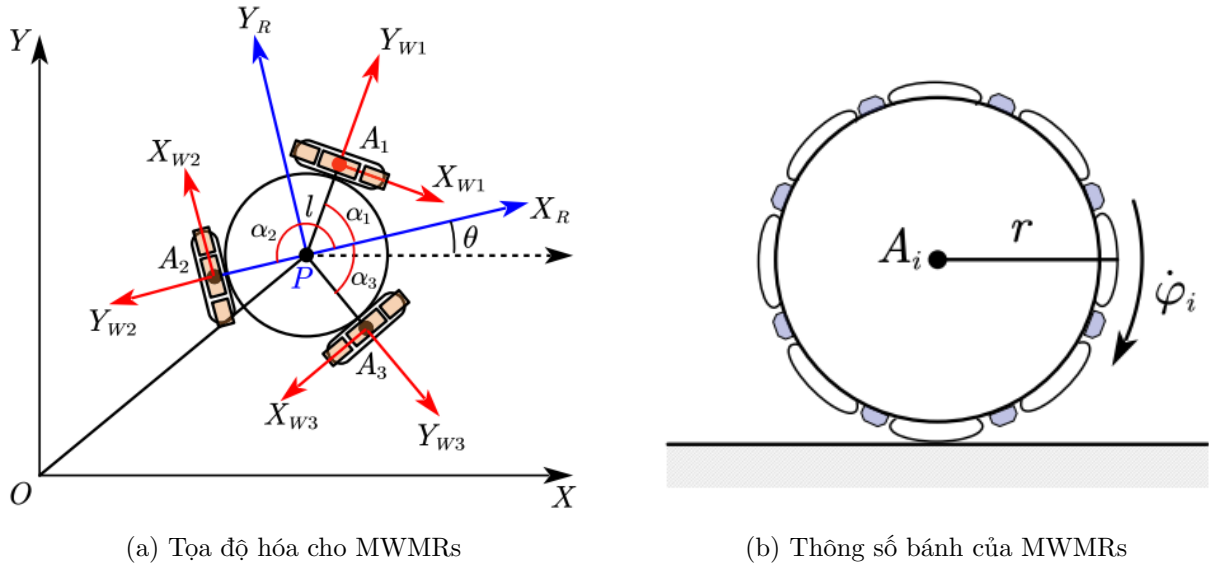


Holonomic mobile robot có khả năng di chuyển linh hoạt hơn non-holonomic mobile robot, sẽ phù hợp hơn khi cần di chuyển trong những không gian hẹp.

## 2.2 Mô hình xe tự hành ba bánh mecanum

### 2.2.1 Mô tả mô hình

Trong đồ án này, chúng ta xem xét xe tự hành có đủ cơ cấu chấp hành omni mobile robot được trang bị ba bánh mecanum được đặt vuông góc với trục gắn với bánh như ở hình 2.2.



Hình 2.2: Xe tự hành MWMRs

Để nghiên cứu chuyển động của robot ba hệ tọa độ của robot được định nghĩa như hình 2.2a. Hệ tọa độ gắn với trái đất  $O - X, Y$  được coi là hệ quy chiếu quán tính cho chuyển động của robot. Hệ tọa độ gắn với thân robot  $P - X_R, Y_R$  được gắn với tâm của vòng tròn robot, trong đó  $X_R$  nằm dọc theo trục quay của bánh  $A_2$ ,  $Y_R$  vuông góc với  $X_R$ .  $A_i - X_{Wi}, Y_{Wi}$  với  $(i = 1, 2, 3)$  là hệ tọa độ gắn với tâm trục của bánh xe thứ  $i$ ,  $X_{Wi}$  nằm dọc theo hướng truyền động của bánh xe thứ  $i$ . Thông số  $l$  xác định khoảng cách của  $PA_i$ ,  $\alpha_i$  là góc giữa  $PA_i$  và  $X_R$ ,  $\beta_i$  là góc giữa  $PA_i$  và  $Y_{Wi}$ ,  $\theta$  (góc giữa  $X_R$  và  $X$ ) là hướng của trọng tâm P của robot trong hệ quy chiếu quán tính,  $r$  là bán kính của bánh xe<sup>[25] [15] [1]</sup>.

Đặt  $\mathbf{q} = (x, y, \theta)^T$  là vị trí  $(x, y)$  của trọng tâm P của robot trong hệ quy chiếu quán tính,  $\dot{\mathbf{q}} = (\dot{x}, \dot{y}, \dot{\theta})^T$  là vận tốc của trọng tâm P trong hệ quy chiếu quán tính,  $\mathbf{v} = (v_x, v_y, \omega)^T$  là vận tốc của trọng P trong hệ quy chiếu gắn với tâm robot.  $\Phi = (\varphi_1, \varphi_2, \varphi_3)^T$  là vị trí góc của tất cả các bánh xe.

Mục đích chính của đồ án này là phát triển bộ điều khiển bám quỹ đạo cho xe tự hành ba bánh mecanum (MWMR) dựa trên thuật toán học tăng cường (RL)<sup>[4]</sup> vì vậy MWMR có thể bám theo quỹ đạo thay đổi theo thời gian  $q_r(t)$ .

Phương trình truyền động bị ảnh hưởng bởi nhiều của MWMR được xây dựng dựa trên mô hình động học cũng như động lực học của xe tự hành ba bánh Mecanum<sup>[22], [26], [14]</sup>.

## 2.2.2 Mô hình động học và động lực học của MWMR

Mối liên hệ giữa vận tốc của  $P$  trong hệ quy chiếu gắn với tâm và hệ quy chiếu quán tính được tính theo công thức dưới đây.

$$\dot{\mathbf{q}} = \mathbf{R}(\theta)v \quad (2.1)$$

Trong đó  $R(\theta)$  là ma trận chuyển từ hệ quy chiếu gắn với tâm sang hệ quy chiếu quán tính

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Tham khảo bài toán động học cho bánh xe Mecanum trong bài<sup>[22]</sup>, trong trường hợp không có trượt thì vận tốc tuyến tính của bánh xe là  $v_{li} = r\dot{\varphi}_i$ . Mối quan hệ giữa vận tốc góc  $\dot{\varphi}_i$  của mỗi bánh xe và vận tốc của robot  $v$  có thể biểu diễn bằng công thức

$$(\sin(\alpha_i + \beta_i), -\cos(\alpha_i + \beta_i), -l \cos(\beta_i))v = v_{li} \quad (2.2)$$

Trong thực tế thì việc bánh xe bị trượt là không thể tránh khỏi có thể do bánh xe bị biến dạng, bánh xe bị lão hóa hay là do mặt đất trơn trượt. Cho nên chúng ta xem xét đến độ trượt của bánh xe, mặt trượt  $\rho$  được xác định dựa trên công thức  $\rho = ((r\dot{\varphi}_i - v_{li})/(r\dot{\varphi}_i))$ , với  $-1 < \rho < 1$ <sup>[13][28]</sup>. Khi đó vận tốc trượt theo hướng của vòng quay bánh xe là  $\rho r\dot{\varphi}_i = r\dot{\varphi}_i - v_{li}$  và (2.2) trở thành

$$(\sin(\alpha_i + \beta_i), -\cos(\alpha_i + \beta_i), -l \cos(\beta_i))v = (1 - \rho)r\dot{\varphi}_i \quad (2.3)$$

Theo hệ quy chiếu gắn với tâm  $P - X_R, Y_R$  ở hình 2 và các thông số  $\alpha_i = (((2i - 1)\pi)/3)$ ,  $\beta_i = 0$ ,  $i = 1, 2, 3$ , từ đó (2.3) trở thành

$$\begin{pmatrix} \sin(\frac{\pi}{3}) & -\cos(\frac{\pi}{3}) & -l \\ \sin \pi & -\cos \pi & -l \\ \sin(-\frac{\pi}{3}) & -\cos(-\frac{\pi}{3}) & -l \end{pmatrix} v = (1 - \rho)r \begin{pmatrix} \dot{\varphi}_1 \\ \dot{\varphi}_2 \\ \dot{\varphi}_3 \end{pmatrix} = (1 - \rho)r\dot{\Phi} \quad (2.4)$$

Kết hợp (2.1) và (2.4) ta có được phương trình động học của MWMR như sau:

$$\dot{\mathbf{q}} = (1 - \rho)r\mathbf{H}(\theta)\dot{\Phi} \quad (2.5)$$

Đặt

$$A = \begin{pmatrix} \sin(\frac{\pi}{3}) & -\cos(\frac{\pi}{3}) & -l \\ \sin \pi & -\cos \pi & -l \\ \sin(-\frac{\pi}{3}) & -\cos(-\frac{\pi}{3}) & -l \end{pmatrix}$$

Kết hợp (1), (4) và (5) ta có:

$$\mathbf{H}(\theta) = \mathbf{R}(\theta)A^{-1} \quad (2.6)$$

Suy ra:

$$\mathbf{H}^{-1}(\theta) = A\mathbf{R}^{-1}(\theta) \quad (2.7)$$

Trong đó,

$$\mathbf{R}^{-1}(\theta) = \begin{pmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Suy ra:

$$\mathbf{H}^{-1}(\theta) \begin{pmatrix} \sin(\theta + \frac{\pi}{3}) & -\cos(\theta + \frac{\pi}{3}) & -l \\ -\sin \theta & \cos \theta & -l \\ \sin(\theta - \frac{\pi}{3}) & -\cos(\theta - \frac{\pi}{3}) & -l \end{pmatrix}$$

Áp dụng phương trình Lagrange<sup>[21][7]</sup>, phương trình động lực học của MWMR được viết như sau

$$M\ddot{\Phi} + C(\Phi, \dot{\Phi})\dot{\Phi} + \mu\dot{\Phi} = \tau \quad (2.8)$$

Trong đó,

$$M = \begin{pmatrix} m1 & m2 & m2 \\ m2 & m1 & m2 \\ m2 & m2 & m1 \end{pmatrix}$$

$$, C(\Phi, \dot{\Phi}) = 0, m_1 = (4mr^2/9) + (I_\omega r^2/9l^2) + I_\varphi, m_2 = (I_\omega r^2/9l^2) - (2mr^2/9).$$

**Tính chất 1** Ma trận quán tính  $M$  là đối xứng và tồn tại số dương  $\delta_1 > 0$  và  $\delta_2 > 0$ , sao cho  $0 < \delta_1 I_{3 \times 3} \leq M \leq \delta_2 I_{3 \times 3}$ .

Đạo hàm cả hai vế theo thời gian của phương trình (2.5) ta được

$$\ddot{\mathbf{q}} = \dot{\theta}\dot{\mathbf{H}}(\theta)\mathbf{H}^{-1}(\theta)\dot{\mathbf{q}} + (1 - \rho)r\mathbf{H}(\theta)\ddot{\Phi} \quad (2.9)$$

Từ công thức (2.5) và (2.9) ta có:

$$\begin{cases} \dot{\Phi} = \frac{1}{(1 - \rho)r}\mathbf{H}^{-1}(\theta)\dot{\mathbf{q}} \\ \ddot{\Phi} = \frac{1}{(1 - \rho)r}\mathbf{H}^{-1}(\theta)\ddot{\mathbf{q}} - \frac{1}{(1 - \rho)r}\dot{\theta}\dot{\mathbf{H}}^{-1}(\theta)\dot{\mathbf{q}} \end{cases} \quad (2.10)$$

Trong đó,

$$\dot{\mathbf{H}}^{-1} = \begin{pmatrix} \cos(\theta + \frac{\pi}{3}) & \sin(\theta + \frac{\pi}{3}) & 0 \\ -\cos \theta & -\sin \theta & 0 \\ \cos(\theta - \frac{\pi}{3}) & \sin(\theta - \frac{\pi}{3}) & 0 \end{pmatrix}$$

Từ (2.8) và (2.10) ta có:

$$\ddot{\mathbf{q}} = -\rho r\mathbf{H}(\theta)M^{-1}\tau + (\dot{\theta}\mathbf{H}(\theta)\dot{\mathbf{H}}^{-1}(\theta) - \mu\mathbf{H}(\theta)M^{-1}\mathbf{H}^{-1}(\theta))\dot{\mathbf{q}} + r\mathbf{H}(\theta)M^{-1}\tau \quad (2.11)$$

Suy ra ta có hệ phương trình động lực học của MWMR:

$$\begin{aligned}\dot{\mathbf{q}} &= v_q \\ \dot{v}_q &= \bar{f}(\mathbf{q})v_q + \bar{g}(\mathbf{q})\tau + \bar{g}(\mathbf{q})d(\mathbf{q}, v_q)\end{aligned}\quad (2.12)$$

Trong đó,

$$\begin{aligned}\bar{f}(\mathbf{q}) &= \dot{\theta}\mathbf{H}(\theta)\dot{\mathbf{H}}^{-1}(\theta) - \mu\mathbf{H}(\theta)M^{-1}\mathbf{H}^{-1}(\theta) \\ \bar{g}(\mathbf{q}) &= r\mathbf{H}(\theta)M^{-1} \\ \bar{f}_d(\mathbf{q}, v_q) &= -\rho r\mathbf{H}(\theta)M^{-1}\tau \\ d(\mathbf{q}, v_q) &= \bar{g}^{-1}(\mathbf{q})\bar{f}_d(\mathbf{q}, v_q)\end{aligned}$$

Mô men xoắn chưa biết  $d(\mathbf{q}, v_q)$  gây ra bởi sự trượt động lực học của MWMR, được coi là như nhiễu bên ngoài<sup>[23]</sup> tác động tới MWMR và được giả định như sau:

**Giả thiết 1** Nhiễu chưa biết  $d(\mathbf{q}, v_q)$  được giới hạn bởi một hàm đã biết  $d_M(\mathbf{q})$  với  $\|d(\mathbf{q}, v_q)\| \leq d_M(\mathbf{q})$ .

### 2.2.3 Xây dựng lại mô hình

Mục đích của đồ án này là xử lý vấn đề bám quỹ đạo cho hệ thống (2.12) sao cho  $q(t)$  có thể bám theo một quỹ đạo khả vi liên tục  $q_r(t)$ .

Vì vậy, đặt  $x = (\mathbf{q}^T, v_q^T)^T \in \mathbf{R}^6$ ,  $f(x) = (v_q^T, (\bar{f}_d(\mathbf{q}, v_q))^T)^T$ ,  $g(x) = (0_{3 \times 3}, \bar{g}(\mathbf{q})^T)^T$ , từ đó (2.12) có thể viết lại như sau:

$$\dot{x} = f(x) + g(x)\tau + g(x)d(\mathbf{q}, v_q)$$

Theo như mô hình trên cùng với việc tham khảo các mô hình đã biết trong các bài báo, chúng ta có thể xác định một số tính chất của mô hình trên tương tự như mô hình trong bài<sup>[20]</sup>.

**Tính chất 2**  $f(x)$  và  $g(x)$  là các ma trận Lipschitz,  $\|f(x)\| \leq b_f\|x\|$  với  $b_f$  là hằng số,  $g(x)$  bị giới hạn bởi hằng số  $b_g$  tức là  $\|g(x)\| \leq b_g$ .

**Tính chất 3** Ma trận  $g(x)$  có hạng cột đầy đủ với mọi  $x \in \mathbf{R}^6$ , và hàm  $g^+ = (g^T g)^{-1} g^T : \mathbf{R}^6 \Rightarrow \mathbf{R}^{3 \times 6}$  là ma trận giả đảo của ma trận  $g(x)$  và được giới hạn bởi một ma trận Lipschitz<sup>[17]</sup>.

Để đạt được mục tiêu bám quỹ đạo, chúng em xem xét quỹ đạo đặt của hệ thống được đặt như sau:

$$\dot{x}_r = h_r(x_r) \quad (2.13)$$

Trong đó,  $x_r = (\mathbf{q}_r^T, v_{qr}^T)^T$  và  $\dot{\mathbf{q}}_r = \dot{v}_r$ ,  $h_r(x_r) : \mathbf{R}^6 \Rightarrow \mathbf{R}^6$  là một hàm liên tục chuyển ma trận Lipschitz. Bằng việc xác định sai lệch bám  $e = x - x_r$ , tốc độ bám quỹ đạo được mô tả như sau:

$$\dot{e} = f(e + x_r) - h_r(x_r) + g(e + x_r)\tau + g(e + x_r)d(\mathbf{q}, v_q) \quad (2.14)$$

Giả sử như không có nhiễu tác động vào hệ thống thì chúng ta có thể xác định được tín hiệu điều khiển trạng thái ổn định  $\tau_r$  tương ứng với quỹ đạo mong muốn  $x_r$  như sau:

$$\tau_r(x_r) = g^+(x_r)(h_r(x_r) - f(x_r)) \quad (2.15)$$

Trong đó,  $g^+(x_r) = (g^T(x_r)g(x_r))^{-1}g^T(x_r)$ .

Đặt  $z = (e^T, x_r^T)^T$  là biến trạng thái mới, từ đó ta có mô hình của hệ thống được mô tả theo  $z$  như sau:

$$\dot{z} = F(z) + G(z)u + \xi \quad (2.16)$$

Trong đó,

$$F(z) = \begin{pmatrix} f(e + x_r) - h_r(x_r) + g(e + x_r)\tau_r \\ h_r(x_r) \end{pmatrix}$$

$$G(z) = \begin{pmatrix} g(e + x_r) \\ 0_{6 \times 3} \end{pmatrix}, u = \tau - \tau_r, \xi = G(z)d(\mathbf{q}, v_q).$$

Ta có thể thấy được tính chất 2 và 3 chỉ ra rằng  $F$  là ma trận Lipschitz và ma trận  $G$  bị chặn. Giả thiết 1 và định nghĩa của  $\xi$  cũng chỉ ra rằng  $\xi$  vẫn bị chặn trên vì  $\|\xi\| = \|G(z)d(\mathbf{q}, v_q)\| \leq d_M(e + x_r)\|g(e + x_r)\| = \lambda_M(z)$ .

Bằng việc chuyển đổi mô hình hệ thống như ở trên, thì bài toán điều khiển bám quỹ đạo của hệ thống (2.12) có thể chuyển thành bài toán tối ưu điều khiển cho hệ thống (2.16) và vì  $\xi$  bị chặn và rất nhỏ cho nên hệ thống (2.16) có thể được viết lại như sau:

$$\dot{z} = F(z) + G(z)u \quad (2.17)$$

Trong phần tiếp theo của đồ án, mục tiêu sẽ là thiết kế một bộ điều khiển tối ưu  $u^*(z)$  cho hệ thống (2.17) để tạo ra các tín hiệu vòng kín của hệ thống (2.16) nằm ở trạng thái ổn định.

## Chương 3

# Thiết kế bộ điều khiển học tăng cường dựa trên mô hình actor-critic cho MWMR

### 3.1 Bộ điều khiển tối ưu và phương trình HJB

Về vấn đề điều khiển tối ưu cho hệ thống (2.17), chúng ta xem xét hàm chi phí như sau:

$$V(z(t)) = \int_t^\infty e^{-\gamma(s-t)} U(z(s), u(s)) ds \quad (3.1)$$

Trong đó,  $\gamma > 0$ ,  $U(z, u) = \lambda_M^2(z) + z^T \bar{Q}z + u^T Ru$ ,  $R \in \mathbf{R}^{3 \times 3}$  là một ma trận hằng số xác định dương, và  $\bar{Q}$  là ma trận được xác định như sau:

$$\bar{Q} = \begin{pmatrix} Q & 0_{6 \times 6} \\ 0_{6 \times 6} & 0_{6 \times 6} \end{pmatrix}$$

Trong đó,  $Q \in \mathbf{R}^{6 \times 6}$  là một ma trận xác định dương.

Trong bài đồ án này, tín hiệu điều khiển mong muốn là  $\tau_r$  và  $\tau_d$  như trong bài<sup>[16]</sup>, cho nên hàm chi phí được xác định trong các bài<sup>[20]</sup> và<sup>[29]</sup> có thể tiến đến vô cực. Do đó, chúng ta sẽ thêm vào hàm  $e$  mũ để tránh việc hàm chi phí tiến tới vô cực.

Chúng ta có phương trình vi phân  $V(z(t))$  như sau:

$$\lambda_M^2(z) + z^T \bar{Q}z + u^T Ru - \gamma V(z) + \nabla V^T(z)(F(z) + G(z)u) = 0$$

Trong đó,  $\nabla V(z) = ((\partial V(z))/\partial z)$ . Phương trình Hamiltonian được cho bởi:

$$H(z, u, \nabla V) = \lambda_M^2(z) + z^T \bar{Q}z + u^T(z)Ru(z) - \gamma V(z) + \nabla V^T(z)(F(z) + G(z)u(z)) \quad (3.2)$$

Từ đó ta có thể xác định được hàm chi phí tối ưu như sau:

$$V^*(z) = \min_{u \in \pi(\Omega)} \int_t^\infty e^{-\gamma(s-t)} (\lambda_M^2(z) + z^T \bar{Q}z + u^T Ru) \quad (3.3)$$

Trong đó  $\pi(\Omega)$  biểu thị một tập luật điều khiển chấp nhận được trên tập  $\Omega \subset \mathbf{R}^{12}$ . Khi đó hàm chi phí tối ưu  $V^*(z)$  thỏa mãn điều kiện của phương trình HJB sau:

$$H^*(z, u^*, \nabla V^*) = \lambda_M^2(z) + z^T \bar{Q}z + u^{*T}(z)Ru^*(z) - \gamma V^*(z) + \nabla V^{*T}(z)(F(z) + G(z)u^*(z)) = 0 \quad (3.4)$$

Từ đó, tín hiệu điều khiển tối ưu được xác định:

$$u^*(z) = \arg \min_{u \in \pi(\Omega)} [H(z, u, \nabla V^*(z))] = -\frac{1}{2}R^{-1}G^T(z) \nabla V^*(z) \quad (3.5)$$

Thay tín hiệu điều khiển tối ưu (3.5) vào (3.4), phương trình HJB (3.4) trở thành:

$$H^*(z, \nabla V^*) = \lambda_M^2(z) + z^T \bar{Q}z - \gamma V^*(z) - \frac{1}{4} \nabla V^{*T}(z)G(z)R^{-1}G^T(z) \nabla V^*(z) + \nabla V^{*T}(z)F(z) = 0 \quad (3.6)$$

Để có được tín hiệu điều khiển tối ưu (3.5), người ta cần giải bài toán cực trị trong phương trình HJB (3.5) là cực khó hoặc có thể nói là không thể. Trong phần tiếp theo, chúng ta giới thiệu thuật toán actor-critic áp dụng vào để giải phương trình này.

## 3.2 Tính toán bộ điều khiển dựa trên thuật toán actor-critic

Trong phần này, chúng ta sử dụng thuật toán AC kết hợp với mạng NN trong thời gian thực<sup>[10]</sup> để thực hiện việc tìm nghiệm gần đúng của phương trình HJB.

Dựa trên định lý xấp xỉ bậc cao Weierstrass<sup>[8]</sup>, hàm chi phí  $V^*(z)$  có thể được xấp xỉ bằng cách sử dụng mạng neuron một lớp như sau:

$$V^*(z) = W^T \phi(z) + \varepsilon_v(z) \quad (3.7)$$

Trong đó  $W \in \mathbf{R}^N$  là một vector trọng số lý tưởng không đổi được giới hạn bởi một vector hằng dương đã biết  $\bar{W}$  sao cho  $\|W\| \leq \bar{W}$ ,  $\phi(z) \in \mathbf{R}^N$  là một vector hàm kích hoạt phù hợp,  $N$  là số neuron,  $\varepsilon_v(z)$  là sai số xấp xỉ. Từ đó chúng ta xác định được gradient của  $V^*(z)$  là

$$\nabla V^*(z) = \nabla \phi^T(z)W + \nabla \varepsilon_v(z) \quad (3.8)$$

Ta biết rằng, trên tập  $\Omega$ , sai số xấp xỉ  $\varepsilon_v$  và gradient  $\nabla \varepsilon_v$  của nó bị chặn sao cho  $\|\varepsilon_v\| \leq b_\varepsilon$  và  $\|\nabla \varepsilon_v\| \leq b_{\varepsilon z}$  [28]. Đối với hàm kích hoạt  $\phi(z)$ , chúng ta có giả thiết sau:

**Giả thiết 2** Hàm kích hoạt  $\phi(z)$  và gradient của nó bị giới hạn  $\|\phi(z)\| \leq b_\phi$ ,  $\|\nabla \phi(z)\| \leq b_{\phi z}$ .  
Thay (3.7) và (3.8) vào (3.6), phương trình Hamiltonian tối ưu sẽ trở thành

$$H^*(z, W) = \lambda_M^2(z) + z^T \bar{Q}z + W^T \nabla \phi F - \gamma W^T \phi - \frac{1}{4} W^T D_1 W + \varepsilon_H = 0 \quad (3.9)$$

Trong đó,  $D_1 = \nabla \phi G R^{-1} G^T \nabla \phi^T$ ,  $D_2 = G R^{-1} G^T$  và  $\varepsilon_H = \nabla \varepsilon_v^T F - \frac{1}{2} \nabla \varepsilon_v^T G R^{-1} G^T \nabla \phi^T W - \frac{1}{4} \nabla \varepsilon_v^T D_2 \nabla \varepsilon_v - \gamma \varepsilon_v(z)$

Tồn tại một giới hạn không đổi  $\varepsilon_h$  sao cho  $\|\varepsilon_H\| \leq \varepsilon_h$  là số đơn vị lớp ẩn  $N$  tăng lên<sup>[27]</sup>. Tín hiệu điều khiển tối ưu (3.5) trở thành:

$$u^*(z) = -\frac{1}{2}R^{-1}G^T(z)(\nabla \phi^T(z)W + \nabla \varepsilon_v(z)) \quad (3.10)$$

Dựa trên công thức (3.7) và (3.10), critic NN xấp xỉ cho hàm chi phí tối ưu và actor NN xấp xỉ cho chính sách tối ưu được đưa ra như sau:

$$\hat{V}(z, \hat{W}_c) = \hat{W}_c^T \phi(z) \quad (3.11)$$

$$\hat{u}(z, \hat{W}_a) = -\frac{1}{2} R^{-1} G^T(z) \nabla \phi^T(z) \hat{W}_a \quad (3.12)$$

Trong đó,  $\hat{W}_c$  và  $\hat{W}_a$  là ước lượng các trọng số lý tưởng  $W$ . Từ (2.13), bộ điều khiển bám quỹ đạo  $\tau$  của MWMR có thể thu được dưới dạng:

$$\tau = -\frac{1}{2} R^{-1} G^T(z) \nabla \phi^T(z) \hat{W}_a + g^+(x_r)(h_r(x_r) - f(x_r))$$

Lấy các giá trị xấp xỉ  $\hat{u}$  và  $\hat{V}$  vào (3.2), ta có thể suy ra được giá trị gần đúng của hàm Hamiltonian như sau:

$$\hat{H}(z, \hat{W}_c, \hat{W}_a) = \lambda_M^2(z) + z^T \bar{Q} z + \frac{1}{4} \hat{W}_a^T D_1 \hat{W}_a - \gamma \hat{W}_c^T \phi + \hat{W}_c^T \nabla \phi (F - \frac{1}{2} G R^{-1} G^T \nabla \phi^T \hat{W}_a) \quad (3.13)$$

Xác định sai lệch Bellman là  $\delta = \hat{H} - H^*$ , từ (3.9) và (3.13), ta có:

$$\delta(z, \hat{W}_c, \hat{W}_a) = \lambda_M^2(z) + z^T \bar{Q} z + \frac{1}{4} \hat{W}_a^T D_1 \hat{W}_a + \hat{W}_c^T (\nabla \phi (F + G \hat{u}) - \gamma \phi) \quad (3.14)$$

Để giải quyết vấn đề tối ưu điều khiển, chúng ta sử dụng phương pháp bình phương tối thiểu được thiết kế để đào tạo cho mạng neural critic để giảm thiểu sai lệch tích phân  $E_c = \int_0^t \delta^2(s) ds$  như sau:

$$\begin{aligned} \dot{\hat{W}}_c &= -\eta_c \Gamma(t) \frac{\sigma(t)}{m_\sigma^2(t)} \delta \\ \dot{\Gamma}(t) &= -\eta_c \left( -\beta \Gamma(t) + \Gamma(t) \frac{\sigma(t) \sigma^T(t)}{m_\sigma^2(t)} \Gamma(t) \right) \end{aligned} \quad (3.15)$$

Trong đó,  $\sigma = \nabla \phi (F + G \hat{u}) - \gamma \phi$ ,  $m_\sigma = (1 + v \sigma^T \Gamma \sigma^{1/2}) / \eta_c$ ,  $v$  là các hằng số dương,  $\beta \in (0, 1)$ ,  $\Gamma(t)$  là ma trận khuếch đại ước lượng.

Xác định sai lệch ước lượng của trọng số critic là  $\tilde{W}_c = W - \hat{W}_c$  và sai lệch ước lượng của trọng số actor là  $\tilde{W}_a = W - \hat{W}_a$ , từ đó sai lệch Bellman có thể được viết lại như sau:

$$\delta = -\tilde{W}_c^T \sigma + \frac{1}{4} \tilde{W}_a^T D_1 \tilde{W}_a - \varepsilon_H \quad (3.16)$$

Đặt  $\bar{\sigma} = (\sigma / m_\sigma)$ , và sử dụng (3.16) trong (3.15), sai lệch động của trọng số critic có thể được thể hiện như sau:

$$\dot{\tilde{W}}_c = \dot{W} - \dot{\hat{W}}_c = -\eta_c \Gamma \bar{\sigma} \bar{\sigma}^T \tilde{W}_c + \eta_c \Gamma \frac{\bar{\sigma}}{m_\sigma} \left( \frac{1}{4} \tilde{W}_a^T D_1 \tilde{W}_a - \varepsilon_H \right) \quad (3.17)$$

Giả thiết rằng điều kiện PE được áp dụng cho  $\bar{\sigma}(t)$  để đảm bảo rằng  $\hat{W}_c$  sẽ tiến tới  $W$ .

**Giả thiết 3** Tồn tại  $T > 0, \beta_1 > 0, \beta_2 > 0$  sao cho với mọi  $t$

$$\beta_1 \mathbf{I} \leq \int_t^{t+T} \bar{\sigma}(s) \bar{\sigma}^T(s) ds \leq \beta_2 \mathbf{I}$$



Từ giả thiết 2 và bài<sup>[20]</sup>, tồn tại  $\alpha_2 > \alpha_1 > 0$  sao cho

$$\alpha_1 \mathbf{I} \leq \Gamma(t) \leq \alpha_2 \mathbf{I} \quad \forall t \in [0, \infty) \quad (3.18)$$

Từ đó, dựa trên (3.18), có thể kết luận rằng:

$$\|\bar{\sigma}(t)\| \leq \frac{1}{\sqrt{v\alpha_1}} \quad \forall t \in [0, \infty) \quad (3.19)$$

Luật điều chỉnh cho trọng số neural của actor được phát triển như sau:

$$\dot{\hat{W}}_a = -\eta_{a1}(\hat{W}_a - \hat{W}_c) - \eta_{a2}\hat{W}_a + \frac{\eta_a}{4}D_1\hat{W}_a\frac{\bar{\sigma}^T}{m_\sigma}\hat{W}_c \quad (3.20)$$

Trong đó  $\eta_{a1} > 0$ ,  $\eta_{a2} > 0$  và  $\eta_a > 0$  là các tham số điều chỉnh.

Vì sai lệch Bellman là phi tuyến đối với chính sách ước tính trọng số, phương pháp bình phương nhỏ nhất không thể được sử dụng để cập nhật trọng số<sup>[20]</sup>. Xác định sai lệch  $\tilde{u} = \hat{u}(z, \hat{W}_a) - \hat{u}_c(z, \hat{W}_c) = -(1/2)R^{-1}G^T(z)\nabla\phi^T(z)(\hat{W}_a - \hat{W}_c)$  và hàm mục tiêu  $E_u = (1/2)\tilde{u}^T R \tilde{u}$ , để giảm thiểu hàm mục tiêu, chúng ta có thể sử dụng thuật toán gradient-descent để điều chỉnh trọng số NN của actor<sup>[16]</sup>,  $\dot{\hat{W}}_a = -\eta_{a1}(E_u/\hat{W}_a) = -\eta_{a1}D_1(\hat{W}_a - \hat{W}_c)$ . Vì vậy, (3.20) có thể xem như là một luật sửa đổi điều chỉnh cho trọng số của actor và số hạng thứ hai của (3.20) là để cải thiện độ bền vững cho hệ thống, còn số hạng thứ ba của (3.20) được thiết kế để đảm bảo sự ổn định của vòng kín của hệ thống theo hàm Lyapunov.

### 3.3 Phân tích tính ổn định và khả năng bám quỹ đạo

#### 3.3.1 Tiêu chuẩn ổn định Lyapunov

Xét hệ dừng, tự trị:

$$\dot{\underline{x}} = f(\underline{x}) \quad (3.21)$$

Hệ (3.21) được gọi là ổn định Lyapunov nếu  $\forall \varepsilon$  bất kỳ luôn tồn tại miền  $\delta(\varepsilon)$  sao cho quỹ đạo trạng thái tự do  $\underline{x}(t)$  là nghiệm của (3.21) tương ứng với điều kiện đầu vào  $\underline{x}(0)$ ,  $\underline{x}(0) = x_0$  thỏa mãn  $\|\underline{x}_0\| < \delta(\varepsilon)$  thì  $\|\underline{x}(t)\| < \varepsilon$ ,  $\forall t > 0$ .  $\Rightarrow$  (3.21) là ổn định tiệm cận Lyapunov nếu có

$$\lim_{t \rightarrow \infty} \underline{x}(t) = \underline{0}$$

Để biết (3.21) có ổn định hay không mà không cần giải tường minh nghiệm  $\underline{x}(t)$ , Lyapunov đã đưa ra ý tưởng.

#### Định lý Lyapunov

Xây dựng một họ đường cong khép kín bao quanh gốc tọa độ bằng cách lấy hình chiếu của các đường đồng mức một hàm nhiều biến  $V(\underline{x})$  thỏa mãn:

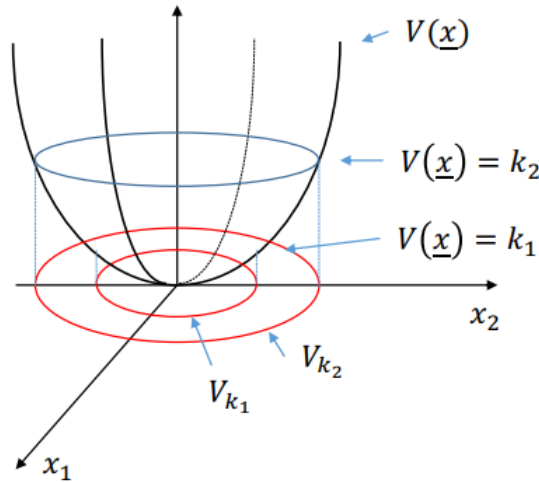
1:  $V(\underline{x})$  là hàm xác định dương

- $V(\underline{x}) > 0$ ,  $\forall \underline{x} \neq 0$

- $V(\underline{0}) = 0$

**2:**  $V(\underline{x})$  là hàm đơn điệu tăng theo  $|\underline{x}|$  và có  $\lim_{|\underline{x}| \rightarrow \infty} V(\underline{x}) = \infty$

**Ví dụ:** Xét hệ bậc 2:  $\underline{x} = [x_1 \ x_2]^T$   
Chọn  $V(\underline{x}) = ax_1^2 + bx_2^2$  với  $a, b > 0$



Hình 3.1: Đồ thị hàm ứng viên Lyapunov và họ các đường cong khép kín

Ta có,

$$\begin{aligned} \text{Grad}V(\underline{x}) &= \left( \frac{\partial V(\underline{x})}{\partial \underline{x}} \right)^T \\ &= \left( \frac{\partial V(\underline{x})}{\partial x_1}, \frac{\partial V(\underline{x})}{\partial x_2} \right)^T \end{aligned}$$

$\Rightarrow \text{Grad}V(\underline{x})$  là vecto luôn vuông góc với đường đồng mức, và có hướng chỉ chiều tăng của  $V(\underline{x})$ .

Để xét (3.21) có ổn định Lyapunov hay không thì chỉ cần xét xem quỹ đạo trạng thái có cắt tất cả các đường đồng mức  $V_k$  theo hướng từ trong ra ngoài hay không.

- Nếu  $\underline{x}(t)$  cắt tất cả các  $V_k$  theo hướng từ ngoài vào trong thì hệ (3.21) ổn định tiệm cận.
- Nếu  $\underline{x}(t)$  không cắt  $V_k$  nào theo hướng từ trong ra ngoài, suy ra  $\underline{x}(t)$  chỉ cắt  $V_k$  từ ngoài vào trong hoặc đi song song với  $V_k$  hoặc trùng với  $V_k$ .

$\Rightarrow \underline{x}(t)$  luôn nằm trong một lân cận nào đó của gốc tọa độ, ta có hệ (3.21) ổn định. Xét:

$$\begin{aligned} -W(\underline{x}) &= \frac{dV(\underline{x})}{dt} \\ &= \frac{\partial V(\underline{x})}{\partial \underline{x}} \cdot \frac{d\underline{x}}{dt} \\ &= \text{Grad}V(\underline{x}) \cdot \frac{d\underline{x}}{dt} \\ &= \left| \frac{\partial V(\underline{x})}{\partial \underline{x}} \right| \cdot |\dot{\underline{x}}| \cdot \cos(\varphi) \end{aligned}$$

nếu  $\varphi > 90^\circ \Rightarrow \cos\varphi < 0$  Khi đó,  $W(\underline{x}) = -\frac{dV(\underline{x})}{dt} > 0$  đồng nghĩa với việc quỹ đạo trạng thái  $\underline{x}(t)$  cắt các đường đồng mức  $V_k$  theo hướng từ ngoài vào trong.

Hệ (3.21) ổn định tiệm cận.

Nếu chỉ có  $W(\underline{x}) = -\frac{dV(\underline{x})}{dt} \geq 0$  thì hệ (3.21) ổn định.

### 3.3.2 Phân tích ổn định hệ thống

**Định lý 5** Xem xét hệ thống (2.17), trọng số NN của critic và tín hiệu điều khiển được cho bởi (3.11) và (3.12). Luật điều chỉnh trọng số cho critic và actor NNs được cung cấp bởi (3.15) và (3.20). Các giả thiết 1-3 đều được giữ lại và  $\bar{\sigma}$  trong (3.17) thỏa mãn điều kiện PE. Sau đó, trạng thái của hệ thống vòng kín, sai lệch trọng số NN critic  $\tilde{W}_c$  và sai lệch trọng số NN actor  $\tilde{W}_a$  đều bị chặn theo điều kiện UB<sup>[11]</sup>. Hơn thế nữa sai số ước lượng trọng số  $\tilde{W}_c$  hội tụ theo cấp số nhân.

**Chứng minh:** Xem xét hàm ứng viên Lyapunov như sau:

$$V(t) = V^*(t) + V_1(t) + V_2(t) = V^*(t) + \frac{1}{2\eta_c} \tilde{W}_c^T(t) \Gamma^{-1}(t) \tilde{W}_c(t) + \frac{1}{2\eta_a} \tilde{W}_a^T(t) \tilde{W}_a(t) \quad (3.22)$$

Trong đó  $V^*(t)$  là hàm chi phí tối ưu và có đạo hàm theo thời gian như sau:

$$\begin{aligned} \dot{V}^*(t) &= \nabla V^{*T}(F + G\hat{u}) = (\nabla \phi^T W + \nabla \varepsilon_v)^T \left( F - \frac{1}{2} GR^{-1} G^T \nabla \phi^T \hat{W}_a \right) \\ &= W^T \nabla - \frac{1}{2} W^T D_1 \hat{W}_a + \nabla \varepsilon_v^T \left( F - \frac{1}{2} GR^{-1} G^T \nabla \phi^T \hat{W}_a \right) \end{aligned}$$

Từ phương trình HJB (3.9), ta có:

$$\dot{V}^*(t) = -\lambda_M^2(z) - z^T \bar{Q}z + \gamma W^T \phi + \frac{1}{4} W^T D_1 W - \frac{1}{2} W^T D_1 \hat{W}_a - \varepsilon_H + \varepsilon_1 \quad (3.23)$$

Trong đó,  $\varepsilon_1 = \dot{\varepsilon}_v = \nabla \varepsilon_v^T (F - (1/2) GR^{-1} G^T \nabla \phi^T \hat{W}_a)$ , Từ (2)  $\varepsilon_1 - \varepsilon_H = \frac{1}{2} \tilde{W}_a^T \nabla \phi D_2 \nabla \varepsilon_v + \frac{1}{4} \nabla \varepsilon_v^T D_2 \nabla \varepsilon_v + \gamma \varepsilon_v(z)$ .

Từ đó, (3.23) có thể được viết lại như sau:

$$\begin{aligned} \dot{V}^*(t) &= -\lambda_M^2(z) - z^T \bar{Q}z + \gamma W^T \phi + \frac{1}{4} W^T D_1 W + \frac{1}{2} \tilde{W}_a^T D_1 W \\ &\quad + \frac{1}{2} \tilde{W}_a^T \nabla \phi D_2 \nabla \varepsilon_v + \frac{1}{4} \nabla \varepsilon_v^T D_2 \nabla \varepsilon_v + \gamma \varepsilon_v(z) \end{aligned}$$

Thay (3.15) và (3.17) vào đạo hàm theo thời gian của  $V_1(t)$  ta có:

$$\begin{aligned} \dot{V}_1(t) &= \frac{1}{\eta_c} \tilde{W}_c^T \Gamma^{-1} \dot{\tilde{W}}_c + \frac{1}{2\eta_c} \tilde{W}_c^T \dot{\Gamma}^{-1} \tilde{W}_c \\ &= \tilde{W}_c^T \Gamma^{-1} \left( -\Gamma \bar{\sigma} \bar{\sigma}^T \tilde{W}_c + \Gamma \frac{\bar{\sigma}}{m} \left( \frac{1}{4} \tilde{W}_a^T D_1 \tilde{W}_a - \varepsilon_H \right) \right) - \frac{1}{2} \tilde{W}_c^T \Gamma^{-1} (\beta \Gamma - \Gamma \bar{\sigma} \bar{\sigma}^T \Gamma) \Gamma^{-1} \tilde{W}_c \quad (3.24) \\ &= -\frac{1}{2} \tilde{W}_c^T \bar{\sigma} \bar{\sigma}^T \tilde{W}_c - \frac{\beta}{2} \tilde{W}_c^T \Gamma^{-1} \tilde{W}_c + \tilde{W}_c^T \frac{\bar{\sigma}}{m} \left( \frac{1}{4} \tilde{W}_a^T D_1 \tilde{W}_a - \varepsilon_H \right) \end{aligned}$$

Với  $V_2(t)$ , sử dụng (3.20), và  $\hat{W}_c = W - \tilde{W}_c$  và  $\hat{W}_a = W - \tilde{W}_a$ , ta có:

$$\begin{aligned} \dot{V}_2(t) = & -\frac{1}{\eta_a} \tilde{W}_a^T \dot{\tilde{W}}_a = -\frac{(\eta_{a1} + \eta_{a2})}{\eta_a} \tilde{W}_a^T \tilde{W}_a + \frac{\eta_{a1}}{\eta_a} \tilde{W}_a^T \tilde{W}_c + \frac{\eta_{a2}}{\eta_a} \tilde{W}_a^T W \\ & - \frac{1}{4} \tilde{W}_a^T D_1 \hat{W}_a \frac{\bar{\sigma}^T}{m_\sigma} \hat{W}_c \end{aligned} \quad (3.25)$$

Thay (3.23), (3.24), (3.25) vào đạo hàm theo thời gian của  $V(t)$  ta có:

$$\begin{aligned} \dot{V}(t) = & -\lambda_M^2(z) - e^T Q e - \frac{1}{4} W^T D_1 W - \frac{\tilde{W}_c^T \bar{\sigma}}{m_\sigma} \varepsilon_H - \tilde{W}_c^T A_1 \tilde{W}_c - \tilde{W}_a^T A_2 \tilde{W}_a + \tilde{W}_a^T A_3 \tilde{W}_c \\ & + \tilde{W}_a^T B_1 + B_2 \end{aligned} \quad (3.26)$$

Trong đó,

$$\begin{aligned} A_1 &= \frac{1}{2} \bar{\sigma} \bar{\sigma}^T + \frac{\beta}{2} \Gamma^{-1}, A_2 = \frac{(\eta_{a1} + \eta_{a2})}{\eta_a} I - \frac{1}{4} D_1 \frac{\bar{\sigma}^T}{m_\sigma} W \\ A_3 &= \frac{1}{4} D_1 W \frac{\bar{\sigma}^T}{m_\sigma} + \frac{\eta_{a1}}{\eta_a} I \\ B_1 &= \frac{1}{2} D_1 W + \frac{1}{2} \nabla \phi D_2 \nabla \varepsilon_v - \frac{1}{4} D_1 W \frac{\bar{\sigma}^T}{m_\sigma} W + \frac{\eta_{a2}}{\eta_a} W \\ B_2 &= \gamma W^T \phi + \frac{1}{4} \nabla \varepsilon_v^T D_2 \nabla \varepsilon_v + \gamma \varepsilon_v(z) \end{aligned}$$

Dựa trên các tính chất 2, 3 và giả thiết 2 cùng với giới hạn của  $\varepsilon_v$  và  $\nabla \varepsilon_v$ , tồn tại các hằng số dương  $\kappa_1, \kappa_2, \kappa_3$  sao cho  $\|B_1\| \leq \kappa_1, \|B_2\| \leq \kappa_2, \|(1/4)D_1(\bar{\sigma}^T/m_\sigma)W\| \leq \kappa_3$ . Áp dụng bất đẳng thức Young, (3.26) trở thành:

$$\begin{aligned} \dot{V}(t) \leq & -\underline{q} \|e\|^2 - \frac{\beta}{4\alpha_2} \|\tilde{W}_c\|^2 - \frac{\eta_{a1} + \eta_{a2}}{2\eta_a} \|\tilde{W}_a\|^2 \\ & - \left( \frac{\beta}{4\alpha_2} - \frac{1}{2\epsilon} \left( \frac{\eta_{a1}}{\eta_a} + \kappa_3 \right) \right) \|\tilde{W}_c\|^2 \\ & - \left( \frac{\eta_{a1} + \eta_{a2}}{2\eta_a} - \kappa_3 - \frac{\epsilon}{2} \left( \frac{\eta_{a1}}{\eta_a} + \kappa_3 \right) \right) \\ & \times \|\tilde{W}_a\|^2 + \frac{\varepsilon_h}{\sqrt{v\alpha_1}} \|\tilde{W}_c\| + \kappa_1 \|\tilde{W}_a\| + \kappa_2 \end{aligned} \quad (3.27)$$

Trong đó,  $\underline{q} = \lambda_{\min}(Q)$ . Chọn các thông số sao cho  $(\beta/4\alpha_2) \geq (1/2\epsilon)((\eta_{a1}/\eta_a) + \kappa_3)$  và  $((\eta_{a1} + \eta_{a2}/2\eta_a) \geq \kappa_3 + (\epsilon/2)((\eta_{a1}/\eta_a) + \kappa_3)$  và đặt  $\zeta = (e^T, \tilde{W}_c^T, \tilde{W}_a^T)^T, \bar{\omega}_1 = \min(\underline{q}, (\beta/4\alpha_2), ((\eta_{a1} + \eta_{a2}/2\eta_a)))$  và  $\bar{\omega}_2 = \max((\varepsilon_h/\sqrt{v\alpha_1}), \kappa_1)$ , (3.27) trở thành:

$$\dot{V}(t) \leq -\bar{\omega}_1 \|\zeta\|^2 + \bar{\omega}_2 \|\zeta\| + \kappa_2$$

Do đó, nếu bất đẳng thức  $\|\zeta\| > (\bar{\omega}_2/2\bar{\omega}_1 + ((\bar{\omega}_2^2/4\bar{\omega}_1^2) + (\kappa_2/\bar{\omega}_1))^{1/2})$  giữ nguyên thì ta có thể suy ra rằng đạo hàm Lyapunov  $\dot{V}(t)$  âm. Sử dụng lý thuyết ổn định Lyapunov ta có thể thấy

rằng trạng thái của hệ thống vòng kín và sai lệch ước tính trọng số  $\tilde{W}_c, \tilde{W}_a$  bị chặn.

Áp dụng (3.5) và (3.12) ta được:

$$u^* - \hat{u} = -\frac{1}{2}R^{-1}G^T(\nabla\phi^T(z)\tilde{W}_a + \nabla\varepsilon_v)$$

Áp dụng định lý 5 và giả thiết 2, ta có:

$$\|u^* - \hat{u}\| \leq \frac{1}{2\lambda_{\min}(R)}b_g(b_{\phi z}b_\zeta + b_{\varepsilon z}) = b_u \quad (3.28)$$

Trong đó  $b_u$  là một hằng số dương.

**Định lý 6** Đối với hệ thống (2.17) với hiệu suất hàm (3.1), luật điều khiển gần đúng tối ưu (3.12) đảm bảo theo dõi sai lệch động lực ở dạng đóng của hệ thống (2.16) cuối cùng bị chặn với điều kiện là  $\lambda_M^2(z) \geq \tau_d^T R \tau_d$ .

**Chứng minh:** Chọn hàm chi phí tối ưu  $V^*(z)$  là hàm ứng viên Lyapunov. Từ (3.5), ta có  $\nabla V^{*T}G = -2u^{*T}R$ . Từ (3.4) ta lại có:

$$\nabla V^{*T}(F + Gu^*) = -\lambda_M^2(z) - z^T \bar{Q}z - u^{*T}Ru^* + \gamma V^*$$

Từ đó, đạo hàm theo thời gian của  $V^*(z)$  dọc theo hệ thống (2.16) với luật điều khiển tối ưu xấp xỉ  $\hat{u}$ , ta có:

$$\begin{aligned} \dot{V}^*(z) &= \nabla V^{*T}(F + G\hat{u} + Gd(q, v_q)) \\ &= \nabla V^{*T}(F + Gu^*) + \nabla V^{*T}G(\hat{u} - u^*) - 2u^{*T}Rd(q, v_q) \\ &= -(\lambda_M^2(z) - d^T(q, v_q)Rd(q, v_q)) + \nabla V^{*T}G(\hat{u} - u^*) \\ &\quad - z^T \bar{Q}z - (u^* + d(q, v_q))^T R(u^* + d(q, v_q)) + \gamma V^* \end{aligned}$$

Bằng việc xem xét (3.7), (3.8), (3.28), và giả thiết 2, ta có:

$$\begin{aligned} \|\nabla V^{*T}G(\hat{u} - u^*)\| &\leq (b_{\phi z}\bar{W} + b_{\varepsilon z})b_g b_u \\ \|\gamma V^*\| &\leq \gamma(\bar{W}b_\phi + b_\varepsilon) \end{aligned}$$

Kết hợp những bất phương trình trên cùng với điều kiện  $\lambda_M^2(z) \geq d^T(q, v_q)Rd(q, v_q)$ , ta có thể thấy rằng

$$\hat{V}^*(z) \leq -\underline{q}\|e\|^2 + \lambda_u$$

Trong đó  $\lambda_u = (b_\phi\bar{W} + b_g b_u + \gamma(\bar{W}b_\phi + b_\varepsilon))$ . Chúng ta có thể kết luận rằng  $\dot{V}^* < 0$  nếu  $\|e\| > (\lambda_u/\underline{q})^{1/2}$ . Vì vậy sai lệch bám động lực của hệ thống không chắc chắn (2.16) cuối cùng bị chặn đồng nhất, điều này cho thấy thêm rằng trạng thái  $q(t)$  có thể theo dõi quy đạo mong muốn  $q_r(t)$ .

# Chương 4

## Kết quả mô phỏng

### 4.1 Kịch bản mô phỏng

Bảng 4.1: Tham số của MWMRs

Tham số	Giá trị
Khối lượng (kg)	$m = 10$
Chiều dài (m)	$l = 0.5, r = 0.05$
Mô men quán tính quay ( $kg \cdot m^2$ )	$I_\theta = 5, I_\varphi = 0.1$
Hệ số ma sát ( $N \cdot m/rad/s$ )	$\mu = 1$

Trong phần này, kết quả mô phỏng được trình bày để đánh giá hiệu suất của thuật toán theo những kịch bản khác nhau. Với các điều kiện ban đầu được đưa ra như sau:

$$x(0) = (1.8 \quad 1.6 \quad 1.4 \quad 0 \quad 0 \quad 0)^T$$

$$\hat{W}_c(0) = 200 \times \mathbf{1}_{51 \times 1}, \hat{W}_a(0) = 6 \times \mathbf{1}_{51 \times 1}, \Gamma(0) = 10 \times \mathbf{I}_{51}.$$

Hàm kích hoạt được chọn như sau:

$$\phi(z) = (1/2)[z_1^2 \ z_2^2 \ z_3^2 \ z_1 z_4 \ z_1 z_5 \ z_1 z_6 \ z_2 z_4 \ z_2 z_5 \ z_2 z_6 \ z_3 z_4 \ z_3 z_5 \ z_3 z_6 \ z_1^2 z_2^2 \ z_1^2 z_3^2 \ z_2^2 z_3^2 \ z_1^2 z_7^2 \ z_1^2 z_8^2 \ z_1^2 z_9^2 \ z_1^2 z_{10}^2 \ z_1^2 z_{11}^2 \ z_1^2 z_{12}^2 \ z_2^2 z_7^2 \ z_2^2 z_8^2 \ z_2^2 z_9^2 \ z_2^2 z_{10}^2 \ z_2^2 z_{11}^2 \ z_2^2 z_{12}^2 \ z_3^2 z_7^2 \ z_3^2 z_8^2 \ z_3^2 z_9^2 \ z_3^2 z_{10}^2 \ z_3^2 z_{11}^2 \ z_3^2 z_{12}^2 \ z_4^2 z_7^2 \ z_4^2 z_8^2 \ z_4^2 z_9^2 \ z_4^2 z_{10}^2 \ z_4^2 z_{11}^2 \ z_4^2 z_{12}^2 \ z_5^2 z_7^2 \ z_5^2 z_8^2 \ z_5^2 z_9^2 \ z_5^2 z_{10}^2 \ z_5^2 z_{11}^2 \ z_5^2 z_{12}^2 \ z_6^2 z_7^2 \ z_6^2 z_8^2 \ z_6^2 z_9^2 \ z_6^2 z_{10}^2 \ z_6^2 z_{11}^2 \ z_6^2 z_{12}^2]^T$$

Các tham số của xe tự hành ba bánh mecanum (MWMRs) được thể hiện trong bảng 4.1.

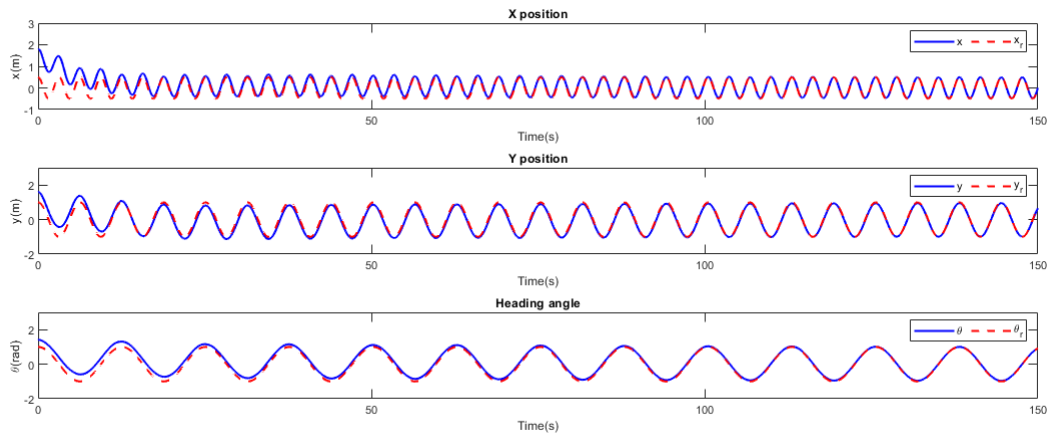
Xem xét nhiễu bên ngoài do trượt gây ra, quỹ đạo đặt của hệ thống, các mô phỏng được thực hiện với các giá trị khác nhau của  $\rho = \pm 0.1$  và các quỹ đạo đặt khác nhau  $q_{r1}(t) = [0.5 \cos(2t) \cos(t) \cos(0.5t)]^T$  và  $q_{r2}(t) = [\sin(t) \cos(t) \cos(0.5t)]^T$ . Với  $\rho = 0.1$ , chúng ta chọn  $Q = \mathbf{I}_6$  và  $R = \mathbf{I}_3$ , và các tham số điều khiển được chọn như sau  $\eta_c = 2, \eta_{a1} = 2, \eta_{a2} = 0.001, \eta_a = 0.001, \beta = 0.001, \gamma = 0.0001, \nu = 0.002$ . Với  $\rho = -0.1$ , các tham số điều khiển là  $\eta_c = 2, \eta_{a1} = 1.2, \eta_{a2} = 0.001, \beta = 0.001, \gamma = 0.0001, \nu = 0.002$ .

Một nhiễu thăm dò  $N(t) = \sin^2(t) \cos(t) + \sin^2(2t) \cos(0.1t) + \sin^2(-1.2t) \cos(0.5t) + \sin^5(t) + \sin^2(1.2t) + \cos(2.4t) \sin^3(2.4t)$  được thêm vào đầu vào điều khiển để đảm bảo thỏa mãn điều kiện PE và nhiễu thăm dò ảnh hưởng đến các trạng thái hệ thống và trọng số NN.

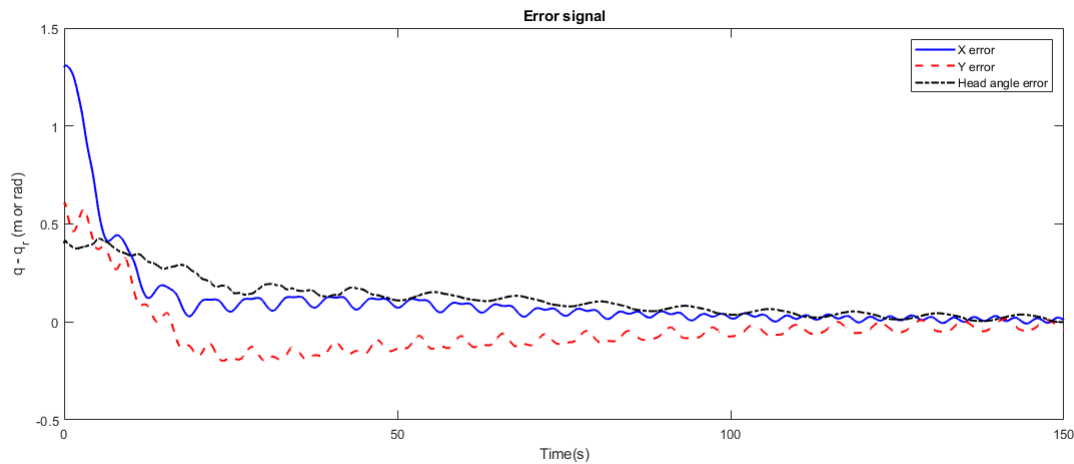
Nhiều tác thăm dò bị tắt ở 50 giây và quá trình điều khiển kéo dài 150s. Sau 50 giây, chúng ta có thể thấy rằng sự hội tụ của trọng số NN đã xảy ra và các trạng thái của hệ thống gần với quỹ đạo đặt. Điều này cho thấy điều kiện PE được đảm bảo hiệu quả bằng việc thêm nhiều thăm dò. Khi hệ thống đã hoạt động ổn định thì điều kiện PE đã không còn cần thiết phải xem xét nữa nên nhiều thăm dò đã bị tắt ở 50 giây. Sau đó trạng thái của hệ thống tiếp tục tiếp cận quỹ đạo đặt.

## 4.2 Kết quả mô phỏng với quỹ đạo nửa cung tròn

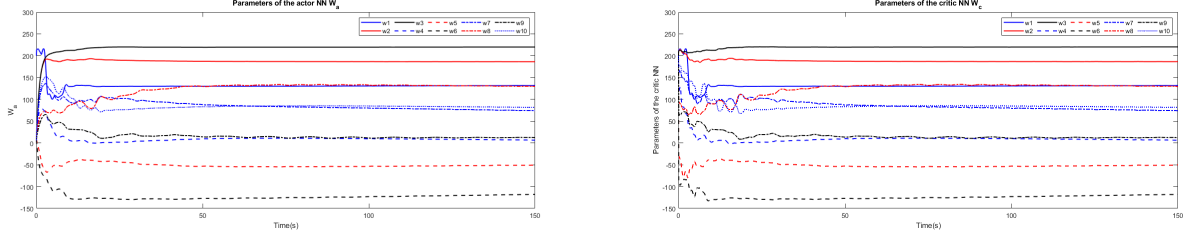
Với quỹ đạo đặt là  $q_{r1}(t) = [0.5 \cos(2t) \cos(t) \cos(0.5t)]^T$  dưới tác động nhiễu với  $\rho = 0.1$  kết quả mô phỏng được thể hiện trong các hình 4.1 - 4.5.



Hình 4.1: Hiệu suất bám của  $x, y, \theta$  với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = 0.1$

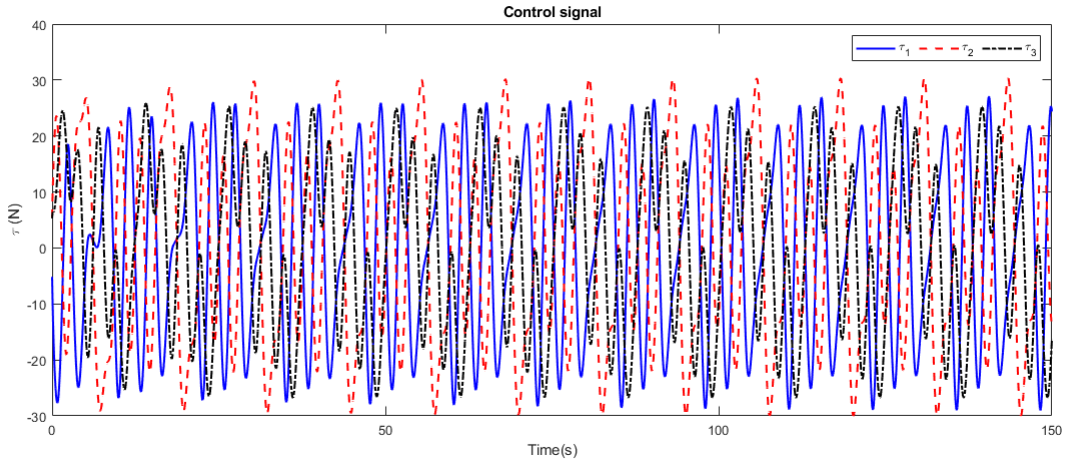


Hình 4.2: Sai lệch bám quỹ đạo  $q_1(t) - q_{r1}(t)$  khi  $\rho = 0.1$

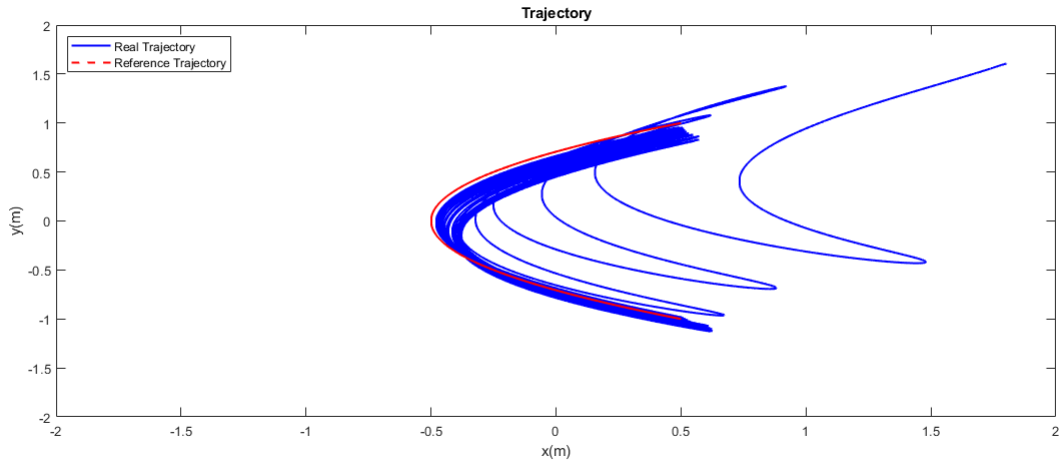


(a) Trọng số NN của actor với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = 0.1$  và (b) Trọng số NN của critic với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = 0.1$

Hình 4.3: Trọng số NN cấu trúc AC với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = 0.1$



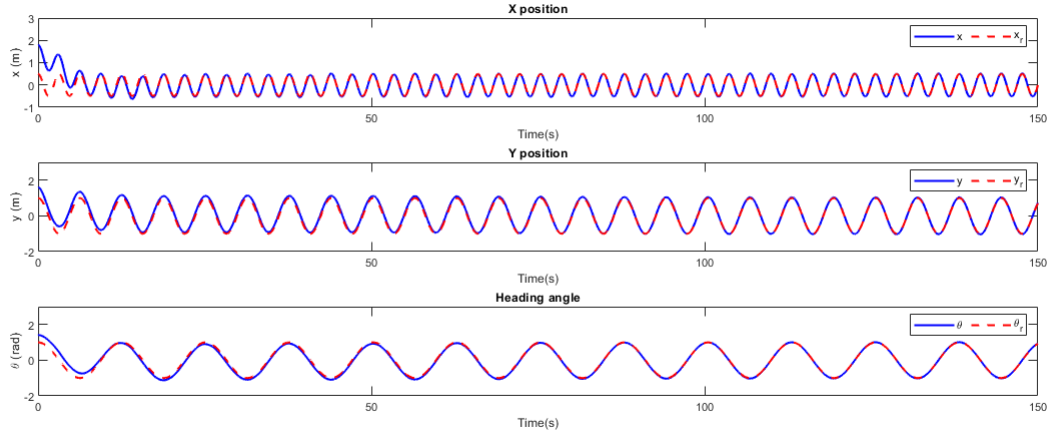
Hình 4.4: Tín hiệu điều khiển với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = 0.1$



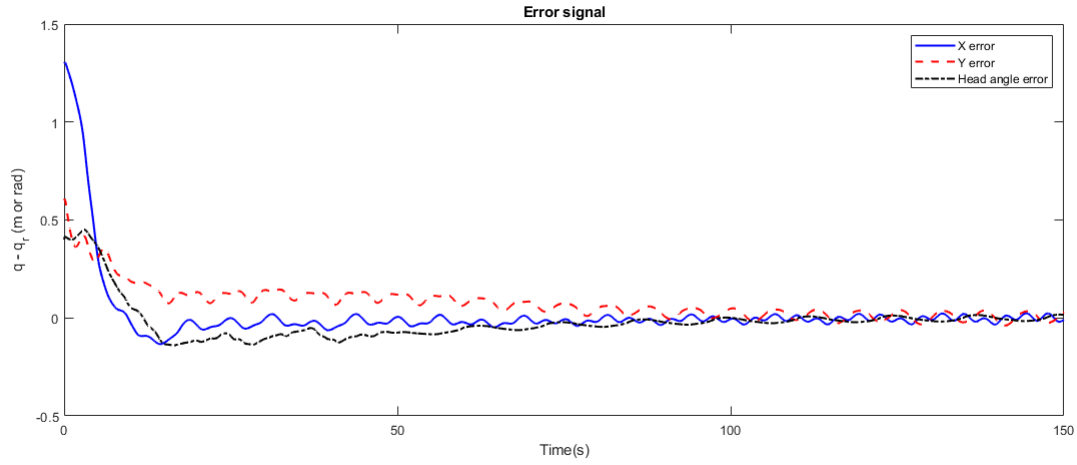
Hình 4.5: Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = 0.1$

Với quỹ đạo đặt là  $q_{r1}(t) = [0.5 \cos(2t) \cos(t) \cos(0.5t)]^T$  dưới tác động nhiễu với  $\rho = -0.1$  kết quả mô phỏng được thể hiện từ hình 4.6 - 4.10.

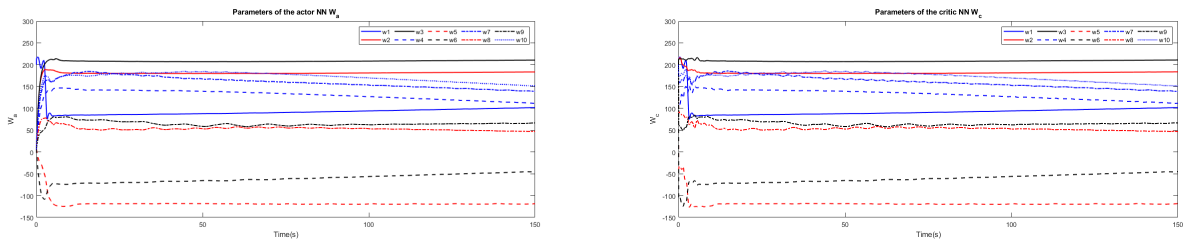




Hình 4.6: Hiệu suất bám của  $x, y, \theta$  với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = -0.1$

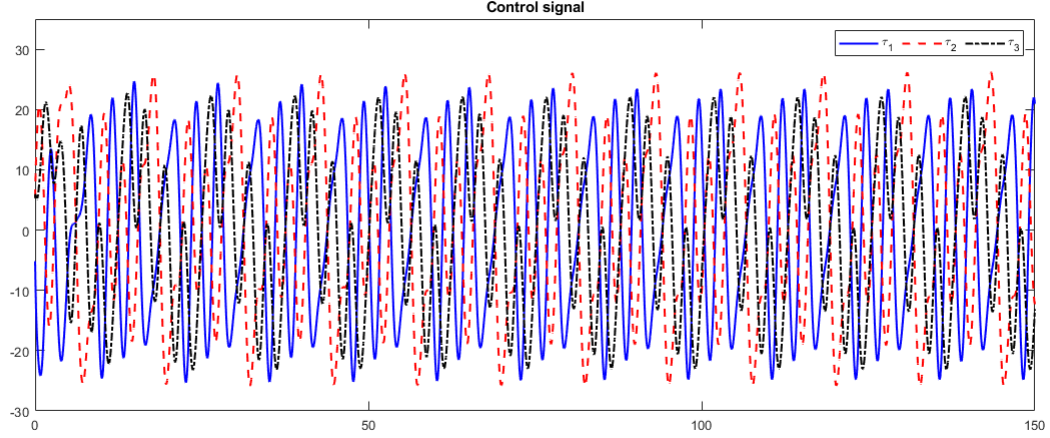


Hình 4.7: Sai lệch bám quỹ đạo  $q_1(t) - q_{r1}(t)$  khi  $\rho = -0.1$

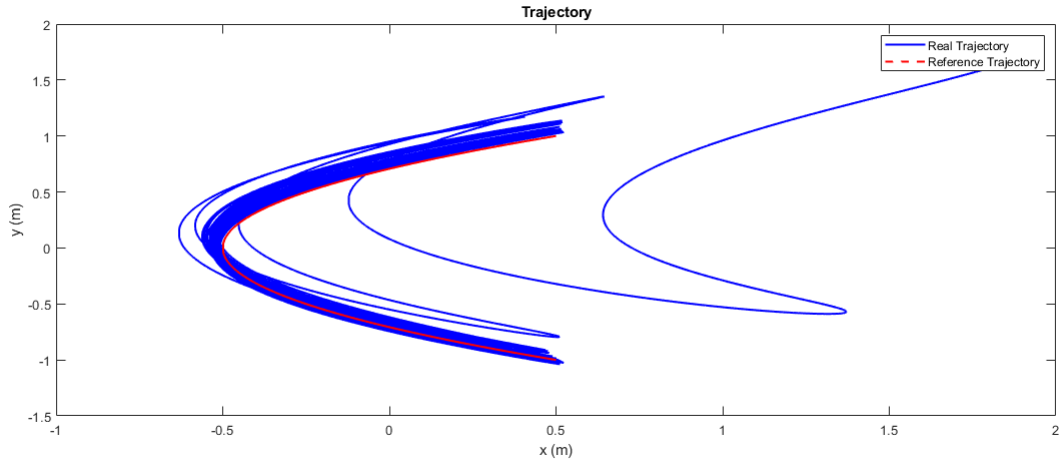


(a) Trọng số NN của actor với quỹ đạo đặt  $q_{r1}(t)$  và (b) Trọng số NN của critic với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = -0.1$

Hình 4.8: Trọng số NN cấu trúc AC với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = -0.1$



Hình 4.9: Tín hiệu điều khiển với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = -0.1$

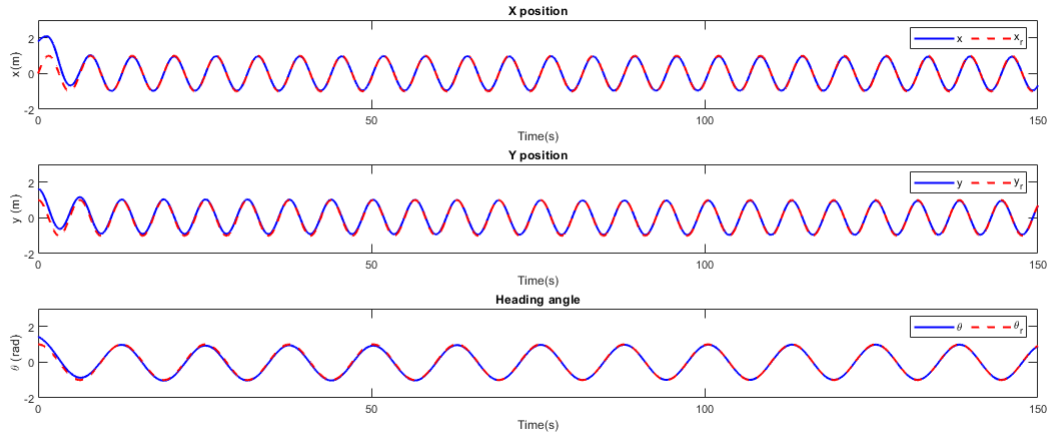


Hình 4.10: Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt  $q_{r1}(t)$  và  $\rho = -0.1$

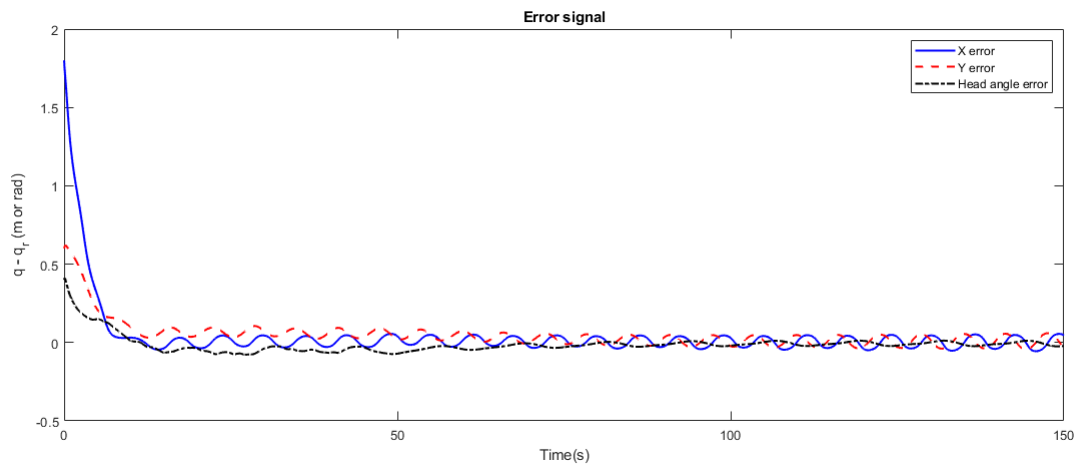
Trong đó, hình (4.2) và hình (4.7) cho thấy được sai lệch bám quỹ đạo của hệ thống trong hai trường hợp thay đổi tín hiệu nhiễu bởi trượt  $\rho = 0.1$  và  $\rho = -0.1$ . Từ đó ta có thể thấy rằng sai lệch bám tiến tới 0 và trạng thái của hệ thống đang tiến tới trạng thái ổn định chúng ta có thể thấy rõ điều này ở quỹ đạo chuyển động của hệ thống ở trong hình(4.5) và (4.10) hoặc hiệu suất bám của hệ thống trên hình (4.1) và (4.6). Các trọng số NN của hệ thống cũng được thể hiện trong các hình (4.3) và (4.8).

### 4.3 Kết quả mô phỏng với quỹ đạo đường tròn

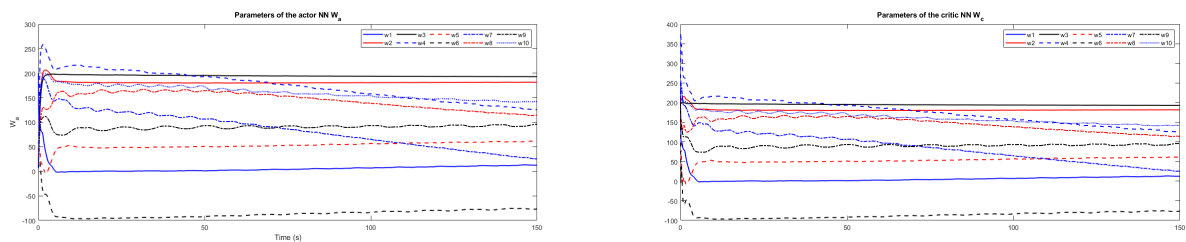
Với quỹ đạo đặt là  $q_{r2}(t) = [\sin(t) \cos(t) \cos(0.5t)]^T$  dưới tác động nhiễu với  $\rho = 0.1$  kết quả mô phỏng được thể hiện từ hình 4.11 - 4.15.



Hình 4.11: Hiệu suất bám của  $x, y, \theta$  với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

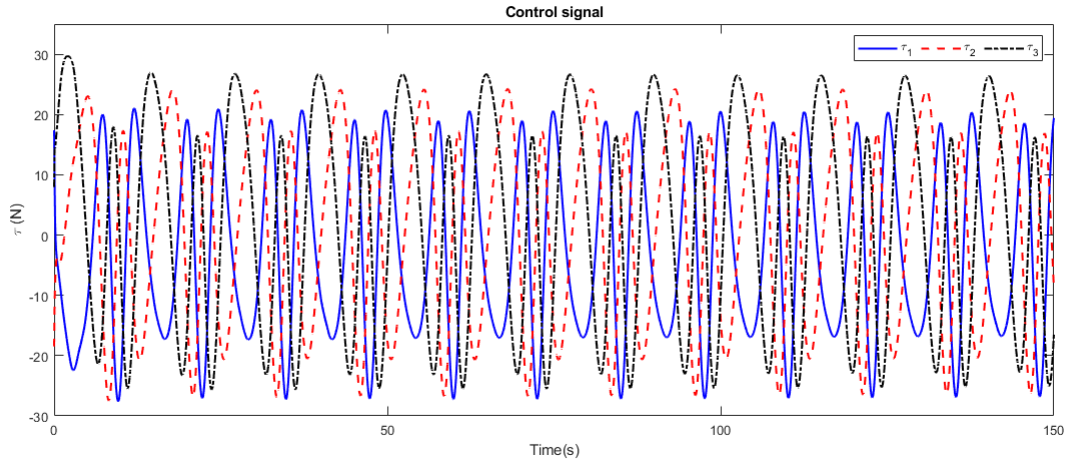


Hình 4.12: Sai lệch bám quỹ đạo  $q_2(t) - q_{r2}(t)$  khi  $\rho = 0.1$

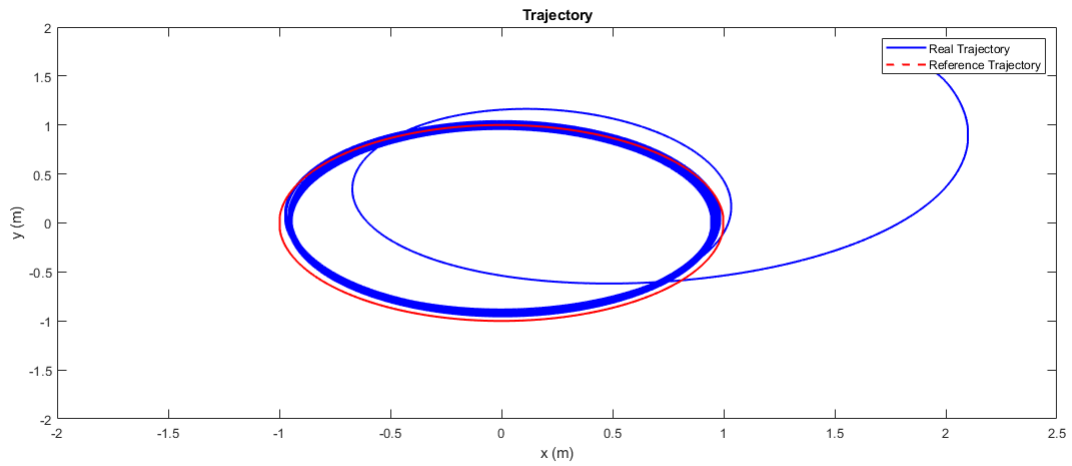


(a) Trọng số NN của actor với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$  và (b) Trọng số NN của critic với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

Hình 4.13: Trọng số NN cấu trúc AC với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

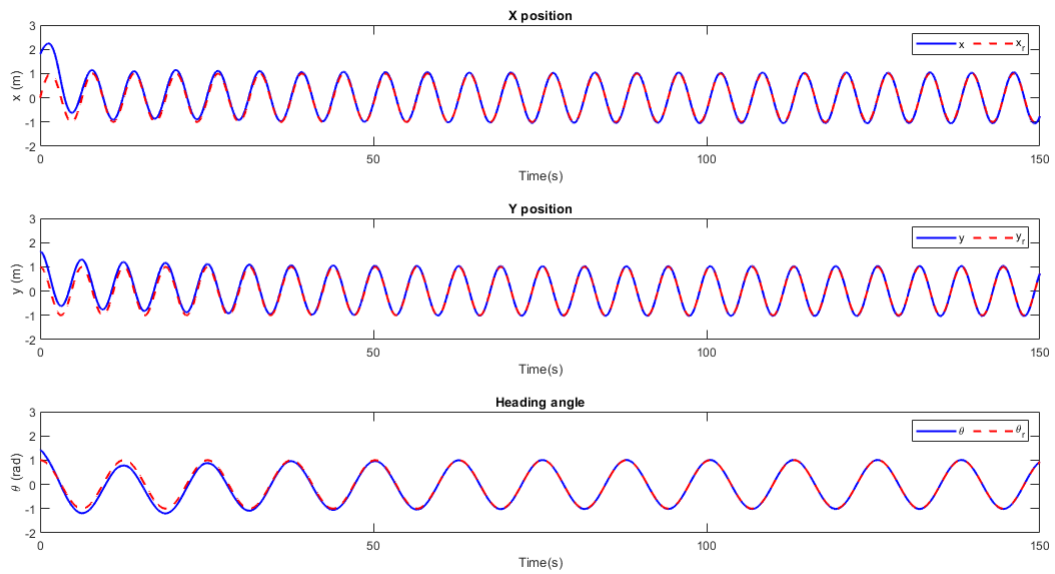


Hình 4.14: Tín hiệu điều khiển với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

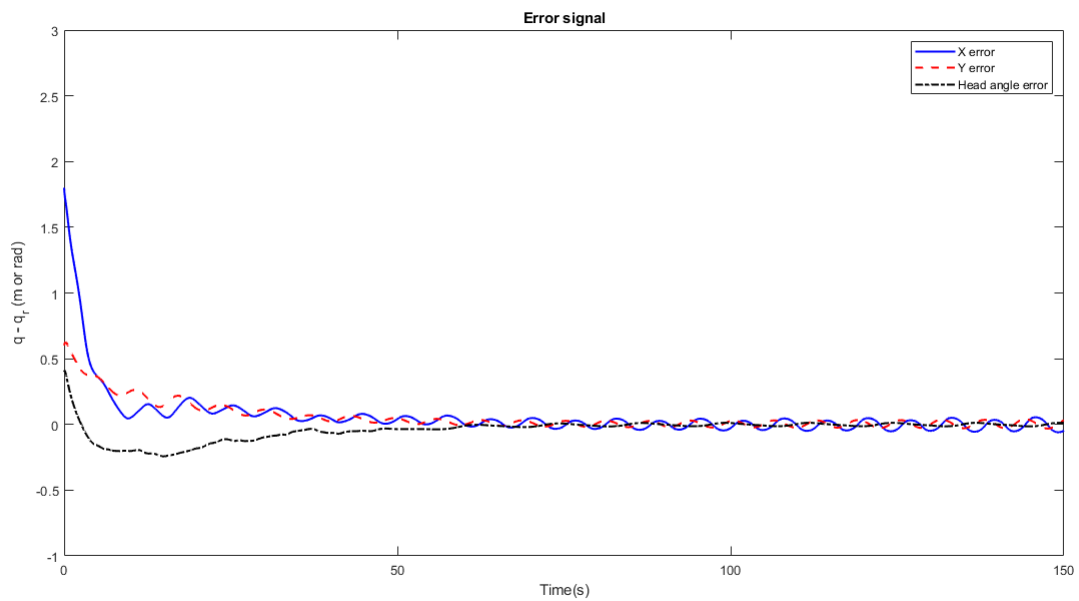


Hình 4.15: Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

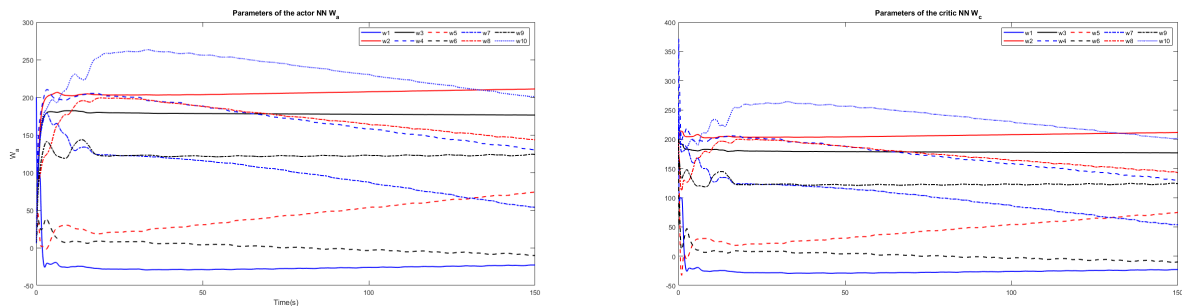
Với quỹ đạo đặt là  $q_{r2}(t) = [\sin(t) \cos(t) \cos(0.5t)]^T$  dưới tác động nhiễu với  $\rho = -0.1$  kết quả mô phỏng được thể hiện từ hình 4.16 - 4.20.



Hình 4.16: Hiệu suất bám của  $x, y, \theta$  với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$

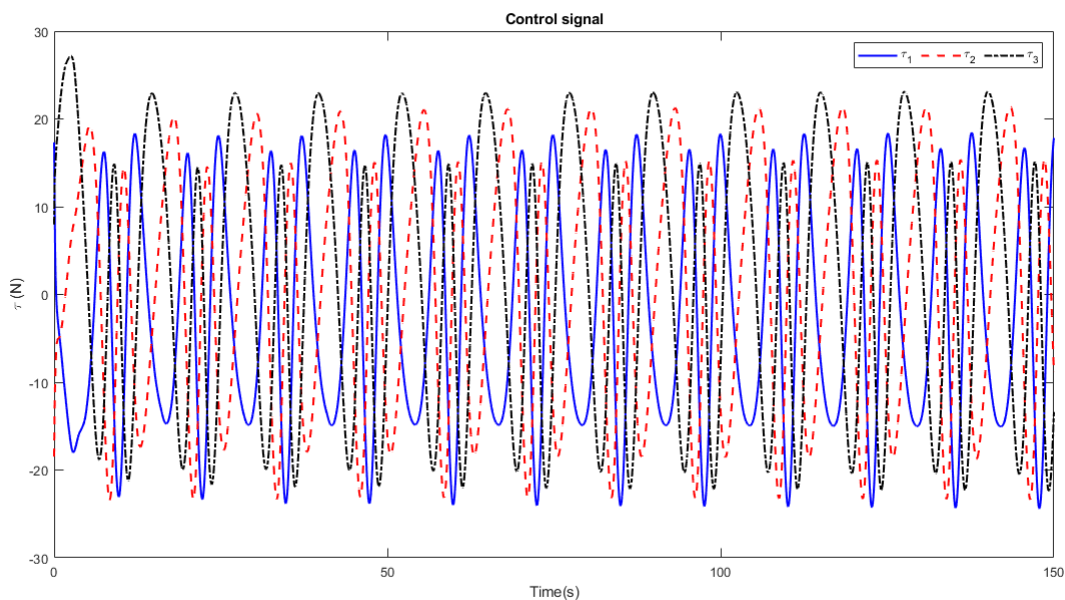


Hình 4.17: Sai lệch bám quỹ đạo  $q_2(t) - q_{r2}(t)$  khi  $\rho = -0.1$

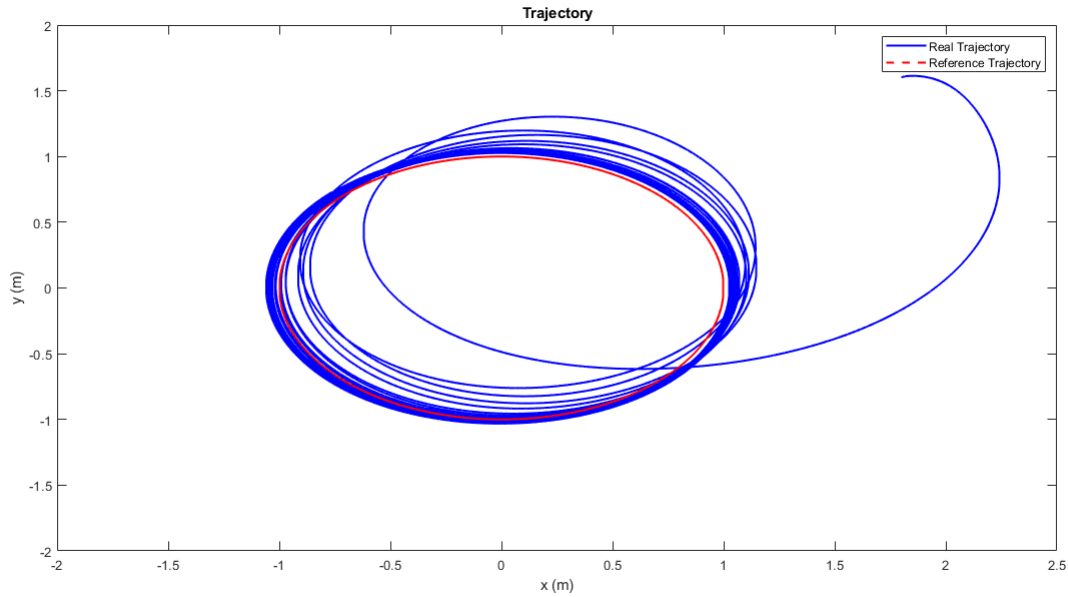


(a) Trọng số NN của actor với quỹ đạo đặt  $q_{r2}(t)$  và (b) Trọng số NN của critic với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$

Hình 4.18: Trọng số NN cấu trúc AC với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$



Hình 4.19: Tín hiệu điều khiển với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$

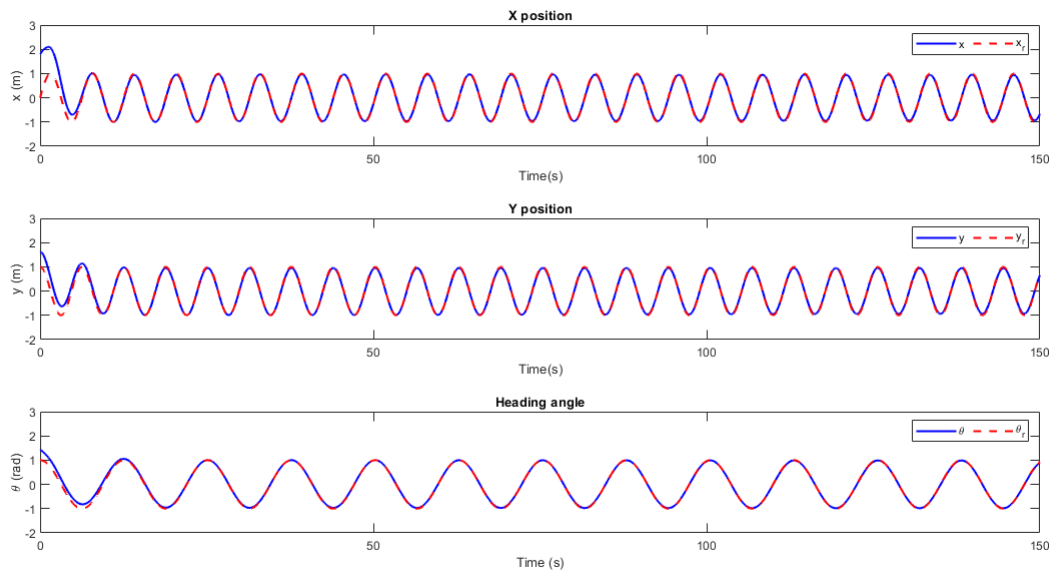


Hình 4.20: Quỹ đạo chuyển động của hệ thống MWMRs với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$

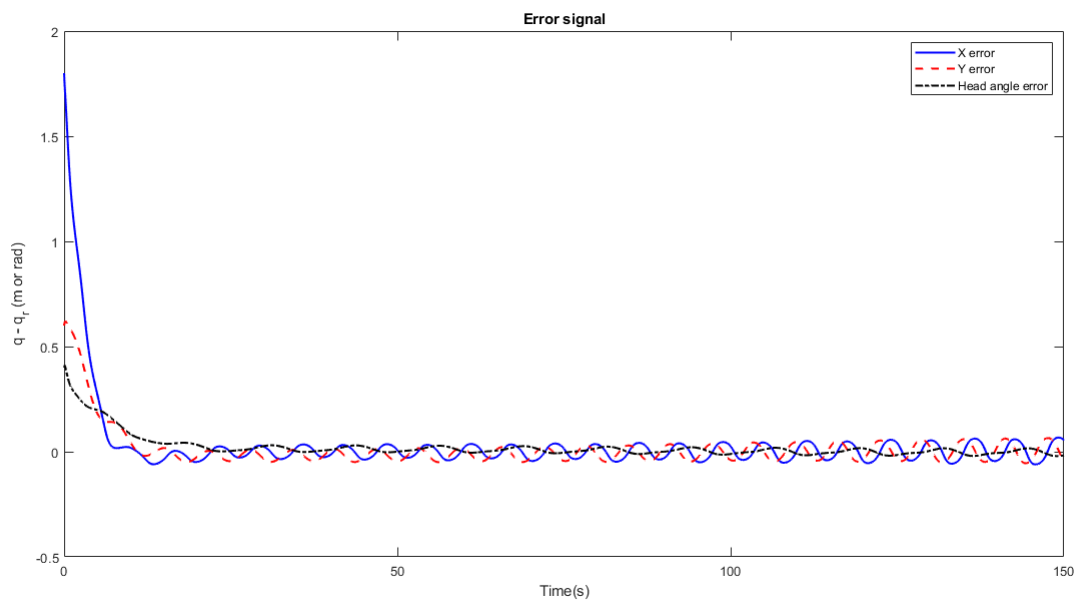
So sánh các hình 4.11 và 4.16, hình 4.12 và 4.17 chúng ta có thể thấy rằng mặc dù chịu tác động nhiễu với các  $\rho$  khác nhau tuy nhiên hệ thống vẫn đảm bảo được hiệu suất bám khá tốt và có sai lệch bám dần tiến tới 0. Tiếp tục nhìn vào các hình 4.13 và 4.18 chúng ta cũng có thể thấy được sự thay đổi trọng số NN đang tiến tới trạng thái ổn định. Quan sát các hình 4.5, 4.10, 4.15, 4.20, chúng ta có thể kết luận rằng với các quỹ đạo khác nhau thì bộ điều khiển học tăng cường dựa trên thuật toán AC với tín hiệu điều khiển như ở hình 4.4, hình 4.9, hình 4.14 và hình 4.19 vẫn đáp ứng được việc bám quỹ đạo đồng thời đưa hệ thống về trạng thái ổn định.

## 4.4 Kết quả mô phỏng với việc không thêm nhiễu thăm dò

Tiếp tục kịch bản mô phỏng với việc không thêm nhiễu thăm dò vào tín hiệu điều khiển với quỹ đạo đặt là  $q_{r2}(t) = [\sin(t) \cos(t) \cos(0.5t)]^T$  dưới tác động nhiễu với  $\rho = 0.1$  kết quả mô phỏng được thể hiện từ hình 4.21 - 4.25.

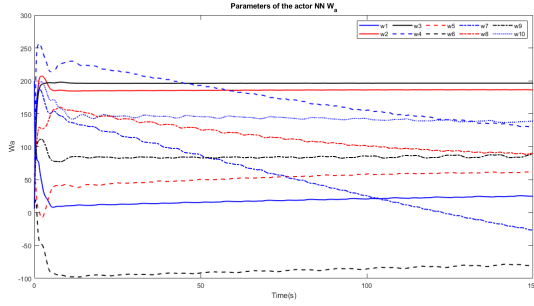


Hình 4.21: Hiệu suất bám của  $x, y, \theta$  khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

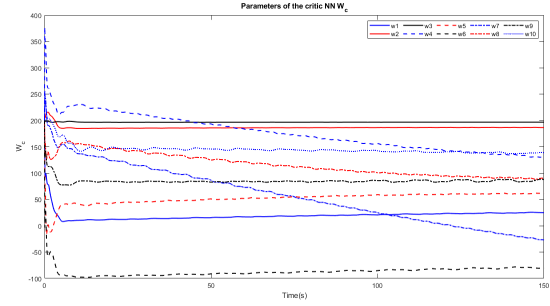


Hình 4.22: Sai lệch bám quỹ đạo  $q(t) - q_r(t)$  khi không có nhiễu thăm dò với  $\rho = 0.1$



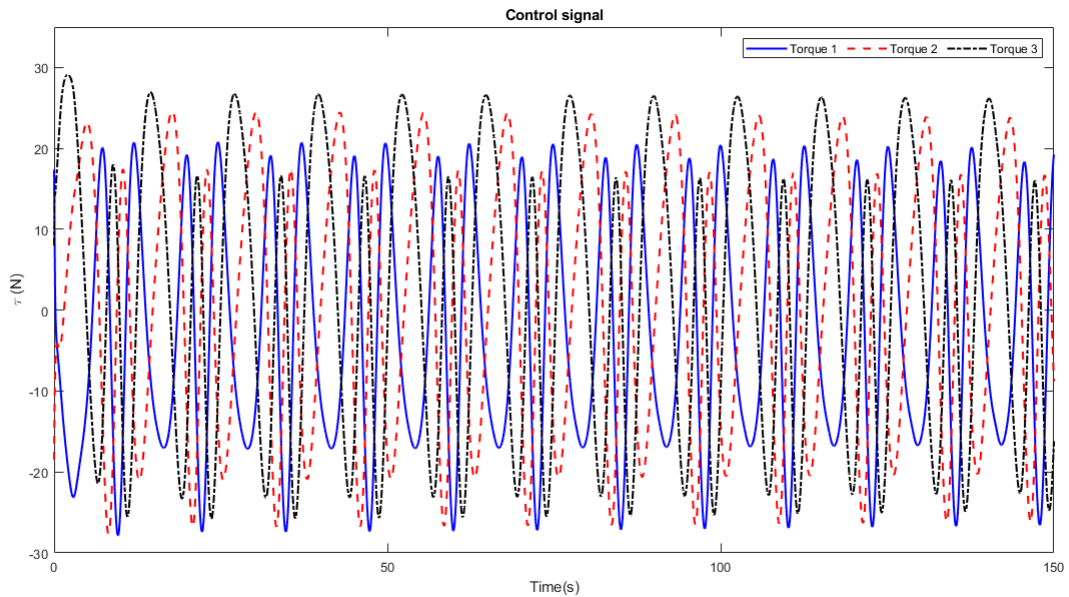


(a) Trọng số NN của actor khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

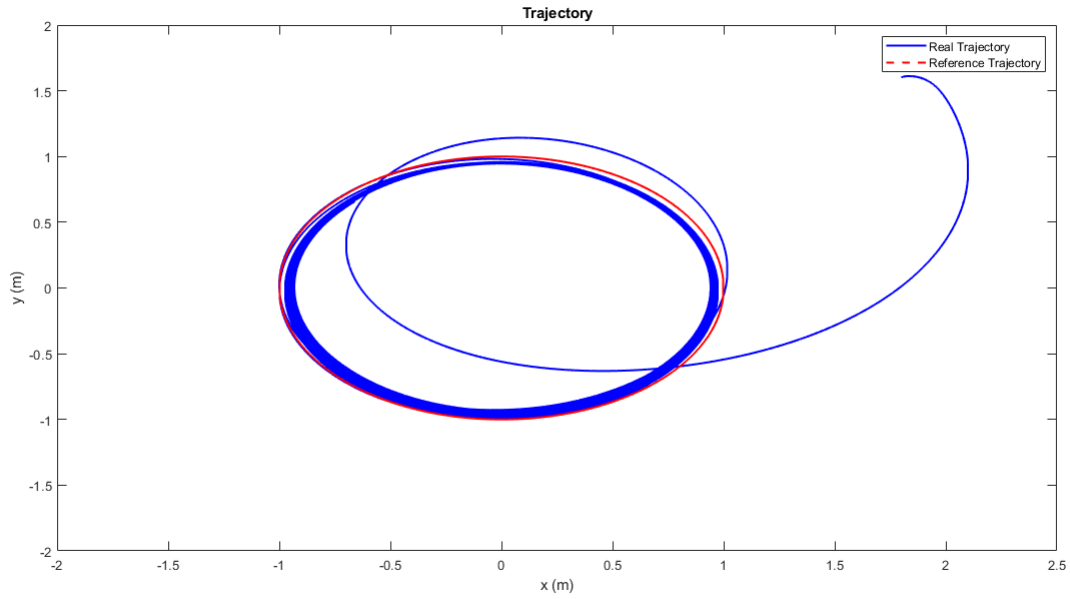


(b) Trọng số NN của critic khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

Hình 4.23: Trọng số NN cấu trúc AC khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

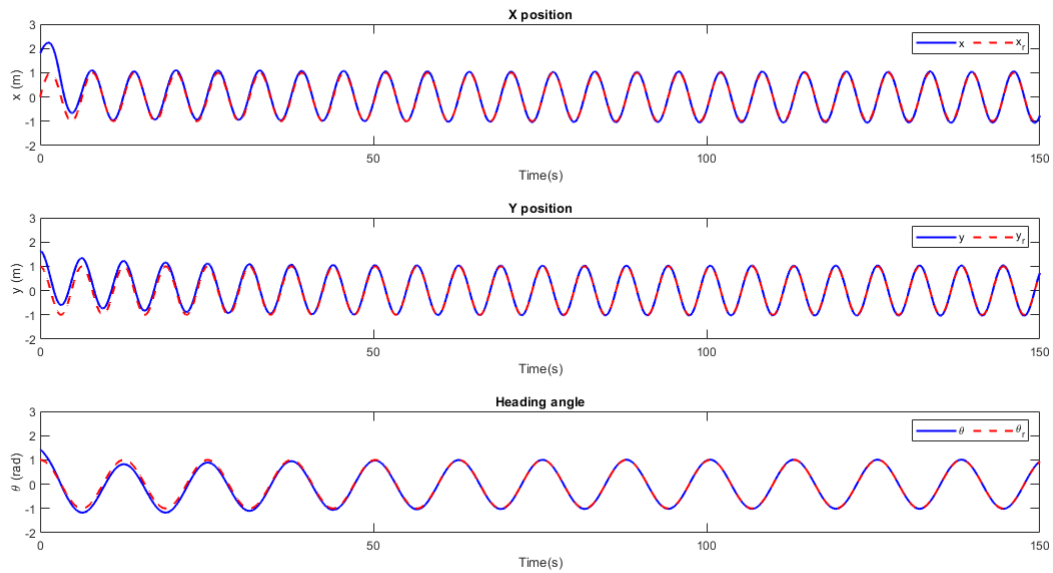


Hình 4.24: Tín hiệu điều khiển khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

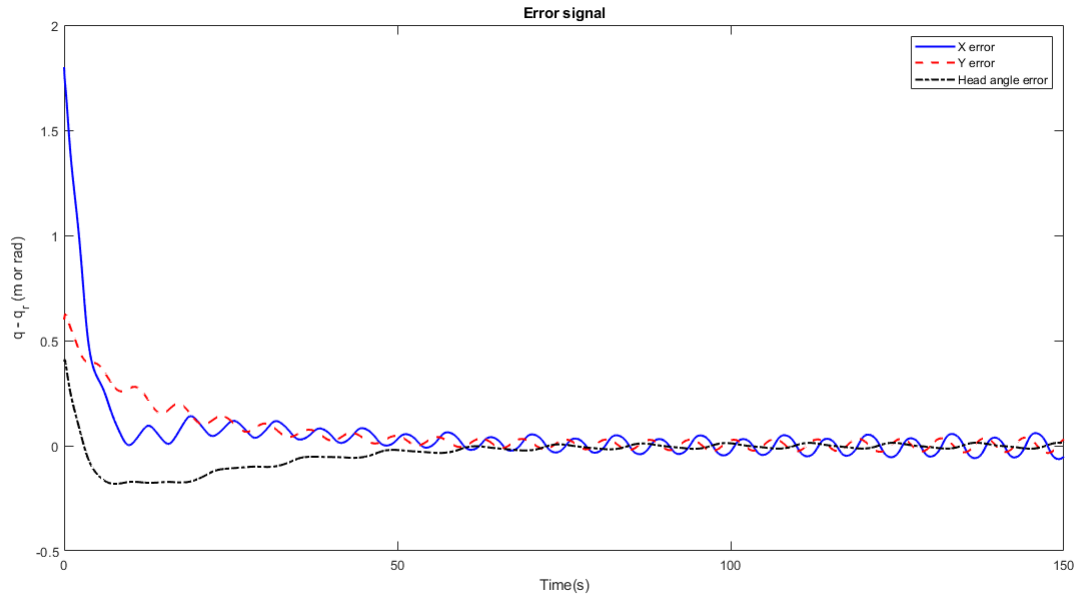


Hình 4.25: Quỹ đạo chuyển động của hệ thống MWMRs khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = 0.1$

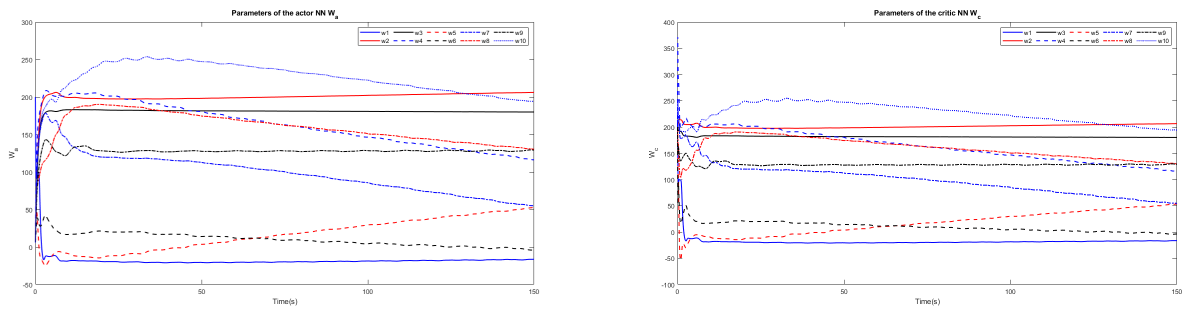
Với việc không có nhiễu thăm dò vào bộ điều khiển và với quỹ đạo đặt là  $q_{r2}(t) = [\sin(t) \cos(t) \cos(0.5t)]^T$  dưới tác động nhiễu với  $\rho = -0.1$  kết quả mô phỏng được thể hiện từ hình 4.26 - 4.30.



Hình 4.26: Hiệu suất bám của  $x, y, \theta$  khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$

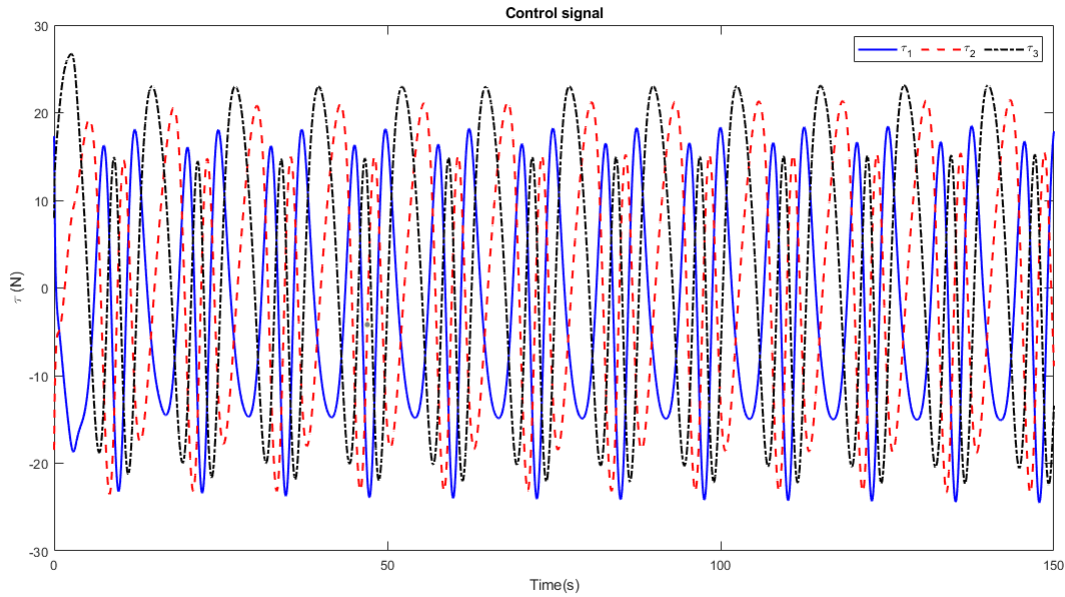


Hình 4.27: Sai lệch bám quỹ đạo  $q_2(t) - q_{r2}(t)$  khi không có nhiễu thăm dò với  $\rho = -0.1$

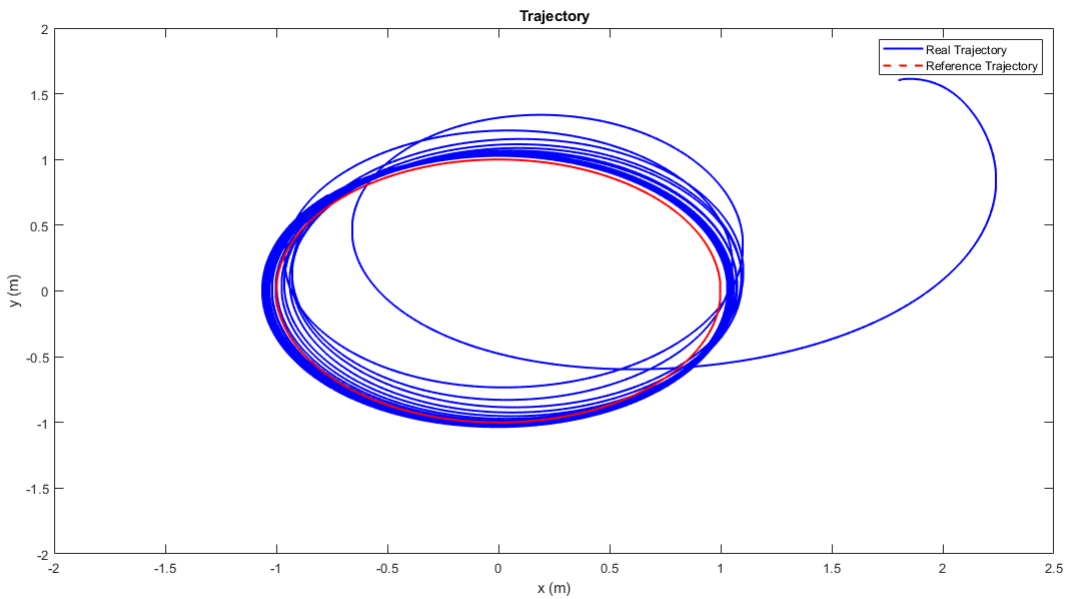


(a) Trọng số NN của actor khi không có nhiễu thăm dò (b) Trọng số NN của critic or khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$

Hình 4.28: Trọng số NN cấu trúc AC khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$



Hình 4.29: Tín hiệu điều khiển khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$



Hình 4.30: Quỹ đạo chuyển động của hệ thống MWMRs khi không có nhiễu thăm dò với quỹ đạo đặt  $q_{r2}(t)$  và  $\rho = -0.1$

Từ các hình 4.21 - 4.30, so sánh với các hình 4.11 - 4.20 ta có thể thấy được rằng khi không có nhiễu thăm dò vào bộ điều khiển thì hệ thống vẫn có thể được điều khiển hoạt động một cách ổn định trong trường hợp này. Thậm chí khi chúng ta nhìn vào các hình 4.12 - 4.22 và 4.17 - 4.27 ta còn có thể thấy được rằng tốc độ hội tụ khi không có nhiễu thăm dò vào bộ điều khiển còn nhanh hơn khi có nhiễu thăm dò. Tuy nhiên để đảm bảo hệ thống làm việc một cách ổn định thì việc cho nhiễu thăm dò vào bộ điều khiển là điều không thể thiếu. Việc này giúp

hệ thống đảm bảo được rằng sẽ thỏa mãn điều kiện PE từ đó giúp chúng ta có thể chắc chắn hệ thống luôn tiến tới trạng thái ổn định.

# Chương 5

## Kết luận

Đồ án đã đưa vào phân tích thuật toán ADP cũng như áp dụng thuật toán cho mô hình MWMRs. Thông qua việc phân tích tính ổn định, kiểm tra chất lượng dựa trên mô hình mô phỏng trong phần mềm MATLAB, có thể thấy rằng quỹ đạo thực tế bám khá tốt so với quỹ đạo mong muốn. Từ đó, có thể thấy tính khả thi và tiềm năng của thuật toán khi áp dụng vào các hệ thống thực tế.

Tuy nhiên, việc sử dụng mô hình mạng nơ ron như hiện tại dẫn đến việc bùng nổ số lượng công việc cần tính toán. Điều này dẫn đến yêu cầu cao về phần cứng trong hệ thống thực tế, nếu không sẽ dẫn đến sự chậm chạp trong tính toán. Từ đây, một hướng phát triển cho thuật toán trong tương lai là thiết kế mạng nơ ron phù hợp giúp giảm khối lượng tính toán cũng như tối ưu về mặt thời gian.

# Tài liệu tham khảo

- [1] V. Alakshendra and S. S. Chiddarwar. “adaptive robust control of mecanum-wheeled mobile robot with uncertainties”. *Nonlinear Dyn.*, vol. 87, no. 4, pp. 2147–2169, Mar. 2017.
- [2] H. Modares B. Kiumarsi, K. G. Vamvoudakis and F. L. Lewis. “optimal and autonomous control using reinforcement learning: A survey”. *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2042–2062, Jun. 2018.
- [3] Y. Sun C. Ren and S. Ma. “passivity-based control of an omnidirectional mobile robot”. *Robot. Biomimetics*, vol. 3, no. 1, pp. 1–9, Jul. 2016.
- [4] X. Yang F. Hu A. Jiang C. Wang, X. Liu and C. Yang. “trajectory tracking of an omnidirectional wheeled mobile robot using a model predictive control strategy”. *Appl. Sci.*, vol. 8, no. 2, p. 231, Feb. 2018.
- [5] D. Wang D. Liu, X. Yang and Q. Wei. “reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints”. *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1372–1385, Jul. 2015.
- [6] D. Liu D. Wang and H. Li. “policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems”. *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 627–632, Apr. 2014.
- [7] D. Wang et al. “formation control of multiple mecanum-wheeled mobile robots with physical constraints and uncertainties”. *Appl. Intell.*, vol. 52, no. 3, pp. 2510–2529, Jun. 2021.
- [8] B. A. Finlayson. *The Method of Weighted Residuals and Variational Principles*, New York, NY, USA: Academic Press, 1990.
- [9] F. L. Lewis H. Modares and M.-B. Naghibi-Sistani. “integral reinforcement learning and experience replay for adaptive optimal control of partially-unknown constrained-input continuous-time systems”. *Automatica*, vol. 50, no. 1, pp. 193–202, 2014.
- [10] Q. Wei H. Zhang and Y. Luo. “a novel infinite-time optimal tracking control scheme for a class of discrete-time nonlinear systems via the greedy hdp iteration algorithm”. *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 4, pp. 937–942, Aug. 2008.
- [11] H.-C. Huang and C.-C. Tsai. “adaptive trajectory tracking and stabilization for omnidirectional mobile robot with dynamic effect and uncertainties”. *IFAC Proc. Volumes*, vol. 41, no. 2, pp. 5383–5388, 2008.

- [12] M. I. Jordan and T. M. Mitchell. “machine learning: Trends, perspectives, and prospects”. *Science*, vol. 349, no. 6245, pp. 255–260, 2015.
- [13] M. Llama V. Santibáñez L. Ovalle, H. Ríos and A. Dzul. “omnidirectional mobile robot robust tracking: Sliding-mode output-based control approaches”. *Control Eng. Pract.*, vol. 85, pp. 50–58, Apr. 2019.
- [14] A. Lazinica. *Introduction to Mobile Robot Control*, Amsterdam, The Netherlands: Elsevier, 2014.
- [15] L.-C. Lin and H.-Y. Shih. “modeling and adaptive control of an omni-mecanum-wheeled robot”. *Intell. Control Autom*, vol. 4, no. 2, pp. 166–179, 2013.
- [16] H. Modares and F. L. Lewis. “optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning”. *Automatica*, vol. 50, no. 7, pp. 1780–1792, Jul. 2014.
- [17] K. L. Moore and N. S. Flann. “a six-wheeled omnidirectional autonomous mobile robot”. *IEEE Control Syst.*, vol. 20, no. 6, pp. 53-66, Dec. 2000.
- [18] Nguyễn Doãn Phước. *Tối ưu hóa trong điều khiển và Điều khiển tối ưu*. Nhà xuất bản Bách khoa Hà Nội, 2016.
- [19] P. Walters R. Kamalapurkar and W. E. Dixon. “model-based reinforcement learning for approximate optimal regulation”. *Automatica*, vol. 64, pp. 94–104, Feb. 2016.
- [20] S. Bhasin R. Kamalapurkar, H. Dinh and W. E. Dixon. “approximate optimal trajectory tracking for continuous-time nonlinear systems”. *Automatica*, vol. 51, pp. 40–48, Jan. 2015.
- [21] B. E. Carter R. L. Williams and G. Giulio. “dynamic model with slip for wheeled omnidirectional robots”. *IEEE Trans. Robot. Autom*, vol. 18, no. 3, pp. 285–293, Jun. 2002.
- [22] I. Nourbakhsh R. Siegwart and D. Scaramuzza. *Introduction to Autonomous Mobile Robots*, (Intelligent Robotics and Autonomous Agents Series). Cambridge, MA, USA: MIT Press, 2011.
- [23] Q. Wu R. Xia and S. Shao. “disturbance observer-based optimal flight control of near space vehicle with external disturbance”. *Trans. Inst. Meas. Control*, vol. 42, no. 2, pp. 272–284, Jan. 2020.
- [24] M. Johnson K. G. Vamvoudakis F. L. Lewis S. Bhasin, R. Kamalapurkar and W. E. Dixon. “a novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems”. *Automatica*, vol. 49, no. 1, pp. 82–92, Jan. 2013.
- [25] N. Hung H. K. Kim T. D. Viet, P. T. Doan and S. B. Kim. “tracking control of a three-wheeled omnidirectional mobile manipulator system with disturbance and friction”. *J. Mech. Sci. Technol*, vol. 26, no. 7, pp. 2197–2211, Jul. 2012.
- [26] S. G. Tzafestas. “geometry and kinematics of the mecanum wheel”. *Comput. Aided Geometric Des.*, vol. 25, no. 9, pp. 784–791, Dec. 2008.



- [27] K. G. Vamvoudakis and F. L. Lewis. “online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem”. *Automatica*, vol. 46, no. 5, pp. 878–888, May 2010.
- [28] D. Wang and C. B. Low. “modeling and analysis of skidding and slipping in wheeled mobile robots: Control design perspective”. *IEEE Trans. Robot.*, vol. 24, no. 3, pp. 676–687, Jun. 2008.
- [29] D. Wang and C. Mu. “adaptive-critic-based robust trajectory tracking of uncertain dynamics and its application to a spring–mass–damper system”. *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 654–663, Jan. 2018.