

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI



ĐỒ ÁN TỐT NGHIỆP

THUẬT TOÁN ADP CHO ĐỔI TƯỢNG TÀU THỦY

PHẠM THÀNH LỘC

loc.pt152308@sis.hust.edu.vn

TRẦN QUANG HUY

huy.tq151711@sis.hust.edu.vn

Ngành KT Điều khiển và Tự động hóa

Chuyên ngành Điều khiển tự động

Giảng viên hướng dẫn: TS. Đào Phương Nam

Chữ ký của GVHD

Bộ môn:

Điều khiển tự động

Viện:

Điện

HÀ NỘI, 06/2020

BỘ GIÁO DỤC và ĐÀO TẠO
TRƯỜNG ĐH BÁCH KHOA HÀ NỘI

CỘNG HÒA XÃ HỘI CHỦ NGHĨA VIỆT NAM
Độc lập – Tự do - Hạnh phúc

**NHIỆM VỤ
ĐỒ ÁN TỐT NGHIỆP**

Họ và tên sinh viên:

Khóa.....Viện: Điện Ngành: CN ĐK và TDH

1. Tên đề tài:

2. Nội dung đề tài:

.....
.....
.....
.....

3. Cán bộ hướng dẫn:

Phần

Họ tên cán bộ

.....
.....
.....
.....

4. Thời gian giao đề tài:.....

5. Thời gian hoàn thành:.....

Ngày..... tháng năm 2020

LÃNH ĐẠO BỘ MÔN

CÁN BỘ HƯỚNG DẪN

SINH VIÊN THỰC HIỆN
(Ký và ghi rõ họ tên)

Lời cảm ơn

Chúng em xin cảm ơn TS. Đào Phương Nam vì đã định hướng và tận tình hướng dẫn không chỉ trong quá trình thực hiện đồ án tốt nghiệp này mà còn trong suốt 5 năm học tập và nghiên cứu tại đại học Bách Khoa Hà Nội. Chúng em cũng xin gửi lời cảm ơn sâu sắc đến gia đình và bạn bè vì đã luôn sát cánh và động viên tinh thần trong suốt quãng thời gian đã qua.

Do thời gian và khả năng còn hạn chế, đồ án không thể tránh khỏi những nhầm lẫn và thiếu sót, chúng em rất mong nhận được sự góp ý của các thầy cô và bạn đọc để giúp luận án trở nên hoàn thiện hơn.

Chúng em xin chân thành cảm ơn!

Tóm tắt nội dung đồ án

Bài toán điều khiển tối ưu cho hệ phi tuyến bị ràng buộc trực tiếp bởi nghiệm của phương trình Hamilton-Jacobi-Bellman (HJB) và bài toán điều khiển tối ưu bền vững bị ràng buộc trực tiếp bởi nghiệm của phương trình Hamilton-Jacobi-Isaacs (HJI). Đây là các phương trình vi phân phi tuyến không có nghiệm giải tích. Từ đó, bài toán xấp xỉ nghiệm HJB và HJI off-line hoặc online được đặt ra. Học củng cố (Reinforcement Learning (RL)) bắt nguồn từ qui hoạch động (Dynamic Programming (DP)), phát triển thành qui hoạch động thích nghi (Adaptive Dynamic Programming (ADP)) trở thành một trong những phương pháp hữu hiệu dùng để xấp xỉ các nghiệm HJB và HJI. Dựa vào cấu trúc điều khiển chuẩn của ADP bao gồm hai hoặc ba xấp xỉ hàm, các giải thuật RL không ngừng được nghiên cứu và phát triển. Ngày nay, các giải thuật điều khiển RL là online, không còn là off-line như những nghiên cứu đã công bố trong những năm đầu của thế kỷ 21. Ví dụ, các giải thuật RL đã được thiết kế để xấp xỉ nghiệm ARE (Algebraic Riccati Equation) cho hệ tuyến tính với các ma trận trạng thái không biết và sau này là xấp xỉ nghiệm HJB và HJI cho hệ phi tuyến với các thành phần động học trong mô hình hệ thống biết hoặc không biết, có nhiều hoặc bỏ qua nhiều.

Đồ án tốt nghiệp này trình bày ứng dụng của giải thuật học củng cố điều khiển thích nghi cho hệ Lagrange, trong đó bao gồm cấu trúc Actor-Critic, online On-Policy IRL và online Off-Policy IRL với hàm phạt. Với cấu trúc AC, hiểu biết về mô hình hệ thống được yêu cầu nhằm giải xấp xỉ nghiệm của phương trình HJB, đây cũng thuật toán được nghiên cứu và phát triển đầu tiên đã được chứng minh rõ ràng trong nhiều tài liệu và là nền tảng để phát triển các thuật toán sau. Tiếp tới là thuật toán online On-Policy IRL, với việc thu thập dữ liệu vào ra của hệ thống nhằm lược bỏ đi một phần yêu cầu cấu trúc của hệ. Xa hơn nữa là thuật toán online Off-Policy IRL với hàm phạt, ngoài việc hoàn toàn không cần biết mô hình hệ thống, giải thuật này đưa ra nhằm xấp xỉ nghiệm phương trình HJI của bài toán tracking. Với từng thuật toán, đồ án này đưa ra đi kèm với các ví dụ cho thấy kết quả quá trình học của giải thuật là chính xác bằng việc so sánh nghiệm lý thuyết và nghiệm của giải thuật. Nhằm cho thấy khả năng ứng dụng của thuật toán, mô phỏng được cho mô hình tàu thủy đủ cơ cấu chấp hành (Surface Vessels) bám theo quỹ đạo cho trước được tiến hành trên phần mềm MATLAB với cả ba thuật

toán đã nêu trên. Cuối cùng là nhận xét ưu nhược điểm của từng giải thuật, tính ứng dụng của chúng và định hướng phát triển trong tương lai.

Hà Nội, 21 tháng 5 năm 2020

Sinh viên

Phạm Thành Lộc

Trần Quang Huy

Một số kí hiệu viết tắt

$\ \cdot\ $	Chuẩn trong không gian Euclid.
RL	Reinforcement Learning.
IRL	Integral Reinforcement Learning.
HJB	Hamilton–Jacobi–Bellman.
PI	Policy Iteration.
ADP	Approximate/Adaptive Dynamic Programming.
NN	Neural Network.
RBF NN	Radial Basis Function Neural Network.
CNN	Critic Neural Network.
ANN	Actor Neural Network.
AC	Actor-Critic.
PE	Persistent Excitation Condition.

Danh sách hình vẽ

1.1	Nguyên tắc chung thuật toán học tăng cường	1
2.1	Mô tả nguyên lý tối ưu Bellman	6
3.1	Cấu trúc ADP sử dụng hai xấp xỉ hàm trong điều khiển tối ưu	16
3.2	Cấu trúc ADP sử dụng ba xấp xỉ hàm trong điều khiển tối ưu \mathbb{H}_∞	17
3.3	Trạng thái của hệ thống với cấu trúc Actor Critic	19
3.4	Sự hội tụ của trọng số W_c với cấu trúc Actor Critic	19
3.5	Sự hội tụ của trọng số W_a với cấu trúc Actor Critic	20
3.6	Trạng thái hệ thống với thuật toán IRL	22
3.7	Sự hội tụ của trọng số W với thuật toán IRL	23
3.8	Trạng thái hệ thống với thuật toán Online Off-Policy IRL . .	26
3.9	Sự hội tụ của trọng số W với thuật toán Online Off-Policy IRL	26
4.1	Các biến chuyển động của phương tiện hàng hải	28
4.2	Các khung tọa độ quy chiếu	29
4.3	Khung tọa độ quy chiếu quán tính gắn với trái đất và khung tọa độ gắn thân	32
4.4	Ôn định khuynh tâm theo chiều ngang tàu	37
4.5	Trạng thái hệ tàu thủy với thuật toán ADP cấu trúc AC . .	44
4.6	Sự hội tụ của trọng số W_c của hệ tàu thủy với thuật toán ADP cấu trúc AC	44
4.7	Sự hội tụ của trọng số W_a của hệ tàu thủy với thuật toán ADP cấu trúc AC	45
4.8	quỹ đạo của hệ tàu thủy với thuật toán ADP cấu trúc AC . .	45
4.9	Trạng thái hệ tàu thủy với thuật toán Online On-Policy IRL .	46
4.10	Sự hội tụ của trọng số W của hệ tàu thủy với thuật toán Online On-Policy IRL	47
4.11	quỹ đạo của hệ tàu thủy với thuật toán Online On-Policy IRL	47
4.12	Trạng thái hệ tàu thủy với thuật toán Online Off-Policy IRL chứa hàm phạt	49
4.13	Quỹ đạo của hệ tàu thủy với thuật toán Online Off-Policy IRL chứa hàm phạt	49

Danh sách bảng biểu

3.1	Một số activation function thường gặp	15
4.1	Các ký hiệu của SNAME	28

Mục lục

Lời cảm ơn	ii
Tóm tắt nội dung đồ án	iii
Một số kí hiệu viết tắt	v
1 Tổng quan thuật toán học tăng cường	1
1.1 Giới thiệu chung về thuật toán	1
1.2 Thuật toán quy hoạch động xấp xỉ/thích nghi	3
2 Thuật toán Policy Iteration cho hệ Affine	6
2.1 Quy hoạch động (Bellman)	6
2.1.1 Bài toán 2.1:	6
2.1.2 Bài toán 2.2:	8
2.2 Thuật toán Policy Iteration cho hệ Affine	10
3 Giải thuật ADP qui hoạch động thích nghi sử dụng xấp xỉ hàm	14
3.1 Xấp xỉ hàm và điều kiện PE	14
3.1.1 Nguyên lý xấp xỉ hàm mạng NN	14
3.1.2 Điều kiện PE	15
3.2 Giải thuật ADP	16
3.2.1 ADP cấu trúc Actor-Critic	17
3.2.2 Online On-Policy Integral Reinforcement Learning	20
3.2.3 Online Off-Policy IRL với hàm phạt	23
4 Ứng dụng thuật toán ADP cho mô hình tàu thủy	27
4.1 Mô hình động lực học của phương tiện hàng hải	27
4.1.1 Phân tích về vị trí và hướng chuyển động của tàu	30
4.1.2 Phương trình chuyển động của phương tiện hàng hải (Dynamics)	31
4.1.3 Mô hình động lực học của tàu thủy ba bậc tự do trên mặt phẳng nằm ngang	38
4.2 Thuật toán ADP cho tàu thủy	40

4.2.1	Sử dụng đổi biến trong ứng dụng ADP cho hệ không dùng	40
4.2.2	Ứng dụng thuật toán ADP cấu trúc Actor-Critic	41
4.2.3	Ứng dụng thuật toán Online On-Policy IRL	45
4.2.4	Ứng dụng thuật toán Online Off-Policy IRL	48
Kết luận		50
Tài liệu tham khảo		54
Phụ Lục 1		55
Phụ Lục 2		57

Chapter 1

Tổng quan thuật toán học tăng cường

1.1 Giới thiệu chung về thuật toán

Học tăng cường là một trong ba thuật toán lớn trong học máy cùng với học giám sát, học không giám sát. Được lấy cảm hứng từ cách con người và động vật thông minh tự học để thay đổi hành vi thông qua trực tiếp tương tác với môi trường và quan sát phản hồi, thuật toán học tăng cường mang đến cho tác tử khả năng tự học để chọn ra chiến lược tốt nhất thông qua quan sát lượng thưởng phạt mà tác tử nhận được. Thuật toán học tăng cường trong quá khứ được phát triển chủ yếu trên nền tảng của chuỗi quyết định Markov (MDP) [25] và được ứng dụng chủ yếu trong lý thuyết trò chơi, vấn đề đưa ra quyết định (Decision Making), v.v... Cốt lõi của thuật toán học tăng cường ứng dụng trong điều khiển được xây dựng từ phương pháp quy hoạch động, phát triển bởi Bellman vào những năm 1950. Quy hoạch động tuy dựa trên nguyên lý đơn giản, tuy nhiên bước đầu đã mang đến nhiều thành công. Tuy nhiên, điểm yếu của phương pháp quy hoạch động là nó yêu cầu thông tin cụ thể về môi trường xung quanh tác tử, điều này trong nhiều trường hợp là không có sẵn. Để giải quyết vấn đề này, phương pháp Monte-Carlo [12] và Temporal-Difference [17] được áp dụng giúp tác tử học chiến lược tối ưu mà không yêu cầu thông tin về môi trường. Thuật toán học tăng cường giúp cho

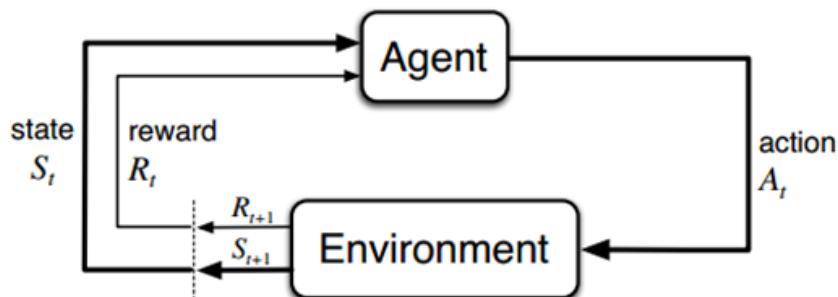


Figure 1.1: Nguyên tắc chung thuật toán học tăng cường

tác tử đưa ra hành động tối ưu trong môi trường bất định, dựa trên nguyên lý quan sát sự phản hồi của môi trường và cố gắng tối ưu hóa hàm mục tiêu là tổng phần thưởng (reward) trong suốt quá trình hoạt động. Nguyên tắc chung của các thuật toán học tăng cường biểu diễn sự tương tác giữa tác tử và môi trường là một vòng khép kín như trên hình 1.1.

Nói rõ hơn, trong đó, với bất kỳ hành động nào mà tác tử đưa ra cũng sẽ ảnh hưởng đến môi trường mà tác tử sống trong. Sự phản hồi có thể là tích cực hoặc tiêu cực được thể hiện qua phần thưởng. Trong thuật toán học tăng cường, để hướng dẫn cho tác tử hoàn thành nhiệm vụ cho trước, người thiết kế không cần phải đưa ra các hướng dẫn cụ thể. Tác tự sẽ tự đưa ra các hành động cụ thể, quan sát lượng phần thưởng được trả về và từ đó đưa ra so sánh và lựa chọn những chiến lược giúp cho tác tử nhận được nhiều phần thưởng hơn. Cách tiếp cận này dựa trên nguyên lý tự học của con người và các loài động vật thông minh. Trong đó, trước khi có tri thức, con người phải thử một loạt các hành động, quan sát liệu hành động đó mang lại lợi ích gì, từ đó đưa ra so sánh và lựa chọn các hành động tốt hơn. Bằng cách tiếp cận này, thuật toán giúp cho tác tử thích nghi với môi trường xung quanh, mặt khác tối ưu hóa chiến lược của tác tử dựa trên quan sát phản hồi của môi trường.

Tuy nhiên, các thuật toán học tăng cường ngày nay gặp khó khăn với sự cân bằng giữa sự tìm tòi (exploration) để tích lũy tri thức cho tác tử, cũng như tận dụng những tri thức đã có để đưa ra quyết định tối ưu (exploitation). Sự cân bằng của hai khía cạnh trên có thể được hiểu như sau. Trong một môi trường mà tác tử chưa có thông tin, tác tử phải đưa ra một loạt các hành động ngẫu nhiên và quan sát phản hồi của môi trường để tích lũy kinh nghiệm cho bản thân, dựa trên những quan sát tác tử có thể nhận xét được chiến lược hiện tại đã tốt hay chưa. Tuy nhiên nếu tập thử của tác tử không đủ lớn, sự ước lượng của tác tử sẽ sai lệch nhiều với thực tế, dẫn đến những quyết định sai lầm. Và vì thuật toán học tăng cường giúp tác tử đưa ra những quyết định tối ưu hơn dựa trên kinh nghiệm tích lũy được của chính tác tử. Do đó nếu như kinh nghiệm của tác tử không nhiều, sự ước lượng sẽ sai có sai lệch lớn hơn. Vì vậy trong một số thuật toán, trong giai đoạn đầu của quá trình huấn luyện, tác tử có xu hướng chọn ngẫu nhiên các hành động và quan sát phản hồi của môi trường nhằm tích lũy dần kinh nghiệm. Qua thời gian, tác tử được lập trình có xu hướng tận dụng kinh nghiệm để đưa ra quyết định tối ưu dựa trên kinh nghiệm hiện tại hơn là thực hiện ngẫu nhiên các hành động. Tuy vậy, tác tử vẫn có một xác suất lớn hơn để đưa ra những quyết ngẫu nhiên. Nhờ có sự luôn thử các hành động mới, thuật toán không bị gấp rơi vào điểm tối ưu cục bộ mà sẽ hội tụ dần về chiến lược tối ưu toàn cục theo thời gian. Đây là một ví dụ điển hình về sự cân bằng giữa sự tìm tòi và sự tận dụng kinh nghiệm. Cũng có thể nhìn vấn đề này như sự kết hợp giữa

vấn đề tối ưu hóa quyết định của tác tử và sự thích nghi của tác tử với môi trường.

Thuật toán học tăng cường trong quá khứ được nghiên cứu chủ yếu trong lĩnh vực học máy và đã có một số thành tựu đáng kể. Tuy nhiên, trong những năm gần đây, lý thuyết điều khiển đã bắt đầu quan tâm được những tiềm năng của thuật toán trên và phát triển thuật toán trong bối cảnh của lý thuyết điều khiển. Thuật toán khi được nghiên cứu trong điều khiển được biết đến với tên Quy Hoạch Động Xấp Xỉ/Thích nghi (Approximate/Adaptive Dynamic Programming - ADP), đã mang đến nhiều cải tiến lớn trong lĩnh vực này. Trong những năm gần đây, Sự phát triển nhanh chóng của thuật toán ADP đã mang đến những thành công nhất định, thuật toán ADP kết hợp giữa yếu tố thích nghi và tối ưu trong điều khiển, giúp giải quyết vấn đề bất định đối tượng cũng như tối ưu hóa hiệu năng của bộ điều khiển.

1.2 Thuật toán quy hoạch động xấp xỉ/thích nghi

Trong bối cảnh điều khiển truyền thống, hai vấn đề lớn của điều khiển là điều khiển thích nghi và điều khiển tối ưu, hai phương pháp điều khiển xử lý hai bài toán lớn khác nhau trong lý thuyết điều khiển. Điều khiển tối ưu đưa ra các phương pháp để tìm luật điều khiển giúp ổn định hệ thống, đồng thời tối ưu hàm mục tiêu cho trước, tuy nhiên tìm ra luật điều khiển tối ưu, các cách tiếp cận cũ đòi hỏi thông tin rõ ràng hệ động học của hệ thống, điều này làm cản trở khả năng của thuật toán khi áp dụng vào thực tế do bất định mô hình. Trong khi đó phương pháp điều khiển thích nghi cho phép thiết kế bộ điều khiển với đối tượng bất định, dựa trên các luật thích nghi cho bộ điều khiển, có thể là gián tiếp thông qua cơ cấu nhận dạng đối tượng hay trực tiếp chỉnh định tham số bộ điều khiển, tuy nhiên điều khiển thích nghi chưa xét đến yếu tố tối ưu chất lượng của luật điều khiển. Dưới góc nhìn của thuật toán học tăng cường, hai cách tiếp cận của hai phương pháp trên được dung hòa làm một, tận dụng điểm mạnh của cả hai phương pháp.

Ban đầu, thuật toán ADP được phát triển để giải xấp xỉ phương trình HJB [1] sử dụng NN xây dựng bộ điều khiển dựa trên cấu trúc Actor-Critic (AC). Việc thực hiện cấu trúc AC có thể dựa trên việc cập nhật tuần tự tham số [1], [26], [16], [21], [28], [29], hoặc cập nhật tham số song song [26], [32], [28]. Trong đó, cấu trúc bộ điều khiển được sử dụng là online AC với tham số của Actor và Critic được cập nhật song song, chứng minh trong định lý đã chỉ ra rằng trạng thái của hệ kín, sai lệch tham số của AC bị giới hạn trong miền xác định. Lớp thuật toán [1], [26], [16], [21] được coi là phương pháp quy hoạch động xấp xỉ, chưa có yếu tố thích nghi, do việc giải phương trình Lyapunov yêu cầu rõ thông tin về động học hệ thống. Để giải quyết vấn đề

về bất định mô hình, [16], [21] sử dụng phương pháp nhận dạng hệ thống với cấu trúc điều khiển Actor-Critic-Identifier (ACI). Việc sử dụng nhận dạng hệ thống làm tăng đáng kể khối lượng tính toán gây ra khó khăn khi thực hiện thuật toán online, hơn nữa sai lệch do nhận dạng đối tượng gây ra có thể ảnh hưởng đến chất lượng điều khiển. Một cách tiếp cận khác do [29] đề xuất, sử dụng dạng tích phân của phương trình Lyapunov, Integral Reinforcement Learning (IRL), do đó loại bỏ được yêu cầu về thông tin động học nội của hệ thống, tuy nhiên vẫn cần một phần thông tin động học của hệ thống. [28] cải tiến thuật toán ở [29] thành cấu trúc AC online, giúp tăng tốc độ hội tụ của thuật toán. Bên cạnh đó, phương pháp Experience Replay (ER) lấy cảm hứng từ [4], được phát triển và áp dụng vào ADP giúp đảm bảo và tăng tốc độ hội tụ cho thuật toán [19], [27], [30].

Một vấn đề lớn trong học tăng cường là việc cân bằng giữa sự tìm tòi (Exploration) và sự tận dụng (Exploitation), dẫn đến hai phương pháp chính để giải quyết vấn đề này là on-policy và off-policy. Trong on-policy, tín hiệu dò được thêm vào bộ điều khiển. [15] đề xuất phương pháp tính toán tín hiệu dò thêm vào bộ điều khiển dựa trên phương pháp IRL giúp thỏa mãn điều kiện Persistent Excitation, đảm bảo sự hội tụ của tham số bộ điều khiển. Mặt khác đối với off-policy, hai luật điều khiển được tách riêng đóng vai trò riêng biệt. Một luật điều khiển có tính thăm dò cao, được áp dụng vào hệ thống để thu thập dữ liệu, trong khi đó, luật điều khiển còn lại được tính toán dựa trên dữ liệu thu thập được và hội tụ đến luật điều khiển tối ưu. Ứng dụng off-policy trong điều khiển được đề xuất trong [11]-[20] đã kiểm chứng sự thành công của thuật toán. Đặc biệt với cách tiếp cận của off-policy IRL, yêu cầu về thông tin động học của hệ được loại bỏ hoàn toàn, giúp thuật toán trở nên linh hoạt và mang tính thực tiễn cao.

Thuật toán ADP tuy được phát triển nhanh chóng, giúp giải quyết bài toán điều khiển thích nghi, tối ưu cho lớp đối tượng affine. Tuy nhiên những công bố cho vấn đề điều khiển bám quỹ đạo chưa thực sự nhận được nhiều quan tâm bởi với bài toán này khi hệ đã tiến được về quỹ đạo mong muốn, vẫn cần có một tín hiệu điều khiển giữ hệ nằm trên quỹ đạo này. Tín hiệu điều khiển này không tiến về 0 nên khi cho toàn bộ tín hiệu điều khiển mà ta sử dụng vào hàm chi phí hay hàm mục tiêu, hàm này sẽ không bị chặn. Một số tài liệu sử dụng cách tiếp cận cũ [13] tách bộ điều khiển thành hai phần, một phần là bộ điều khiển truyền thẳng được tiền xử lý, và một bộ điều khiển phản hồi được thiết kế dựa trên nguyên lý tối ưu. Tuy nhiên cách thiết kế như vậy chỉ được coi là bán tối ưu và đòi hỏi thông tin về đối tượng để thiết kế bộ điều khiển truyền thẳng. Một cách tiếp cận khác tổng quát hơn được giới thiệu trong [20], [33], [19], [18] giúp thiết kế bộ điều khiển bám một cách tổng quát, không yêu cầu tách bộ điều khiển và thực hiện thiết kế tối ưu với

hàm mục tiêu có thành phần hàm phạt. Thuật toán ADP sau đó được áp dụng vào giúp giải xấp xỉ luật điều khiển tối ưu với hàm tối ưu được sửa đổi.

Bố cục phần còn lại của luận án được trình bày như sau. Chương 2 sẽ giới thiệu nền tảng kiến thức cho nội dung đồ án. Cụ thể, phần 2.1 sẽ giới thiệu phương pháp quy hoạch động và phương trình HJB, HJI. Phần 2.2 giới thiệu thuật toán Policy Iteration cho hệ Affine. Ở chương 3.1 đưa ra vấn đề xấp xỉ hàm và tích hợp vấn đề này vào các giải thuật nêu trong chương 2 nhằm đưa ra thuật toán xấp xỉ tối ưu (ADP) trong chương 3.2. Đi cùng với đó là một số mô phỏng kiểm chứng cho tính hội tụ của thuật toán qua các ví dụ so sánh. Chương 4 trình bày ứng dụng của thuật toán ADP để xuất cho điều khiển bám quỹ đạo cho đối tượng tàu thủy và các kết quả mô phỏng đi kèm. Cuối cùng là kết luận và đưa ra các hướng phát triển trong tương lai.

Chapter 2

Thuật toán Policy Iteration cho hệ Affine

2.1 Quy hoạch động (Bellman)

Nguyên lý tối ưu:

Nguyên lý tối ưu của Bellman có nội dung như sau: "Mỗi đoạn cuối của quỹ đạo trạng thái tối ưu cũng sẽ là một quỹ đạo trạng thái tối ưu".

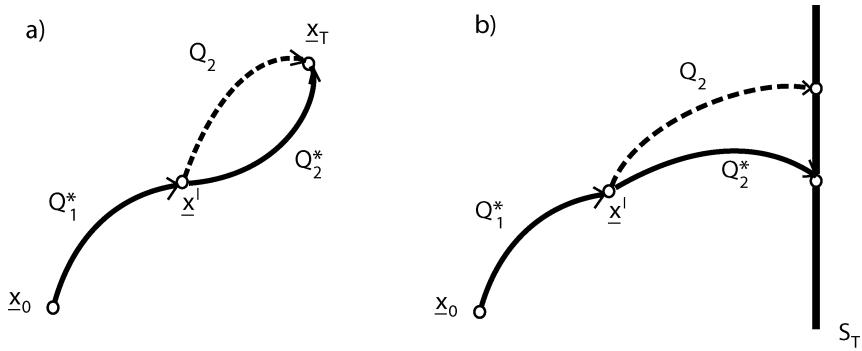


Figure 2.1: Mô tả nguyên lý tối ưu Bellman

Có thể kiểm chứng được ngay tính đúng đắn của nguyên lý Bellman nhờ hình minh họa (2.1). Giả sử quỹ đạo liền nét đi từ điểm \underline{x}_0 qua \underline{x}' đến \underline{x}_T là tối ưu, gồm hai đoạn 1 và 2, tương ứng với $V^* = V_1^* + V_2^*$, trong đó phần quỹ đạo cuối là 2 đi từ \underline{x}' đến \underline{x}_T có V_2^* , lại không phải tối ưu. Vậy thì phải tồn tại đoạn tối ưu từ \underline{x}' đến \underline{x}_T là đoạn 2' trên hình với $V_2 < V_2^*$. Suy ra, dọc theo đoạn 1 – 2', hàm $V = V_1^* + V_2$ sẽ có giá trị nhỏ hơn $V^* = V_1^* + V_2^*$ tính theo 1 – 2. Điều này trái với giả thiết rằng đoạn 1 – 2 là tối ưu.

2.1.1 Bài toán 2.1:

Xét bài toán với hệ liên tục không dừng:

$$\dot{\underline{x}} = f(\underline{x}, \underline{u}, t) \quad (2.1)$$

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

- Tập $U \subseteq \mathbb{R}^m$ là một tập con (hở hoặc đóng) trong không gian điều khiển \mathbb{R}^m .
- Khoảng thời gian T xảy ra quá trình tối ưu là cố định cho trước.
- Điểm đầu $\underline{x}(0) = \underline{x}_0$ là tùy ý, nhưng phải cho trước.
- Điểm cuối $\underline{x}(T) = \underline{x}_T$ là bất kỳ. Mục tiêu là xác định bộ điều khiển phản hồi trạng thái tối ưu $\underline{u}^* = \underline{u}(\underline{x}, t) \in \mathbb{U}$ đưa hệ đi từ \underline{x}_0 tới \underline{x}_T trong khoảng thời gian T sao cho hàm chi phí V cho bởi:

$$V = \int_0^T r(\underline{x}, \underline{u}, t) dt \rightarrow \min \quad (2.2)$$

đạt giá trị nhỏ nhất.

Trước khi thiết kế thuật toán ADP, ta đưa ra định nghĩa về tập tín hiệu điều khiển chấp nhận được (admissible control).

Định nghĩa 2.1.1. Một luật điều khiển $u(x) \in \Psi(x)$ được coi là tập các tín hiệu điều khiển chấp nhận được (admissible control) nếu như $u(x)$ khiến cho hệ thống (2.1) ổn định trong miền $x \in \Omega$ và hàm mục tiêu (2.2) ứng với luật điều khiển $u(x)$ là hữu hạn.

Nội dung phương pháp Trước tiên, từ nội dung nguyên lý tối ưu, ta định nghĩa hàm Bellman:

$$V^*(\underline{x}, t) = \min_{\underline{u} \in \mathbb{U}} \int_t^T r(\underline{x}, \underline{u}, \tau) d\tau \quad (2.3)$$

hay hàm Bellman chính là phần giá trị tối ưu của (2.3) 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 $\underline{x}(t) = \underline{x}'$ tới \underline{x}_T bất kỳ (hình 2.1).

Khi đó, theo nguyên lý tối ưu Bellman thì:

$$\begin{aligned} V^*(\underline{x}, t) &= \min_{\underline{u} \in \mathbb{U}} \int_t^T r(\underline{x}, \underline{u}, \tau) d\tau \\ &= \min_{\underline{u} \in \mathbb{U}} \left[\int_t^{t+\delta} r(\underline{x}, \underline{u}, \tau) d\tau + V^*(\underline{x}(t+\delta), t+\delta) \right] \end{aligned} \quad (2.4)$$

Định lí 2.1.2. Nếu $\underline{u}^* = \underline{u}(\underline{x}, t) \in \mathbb{U}$ là nghiệm của bài toán tối ưu (2.1.1) thì hàm Bellman (2.3) phải thỏa mãn:

$$a) \quad \begin{cases} \frac{\partial V^*(\underline{x}, t)}{\partial t} + \frac{\partial V^*(\underline{x}, t)}{\partial \underline{x}} \underline{f}(\underline{x}, \underline{u}^*, t) + r(\underline{x}, \underline{u}^*, t) = 0, \\ V^*(\underline{x}_T, T) = 0 \end{cases} \quad (2.5)$$

$$b) \quad \underline{u}^* = \arg \min_{\underline{u} \in \mathbb{U}} \left[\frac{\partial V^*(\underline{x}, t)}{\partial \underline{x}} \underline{f}(\underline{x}, \underline{u}, t) + r(\underline{x}, \underline{u}, t) \right] \quad (2.6)$$

Phương trình vi phân đạo hàm riêng (2.5) có tên gọi là phương trình Hamilton-Jacobi-Bellman, viết tắt là phương trình HJB.

Định lý đã chứng minh trong [23]. Từ định lý (2.1.2), ta có các bước tìm nghiệm của bài toán (2.1) như sau:

Thuật toán 1. Thuật toán quy hoạch động cho hệ liên tục:

1. Từ điều kiện (2.6) ta xác định quan hệ phải có của tín hiệu điều khiển tối ưu \underline{u}^* với \underline{x} và $\frac{\partial V(\underline{x}, t)}{\partial \underline{x}}$, nói cách khác là xác định quan hệ:

$$\underline{u}^* = \underline{u}(\underline{x}, V(\underline{x}, t)) \quad (2.7)$$

2. Thay quan hệ (2.7) vừa tìm được vào phương trình HJB (2.5) và xác định nghiệm $V(\underline{x}, t)$ của nó thỏa mãn điều kiện $V(\underline{x}_T, T) = 0$.
 3. Thay nghiệm $V(\underline{x}, t)$ tìm được vào (2.7) để có bộ điều khiển tối ưu $\underline{u}^* = \underline{u}(\underline{x}, t)$.
-

2.1.2 Bài toán 2.2:

Xét hệ thống phi tuyến:

$$\dot{\underline{x}} = \underline{f}(\underline{x}, \underline{u}, \underline{d}, t) \quad (2.8)$$

với $\underline{x} \in \mathbb{R}^n$ là trạng thái, $\underline{u} = [u_1, \dots, u_m] \in \mathbb{R}^m$ là tín hiệu điều khiển, $\underline{d} = [d_1, \dots, d_q] \in \mathbb{R}^q$ là nhiễu ngoài, $\underline{f}(\underline{x}, \underline{u}, \underline{d}, t) \in \mathbb{R}^n$ động học hệ.

Mục đích điều khiển: Xác định tín hiệu điều khiển \underline{u} để hệ phi tuyến (2.8) có L_2 -Gain nhỏ hơn hoặc bằng γ , tức là với mọi $\underline{d} \in L_2[0, \infty)$:

$$\frac{\int_t^\infty e^{-\alpha(\tau-t)} \|z(\tau)\|^2 d\tau}{\int_t^\infty e^{-\alpha(\tau-t)} \|\underline{d}(\tau)\|^2 d\tau} \leq \gamma^2 \quad (2.9)$$

với $\alpha > 0$ là hệ số suy giảm và γ là sự suy giảm tác động của nhiễu đầu vào $\underline{d}(t)$ vào hiệu năng đầu ra $z(t)$ được định nghĩa:

$$\|z(t)\|^2 = \underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u} \quad (2.10)$$

Sử dụng (2.10) vào (2.9), chúng ta có:

$$\int_t^\infty e^{-\alpha(\tau-t)} (\underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u}) d\tau \leq \gamma^2 \int_t^\infty e^{-\alpha(\tau-t)} (\underline{d}^T \underline{d}) d\tau \quad (2.11)$$

Điểm khác nhau cơ bản giữa định nghĩa 2 và định nghĩa tiêu chuẩn của điều khiển bám H_∞ ([2], định nghĩa 5.2.1]) là điều kiện suy giảm nhiễu được đưa ra. Từ việc toàn bộ tín hiệu điều khiển và sai lệch bám bị phạt trong điều kiện suy giảm nhiễu (2.11), vấn đề trong định nghĩa 1 đưa ra một phương pháp tối ưu, ngược lại với định nghĩa tiêu chuẩn kết quả là một phương pháp bán tối ưu, được nêu ra trong [2]

Chú ý 2.1.3. *Hàm hiệu năng bên trái của điều kiện suy giảm (2.11) bao gồm một hàm phạt dương trên sai lệch bám và một hàm phạt dương cho tín hiệu điều khiển. Sử dụng hệ số suy giảm là cần thiết bởi thành phần truyền thẳng của đầu vào điều khiển nhìn chung không hội tụ về không, do đó hàm phạt tín hiệu điều khiển trong hàm hiệu năng không có hệ số suy giảm sẽ khiến cho nó không bị chặn.*

Chú ý 2.1.4. *Những nghiên cứu trước về bám tối ưu H_∞ chia tín hiệu điều khiển thành phần phản hồi và phần truyền thẳng. Đầu tiên, phần truyền thẳng thu được riêng biệt không phụ thuộc vào bất kì yêu cầu tối ưu nào. Sau đó, vấn đề thiết kế phần phản hồi được giảm xuống một vấn bài toán tối ưu H_∞ thông dụng. Ngược lại, trong đề xuất mới này, cả phần truyền thẳng và phản hồi của tín hiệu điều khiển được thu cùng lúc và tối ưu như là kết quả của L_2 -gain với một hệ số suy giảm trong (2.11).*

Phương pháp điều khiển cho vấn đề bám H_∞ với điều kiện suy giảm được đề xuất trong (2.11) được nêu ra trong các phần sau. Chúng tôi sẽ cho thấy điều kiện suy giảm nhiều cho phép chúng ta tìm được cả phần truyền thẳng và phản hồi của tín hiệu điều khiển một cách đồng thời, do đó mở rộng phương pháp off-policy RL để giải quyết các bài toán khó mà không yêu cầu hiểu biết về hệ thống động học của hệ.

Phương trình HJI Trong phần này, một phương trình HJI được đưa ra nhằm đưa ra lời giải cho bài toán điều khiển H_∞ .

Dựa trên (2.11), ta đưa ra hàm hiệu năng

$$J(\underline{u}, \underline{d}) = \int_t^\infty e^{-\alpha(\tau-t)} (\underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u} - \gamma^2 \underline{d}^T \underline{d}) d\tau \quad (2.12)$$

Chú ý 2.1.5. *Chú ý rằng bài toán tìm luật điều khiển thỏa mãn điều kiện L_2 -gain bị chặn cho vấn đề bám tối ưu tương đương với cực tiểu hàm hiệu năng đã được suy giảm (2.12) trên hệ thống (2.8).*

Vấn đề điều khiển H_∞ liên quan gần với two-player zero-sum differential game theory [3]. Rõ ràng rằng vấn đề điều khiển H_∞ tương đương với giải quyết bài toán zero-sum game dưới đây [3]

$$V^*(\underline{x}(t)) = J(\underline{u}^*, \underline{d}^*) = \min_{\underline{u}} \max_{\underline{d}} J(\underline{u}, \underline{d}) \quad (2.13)$$

với J được định nghĩa trong (2.12) và $V^*(\underline{x}(t))$ được định nghĩa là hàm giá trị tối ưu. Vấn đề điều khiển two-player zero-sum game này có một nghiệm duy nhất nếu một điểm yên ngựa trong lý thuyết trò chơi tồn tại, nếu điều kiện Nash sau đây thỏa mãn:

$$V^*(\underline{x}(t)) = \min_{\underline{u}} \max_{\underline{d}} J(\underline{u}, \underline{d}) = \max_{\underline{d}} \min_{\underline{u}} J(\underline{u}, \underline{d}) \quad (2.14)$$

Đạo hàm từng phần phương trình (2.12) theo t và x và với $V(\underline{x}(t)) = J(\underline{u}(t), \underline{d}(t))$ thu được phương trình Hamilton-Jacobi-Isaacs (HJI) như sau:

$$H(V^*, \underline{u}^*, \underline{d}^*) \triangleq \underline{x}^T Q \underline{x} + \underline{u}^{*T} R \underline{u}^* - \gamma^2 \underline{d}^{*T} \underline{d}^* - \alpha V^* + V_{\underline{x}}^{*T} \underline{f}(\underline{x}, \underline{u}^*, \underline{d}^*, t) = 0 \quad (2.15)$$

với $V_{\underline{x}}^* = \partial V^* / \partial \underline{x}$.

2.2 Thuật toán Policy Iteration cho hệ Affine

Bài toán 2.1: Xét hệ Affine liên tục theo thời gian:

$$\dot{\underline{x}} = \underline{f}(\underline{x}) + \underline{g}(\underline{x}) \underline{u} \quad (2.16)$$

với $\underline{x} \in \mathbb{R}^n, \underline{u} \in \mathbb{R}^m, \underline{f}(\underline{0}) = \underline{0}$ và $\underline{f}(\underline{x}) + \underline{g}(\underline{x}) \underline{u}$ thỏa mãn tích chất liên tục Lipschitz trong tập $\Omega \in \mathbb{R}^n$.

Xét hàm mục tiêu:

$$V(x, u) = \int_t^\infty r(\underline{x}, \underline{u}) d\tau \quad (2.17)$$

Trong đó, $r(\underline{x}, \underline{u}) = Q(\underline{x}) + \underline{u}^T R \underline{u}$. Với $Q(\underline{x})$ là hàm xác định dương của \underline{x} , R là ma trận đối xứng xác định dương.

Mục tiêu thiết kế là tìm luật điều khiển $\underline{u}(\underline{x})$ giúp ổn định hệ thống (2.1) và tối thiểu hóa hàm mục tiêu (2.2). Áp dụng (2.6) cho hệ affine nêu trên ta thu được tín hiệu điều khiển tối ưu:

$$\underline{u}^*(x) = -\frac{1}{2} R^{-1} g^T(x) \frac{\partial V^*(\underline{x})^T}{\partial \underline{x}} \quad (2.18)$$

Từ phương trình (2.5) ta thu được phương trình HJB cho hệ affine với hàm mục tiêu (2.17) và $\underline{x}_T = 0$ như sau:

$$\begin{cases} H^*(\underline{x}, \underline{u}^*, V_{\underline{x}}^*) = \frac{\partial V^*(\underline{x})}{\partial \underline{x}} (\underline{f}(\underline{x}) + \underline{g}(\underline{x}) \underline{u}^*) + Q(\underline{x}) + \underline{u}^{*T} R \underline{u}^* = 0 \\ V^*(\underline{0}) = 0 \end{cases} \quad (2.19)$$

Tuy nhiên, việc giải trực tiếp phương trình HJB (2.19) với tín hiệu điều khiển tối ưu (2.18) là vô cùng khó khăn bởi đây là phương trình vi phân phi tuyến. Đó là lý do thuật toán xấp xỉ để quy được phát triển bởi Saridis và Lee (1979) [24].

Kĩ thuật xấp xỉ để quy bấy giờ được áp dụng vào (2.18) và (2.19). Ta có bỗng đẽ:

Bố đề 2.2.1. Nếu $u^{(i)} \in \Psi(\Omega)$, và $V^{(i)} \in C^1(\Omega)$ thỏa mãn phương trình HJB $H(\underline{x}, \underline{u}^{(i)}, V^{(i)}) = 0$ với điều kiện $V^{(i)}(0) = 0$, thì tín hiệu điều khiển mới như sau:

$$u^{(i+1)}(x) = -\frac{1}{2} R^{-1} g^T(x) \frac{\partial V^{(i)}(\underline{x})^T}{\partial \underline{x}} \quad (2.20)$$

là một tín hiệu điều khiển chấp nhận được cho hệ (2.16) trên miền Ω . Hơn nữa, nếu $V^{(i+1)}$ là hàm xác định dương duy nhất thỏa mãn phương trình $H(\underline{x}, \underline{u}^{(i)}, V^{(i)}) = 0$, với điều kiện biên $V^{(i+1)}(0) = 0$, thì $V^*(\underline{x}) \leq V^{(i+1)}(\underline{x}) \leq V^{(i)}(\underline{x}) \quad \forall \underline{x} \in \Omega$.

đã được chứng minh trong [1].

Từ bối cảnh (2.2.1) ta đưa ra thuật toán PI (Policy Iteration) cho hệ affine (2.16) và hàm mục tiêu (2.17) như sau:

Thuật toán 2. PI cho bài toán điều khiển tối ưu

Bước 1: $\forall \underline{x} \in \Omega_{\underline{x}}$, khởi tạo luật điều khiển chấp nhận được $\underline{u}^{(0)}(\underline{x})$ và giá trị $V^{(0)}(\underline{x}) = 0$.

- $i \leftarrow 0$

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

- Xác định $V^{(i+1)}(\underline{x})$ từ hệ phương trình:

$$\begin{cases} \frac{\partial V^{(i+1)}(\underline{x})}{\partial \underline{x}} (\underline{f}(\underline{x}) + \underline{g}(\underline{x})\underline{u}^{(i)}) + Q(\underline{x}) + (\underline{u}^{(i)})^T R \underline{u}^{(i)} = 0 \\ V^{(i+1)}(0) = 0 \end{cases} \quad (2.21)$$

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

- Cập nhật:

$$u^{(i+1)}(\underline{x}) = -\frac{1}{2}R^{-1}g^T(\underline{x}) \frac{\partial V^{(i+1)}(\underline{x})^T}{\partial \underline{x}} \quad (2.22)$$

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

Định lí 2.2.2. Nếu $u^{(0)} \in \Psi(\Omega)$, thì với thuật toán 2 có $u^{(i)} \in \Psi(\Omega), \forall i \geq 0$, hơn nữa, $V^{(i)} \rightarrow V^*$, $u^{(i)} \rightarrow u^*$ trong Ω .

đã chứng minh trong [1].

Do đó thích nghi/xấp xỉ được ứng dụng vào thuật toán quy hoạch động nhằm giải quyết vấn đề này. Ta xấp xỉ giá trị của $u^*(\underline{x})$ bởi $\hat{u}(\underline{x})$ (actor) và

$V_{\underline{x}}^*(\underline{x})$ bởi $\hat{V}_{\underline{x}}(\underline{x})$ (critic) với $V_{\underline{x}}^*(\underline{x}) = \frac{\partial V^*(\underline{x})}{\partial \underline{x}} \in \mathbb{R}^{1 \times n}$ phương trình HJB được xấp xỉ thành:

$$\hat{H}^*(\underline{x}, \hat{u}, \hat{V}_{\underline{x}}) = \frac{\partial \hat{V}(\underline{x})}{\partial \underline{x}} (\underline{f}(\underline{x}) + \underline{g}(\underline{x})\hat{u}) + Q(\underline{x}) + \hat{u}^T R \hat{u} \quad (2.23)$$

Bài toán 2.2: Xét hệ Affine liên tục theo thời gian:

$$\dot{\underline{x}} = \underline{f}(\underline{x}) + \underline{g}(\underline{x})\underline{u} + \underline{k}(\underline{x})\underline{d} \quad (2.24)$$

với $\underline{x} \in \mathbb{R}^n$, $\underline{u} \in \mathbb{R}^m$ và $\underline{f}(\underline{x}) + \underline{g}(\underline{x})\underline{u} + \underline{k}(\underline{x})\underline{d}$ thỏa mãn tính chất liên tục Lipschitz trong tập $\Omega \in \mathbb{R}^n$.

Xét hàm mục tiêu:

$$V(\underline{x}, \underline{u}, \underline{d}) = \int_t^\infty e^{-\alpha(\tau-t)} (\underline{x}^T Q \underline{x} + \underline{u}^T R \underline{u} - \gamma^2 \underline{d}^T \underline{d}) d\tau \quad (2.25)$$

Với Q, R là các ma trận đối xứng xác định dương.

Mục tiêu thiết kế là tìm luật điều khiển $\underline{u}(\underline{x})$ giúp ổn định hệ thống (2.24) và tối thiểu hóa hàm mục tiêu (2.25). Áp dụng (2.20) cho hệ affine nêu trên ta thu được tín hiệu điều khiển tối ưu:

$$\underline{u}^*(\underline{x}) = -\frac{1}{2} R^{-1} \underline{g}^T(\underline{x}) \frac{\partial V^*(\underline{x})^T}{\partial \underline{x}} \quad (2.26)$$

và nhiều:

$$\underline{d}^*(\underline{x}) = \frac{1}{2\gamma^2} K^{-1} \frac{\partial V^*(\underline{x})^T}{\partial \underline{x}} \quad (2.27)$$

Từ phương trình (2.15) ta thu được phương trình HJI cho hệ affine với hàm mục tiêu (2.25) và $\underline{x}_T = 0$ như sau:

$$\begin{cases} H(V^*, \underline{u}^*, \underline{d}^*) \triangleq \underline{x}^T Q \underline{x} + \underline{u}^{*T} R \underline{u}^* - \gamma^2 \underline{d}^{*T} \underline{d}^* \\ \quad - \alpha V^* + V_{\underline{x}}^{*T} (\underline{f}(\underline{x}) + \underline{g}(\underline{x})\underline{u}^* + \underline{k}(\underline{x})\underline{d}^*) = 0 \\ V^*(0) = 0 \end{cases} \quad (2.28)$$

Tương tự với bài toán 2.1, việc giải trực tiếp phương trình HJI nêu trên là vô cùng khó khăn bởi đây là phương trình vi phân phi tuyến. Do đó kĩ thuật xấp xỉ đê quy được áp dụng. [20] đưa ra thuật toán PI (Policy Iteration) cho bài toán 2.2 như sau:

Thuật toán 3. PI cho bài toán \mathbb{H}_∞

Bước 1: $\forall \underline{x} \in \Omega_{\underline{x}}$, khởi tạo luật điều khiển chấp nhận được $\underline{u}^{(0)}(\underline{x})$, nhiều $\underline{d}^{(0)}(\underline{x})$ và giá trị $V^{(0)}(\underline{x}) = 0$.

- $i \leftarrow 0$

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

- Xác định $V^{(i+1)}(\underline{x})$ từ hệ phương trình:

$$\begin{cases} -\alpha V^{(i+1)} + \frac{\partial V^{(i+1)}(\underline{x})}{\partial \underline{x}} \left(\underline{f}(\underline{x}) + \underline{g}(\underline{x})\underline{u}^{(i)} + \underline{k}(\underline{x})\underline{d}^{(i)} \right) \\ \quad + \underline{x}^T Q \underline{x} + (\underline{u}^{(i)})^T R \underline{u}^{(i)} - \gamma^2 (\underline{d}^{(i)})^T \underline{d}^{(i)} = 0 \\ V^{(i+1)}(\underline{0}) = 0 \end{cases} \quad (2.29)$$

Bước 3: Cập nhật luật điều khiển và nhiễu cho vòng lặp kế tiếp theo.

- Cập nhật:

$$u^{(i+1)}(\underline{x}) = -\frac{1}{2} R^{-1} g^T(\underline{x}) \frac{\partial V^{(i+1)}(\underline{x})^T}{\partial \underline{x}} \quad (2.30)$$

$$d^{(i+1)}(\underline{x}) = \frac{1}{2\gamma^2} K^{-1} \frac{\partial V^{(i+1)}(\underline{x})^T}{\partial \underline{x}} \quad (2.31)$$

- Nếu thỏa mãn tiêu chuẩn hời tu sao cho $\|V^{(i+1)} - V^{(i)}\| \leq v$ với v là số dương đủ nhỏ thì gán $\underline{u}^*(\underline{x}) = \underline{u}^{(i+1)}(\underline{x})$, $\underline{d}^*(\underline{x}) = \underline{d}^{(i+1)}(\underline{x})$ và $V^*(\underline{x}) = V^{(i+1)}(\underline{x})$, kết thúc giải thuật.
- Nếu không thỏa mãn, gán $i \leftarrow i + 1$ và quay lại bước 2.

Do đó thích nghi/xấp xỉ được ứng dụng vào thuật toán quy hoạch động nhằm giải quyết vấn đề này. Ta xấp xỉ giá trị của $\underline{u}^*(\underline{x})$ bởi $\hat{\underline{u}}(\underline{x})$, $\underline{d}^*(\underline{x})$ bởi $\hat{\underline{d}}(\underline{x})$ (actor) và $V_x^*(\underline{x})$ bởi $\hat{V}_x(\underline{x})$ (critic) với $V_x^*(\underline{x}) = \frac{\partial V^*(\underline{x})}{\partial \underline{x}} \in \mathbb{R}^{1 \times n}$ phương trình HJB được xấp xỉ thành:

$$\hat{H}^*(\underline{x}, \hat{\underline{u}}, \hat{\underline{d}}, \hat{V}_x) = -\alpha \hat{V}(\underline{x}) + \frac{\partial \hat{V}(\underline{x})}{\partial \underline{x}} \left(\underline{f}(\underline{x}) + \underline{g}(\underline{x})\hat{\underline{u}} + \underline{k}(\underline{x})\hat{\underline{d}} \right) + \underline{x}^T Q \underline{x} + \hat{\underline{u}}^T R \hat{\underline{u}} - \gamma^2 \hat{\underline{d}}^T \hat{\underline{d}} \quad (2.32)$$

Chapter 3

Giải thuật ADP qui hoạch động thích nghi sử dụng xấp xỉ hàm

3.1 Xấp xỉ hàm và điều kiện PE

Trong chương này ta sẽ tìm hiểu lý thuyết xấp xỉ hàm, cũng như định nghĩa, tính chất của điều kiện PE để có thể áp dụng trong việc điều khiển và chứng minh tính ổn định.

3.1.1 Nguyên lý xấp xỉ hàm mạng NN

Định lí 3.1.1. (*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.*

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 lí 3.1.2. *Mạng NN có thể xấp xỉ giá trị của các hàm liên tục trên tập con của \mathbb{R}^n , với độ chính xác tùy ý.*

Định lí 3.1.3. *Gọi $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ là một hàm liên tục bất kì (còn gọi là activation function). Gọi $K \subseteq \mathbb{R}^n$ là tập compact. Không gian các hàm liên tục trên tập K kí hiệu là $C(K)$. Gọi \mathbb{M} là không gian các hàm có dạng:*

$$F(X) = \sum_{i=1}^N v_i \varphi(\omega_i^T x + b_i) \quad (3.1)$$

cho tất cả số nguyên $N \in \mathbb{N}$, hằng số $v_i, b_i \in \mathbb{R}$, vector $\omega_i \in \mathbb{R}^m$ với $i = 1, \dots, N$. thì, khi và chỉ khi φ 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 $F \in \mathbb{M}$ để:

$$|F(x) - f(x)| < \epsilon \quad (3.2)$$

với mọi $x \in K$

Table 3.1: 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	$f(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x > 0 \end{cases}$
4	Gaussian	$f(x) = e^{-x^2}$
5	Sinc	$f(x) = \begin{cases} 1 & \text{for } x = 0 \\ \frac{\sin(x)}{x} & \text{for } x \neq 0 \end{cases}$

Sử dụng các định lý nêu trên ta có giả thiết sau:

Giả thiết 1. Cho hàm liên tục $\Upsilon : \mathbb{S} \rightarrow \mathbb{R}^n$ với \mathbb{S} là một tập liên thông, tồn tại các trọng số lý tưởng W, V sao cho hàm này có thể biểu diễn bởi mạng NN:

$$\Upsilon(x) = W^T \phi(x) + \epsilon(x) \quad (3.3)$$

với ϕ là activation function, ϵ là sai lệch tái cấu trúc.

Giả thiết 2. Các trọng số lý tưởng W là bị chặn, $\|W\| \leq \bar{W}$

Giả thiết 3. Hàm hoạt động $\phi(X)$ và đạo hàm riêng theo X là bị chặn

Giả thiết 4. Sai lệch tái cấu trúc bị chặn tức là $\|\epsilon\| \leq \bar{\epsilon}$ và đạo hàm riêng theo X của nó cũng bị chặn.

3.1.2 Điều kiện PE

Một hàm số S liên tục từng phần, bị chặn toàn cục, đi từ $[0, \infty) \rightarrow \mathbb{R}^m$ được cho là thỏa mãn điều kiện PE nếu tồn tại các hằng số dương α_1, α_2 và T_0 sao cho:

$$\alpha_1 I \geq \int_{t_0}^{t_0+T_0} S(\tau) S(\tau)^T d\tau \geq \alpha_2 I, \forall t_0 \geq 0 \quad (3.4)$$

Trong đó $I \in \mathbb{R}^{m \times m}$ là ma trận đơn vị. Theo như định nghĩa trên đây, 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 .

Cần chú ý rằng nếu S thỏa mãn điều kiện PE trong khoảng thời gian $[t_0, t_0 + T_0]$, thì nó thỏa mãn điều kiện PE với mọi khoảng có độ lớn bất kì $T_1 > T_0$.

3.2 Giải thuật ADP

Để xấp xỉ luật điều khiển online trong giải thuật PI, các nghiên cứu [10], [14], [22], [31] (xem thêm các tài liệu tham khảo trong đó) đề xuất cấu trúc ADP (còn gọi là cấu trúc AC) sử dụng một, hai hoặc ba xấp xỉ hàm (Hình 3.1, 3.2). Các xấp xỉ hàm trong ADP chủ yếu là các NN truyền thẳng một lớp. NN thứ nhất đóng vai trò critic (Critic Neural Network (CNN)) dùng để xấp xỉ online hàm đánh giá tối ưu (2.21) trong bước 2 của thuật toán (2), các NN còn lại đóng vai trò actor (Actor Neural Network (ANN)) xấp xỉ luật điều khiển tối ưu (2.22). Luật cập nhật tham số của các NN phụ thuộc lẫn nhau. ANN cập nhật trọng số sử dụng tín hiệu từ CNN. Tới nay rất nhiều hình thức xấp xỉ nhằm giải quyết bài toán đã đưa ra, đồng thời cũng có rất nhiều hình thức để phân loại các thuật toán đó theo hình thức cập nhật hay số lượng NN, nhưng trong phạm trù đồ án này, chúng em đề cập tới hai dạng

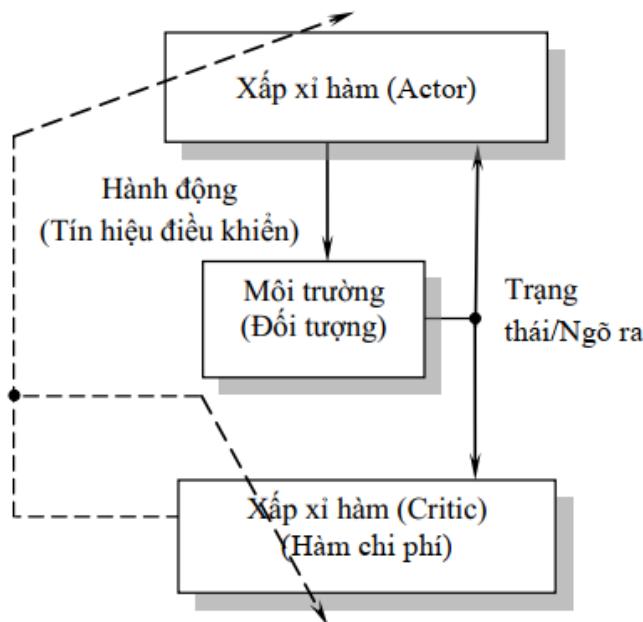
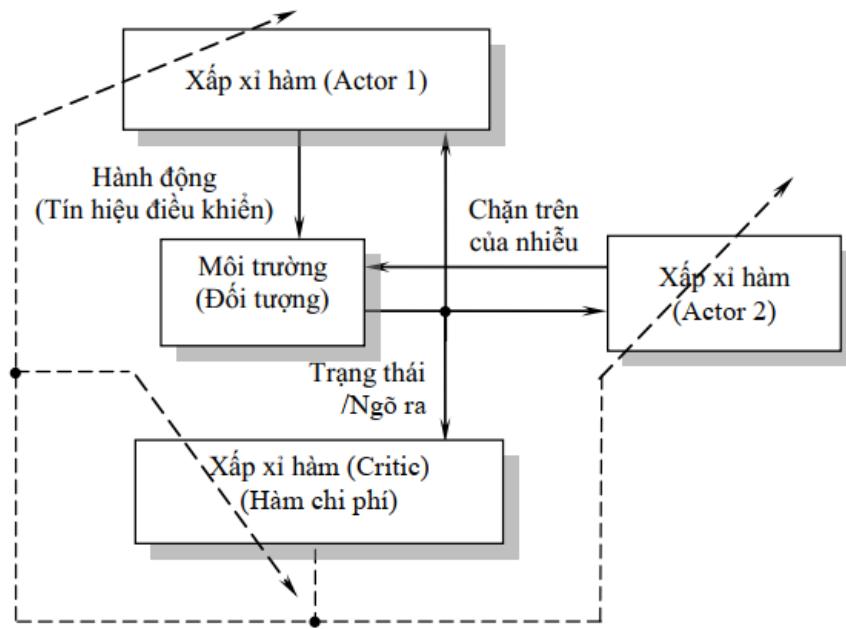


Figure 3.1: Cấu trúc ADP sử dụng hai xấp xỉ hàm trong điều khiển tối ưu

thích nghi cơ bản của giải thuật ADP: Actor-Critic hai NN truyền thống yêu cầu thông tin mô hình hệ thống (cả $f(\underline{x})$ và $g(\underline{x})$ trong hệ thống (2.16)) và Integral Reinforcement Learning một NN yêu cầu một phần thông tin mô hình hệ thống ($g(\underline{x})$ trong hệ thống (2.16)).

Để thuận tiện thì từ chương này trở đi nhằm đơn giản hóa các ký tự, chúng em sẽ bỏ đi dấu gạch dưới các biến (kí hiệu này thể hiện đó là giá trị vector).

Figure 3.2: Cấu trúc ADP sử dụng ba xấp xỉ hàm trong điều khiển tối ưu H_∞

3.2.1 ADP cấu trúc Actor-Critic

Áp dụng khả năng xấp xỉ của mạng NN đã nêu trong chương (3.1.1), theo công thức (3.3), hàm giá trị tối ưu và tín hiệu điều khiển tối ưu trong chương 2 có thể được xấp xỉ dưới mô hình mạng NN như sau:

$$\begin{aligned} V^*(x) &= W^T \phi(x) + \varepsilon_\nu(x) \\ u^*(x) &= -\frac{1}{2} R^{-1} g^T(x) \left(\frac{\partial \phi(x)}{\partial x}^T W + \frac{\partial \varepsilon_\nu(x)}{\partial x}^T \right) \end{aligned} \quad (3.5)$$

Trong đó $W \in \mathbb{R}^N$ là ma trận trọng số lý tưởng, N là số neuron, $\phi = [\phi_1(x), \phi_2(x), \dots, \phi_N(x)]^T \in \mathbb{R}^N$ là một activation function trơn có $\phi_i(0) = 0, \phi'_i(0) = 0, \forall i = 1, \dots, N$ và $\varepsilon_\nu(\cdot) \in \mathbb{R}$ là sai số do cấu trúc mạng.

Giả thiết 5. *Những activation function $\phi_i(x) : i = 1 \dots N$ được lựa chọn sao cho khi $N \rightarrow \infty, \phi(x)$ chứa tất cả hàm cơ bản độc lập với nhau của $V^*(x)$.*

Áp dụng giả thiết 5 và định lý xấp xỉ bậc cao Weierstrass, cả $V^*(x)$ và $\frac{\partial V^*(x)}{\partial x}$ có thể được xấp xỉ toàn cục bởi NNs trong (3.5), đó là khi $N \rightarrow \infty$, sai số xấp xỉ $\varepsilon_\nu(x), \varepsilon'_\nu(x) \rightarrow 0$ [1]. Thành phần Critic $\hat{V}(x)$ và Actor $\hat{u}(x)$ xấp xỉ hàm giá trị tối ưu và tín hiệu điều khiển tối ưu trong (3.5) sẽ cho bởi:

$$\hat{V}(x) = \hat{W}_c^T \phi(x); \quad \hat{u}(x) = -\frac{1}{2} R^{-1} g^T(x) \phi'^T(x) \hat{W}_a \quad (3.6)$$

với $\hat{W}_c(t) \in \mathbb{R}^N$ và $\hat{W}_a(t) \in \mathbb{R}^N$ là các xấp xỉ trọng số lý tưởng của NN Critic và Actor. Theo chương 12 [16], luật cập nhật của hai trọng số dựa theo tối thiểu hóa sai số hàm Bellman $\delta_{hjb}(\cdot) = \hat{H} - H^*$ với H^* trong (2.22) và \hat{H} là xấp xỉ của nó:

$$\delta_{hjb} = \hat{W}_c^T \omega + r(x, \hat{u}) \quad (3.7)$$

với $\omega(x, \hat{u}) = \phi'(x)F_{\hat{u}}(x, \hat{u}) \in R^N$. Từ đó ta có luật cập nhật NNs của Critic và Actor dựa vào recursive LS như sau:

$$\dot{\hat{W}}_c = \eta_c \Gamma \frac{\omega}{1 + \gamma \omega^T \Gamma \omega} \delta_{hjb} \quad (3.8)$$

với $\gamma, \eta_c \in \mathbb{R}$ là các hằng số dương, và $\Gamma(t) \in \mathbb{R}^{N \times N}$ được cập nhật như sau:

$$\dot{\Gamma} = \eta_c \Gamma \frac{\omega \omega^T}{1 + \gamma \omega^T \Gamma \omega} \Gamma \quad (3.9)$$

và Actor:

$$\dot{\hat{W}}_a = -\frac{\eta_{a1}}{\sqrt{1 + \omega^T \omega}} \left(\hat{W}_c^T \phi' \frac{\partial F_{\hat{u}}}{\partial \hat{u}} \frac{\partial \hat{u}}{\partial \hat{W}_a} + \hat{W}_a^T \phi' G \phi'^T \right)^T \delta_{hjb} - \eta_{a2} (\hat{W}_a - \hat{W}_c) \quad (3.10)$$

với $G(x) = g(x)R^{-1}g(x)^T \in \mathbb{R}^{n \times n}$, $\eta_{a1}, \eta_{a2} \in \mathbb{R}$ là các hằng số thích hợp.

Chú ý 3.2.1. Ta thấy tín hiệu điều khiển được xác định dựa theo gradient của hàm giá trị tối ưu (3.5), NN Critic trong (3.6) có thể sử dụng để xác định Actor mà không cần sử dụng NN Actor. Tuy nhiên, nhằm đơn giản luật cập nhật và quá trình phân tích tính ổn định hệ thống, hai NN riêng biệt đã được sử dụng cho Critic và Actor [26].

Chứng minh ổn định và sự hội tụ về giá trị tối ưu của thuật toán đã được trình bày đầy đủ trong [1]. Sau đây là ví dụ cho thuật toán này:

Ví dụ 1. Xét hệ phi tuyến affine:

$$\dot{x} = \begin{bmatrix} -x_1 + x_2 \\ -0.5x_1 - 0.5x_2(1 - (\cos(2x_1) + 2)^2) \end{bmatrix} + \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix} u \quad (3.11)$$

với hàm mục tiêu:

$$\begin{aligned} J(x, u) &= \int_0^\infty (x^T Q x + u^T R u) d\tau \\ Q(x) &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad R = 1. \end{aligned} \quad (3.12)$$

Giải phương trình HJB ta thu được hàm Bellman và tín hiệu điều khiển tối ưu:

$$\begin{aligned} V^*(x) &= \frac{1}{2}x_1^2 + x_2^2 \\ u^*(x) &= -(\cos(2x_1) + 2)x_2 \end{aligned} \quad (3.13)$$

Nhằm chứng minh tính đúng đắn thuật toán, ta chọn activation function của hai mạng NNs có dạng:

$$\phi(x) = \begin{bmatrix} x_1^2 & x_1 x_2 & x_2^2 \end{bmatrix}^T \quad (3.14)$$

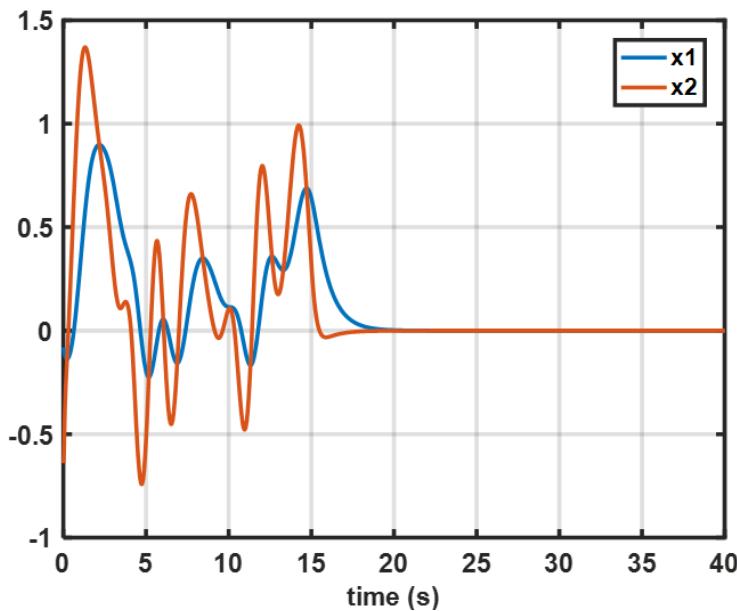
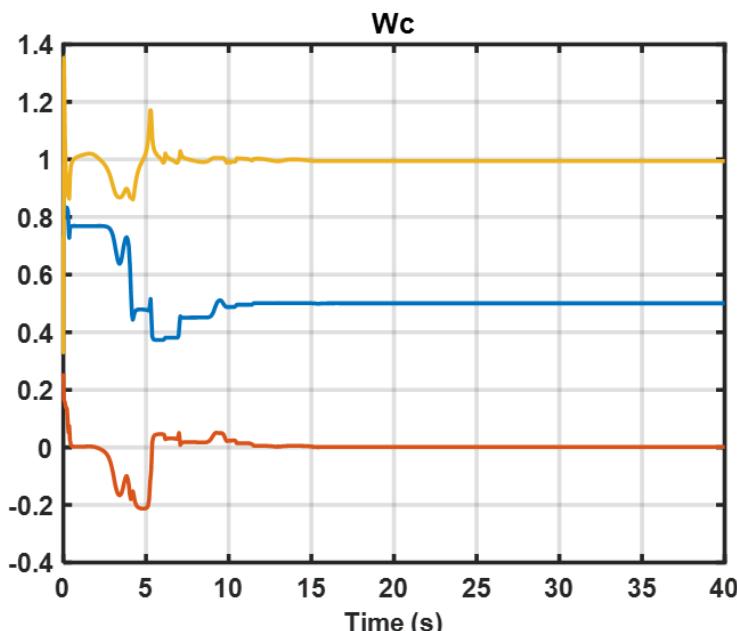


Figure 3.3: Trạng thái của hệ thống với cấu trúc Actor Critic

Tại thời điểm đầu trọng số \hat{W}_a và \hat{W}_c của hai mạng NNs được khởi tạo ngẫu nhiên trong khoảng $[-2, 2]$, từ phương trình (3.13) và (3.14) ta có $W^* = [0.5 \ 0 \ 1]^T$. Với các hệ số điều khiển $\eta_{a1} = 10, \eta_{a2} = 50, \eta_c = 20, \gamma = 0.005$ và nhằm đảm bảo điều kiện PE cho ω ta thêm tín hiệu thăm dò như sau vào hệ thống trong khoảng thời gian đầu:

$$n(t) = \sin^2(t) \cos(t) + \sin^2(2t) \cos(0.1t) + \sin^2(-1.2t) \cos(0.5t) + \sin^5(t)$$

Figure 3.4: Sự hội tụ của trọng số W_c với cấu trúc Actor Critic

Ta thu được kết quả mô phỏng hệ thống ổn định với tín hiệu điều khiển

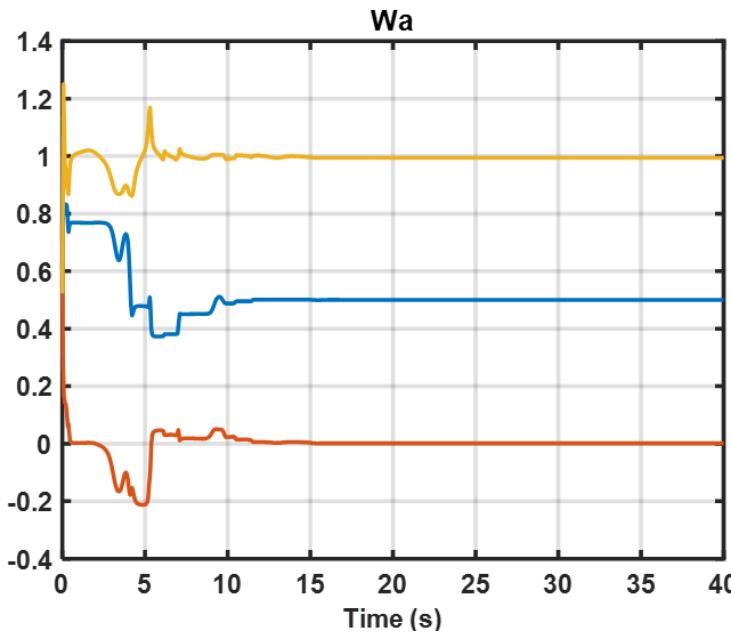


Figure 3.5: Sự hội tụ của trọng số W_a với cấu trúc Actor Critic

của thuật toán, sự hội tụ về đúng giá trị lý tưởng W^* của trong số \hat{W}_c và \hat{W}_a miêu tả trong hình (3.3-3.5). Qua đó cho thấy tính đúng đắn của thuật toán.

3.2.2 Online On-Policy Integral Reinforcement Learning

Thuật toán đã được trình bày ở chương 3.2.1 sử dụng cách tiếp cận giải phương trình HJB (2.5), trong đó phương trình yêu cầu thông tin về động học của hệ thống. Trong tài liệu chương 12 [16], [21], tác giả loại bỏ yêu cầu về động học hệ thống bằng việc sử dụng neuron để xấp xỉ hàm $f(x)$ chương 12 [16] và $g(x)$ [21]. Tuy nhiên, cách tiếp cận này bị hạn chế do yêu cầu khối lượng tính toán lớn, gây trở ngại khi thực hiện thuật toán trong thời gian thực (online), bên cạnh đó sai lệch xấp xỉ hàm có thể ảnh hưởng đến chất lượng của vòng kín. Trong phần này, một cách tiếp cận mới được giới thiệu trong [29] giúp loại bỏ yêu cầu về thông tin động học nội $f(x)$ mà không sử dụng xấp xỉ hàm.

Thay vì chuyển về dạng phương trình vi phân như (2.5), ta để phương trình HJB ở dạng như phương trình (2.4) ta có:

$$\begin{aligned}
 V^*(x(t)) &= \min_{u \in \mathbb{U}} \int_t^\infty r(x, u) d\tau \\
 &= \min_{u \in \mathbb{U}} \left[\int_t^{t+T} r(x, u) d\tau + V^*(x(t+T)) \right] \\
 &= \int_t^{t+T} r(x, u^*) d\tau + V^*(x(t+T))
 \end{aligned} \tag{3.15}$$

Từ đó thuật toán PI (2) có thể biến đổi thành dạng IRL như sau:

Thuật toán 4. PI (*Online On-Policy IRL*)

Bước 1: $\forall \underline{x} \in \Omega_{\underline{x}}$, khởi tạo luật điều khiển chấp nhận được $\underline{u}^{(0)}(\underline{x})$ và giá trị $V^{(0)}(\underline{x}) = 0$.

- Cho tín hiệu điều khiển $\underline{u}^{(0)}$ vào hệ thống và thu thập thông tin cần thiết của hệ thống về trạng thái, tín hiệu điều khiển tại N trích mẫu khác nhau trong khoảng thời gian T .
- $i \leftarrow 0$

Bước 2: Sử dụng các thông tin đã thu thập được về hệ thống nhằm xấp xỉ hàm $V^{(i+1)}(\underline{x})$ ở bước $i + 1$ với tín hiệu điều khiển vào hệ thống là $\underline{u}^{(i)}$:

- Xác định $V^{(i+1)}(\underline{x})$ từ hệ phương trình:

$$\begin{cases} V^{(i+1)}(\underline{x}) = \int_t^{t+T} r(\underline{x}, \underline{u}^{(i)}) d\tau + V^{(i+1)}(\underline{x}(t+T)) \\ V^{(i+1)}(\underline{0}) = 0 \end{cases} \quad (3.16)$$

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

- Cập nhật:

$$u^{(i+1)}(\underline{x}) = -\frac{1}{2} R^{-1} g^T(\underline{x}) \frac{\partial V^{(i+1)}(\underline{x})^T}{\partial \underline{x}} \quad (3.17)$$

- Nếu thỏa mãn tiêu chuẩn hội tụ sao cho $\|V^{(i+1)} - V^{(i)}\| \leq v$ với v là số dương đủ nhỏ thì gán $\underline{u}^*(\underline{x}) = \underline{u}^{(i+1)}(\underline{x})$ và $V^*(\underline{x}) = V^{(i+1)}(\underline{x})$, kết thúc giải thuật.
 - Nếu không thỏa mãn, gán $i \leftarrow i + 1$, cho tín hiệu $\underline{u}^{(i)}$ vào hệ thống và thu thập thông tin cần thiết của hệ thống về trạng thái, tín hiệu điều khiển tại N trích mẫu khác nhau trong khoảng thời gian T rồi quay lại bước 2.
-

Chứng minh phương trình (3.16) và (2.21) có chung nghiệm duy nhất được trình bày trong phụ lục 2. Áp dụng khả năng xấp xỉ của mạng NN đã nêu trong chương (3.1), đồng thời nhằm giảm khối lượng tính toán so với cấu trúc Actor-Critic như đã nêu trong phần trước, giải thuật này chỉ sử dụng một NN nhằm xấp xỉ hàm giá trị tối ưu như sau:

$$V^{(i+1)}(\underline{x}) = \hat{W}^T \phi(\underline{x}) \quad (3.18)$$

với $\phi(x) \in \mathbb{R}^N$ là vector các hàm cơ bản phù hợp, $\hat{W} \in \mathbb{R}^N$ là vector trọng số, N là số neurons của mạng. Thay (3.18) vào (3.16) thu được:

$$e(t) = \hat{W}^T(\phi(x(t+T)) - \phi(x(t))) = - \int_t^{t+T} r(x, u) d\tau \quad (3.19)$$

với $e(t)$ là sai số xấp xỉ của hàm Bellman. Nhằm cực tiểu hóa sai số, phương pháp Least Square được sử dụng. Ta viết lại phương trình (3.19) như sau:

$$y(t) + e(t) = \hat{W}^T h(t) \quad (3.20)$$

với

$$\begin{aligned} h(t) &= \phi(x(t+T)) - \phi(x(t)) \\ y(t) &= \int_t^{t+T} r(x, u) d\tau \end{aligned} \quad (3.21)$$

Thông tin của hệ thống được thu thập N trích mẫu khác nhau trong khoảng thời gian T do đó ta tính toán (3.21) tại N điểm $t_1 \rightarrow t_N$ thu được:

$$\begin{aligned} H &= [h(t_1), \dots, h(t_N)] \\ Y &= [y(t_1), \dots, y(t_N)]^T \end{aligned} \quad (3.22)$$

Lời giải Least-Squares cho phương trình (3.20) như sau:

$$\hat{W} = (HH^T)^{-1} HY \quad (3.23)$$

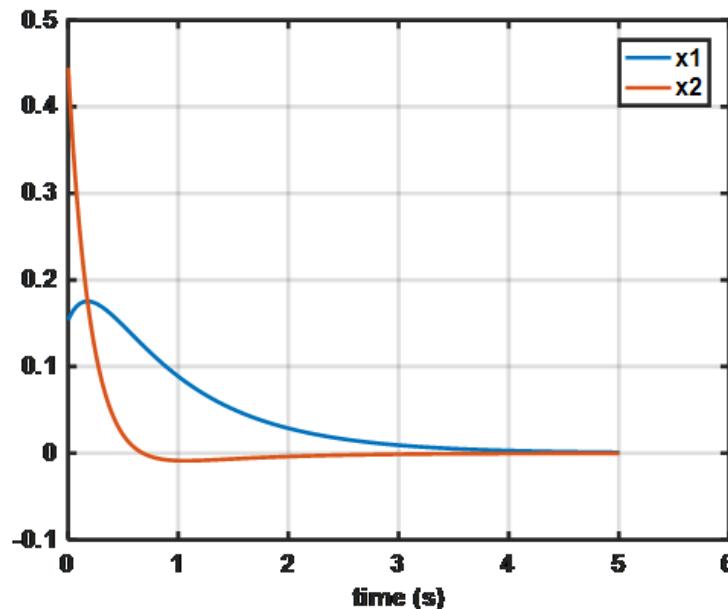
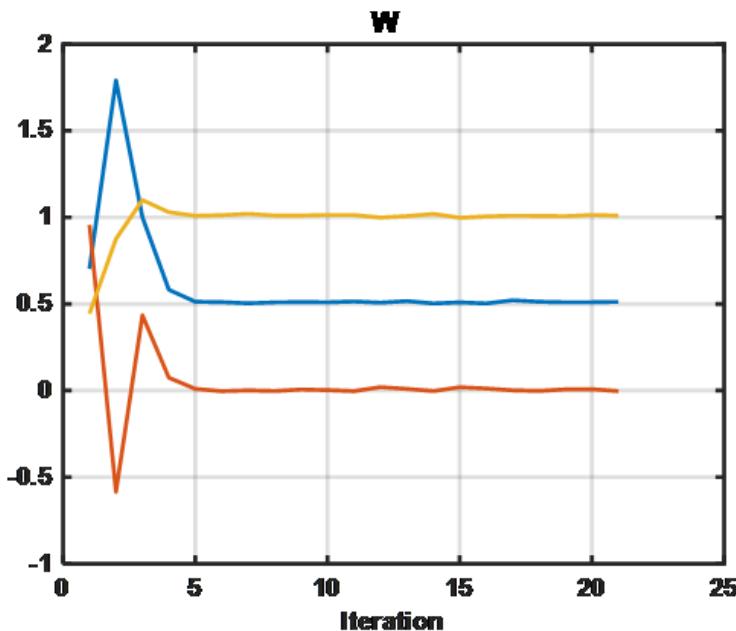


Figure 3.6: Trạng thái hệ thống với thuật toán IRL

Ví dụ 2. Ta áp dụng thuật toán cho hệ thống đã nêu trong ví dụ (1), ta chọn mạng NN tương tự trong (3.15), tín hiệu thăm dò tương tự. Ta thu được kết

Figure 3.7: Sự hội tụ của trọng số W với thuật toán IRL

quả mô phỏng như hình (3.6) và (3.7)

Như đồ thị ta thấy trọng số W hội tụ về chính xác giá trị tối ưu của nó, trong khi đó tín hiệu điều khiển vẫn giúp cho hệ ổn định với tốc độ khá tốt. Qua đó cho thấy tính đúng đắn của thuật toán.

3.2.3 Online Off-Policy IRL với hàm phạt

Tương tự với thuật toán (4) Online On-Policy IRL ở trên, thay vì chuyển về dạng phương trình vi phân như (2.29), ta viết lại phương trình HJI dưới dạng tích phân như sau:

$$V^*(x(t)) = \int_t^{t+T} e^{-\alpha(\tau-t)} (x^T Q_T x + u^{*T} R u^* - \gamma^2 d^{*T} d^*) d\tau + e^{-\alpha T} V^*(x(t+T)) \quad (3.24)$$

Từ đó thuật toán PI (2.3) và theo [20] ta đưa ra thuật toán Online Off-Policy IRL với hàm phạt như sau:

Thuật toán 5. PI (Online Off-Policy IRL với hàm phạt) giải quyết phương trình HJI

Pha 1: (thu thập dữ liệu sử dụng một luật điều khiển cố định): Áp dụng luật điều khiển u vào hệ thống và thu thập thông tin hệ thống yêu cầu về trạng thái, tín hiệu điều khiển và nhiễu tại N khoảng thời gian trích mẫu khác nhau.
 Pha 2: (sử dụng lặp đi lặp lại các dữ liệu đã được thu thập một cách tuần tự nhằm tìm ra một luật điều khiển tối ưu): với tín hiệu điều khiển $u^{(i)}$ và

$d^{(i)}$, sử dụng các thông tin đã thu thập được từ pha 1 nhằm giải phương trình Bellman cho $V^{(i)}$, $u^{(i+1)}$ và $d^{(i+1)}$ một cách đồng thời từ phương trình:

$$\begin{aligned} & e^{-\alpha T} V_i(X(t+T)) - V_i(X(t)) \\ &= \int_t^{t+T} e^{-\alpha(\tau-t)} (-X^T Q_T X - u_i^T R u_i + \gamma^2 d_i^T d_i) d\tau \\ &+ \int_t^{t+T} e^{-\alpha(\tau-t)} (-2u_{i+1}^T R(u - u_i) + 2\gamma^2 d_{i+1}^T (d - d_i)) d\tau \end{aligned} \quad (3.25)$$

Dùng nếu điều kiện dùng được thỏa mãn, ngược lại đặt $i = i + 1$ và chạy lại pha 2.

Áp dụng khả năng xấp xỉ của mạng NN đã nêu trong phần 3.1, đồng thời nhằm loại bỏ yêu cầu biết thông tin hàm $g(x)$ và $k(x)$ ở cấu trúc Actor-Critic hay On-Policy IRL như đã nêu trong phần trước, giải thuật này chỉ sử dụng ba NN nhằm xấp xỉ hàm giá trị tối ưu như sau:

$$\hat{V}_i(X) = \hat{W}_1^T \sigma(X) \quad (3.26)$$

$$\hat{u}_{i+1}(X) = \hat{W}_2^T \phi(X) \quad (3.27)$$

$$\hat{d}_{i+1}(X) = \hat{W}_3^T \varphi(X) \quad (3.28)$$

với $\sigma = [\sigma_1, \dots, \sigma_{l_1}] \in \mathbb{R}^{l_1}$, $\phi = [\phi_1, \dots, \phi_{l_2}] \in \mathbb{R}^{l_2}$, và $\varphi = [\varphi_1, \dots, \varphi_{l_3}] \in \mathbb{R}^{l_3}$ là các vector hàm cơ bản phù hợp, $\hat{W}_1 \in \mathbb{R}^{l_1}$, $\hat{W}_2 \in \mathbb{R}^{m \times l_2}$, và $\hat{W}_3 \in \mathbb{R}^{q \times l_3}$ là các vector trọng số hằng số, và l_1, l_2 và l_3 là số lượng neuron. Định nghĩa $v^1 = [v_1^1, \dots, v_1^m]^T = u - u_i$, $v^2 = [v_1^2, \dots, v_q^2]^T = d - d_i$ và giả thuyết rằng $R = diag(r, \dots, r_m)$. Sau đó, thay (3.26)-(3.28) vào (3.25) ta thu được

$$\begin{aligned} e(t) &= \hat{W}_1^T (e^{-\alpha T} \sigma(X(t+T)) - \sigma(X(t))) \\ &- \int_t^{t+T} e^{-\alpha(\tau-t)} (-X^T Q_T X - u_i^T R u_i + \gamma^2 d_i^T d_i) d\tau \\ &+ 2 \sum_{l=1}^m r_l \int_t^{t+T} e^{-\alpha(\tau-t)} \hat{W}_{2,l}^T \phi(X(t)) v_l^1 d\tau \\ &- 2\gamma^2 \sum_{k=1}^q \int_t^{t+T} e^{-\alpha(\tau-t)} \hat{W}_{3,k}^T \varphi(X(t)) v_k^2 d\tau \end{aligned} \quad (3.29)$$

với $e(t)$ là sai lệch xấp xỉ Bellman, $\hat{W}_{2,l}$ là cột thứ l của \hat{W}_2 , và $\hat{W}_{3,k}$ là cột thứ k của \hat{W}_3 . Sai lệch xấp xỉ Bellman là the continuous-time counter-part of the temporal difference (TD) [25]. Nhằm đưa sai lệch TD tới giá trị nhỏ nhất của nó, phương pháp least-squares được sử dụng. Viết lại phương trình (3.29) như sau

$$y(t) + e(t) = \hat{W}^T h(t) \quad (3.30)$$

với

$$\hat{W} = [\hat{W}_1^T, \hat{W}_{2,l}^T, \dots, \hat{W}_{2,m}^T, \hat{W}_{3,1}^T, \dots, \hat{W}_{3,q}^T]^T \in \mathbb{R}^{l_1+m \times l_2+q \times l_3} \quad (3.31)$$

$$h(t) = \begin{bmatrix} e^{-\alpha T} \sigma(X(t+T)) - \sigma(X(t)) \\ 2r_1 \int_t^{t+T} e^{-\alpha(\tau-t)} \phi(X(\tau)) v_1^1 d\tau \\ \vdots \\ 2r_m \int_t^{t+T} e^{-\alpha(\tau-t)} \phi(X(\tau)) v_m^1 d\tau \\ -2\gamma^2 \int_t^{t+T} e^{-\alpha(\tau-t)} \varphi(X(\tau)) v_1^2 d\tau \\ \vdots \\ -2\gamma^2 \int_t^{t+T} e^{-\alpha(\tau-t)} \varphi(X(\tau)) v_q^2 d\tau \end{bmatrix} \quad (3.32)$$

$$y(t) = \int_t^{t+T} e^{-\alpha(\tau-t)} (-X^T Q_T X - u_i^T R u_i + \gamma^2 d_i^T d_i) d\tau \quad (3.33)$$

Vector tham số \hat{W} đưa ra hàm giá trị được xấp xỉ, actor và disturbance (3.26)-(3.28), được xác định bởi quá trình tối thiểu hóa sai lệch Bellman (3.30) qua least-square. Giả thuyết rằng thông tin trạng thái, đầu vào và nhiễu được thu thập tại $N \geq l_1 + m \times l_2 + q \times l_3$ (số lượng phần tử độc lập trong \hat{W}) điểm t_1 tới t_N trong không gian trạng thái, trên khoảng thời gian T trong pha 1. Sau đó, với u_i và d_i đã có, ta sử dụng thông tin này đánh giá (3.32) và (3.33) tại N điểm để đưa ra:

$$H = [h(t_1), \dots, h(t_N)] \quad (3.34)$$

$$Y = [y(t_1), \dots, y(t_N)]^T. \quad (3.35)$$

Phương pháp least-square cho (3.30) được sử dụng thu được

$$\hat{W} = (HH^T)^{-1}HY \quad (3.36)$$

cho V_i, u_{i+1} và d_{i+1} .

Chú ý 12: chú ý rằng, mặc dù $X(t+T)$ trong (3.29), phương trình này được giải quyết theo least-square sau khi quan sát N mẫu $X(t), X(t+T), \dots, X(t+NT)$. Do đó, hiểu biết về hệ thống không yêu cầu để dự đoán trạng thái tương lai $X(t+T)$ tại thời điểm t để giải (3.29).

Ví dụ 3. Ta áp dụng thuật toán cho hệ thống:

$$\dot{x} = x + u + d \quad (3.37)$$

với $\alpha = 0, \gamma = 2, Q = 1, R = 1$.

Ta thấy bài toán điều khiển \mathbb{H}_∞ cho hệ tuyến tính (3.37) nếu trên có $V^* = px^2$, $u^* = -px$, $d^* = \frac{1}{4}px$ với p là nghiệm của phương trình:

$$1 + p^2 - \frac{1}{4}p^2 + 2p(1 - p + \frac{1}{4}p) = 0 \quad (3.38)$$

thu được $p = 3.0972$. Bên cạnh đó, trong mô phỏng, chọn $\sigma(x) = x^2, \phi(x) = x, \varphi(x) = x$ ta thu được kết quả mô phỏng như hình (3.8-3.9):

Nhận thấy W hội tụ về 3.0972 đúng với lý thuyết. Qua đó cho thấy tính đúng đắn của thuật toán.

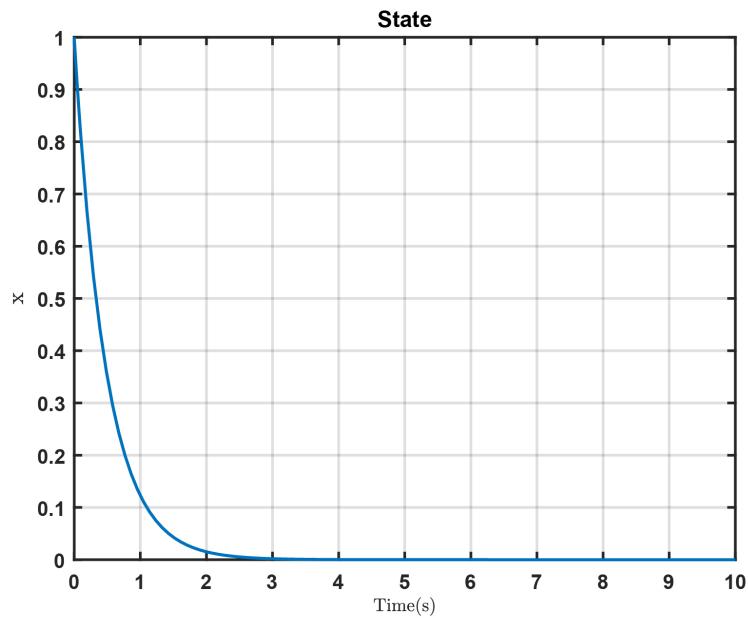
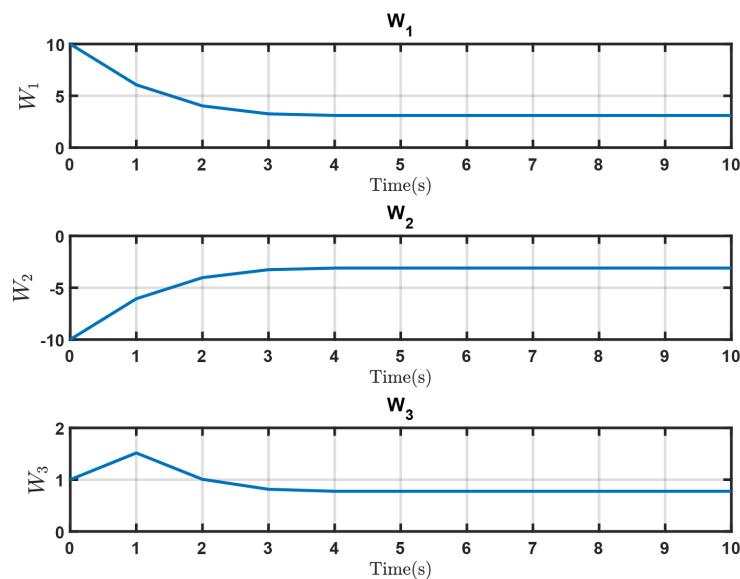


Figure 3.8: Trạng thái hệ thống với thuật toán Online Off-Policy IRL

Figure 3.9: Sự hội tụ của trọng số W với thuật toán Online Off-Policy IRL

Chapter 4

Ứng dụng thuật toán ADP cho mô hình tàu thủy

4.1 Mô hình động lực học của phương tiện hàng hải

Mô hình động lực học của phương tiện hàng hải được xây dựng dựa trên lý thuyết cơ học, những nguyên lý của động học và tĩnh học. Mô hình động lực học của phương tiện hàng hải được sử dụng để thiết kế các hệ thống điều khiển cho phương tiện này đáp ứng những mục tiêu cụ thể. Trong chương này, tóm tắt cách biểu diễn mô hình động lực học của phương tiện hàng hải dựa trên những kết quả trong [7], [6]. Các tính chất vật lý đặc trưng và các yêu cầu về điều khiển hệ thống lái tàu cũng được phân tích kỹ lưỡng để phục vụ cho phần tổng hợp bộ điều khiển và khảo sát tính ổn định cho hệ thống kín.

Các ký hiệu được sử dụng như chiều chuyển động, lực và momen tác động, tốc độ và vị trí cho các phương tiện hàng hải [6] được biểu diễn trong Bảng 4.1 và Hình 4.1 tuân thủ theo hiệp hội SNAME.

Đối với chuyển động của phương tiện hàng hải, 6 tọa độ độc lập là đủ để xác định vị trí và hướng của phương tiện. Sáu biến chuyển động khác nhau của phương tiện hàng hải gồm chuyển động tiến (surge), chuyển động dạt (sway), chuyển động lên xuống (heave), chuyển động quay lắc (roll), chuyển động quay lật (pitch) và chuyển động quay hướng (yaw) được định nghĩa như trong Hình 4.1 và Bảng 4.1.

Ba tọa độ đầu tiên (x, y, z) là vị trí của phương tiện hàng hải và đạo hàm của chúng theo thời gian (u, v, ω) là vận tốc chuyển động tịnh tiến của phương tiện dọc theo trục x, y, z . Ba tọa độ cuối (ϕ, θ, Ψ) là các góc miêu tả hướng của phương tiện quanh các trục x, y, z và đạo hàm theo thời gian của chúng (p, q, r) là tốc độ quay xung quanh các trục này.

Để phân tích chuyển động của phương tiện hàng hải trong 6 bậc tự do, các khung tọa độ sau cần được xét: các khung tọa độ gắn tâm trái đất và các

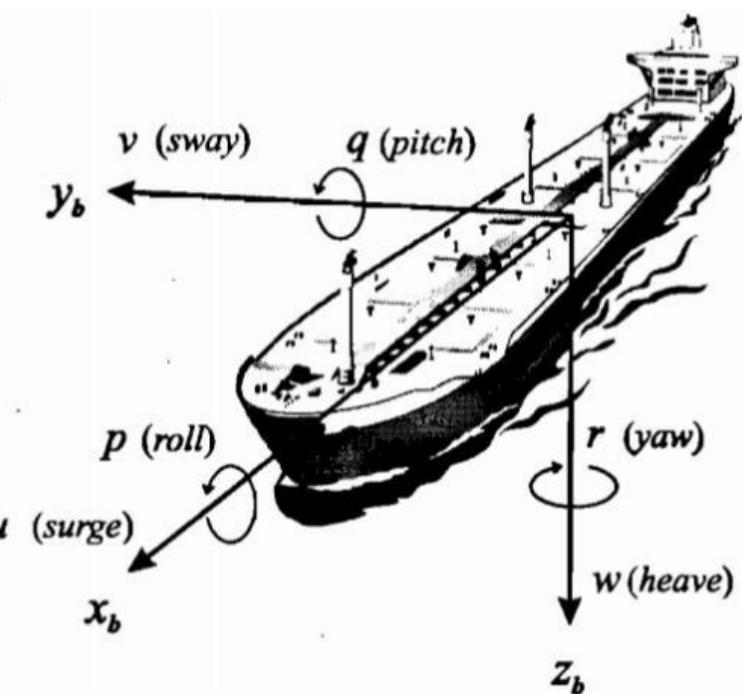


Figure 4.1: Các biến chuyển động của phương tiện hàng hải

Table 4.1: Các ký hiệu của SNAME

Bậc tự do	Chuyển động	Lực và moment	Tốc độ dài và tốc độ góc
1	Chuyển động tiến theo trục x	X	u
2	Chuyển động tiến theo trục y	Y	v
3	Chuyển động tiến theo trục z	Z	ω
4	Chuyển động quay quanh trục x	K	p
5	Chuyển động quay quanh trục y	M	q
6	Chuyển động quay quanh trục z	N	r

khung tọa độ quy chiếu cần để miêu tả chuyển động của phương tiện hàng hải.

Khung tọa độ quy chiếu gắn tâm trái đất (Earth-Centered Reference Frames) gồm:

ECI (i-frame) là khung tọa độ quán tính để định vị trái đất (ứng với khung quy chiếu không gia tốc trong định luật Newton để ứng dụng xét các chuyển động). Gốc của khung tọa độ ECI $x_{eye}z_e$ được đặt tại tâm của trái đất với các trục được chỉ ra ở Hình 4.2.

ECEF (e-frame) $x_{eye}z_e$ có gốc gắn với thân trái đất nhưng trục quay so với khung quán tính ECI, với tốc độ quay là $\omega_e = 7.2921 \cdot 10^{-5} rad/s$. Đối với những phương tiện hàng hải, sự quay của trái đất có thể được bỏ qua và do đó khung e-frame có thể xem như là khung quán tính. Khung tọa độ e-frame được sử dụng cho việc dẫn đường, định vị và điều khiển nói chung. Ví dụ khi phải miêu tả chuyển động và vị trí con tàu qua cảnh giữa các đại dương.

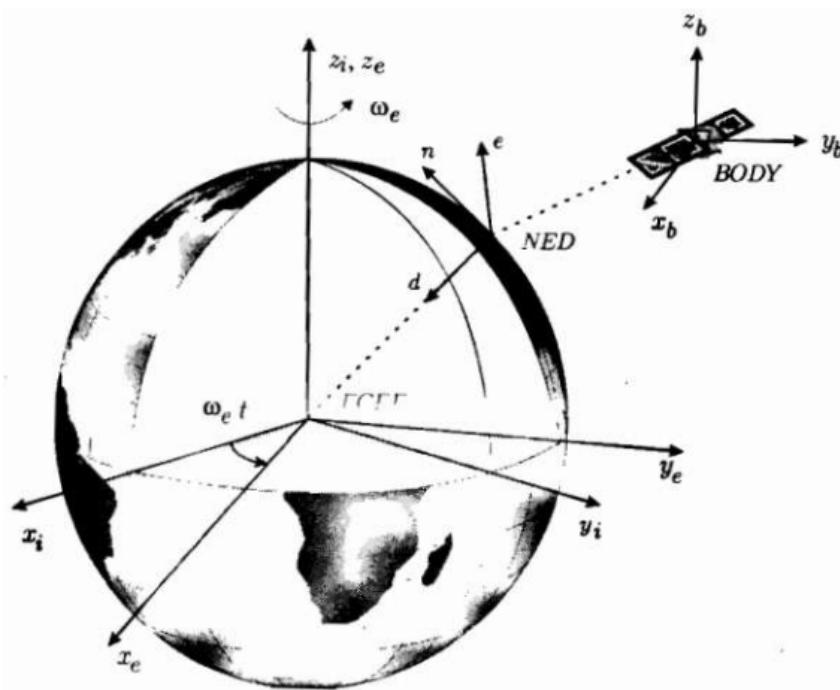


Figure 4.2: Các khung tọa độ quy chiếu

Khung tọa độ quy chiếu địa lý (Geographic Reference Frames) gồm:

NED (n-frame) hệ tọa độ North-East-Down $x_n y_n z_n$. Đó là hệ trục tọa độ chúng ta thường đề cập đến trong cuộc sống hàng ngày. Nó thường được định nghĩa như mặt phẳng tiếp tuyến trên bề mặt của trái đất và chuyển động cùng với phương tiện, trục x chỉ theo hướng bắc, trục y chỉ theo hướng đông, trục z chỉ theo hướng tới bề mặt trái đất. Vị trí của n-frame so với e-frame được xác định bằng hai góc l (kinh độ) và μ (vĩ độ).

Đối với những phương tiện hàng hải hoạt động trong vùng cục bộ, kinh độ và vĩ độ gần như không đổi, có thể sử dụng mặt phẳng tiếp tuyến trên bề mặt trái đất để định vị cho phương tiện. Khi đó trái đất được coi như một mặt phẳng định vị và để đơn giản nó được ký hiệu là n-frame. Khi định vị trái đất là mặt phẳng và n-frame là khung tọa độ quán tính thì định luật Newton vẫn được áp dụng.

BODY (b-frame) khung quy chiếu gắn thân $x_b y_b z_b$ là khung tọa độ được gắn với phương tiện, di chuyển cùng phương tiện.

Vị trí và hướng của phương tiện được miêu tả trong khung tọa độ quy chiếu quán tính n-frame (vì khung tọa độ e-frame và n-frame xấp xỉ bằng nhau đối với phương tiện hàng hải), trong khi vận tốc góc và vận tốc dài của phương tiện thường được biểu diễn trong khung tọa độ gắn thân b-frame.

Với tàu đại dương nói chung, vị trí thông dụng nhất của khung tọa độ gắn thân là tạo ra sự đối xứng xung quanh mặt phẳng $o_b x_b z_b$ và sự xấp xỉ đối xứng xung quanh mặt phẳng $O_b y_b z_b$. Theo nghĩa này, trục gắn thân x_b, y_b và

z_b được chọn trùng với trục chính của quán tính và chúng thường được xác định như Hình 4.1.

x_b – longitudinal axis – trục dọc (hướng từ đuôi tới mũi tàu).

y_b – transverse axis – trục ngang (hướng sang mạn phải của tàu)

z_b – normal axis – trục thẳng đứng (hướng từ đỉnh tới đáy tàu).

Dựa trên những ký hiệu trong Bảng 4.1, chuyển động của phương tiện hàng hải được miêu tả bởi những véc-tơ sau:

$$\begin{aligned}\eta &= [\eta_1^T, \eta_2^T]^T \in \mathbb{R}^6; & \eta_1 &= [x, y, z]^T \in \mathbb{R}^3; \eta_2 &= [\phi, \theta, \Psi]^T \in \mathbb{R}^3 \\ v &= [v_1^T, v_2^T]^T \in \mathbb{R}^6; & v_1 &= [u, v, \omega]^T \in \mathbb{R}^3; v_2 &= [p, q, r]^T \in \mathbb{R}^3 \\ \tau &= [\tau_1^T, \tau_2^T]^T \in \mathbb{R}^6; & \tau_1 &= [x, y, z]^T \in \mathbb{R}^3; \tau_2 &= [K, M, N]^T \in \mathbb{R}^3\end{aligned}$$

trong đó η ký hiệu véc-tơ vị trí và hướng với khung tọa độ gắn trái đất e-frame, v ký hiệu véc-tơ vận tốc dài và vận tốc góc với hệ tọa độ gắn thân b-frame, τ ký hiệu lực và momen tác động lên tàu trong khung tọa độ gắn thân.

Nghiên cứu về động lực học của phương tiện hàng hải có thể được chia thành hai phần [6]:

- Phân tích về vị trí và hướng của chuyển động (kinematic).
- Phân tích về những lực gây ra chuyển động (dynamic).

4.1.1 Phân tích về vị trí và hướng chuyển động của tàu

Đạo hàm bậc nhất theo thời gian của véc-tơ vị trí η_1 có mối liên hệ với vector v_1 thông qua sự chuyển đổi sau:

$$\dot{\eta}_1 = J_1(\eta_2)v_1 \quad (4.1)$$

trong đó $J_1(\eta_2)$ là một ma trận chuyển đổi, gồm các hàm của các góc (ϕ, θ, Ψ) . Ma trận này được biểu diễn như sau:

$$J_1(\eta_2) = \begin{bmatrix} c\Psi c\theta & -s\Psi c\phi + c\Psi s\theta s\phi & s\Psi s\phi + c\Psi c\phi s\theta \\ s\Psi c\theta & -c\Psi c\phi + s\Psi s\theta s\phi & -c\Psi s\phi + s\Psi s\phi c\theta \\ -s\theta & c\theta s\phi & c\theta c\phi \end{bmatrix} \quad (4.2)$$

với $s = \sin(\cdot)$, $c = \cos(\cdot)$, $t = \tan(\cdot)$.

Ma trận $J_1(\eta_2)$ là ma trận trực giao $J_1^{-1}(\eta_2) = J_1^T(\eta_2)$.

Mặt khác, đạo hàm bậc nhất theo thời gian của vector góc η_2 có mối liên hệ với vector v_2 thông qua sự chuyển đổi sau:

$$\dot{\eta}_2 = J_2(\eta_2)v_2 \quad (4.3)$$

trong đó ma trận chuyển đổi

$$J_2(\eta_2) = \begin{bmatrix} 1 & \sin(\phi)\tan(\theta) & \cos(\phi)\tan(\theta) \\ 0 & \cos(\phi) & -\sin(\phi) \\ 0 & \frac{\sin(\phi)}{\cos(\theta)} & \frac{\cos(\phi)}{\cos(\theta)} \end{bmatrix} \quad (4.4)$$

Chú ý rằng ma trận chuyển đổi $J_2(\eta_2)$ không xác định đối với góc quay lật $\theta = \pm 90^\circ$ và $J_2(\eta_2)$ không thỏa mãn tính chất của ma trận trực giao. Đối với phương tiện trên bề mặt biển không hoạt động ở góc quay lật $\theta = \pm 90^\circ$, tuy nhiên tàu ngầm và máy bay đều có thể hoạt động tại điểm đặc biệt này, chi tiết được trình bày trong tài liệu [6].

Kết hợp (4.1) và (4.3) tạo ra phương trình mô tả vị trí và hướng của phương tiện hàng hải:

$$\begin{bmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \end{bmatrix} = \begin{bmatrix} J_1(\eta_2) & 0_{3 \times 3} \\ 0_{3 \times 3} & J_2(\eta_2) \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \Leftrightarrow \dot{\eta} = J(\eta)v \quad (4.5)$$

4.1.2 Phương trình chuyển động của phương tiện hàng hải (Dynamics)

Phương trình chuyển động của vật rắn

Để xét chuyển động và các yếu tố ảnh hưởng đến chuyển động của tàu, trước tiên ta xét chuyển động của tàu như chuyển động của vật rắn với khung quy chiếu gắn thân $x_b y_b z_b$, gốc tọa độ O (Hình 4.3). Ứng dụng công thức Newton-Euler cho vật rắn có khối lượng m , phương trình cân bằng lực và momen tác động lên tàu như sau [8]:

$$m [\ddot{v}_1 + v_2 \times v_1 + \dot{v}_2 \times r_g + v_2 \times (v_2 \times r_g)] = \tau_1 \quad (4.6)$$

$$I_o \ddot{v}_2 + v_2 \times (I_o v_2) + m r_g \times (\dot{v}_1 + v_2 \times v_1) = \tau_2 \quad (4.7)$$

trong đó $r_g = [x_g, y_g, z_g]^T$ là tọa độ của trọng tâm của vật rắn trong khung tọa độ gắn thân và I_o là ma trận quán tính hệ thống xung quanh điểm O (Hình 4.1-4.3).

$$I_o := \begin{bmatrix} I_x & -I_{xy} & -I_{xz} \\ -I_{yx} & I_y & -I_{yz} \\ -I_{zx} & -I_{zy} & I_z \end{bmatrix} \quad I_o = I_o^T > 0; \quad \dot{I}_o = 0 \quad (4.8)$$

trong đó

I_x, I_y và I_z là những momen quán tính xung quanh trục x_b, y_b, z_b và $I_{xy} = I_{yx}, I_{xz} = I_{zx}$ và $I_{yz} = I_{zy}$ là những tương tác của momen quán tính trục này lên trục khác.

Phương trình chuyển động của vật rắn được biểu diễn bằng tập các vector:

$$M_{RB}\ddot{v} + C_{RB}(v)v = \tau_{RB} \quad (4.9)$$

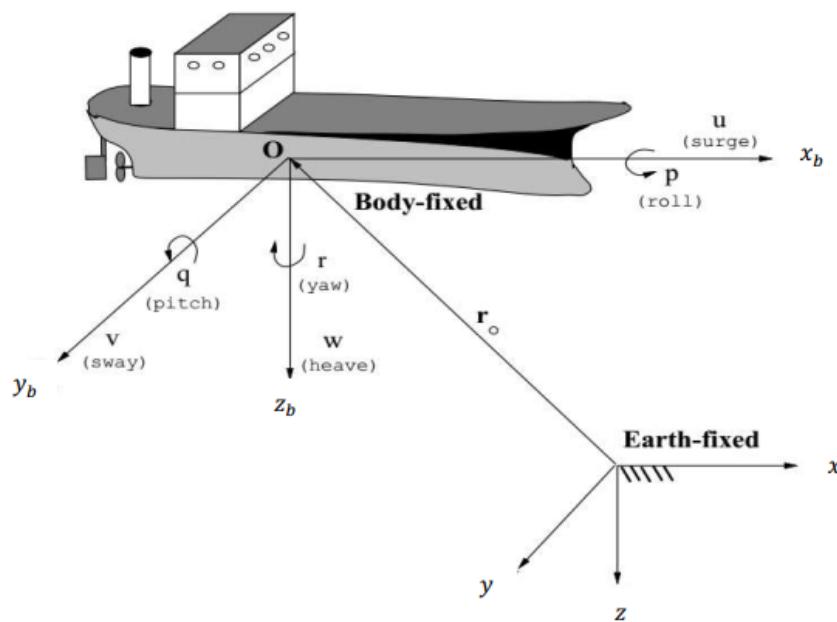


Figure 4.3: Khung tọa độ quy chiếu quán tính gắn với trái đất và khung tọa độ gắn thân

trong đó:

$v = [u, v, \omega, p, q, r]^T$ là véc-tơ vận tốc tổng quát được phân tích trong khung b-frame.

$\tau_{RB} = [X, Y, Z, K, M, N]^T$ là vector tổng quát của lực và momen ngoài được phân tích trong khung b-frame.

M_{RB} là ma trận quán tính hệ thống vật rắn.

$C_{RB}(v)$ là ma trận Coriolis và lực hướng tâm vật rắn.

Ma trận quán tính hệ thống của vật rắn

$$M_{RB} = \begin{bmatrix} m & 0 & 0 & 0 & mz_g & -my_g \\ 0 & m & 0 & -mz_g & 0 & mx_g \\ 0 & 0 & m & my_g & -mx_g & 0 \\ 0 & -mz_g & my_g & I_x & -I_{xy} & -I_{xz} \\ mz_g & 0 & -mx_g & -I_{yx} & I_y & -I_{yz} \\ -my_g & mx_g & 0 & -I_{zx} & -I_{zy} & I_z \end{bmatrix} \quad (4.10)$$

Ma trận Coriolis và lực hướng tâm của vật rắn

$$C_{RB}(v) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ -m(y_gq + z_gr) & m(y_gp + \omega) & m(z_gp - v) \\ -m(x_gq - \omega) & -m(z_gr + x_gp) & m(z_gp + y_gq) \\ m(x_gr + v) & m(y_gr - u) & -m(x_gp + y_gq) \\ \\ m(y_gq + z_gr) & -m(x_gq - \omega) & -m(x_gr + v) \\ -m(y_gp + \omega) & m(z_gr + x_gp) & -m(y_gr - u) \\ -m(z_gp - v) & -m(z_gq + u) & m(x_gp + y_gq) \\ 0 & -I_{yz}q - I_{xz}p + I_zr & I_{yz}r + I_{xy}p - I_yq \\ I_{yz}r + I_{xz}p - I_zr & 0 & -I_{xz}r - I_{xy}q + I_xp \\ -I_{yz}r - I_{xy}p + I_yq & I_{xz}r + I_{xy}q - I_xp & 0 \end{bmatrix} \quad (4.11)$$

Vector tổng quát của lực và momen ngoài τ_{RB} là tổng của véc-tơ lực và momen thủy động lực học τ_H , véc-tơ lực và momen nhiễu từ môi trường ω , véc-tơ lực và momen đẩy của tàu.

Lực và momen thủy động lực học

Khi phương tiện hàng hải chuyển động trên biển, phương tiện chịu tác động của các lực và momen thủy động lực học như sau:

- **Lực cảm ứng bức xạ (RIF)**

Theo Faltinsen [5] "RIF là những lực tác động lên vật khi vật bị buộc dao động với tần số sóng kích thích và không có sóng bất thường".

Lực và momen cảm ứng bức xạ có thể được định nghĩa như tổng của 3 thành phần mới:

1. Khối lượng nước kèm (added mass) do quán tính của chất lỏng xung quanh.
2. Sự suy giảm thế năng cảm ứng-bức xạ do năng lượng bị mất bởi sóng trên bề mặt.
3. Lực phục hồi theo Archimedes (trọng lượng và lực đẩy).

Ba thành phần này tạo thành các lực và momen có thể được biểu diễn toán học như sau:

$$\tau_R = -\underbrace{M_A \dot{v} - C_A(v)v}_{\text{khối lượng nước kèm}} - \underbrace{D_P(v)v}_{\text{suy giảm thế năng}} - \underbrace{g(\eta) + g_o}_{\text{lực phục hồi}} \quad (4.12)$$

Khối lượng nước kèm

Khi phương tiện chuyển động sẽ ép chất lỏng bao quanh tàu dao động với các biên độ lớp bao chất lỏng khác nhau, đồng bộ với chuyển động điều hòa cưỡng bức của phương tiện. Khối lượng nước kèm được hiểu như lực và momen cảm ứng áp suất sinh ra từ chuyển động điều hòa cưỡng bức của vật rắn và tỉ lệ với gia tốc của vật rắn. Trong đó M_A là ma trận quán tính hệ thống của khối lượng nước kèm, $C_A(v)$ là ma trận Coriolis và lực hướng tâm thủy động lực học.

Ma trận M_A là ma trận vuông 6x6 được định nghĩa như sau:

$$M_A = - \begin{bmatrix} X_{\dot{u}} & X_{\dot{v}} & X_{\dot{\omega}} & X_{\dot{p}} & X_{\dot{q}} & X_{\dot{r}} \\ Y_{\dot{u}} & Y_{\dot{v}} & Y_{\dot{\omega}} & Y_{\dot{p}} & Y_{\dot{q}} & Y_{\dot{r}} \\ Z_{\dot{u}} & Z_{\dot{v}} & Z_{\dot{\omega}} & Z_{\dot{p}} & Z_{\dot{q}} & Z_{\dot{r}} \\ K_{\dot{u}} & K_{\dot{v}} & K_{\dot{\omega}} & K_{\dot{p}} & K_{\dot{q}} & K_{\dot{r}} \\ M_{\dot{u}} & M_{\dot{v}} & M_{\dot{\omega}} & M_{\dot{p}} & M_{\dot{q}} & M_{\dot{r}} \\ N_{\dot{u}} & N_{\dot{v}} & N_{\dot{\omega}} & N_{\dot{p}} & N_{\dot{q}} & N_{\dot{r}} \end{bmatrix} \quad (4.13)$$

Ký hiệu của SNAME (1950) được sử dụng trong biểu thức này như sau, ví dụ lực khối lượng nước kèm thủy động lực học Y dọc theo trục y do gia tốc \dot{u} (hướng x) tạo ra được viết như sau [19,24]:

$$Y = -Y_{\dot{u}}\dot{u} \quad Y_{\dot{u}} := \frac{\partial Y}{\partial \dot{u}} \quad (4.14)$$

Với những ứng dụng điều khiển, có thể giả thiết rằng $M_A > 0$ là hằng số dương. Điều này dựa trên giả thiết là M_A là độc lập với tần số sóng, đây là một giả thiết tốt đối với những ứng dụng điều khiển tần số thấp. Với những ứng dụng như hoạt động của phương tiện ngầm ở độ sâu lớn và tàu thủy ở vị trí cố định, chúng ta có thể giả thiết M_A là đối xứng. Tuy nhiên giả thiết M_A là đối xứng không đúng đắn với tàu thủy đang chuyển động ở tốc độ nào đó [6], [8].

Ma trận Coriolis và lực hướng tâm thủy động lực học $C_A(v)$ có thể được viết chi tiết như sau:

$$C_A(v) = \begin{bmatrix} 0 & 0 & 0 & 0 & -a_3 & a_2 \\ 0 & 0 & 0 & a_3 & 0 & -a_1 \\ 0 & 0 & 0 & -a_2 & a_1 & 0 \\ 0 & -a_3 & a_2 & 0 & -b_3 & b_2 \\ a_3 & 0 & -a_1 & b_3 & 0 & -b_1 \\ -a_2 & a_1 & 0 & -b_2 & b_1 & 0 \end{bmatrix} \quad (4.15)$$

trong đó

$$\begin{aligned}
 a_1 &= X_{\dot{u}}u + X_{\dot{v}}v + X_{\dot{\omega}}\omega + X_{\dot{p}}p + X_{\dot{q}}q + X_{\dot{r}}r \\
 a_2 &= Y_{\dot{u}}u + Y_{\dot{v}}v + Y_{\dot{\omega}}\omega + Y_{\dot{p}}p + Y_{\dot{q}}q + Y_{\dot{r}}r \\
 a_3 &= Z_{\dot{u}}u + Z_{\dot{v}}v + Z_{\dot{\omega}}\omega + Z_{\dot{p}}p + Z_{\dot{q}}q + Z_{\dot{r}}r \\
 b_1 &= K_{\dot{u}}u + K_{\dot{v}}v + K_{\dot{\omega}}\omega + K_{\dot{p}}p + K_{\dot{q}}q + K_{\dot{r}}r \\
 b_2 &= M_{\dot{u}}u + M_{\dot{v}}v + M_{\dot{\omega}}\omega + M_{\dot{p}}p + M_{\dot{q}}q + M_{\dot{r}}r \\
 b_3 &= N_{\dot{u}}u + N_{\dot{v}}v + N_{\dot{\omega}}\omega + N_{\dot{p}}p + N_{\dot{q}}q + N_{\dot{r}}r
 \end{aligned} \tag{4.16}$$

- Lực và momen suy giảm do ma sát bề mặt, độ trôi của sóng và dòng xoáy**

Ngoài sự suy giảm thể năng cảm ứng bức xạ, cần phải xét cả những tác động suy giảm khác như là ma sát bề mặt, sự suy giảm độ trôi của sóng và sự suy giảm do dòng xoáy đó là:

$$\tau_D = - \underbrace{D_S(v)v}_{\text{ma sát bề mặt}} - \underbrace{D_W(v)v}_{\text{suy giảm độ trôi sóng}} - \underbrace{D_M(v)v}_{\text{suy giảm do dòng xoáy}} \tag{4.17}$$

Từ đó ma trận suy giảm thủy động lực toàn phần được định nghĩa là:

$$D(v) := D_P(v) + D_S(v) + D_W(v) + D_M(v) \tag{4.18}$$

có nghĩa là lực và momen thủy động lực τ_H có thể được viết như là tổng của τ_R và τ_D :

$$\tau_H = -M_A\dot{v} - C_A(v)v - D(v)v - g(\eta) + g_o \tag{4.19}$$

Hệ phương trình chuyển động 6 bậc tự do của phương tiện hàng hải

Động học vật rắn được biểu diễn là (xem mục 4.1.2):

$$M_{RB}\ddot{v} + C_{RB}(v)v = \tau_{RB} \tag{4.20}$$

trong đó:

$$\tau_{RB} = \tau_H + \omega + \tau$$

vector τ biểu diễn lực và momen đẩy do tàu sinh ra. Mô hình kết quả được đưa ra như sau:

$$M\ddot{v} + C(v)v + D(v)v + g(\eta) = \tau + g_o + \omega \tag{4.21}$$

trong đó:

$$\begin{aligned}
 M &= M_{RB} + M_A \\
 C(v) &= C_{RB}(v) + C_A(v) \\
 D(v) &= D_P(v) + D_S(v) + D_W(v) + D_M(v)
 \end{aligned}$$

- Ma trận $D(v)$ là ma trận suy giảm thủy động lực học và là ma trận không đối xứng.
- $g(\eta)$ là vector lực đẩy và lực trọng trường.
- g_o là vector được sử dụng khi có điều khiển cân bằng trong trường hợp không tải.
- ω là các nhiễu loạn từ môi trường.

Ma trận suy giảm thủy động lực học $D(v)$

Trong nhiều trường hợp, để thuận lợi khi biểu diễn, ma trận suy giảm thủy động lực được biểu diễn như sau:

$$D(v) = D + D_n(v) \quad (4.22)$$

trong đó: D là ma trận suy giảm tuyến tính $D_n(v)$ là ma trận suy giảm phi tuyến.

Với tàu thủy (ba bậc tự do) ở tốc độ thấp hoặc trường hợp vị trí động của tàu, ma trận suy giảm phi tuyến được bỏ qua, khi đó:

$$D(v) = D = - \begin{bmatrix} X_u & 0 & 0 \\ 0 & Y_u & Y_r \\ 0 & N_v & N_r \end{bmatrix} \quad (4.23)$$

Khi tàu thủy chạy ở tốc độ cao, ma trận suy giảm thủy động lực học bao gồm cả ma trận suy giảm phi tuyến. Mô hình của ma trận suy giảm phi tuyến được Blanke [6] đề xuất như sau:

$$D_n(v) = \begin{bmatrix} -X_{|u|u}|u| & 0 & 0 \\ 0 & -Y_{|v|v}|v| - Y_{|r|r}|r| & -Y_{|vr|v}|v| - Y_{|r|r}|r| \\ 0 & -N_{|v|v}|v| - N_{|r|r}|r| & -N_{|v|r}|v| - N_{|r|r}|r| \end{bmatrix} \quad (4.24)$$

Trong thực tế, các thành phần của ma trận suy giảm phi tuyến rất khó xác định.

Lực đẩy và lực trọng trường $g(\eta)$

Bên cạnh lực do khối lượng nước kèm và lực suy giảm, tàu ngầm và tàu thủy còn chịu tác động bởi lực đẩy và trọng lực. Trong thuật ngữ thủy động lực, lực đẩy và lực trọng trường được gọi là lực phục hồi, và chúng tương đương với lực đòn hồi trong hệ thống đòn hồi–suy giảm–tăng thêm. Lực và momen phục hồi tác động lên tàu ngầm và tàu thủy là khác nhau.

Trong tài liệu thủy tĩnh, sự ổn định tĩnh do lực phục hồi thường được đề cập đến như sự ổn định khuynh tâm. Một tàu ổn định khuynh tâm sẽ chống lại sự nghiêng so với trạng thái đứng hoặc điểm cân bằng trong chuyển động lên xuống (ω), chuyển động quay lắc (p) và chuyển động quay lật (q). Đối

với phương tiện trên mặt nước, lực phục hồi sẽ phụ thuộc vào độ cao khuynh tâm của tàu, vị trí của CG (trọng tâm) và CB (tâm nổi) cũng như hình dạng và kích thước của mặt phẳng nước. Đặt A_{wp} là ký hiệu của diện tích mặt phẳng nước (là một hàm của vị trí heave), ∇ là thể tích chất lỏng bị chiếm bởi phương tiện, g là gia tốc trọng trường (chiều dương hướng xuống dưới), ρ là tỷ trọng của nước và:

\overline{GM}_T = độ cao khuynh tâm theo chiều ngang tàu (m).

\overline{GM}_L = độ cao khuynh tâm theo chiều dọc tàu (m).

Biểu thức lực và momen phục hồi được biểu diễn như sau:

$$g(\eta) = \begin{bmatrix} -\rho g \int_0^z A_{wp}(\zeta) d\zeta \sin(\theta) \\ \rho g \int_0^z A_{wp}(\zeta) d\zeta \cos(\theta) \sin(\phi) \\ \rho g \int_0^z A_{wp}(\zeta) d\zeta \cos(\theta) \cos(\phi) \\ \rho g \nabla \overline{GM}_T \sin(\phi) \cos(\theta) \cos(\phi) \\ \rho g \nabla \overline{GM}_L \sin(\theta) \cos(\theta) \cos(\phi) \\ \rho g \nabla (-\overline{GM}_L \cos(\theta) + \overline{GM}_T) \sin(\phi) \sin(\theta) \end{bmatrix} \quad (4.25)$$

Các nhiễu loạn từ môi trường

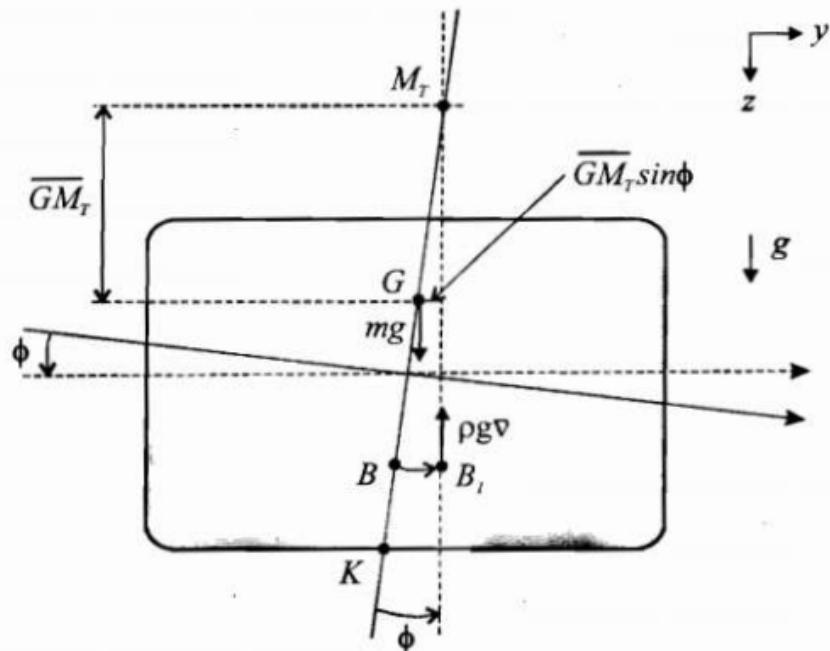


Figure 4.4: Ở định khuynh tâm theo chiều ngang tàu

Khi hoạt động trong môi trường đại dương, tàu thủy chịu tác động lớn từ gió, sóng và dòng chảy đại dương. Đối với hầu hết những ứng dụng thiết kế hệ thống điều khiển tàu biển, khi xem xét nhiễu loạn của sóng và gió thường áp dụng nguyên tắc xếp chồng. Nguyên tắc xếp chồng giả thiết rằng nhiễu loạn gồm sóng, gió và dòng chảy được định nghĩa:

$$\omega = \omega_{wind} + \omega_{wave} + \omega_{current} \quad (4.26)$$

Biểu thức của các nhiễu tác động này được trình bày trong tài liệu [19] và được trình bày tóm tắt trong phụ lục.

Như vậy nhiễu từ môi trường đại dương tác động đến mô hình động lực học của tàu thủy là nhiễu tổng hợp của cả ba loại nhiễu.

4.1.3 Mô hình động lực học của tàu thủy ba bậc tự do trên mặt phẳng nằm ngang

Khi tàu thủy chuyển động trên đại dương giống như tàu chuyển động trên mặt phẳng nằm ngang, tiếp tuyến với bề mặt trái đất. Khi đó chuyển động của tàu thủy thường được mô tả bởi các thành phần chuyển động tiến, chuyển động dạt và chuyển động quay hướng, các chuyển động lên xuống, chuyển động quay lắc và chuyển động quay lật bị bỏ qua. Do đó từ mô hình chuyển động sáu bậc tự do của phương tiện hàng hải, phương trình chuyển động của tàu thủy chỉ còn ba bậc tự do gồm $v = [u, v, r]^T$ và $\eta = [x, y, \Psi]^T$ các thành phần $\omega = p = q = 0$.

Giả thiết 6. Nhằm đơn giản quá trình tính toán và phân tích:

1. Tàu có sự phân bố khối lượng đồng đều và đối xứng qua mặt phẳng xz vì vậy:

$$I_{xy} = I_{yz} = 0 \quad (4.27)$$

2. Đặt gốc tọa độ khung b vào đường trung tâm của con tàu, sao cho $y_g = 0$, tâm của khối lượng gia tăng trùng với trọng tâm của tàu thủy.

Với Giả thiết (6), mô hình động lực học phi tuyến của tàu thủy ba bậc tự do như sau (theo [6]):

$$\begin{cases} \dot{\eta} = J(\eta)v \\ M\ddot{v} + C(v)v + D(v)v + g(\eta) = \tau + \Delta(\eta, v) \end{cases} \quad (4.28)$$

trong đó các ma trận quay xung quanh trục z được biểu diễn:

$$J(\eta) = \begin{bmatrix} \cos(\Psi) & -\sin(\Psi) & 0 \\ \sin(\Psi) & \cos(\Psi) & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.29)$$

Ma trận quán tính hệ thống:

$$M = \begin{bmatrix} m - X_{\dot{u}} & 0 & 0 \\ 0 & m - Y_{\dot{v}} & mx_g - Y_{\dot{r}} \\ 0 & mx_g - N_{\dot{v}} & I_z - N_{\dot{r}} \end{bmatrix} \quad (4.30)$$

Ma trận Coriolis và lực hướng tâm hệ thống:

$$C(v) = \begin{bmatrix} 0 & 0 & -(m - Y_v)v - (mx_g - Y_r)r \\ 0 & 0 & (m - X_u)u \\ (m - Y_v)v + (mx_g - Y_r)r & -(m - X_u)u & 0 \end{bmatrix} \quad (4.31)$$

Ma trận suy giảm thủy động lực học:

$$D(v) = D + D_n(v) \quad (4.32)$$

với

$$D = \begin{bmatrix} -X_u & 0 & 0 \\ 0 & -Y_v & -Y_r \\ 0 & -N_v & -N_r \end{bmatrix} \quad (4.33)$$

$$D_n(v) = \begin{bmatrix} -X_{|u|u}|u| & 0 & 0 \\ 0 & -Y_{|v|v}|v| - Y_{|r|v}|r| & -Y_{|v|r}|v| - Y_{|r|r}|r| \\ 0 & -N_{|v|v}|v| - N_{|r|v}|r| & -N_{|v|r}|v| - N_{|r|r}|r| \end{bmatrix}$$

$g(\eta)$ là véc-tơ lực đẩy và lực trọng trường, tàu thủy ba bậc tự do với Giả thiết (6) có thể coi $g(\eta) = 0$. Tuy nhiên nhiều từ môi trường có thể tác động làm nghiêng tàu, khi đó sẽ xuất hiện lực và momen đẩy để đưa tàu về vị trí cân bằng. Vì vậy không mất đi tính tổng quát khi trong công thức (4.28) vẫn có thành phần $g(\eta)$.

$\Delta(\eta, v)$ gồm các véc-tơ lực và momen nhiều từ môi trường và các thành phần không xác định của mô hình tàu.

Hệ thống (4.28) thỏa mãn những tính chất sau:

1. $M = M^T > 0$ (đối xứng xác định dương)
2. $C(v) = -C^T(v)$ (đối xứng lệch)
3. $D(v) > 0$ (xác định dương)
4. $J(\eta)$ là ma trận quay xung quanh trục z và là ma trận trực giao $J^{-1}(\eta) = J^T(\eta)$

Giả thiết (1) có thể sử dụng cho cả tàu chuyển động với tốc độ cao có ma trận quán tính không đối xứng ($M \neq M^T$) nhờ sử dụng phản hồi gia tốc như Fossen đã trình bày trong các tài liệu [6], [9].

Chuyển động của tàu thủy trên bề mặt đại dương là đối tượng nghiên cứu chính của luận án. Mô hình động lực học phi tuyến ba bậc tự do của tàu thủy được sử dụng trong suốt luận án để tổng hợp bộ điều khiển và chứng minh tính ổn định của hệ thống.

Trong phạm vi đồ án này chúng em sử dụng mô hình với các tham số hệ thống như sau:

$$M = \begin{bmatrix} 20 & 0 & 0 \\ 0 & 19 & 0.72 \\ 0 & 0.72 & 2.7 \end{bmatrix}$$

$$C(v) = \begin{bmatrix} 0 & 0 & -19v_y - 0.72v_z \\ 0 & 0 & 20v_x \\ 19v_y + 0.72v_z & -20v_x & 0 \end{bmatrix}$$

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

$$g(\eta) = 0$$

Với quỹ đạo mong muốn $\eta_d(t) = [12 \sin(0.2t), -12 \cos(0.2t), 0.2t]^T$.

4.2 Thuật toán ADP cho tàu thủy

4.2.1 Sử dụng đổi biến trong ứng dụng ADP cho hệ không dừng

Từ phương trình (4.28) ta có mô hình tàu thủy với 3 bậc tự do:

$$\begin{aligned} \dot{\eta}(t) &= J(\eta)v(t) \\ M\dot{v}(t) &= -C(v)v(t) - D(v)v(t) - g(\eta) + \tau + \Delta(\eta, v) \end{aligned} \tag{4.34}$$

Bài toán đặt ra là thiết kế bộ điều khiển ứng dụng thuật toán ADP làm cho hệ (4.34) thỏa mãn:

- Tất cả các tín hiệu đều bị chặn.
- Tọa độ $\eta(t)$ bám theo quỹ đạo đặt $\eta_d(t) = [\eta_{dx}(t), \eta_{dy}(t), \eta_{dz}(t)]^T$ với tọa độ chính xác mong muốn.

Định nghĩa vector sai lệch bám $z_\eta = \eta - \eta_d$ ta có phương trình:

$$\dot{z}_\eta = -\dot{\eta}_d + J(\eta)v \tag{4.35}$$

Coi v là bộ điều khiển ảo của hệ (4.35). Để dàng thiết kế $v = v_d$ làm hệ (4.35) ổn định tiệm cận.

$$v_d(z_\eta, \eta_d) = J^{-1}(\eta)(\dot{\eta}_d - \beta_\eta z_\eta) \tag{4.36}$$

với $\beta_\eta > 0$ là ma trận điều khiển xác định dương. Định nghĩa $z_v = v - v_d$ ta có phương trình:

$$\dot{z}_v = -M^{-1}C(v)v - M^{-1}D(v)v - M^{-1}g(\eta) - \dot{v}_d + M^{-1}\tau + M^{-1}\Delta(\eta, v) \tag{4.37}$$

Ta cần tìm τ làm cho $z_v \rightarrow 0$, định nghĩa luật điều khiển feedforward khi hệ đã đạt tới trạng thái dừng $z_v = 0$ như sau:

$$\tau_d = M\dot{v}_d + C(v_d)v_d + D(v_d)v_d + g(\eta) \quad (4.38)$$

Định nghĩa hàm $l(x) = C(x)x + D(x)x$, biến $X = [z_v^T, z_\eta^T, \eta_d^T]^T$, $u = \tau - \tau_d$ và giả sử có quan hệ $\dot{\eta}_d = h_1(\eta_d)$, $\ddot{\eta}_d = h_2(\eta_d)$, ta kết hợp với (4.35) và (4.37) thu được:

$$\begin{aligned} \dot{X} &= \begin{bmatrix} -M^{-1}l(z_v + v_d(z_\eta, \eta_d)) + M^{-1}l(v_d(z_\eta, \eta_d)) \\ J(z_\eta + \eta_d)z_v - \beta_\eta z_\eta \\ h_1(\eta_d) \end{bmatrix} + \begin{bmatrix} M^{-1} \\ 0 \\ 0 \end{bmatrix} u + \begin{bmatrix} M^{-1} \\ 0 \\ 0 \end{bmatrix} \Delta(\eta, v) \\ &= F(X) + G(X)(u + \Delta(\eta, v)) \end{aligned} \quad (4.39)$$

Bộ điều khiển u được thiết kế để làm tối thiểu hàm mục tiêu

$$V = \int_0^\infty (z_v^T Q z_v + u^T R u) ds = \int_0^\infty (X^T Q_T X + u^T R u) ds \quad (4.40)$$

với $Q_T = \begin{bmatrix} Q & 0_{3 \times 6} \\ 0_{6 \times 3} & 0_{6 \times 6} \end{bmatrix}$, Q, R là các ma trận xác định dương.

4.2.2 Ứng dụng thuật toán ADP cấu trúc Actor-Critic

Áp dụng thuật toán ADP với cấu trúc Actor-Critic đã nêu trong chương (3.2.1) cho hệ (4.39) và hàm mục tiêu (4.40):

$$\begin{aligned} \hat{V}(X) &= \hat{W}_c^T \phi(X) \\ \hat{u}(X) &= -\frac{1}{2} R^{-1} G^T(X) \left(\frac{\partial \phi}{\partial X} \right)^T \hat{W}_a \end{aligned} \quad (4.41)$$

Viết lại sai số Bellman (3.7) ta có:

$$\delta_{hjb} = \hat{W}_c^T \omega + X^T Q_T X + \hat{u}^T R \hat{u} \quad (4.42)$$

với $\omega \triangleq \frac{\partial \phi}{\partial X} (F(X) + G(X)\hat{u})$ Luật cập nhật cho \hat{W}_c được thiết kế như sau:

$$\dot{\hat{W}}_c = -\eta_c \Gamma \frac{\omega}{1 + \nu \omega^T \Gamma \omega} \delta_{hjb} \quad (4.43)$$

trong đó ν, η_c là các hệ số dương, $\Gamma \in \mathbb{R}^{N \times N}$ là ma trận đối xứng được cập nhật như sau

$$\dot{\Gamma} = -\eta_c \Gamma \frac{\omega \omega^T}{1 + \nu \omega^T \Gamma \omega} \Gamma \quad (4.44)$$

$$\Gamma(t_r^+) = \Gamma(0) = \varphi_0 I \quad (4.45)$$

trong đó t_r^+ là thời điểm tại đó $\lambda_{min}(\Gamma) \leq \varphi_1, \varphi_0 > \varphi_1 > 0$. Việc reset lại giúp tránh khỏi covariance wind-up problem, tức là việc thích nghi chậm ở một số tham số. Như vậy Γ bị chặn bởi:

$$\varphi_1 I \leq \Gamma(t) \leq \varphi_0 I \quad (4.46)$$

Luật cập nhật cho \hat{W}_a được thiết kế theo phương pháp gradient như sau:

$$\begin{aligned} \dot{\hat{W}}_a = & -\frac{\eta_{a1}}{\sqrt{1+\omega^T\omega}} \frac{\partial\phi}{\partial X} G(X) R^{-1} G^T(X) \left(\frac{\partial\phi}{\partial X} \right)^T (\hat{W}_a - \hat{W}_c) \delta_{hjb} \\ & - \eta_{a2} (\hat{W}_a - \hat{W}_c) \end{aligned} \quad (4.47)$$

Để phân tích tính ổn định của thuật toán, ta đưa ra các giả thiết sau:

Giả thiết 7. *Ma trận $G(X)$ là bị chặn, tức là $0 < \|G(X)\| \leq \bar{G}$.*

Giả thiết 8. *Thành phần nhiễu Δ bị chặn, tức là $\|\Delta\| \leq \bar{\Delta}$*

Sai số Bellman được viết dưới dạng:

$$\begin{aligned} \delta_{hjb} = & \hat{W}_c^T \omega - \frac{\partial\phi}{X} (F(X) + G(X)u^*) - u^{*T} Ru^* + \hat{u}^T R \hat{u} \\ & - \frac{\partial\epsilon}{\partial X} (F(X) + G(X)u^*) \\ = & -\tilde{W}_c^T \omega + \frac{1}{4} \tilde{W}_a^T \left(\frac{\partial\phi}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial\phi}{\partial X} \right)^T \tilde{W}_a \\ & - \frac{1}{4} \left(\frac{\partial\epsilon}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial\epsilon}{\partial X} \right)^T - \frac{\partial\epsilon}{\partial X} (F(X) + G(X)u^*) \end{aligned} \quad (4.48)$$

Thay (4.48) vào (4.43) ta có:

$$\begin{aligned} \dot{\tilde{W}}_c = & -\eta_c \Gamma \psi \psi^T \tilde{W}_c + \eta_c \Gamma \frac{\omega}{1 + \nu \omega^T \Gamma \omega} \left(\frac{1}{4} \tilde{W}_a^T \left(\frac{\partial\phi}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial\phi}{\partial X} \right)^T \tilde{W}_a \right. \\ & \left. - \frac{1}{4} \left(\frac{\partial\epsilon}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial\epsilon}{\partial X} \right)^T - \frac{\partial\epsilon}{\partial X} (F(X) + G(X)u^*) \right) \end{aligned} \quad (4.49)$$

với $\psi(t) = \frac{\omega}{\sqrt{1 + \nu \omega^T \Gamma \omega}}$, bị chặn bởi

$$\|\psi\| \leq \frac{1}{\sqrt{\nu \varphi_1}} \quad (4.50)$$

Hệ nominal

$$\dot{\tilde{W}}_c = -\eta_c \Gamma \psi \psi^T \tilde{W}_c \quad (4.51)$$

ổn định hàm mũ, nếu tín hiệu $\psi(t)$ thỏa mãn điều kiện PE, tức là

$$\mu_2 I \geq \int_{t_0}^{t_0+\delta} \psi(s) \psi(s)^T ds \geq \mu_1 I, \forall t_0 \geq 0 \quad (4.52)$$

Định lý đảo Lyapunov chỉ ra rằng tồn tại hàm $V_c : \mathbb{R}^N \times [0, \infty) \rightarrow \mathbb{R}$ thỏa mãn các bất đẳng thức sau:

$$\begin{aligned} c_1 \|\tilde{W}_c\|^2 &\leq V_c(\tilde{W}_c, t) \leq c_2 \|\tilde{W}_c\|^2 \\ \frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_c} (-\eta_c \Gamma \psi \psi^T \tilde{W}_c) &\leq -c_3 \|\tilde{W}_c\|^2 \\ \left\| \frac{\partial V_c}{\partial \tilde{W}_c} \right\| &\leq c_4 \|\tilde{W}_c\| \end{aligned} \quad (4.53)$$

với $c_1, c_2, c_3, c_4 \in \mathbb{R}$ là các hằng số dương. Từ các giả thiết 2,3,4,7 ta có các đại lượng sau bị chặn:

$$\begin{aligned} \|\tilde{W}_a\| &\leq \kappa_1 \\ \left\| \frac{\partial \phi}{\partial X} G R^{-1} G^T \left(\frac{\partial \phi}{\partial X} \right) \right\| &\leq \kappa_2 \\ \left\| \frac{1}{4} \tilde{W}_a^T \left(\frac{\partial \phi}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \phi}{\partial X} \right)^T \tilde{W}_a - \frac{1}{4} \left(\frac{\partial \epsilon}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \epsilon}{\partial X} \right)^T \right. \\ \left. - \frac{\partial \epsilon}{\partial X} (F(X) + G(X)u^*) \right\| &\leq \kappa_3 \\ \left\| \frac{1}{2} W^T \left(\frac{\partial \phi}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \epsilon}{\partial X} \right)^T + \frac{1}{2} \left(\frac{\partial \epsilon}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \epsilon}{\partial X} \right)^T \right. \\ \left. + \frac{1}{2} W^T \left(\frac{\partial \phi}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \phi}{\partial X} \right)^T \tilde{W}_a \right. \\ \left. + \frac{1}{2} \left(\frac{\partial \epsilon}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \phi}{\partial X} \right)^T \tilde{W}_a \right\| &\leq \kappa_4 \end{aligned} \quad (4.54)$$

Định lí 4.2.1. Giả sử các giả thiết trên thỏa mãn, vector tín hiệu $\psi(t)$ thỏa mãn điều kiện PE, và điều kiện sau thỏa mãn

$$\frac{c_3}{\eta_{a1}} > \kappa_1 \kappa_2 \quad (4.55)$$

thì bộ điều khiển (4.41), với các luật cập nhật (4.43), (4.47) sẽ làm cho sai lệch bám của đối tượng (4.39) và sai lệch ước lượng \tilde{W}_a, \tilde{W}_c là UUB.

Chứng minh được trình bày trong phụ lục 1.

Dưới đây là kết quả mô phỏng thuật toán ADP cấu trúc Actor-Critic cho hệ tàu thủy (4.34) với các tham số điều khiển như hình (4.5-4.8):

$$\beta_\eta = 2 \quad \eta_c = 10 \quad \eta_{a1} = 10 \quad \eta_{a2} = 20$$

$$\nu = 0.01 \quad \varphi_0 = 20 \quad \varphi_1 = 12$$

$$Q = I_3 \quad R = 1.$$

Ta thu được kết quả mô phỏng như sau: Ta sử dụng mạng NN cho cả Actor

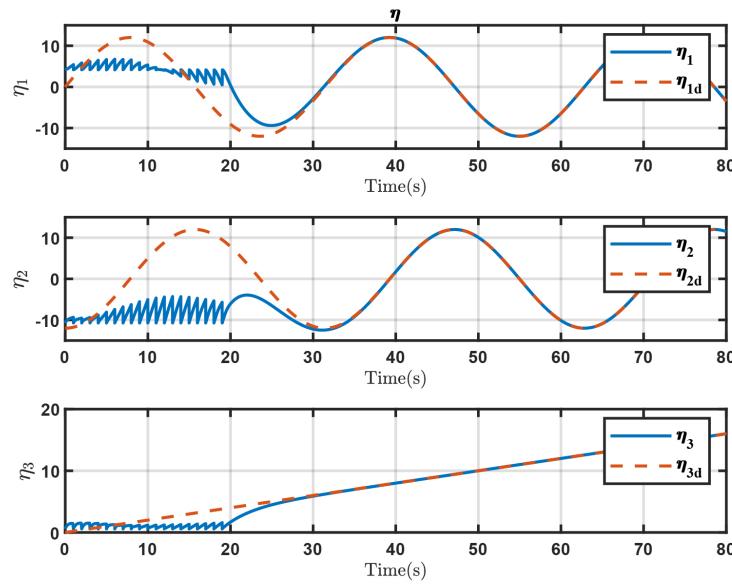
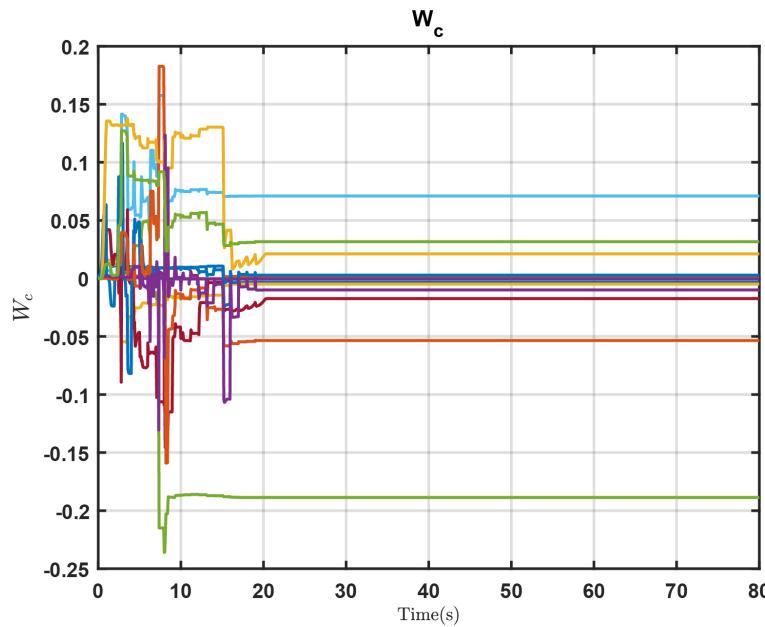


Figure 4.5: Trạng thái hệ tàu thủy với thuật toán ADP cấu trúc AC

Figure 4.6: Sự hội tụ của trọng số W_c của hệ tàu thủy với thuật toán ADP cấu trúc AC

và Critic chứa 12 node mạng với activation function như sau:

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

W_a và W_c hội tụ về giá trị:

$$\begin{bmatrix} -0.0033 & 0.0018 & -0.0049 & -0.0005 & -0.1886 & 0.0709 \\ -0.0174 & 0.0027 & -0.0534 & 0.0211 & -0.0099 & 0.0316 \end{bmatrix}^T$$

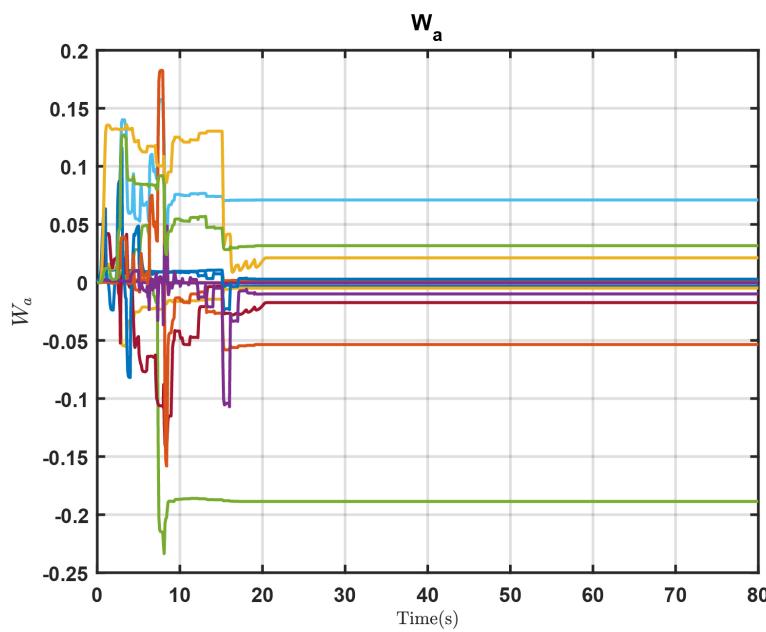


Figure 4.7: Sự hội tụ của trọng số W_a của hệ tàu thủy với thuật toán ADP cấu trúc AC

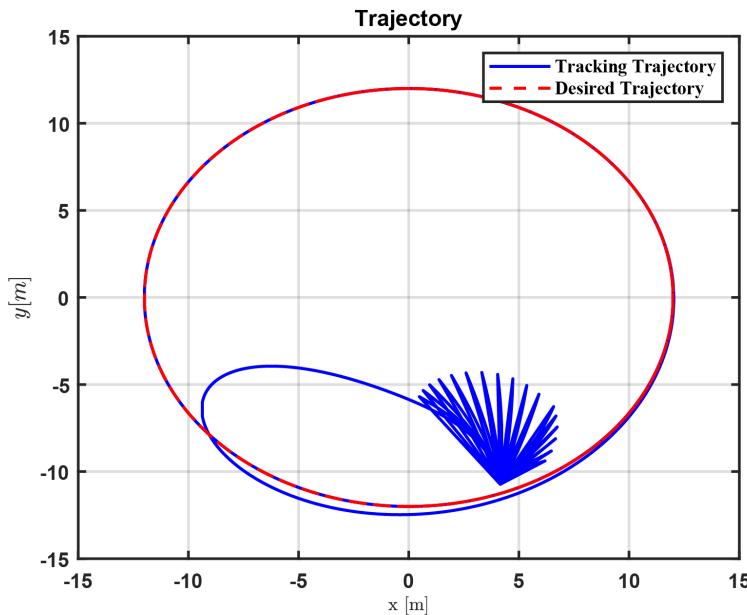


Figure 4.8: quỹ đạo của hệ tàu thủy với thuật toán ADP cấu trúc AC

4.2.3 Ứng dụng thuật toán Online On-Policy IRL

Như đã trình bày ở trên thì thuật toán IRL hoạt động dựa trên thông tin đầu vào và ra của hệ thống, nhờ đó việc hiểu biết thông tin hệ thống được giảm tải. Tuy nhiên việc này cũng ảnh hưởng tới việc đổi chiều kết quả hai thuật toán AC và IRL ở chỗ: khi hệ thống có nhiều $\Delta(\eta, v)$ như nêu trong mô hình, với thuật toán IRL, nó sẽ coi biến đổi do nhiều gây ra tương đương với biến đổi của mô hình, hay rõ hơn là coi nhiều như một phần của mô hình. Do đó để so sánh kiểm chứng giữa hai thuật toán AC và IRL có cho

cùng kết quả, trong phần ứng dụng IRL này sẽ bỏ qua tác động của nhiễu ($\Delta(\eta, v) = 0, \forall \eta, v$).

Như vậy ta viết lại hệ thống (4.39) như sau:

$$\begin{aligned}\dot{X} &= \begin{bmatrix} -M^{-1}l(z_v + v_d(z_\eta, \eta_d)) + M^{-1}l(v_d(z_\eta, \eta_d)) \\ J(z_\eta + \eta_d)z_v - \beta_\eta z_\eta \\ h_1(\eta_d) \end{bmatrix} + \begin{bmatrix} M^{-1} \\ 0 \\ 0 \end{bmatrix} u \\ &= F(X) + G(X)u\end{aligned}\quad (4.56)$$

và hàm mục tiêu

$$V = \int_0^\infty (z_v^T Q z_v + u^T R u) ds = \int_0^\infty (X^T Q_T X + u^T R u) ds \quad (4.57)$$

với $Q_T = \begin{bmatrix} Q & 0_{3 \times 6} \\ 0_{6 \times 3} & 0_{6 \times 6} \end{bmatrix}$, Q, R là các ma trận xác định dương. Ta nhận thấy hệ (4.56) thỏa mãn các yêu cầu thuật toán IRL, do đó áp dụng thuật toán IRL nêu trong chương 3.3 cho hệ (4.56) và hàm mục tiêu (4.57) với các tham số điều khiển $Q = I_3, R = 1..$. Mạng NN được sử dụng chứa 12 node với activation function chọn tương tự thuật toán cấu trúc AC.

Ta thu được kết quả mô phỏng như hình (4.9-4.11):

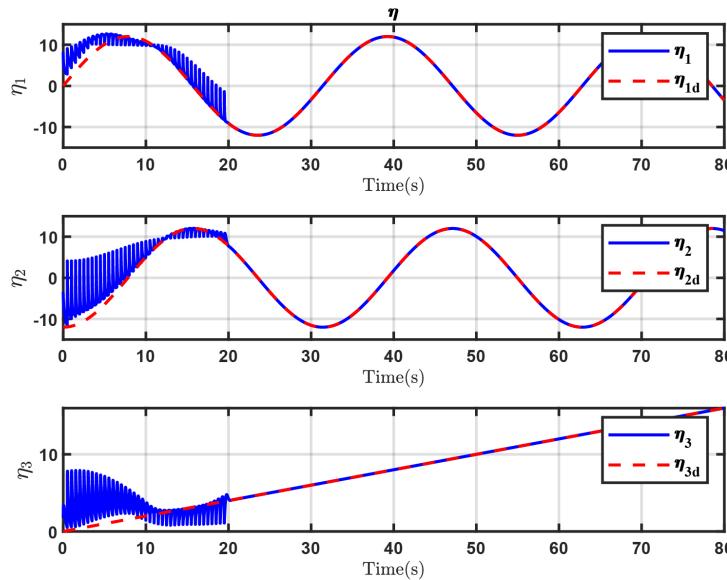


Figure 4.9: Trạng thái hệ tàu thủy với thuật toán Online On-Policy IRL

W hội tụ về giá trị:

$$\begin{bmatrix} -0.0034 & 0.0019 & -0.0049 & -0.0005 & -0.1890 & 0.0715 \\ -0.0180 & 0.0028 & -0.0533 & 0.0201 & -0.0099 & 0.0311 \end{bmatrix}^T$$

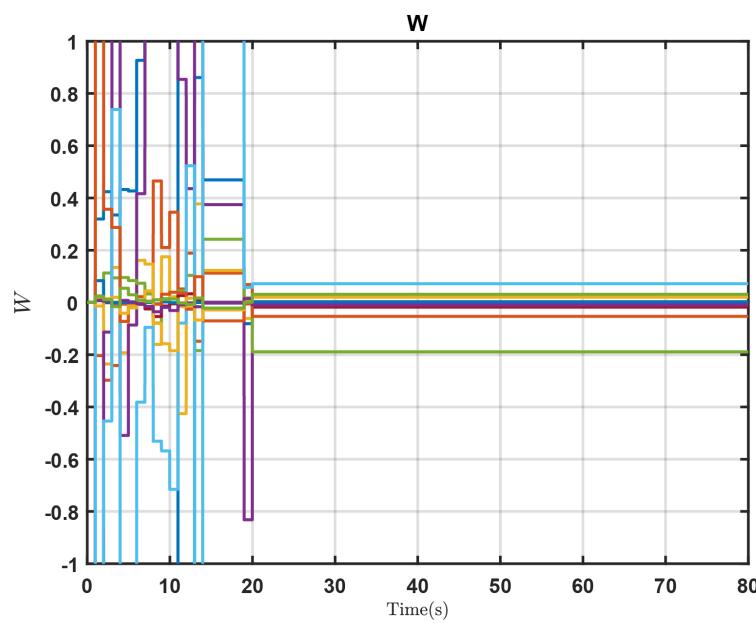


Figure 4.10: Sự hội tụ của trọng số W của hệ tàu thủy với thuật toán Online On-Policy IRL

Nhận thấy kết quả hội tụ của giá trị W trong thuật toán AC và Online On-Policy IRL là khá sát nhau. Qua đó cho thấy tính đúng đắn của thuật toán.

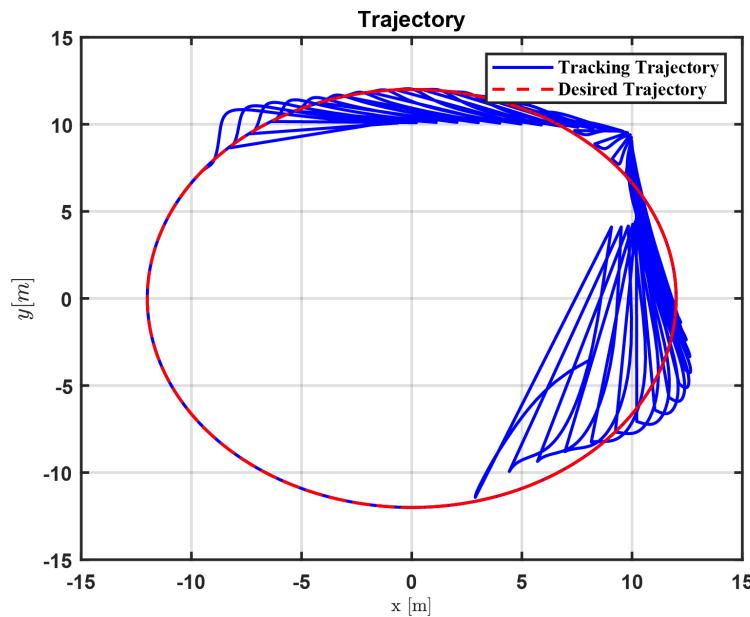


Figure 4.11: quỹ đạo của hệ tàu thủy với thuật toán Online On-Policy IRL

4.2.4 Ứng dụng thuật toán Online Off-Policy IRL

Như đã thấy trong hai thuật toán trên, với bài toán tracking, yêu cầu một tín hiệu điều khiển không tiến về 0 khi hệ đạt trạng thái dừng. Do đó hai thuật toán trên tách tín hiệu điều khiển thành hai phần: một tín hiệu điều khiển giúp giữ hệ thống trên quỹ đạo mong muốn khi hệ thống đã nằm trên quỹ đạo này (có thể một bộ điều khiển feed forward phụ thuộc vào mô hình hệ thống và quỹ đạo mong muốn như đã trình bày ở trên, tín hiệu này trong đa số trường hợp không tiến về 0), tín hiệu điều khiển thứ hai nhằm đưa hệ thống về quỹ đạo mong muốn (được xây dựng dựa trên thuật toán RL, tín hiệu này sẽ tiến về 0 khi hệ đạt quỹ đạo mong muốn). Trong hàm chi phí hay hàm mục tiêu thì ta chỉ xét tới tín hiệu điều khiển thứ hai bởi nếu xét cả tín hiệu thứ nhất thì hàm này sẽ không bị chặn và tiến ra vô cùng. Qua đó ta thấy được việc tối ưu như nêu trong hai thuật toán trên là chưa triệt để. Ngoài ra việc vẫn phải biết thông tin hệ thống, với AC là biết hoàn toàn và với IRL là một phần, đó cũng là một nhược điểm của hai thuật toán trên. Đó là lý do thuật toán Off-Policy IRL với sự xuất hiện của hàm phạt được nghiên cứu như trong chương (2.1.2).

Ta biến đổi hệ (4.34) với việc bỏ đi phần tín hiệu điều khiển feed forward τ_d (4.38) như sau: Định nghĩa hàm $l(x) = C(x)x + D(x)x$, biến $X = [z_v^T, z_\eta^T, \eta_d^T]^T$, $u = \tau$, $d = \Delta(\eta, v)$, $v_d(z_\eta, \eta_d) = J^T(z_\eta + \eta_d)(\dot{\eta}_d - \beta_\eta z_\eta)$ và giả sử có quan hệ $\dot{\eta}_d = h_1(\eta_d)$, $\ddot{\eta}_d = h_2(\eta_d)$, ta kết hợp với (4.35) và (4.37) thu được:

$$\begin{aligned} \dot{X} &= \begin{bmatrix} -\dot{v}_d(z_\eta, \eta_d) - M^{-1}l(z_v + v_d(z_\eta, \eta_d)) - M^{-1}g(z_\eta + \eta_d) \\ J(z_\eta + \eta_d)z_v - \beta_\eta z_\eta \\ h_1(\eta_d) \end{bmatrix} + \begin{bmatrix} M^{-1} \\ 0 \\ 0 \end{bmatrix} u + \begin{bmatrix} M^{-1} \\ 0 \\ 0 \end{bmatrix} d \\ &= F(X) + G(X)(u + d) \end{aligned} \quad (4.58)$$

Với hàm mục tiêu:

$$V(x, u, d) = \int_t^\infty e^{-\alpha(\tau-t)}(X^T Q X + u^T R u - \gamma^2 d^T d) d\tau \quad (4.59)$$

Nhận thấy hệ (4.58) thỏa mãn các yêu cầu của thuật toán Off-Policy IRL với hàm phạt, ta áp dụng thuật toán đã nêu trong chương 3.4 cho hệ (4.58) và hàm mục tiêu (4.59) ta thu được kết quả mô phỏng như hình (4.12-4.13):

Kết quả cho thấy tốc độ bám của hệ thống là khá cao bởi đã trải qua quá trình tối ưu.

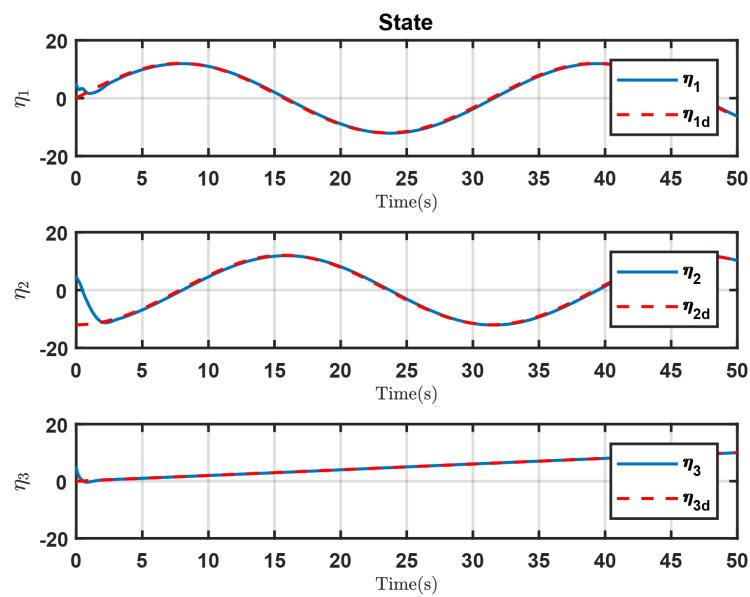


Figure 4.12: Trạng thái hệ tàu thủy với thuật toán Online Off-Policy IRL chưa hàm phạt

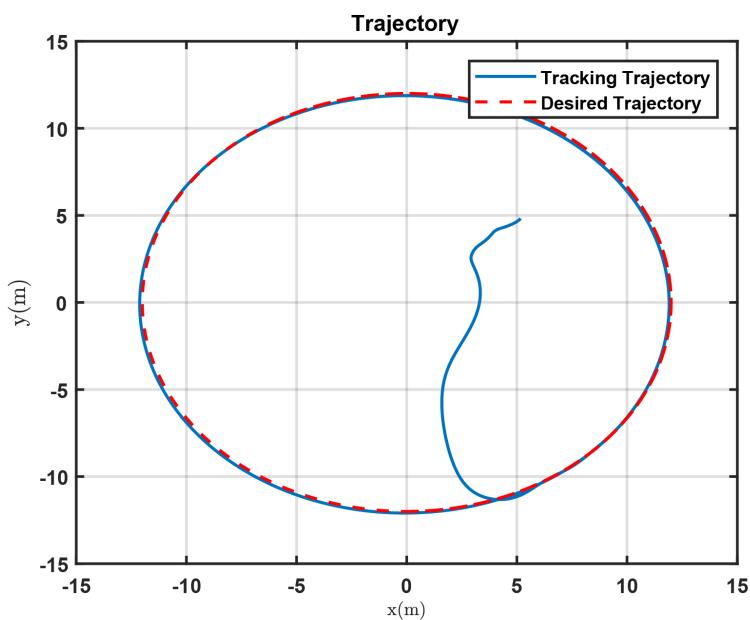


Figure 4.13: Quỹ đạo của hệ tàu thủy với thuật toán Online Off-Policy IRL chưa hàm phạt

Kết luận

Thông qua phân tích, có thể thấy được tiềm năng của thuật toán ADP trong việc giải quyết đồng thời bài toán tối ưu và thích nghi trong điều khiển. Qua các định lý và mô phỏng, sự chặt chẽ về tính hội tụ của thuật toán được thể hiện rõ ràng. Việc ứng dụng thuật toán vào hệ điều khiển tối ưu bám quy đạo cho tàu thủy cũng cho thấy tính khả thi của thuật toán khi áp dụng vào những hệ thống trong thực tế.

Tuy nhiên, việc ứng dụng NN tuyến tính vào thuật toán ADP còn gặp nhiều hạn chế. NN tuyến tính tuy phù hợp với phân tích toán học chặt chẽ, nhưng có khả năng xấp xỉ hàm yếu, cùng với vấn đề về bùng nổ kích thước NN khi số lượng đầu vào mạng tăng. Do đó hướng phát triển tương lai của đề tài là áp dụng NN nhân tạo hay NN phi tuyến vào thuật toán thay thế NN tuyến tính, giúp thuật toán phù hợp hơn với bối cảnh thực tế.

Bibliography

- [1] Murad Abu-Khalaf and Frank L Lewis. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network hjb approach. *Automatica*, 41(5):779–791, 2005.
- [2] MDS Aliyu. *Nonlinear H-infinity control, Hamiltonian systems and Hamilton-Jacobi equations*. CRC Press, 2017.
- [3] Tamer Basar and Pierre Bernhard. $H\infty$ -optimal control and related minimax design problems. *systems & control: Foundations & applications*, 1995.
- [4] Girish Chowdhary and Eric Johnson. Concurrent learning for convergence in adaptive control without persistency of excitation. In *49th IEEE Conference on Decision and Control (CDC)*, pages 3674–3679. IEEE, 2010.
- [5] Odd Faltinsen. *Sea loads on ships and offshore structures*, volume 1. Cambridge university press, 1993.
- [6] Thor I Fossen. Marine control systems—guidance, navigation, and control of ships, rigs and underwater vehicles. *Marine Cybernetics, Trondheim, Norway, Org. Number NO 985 195 005 MVA, www. marinecybernetics.com, ISBN: 82 92356 00 2*, 2002.
- [7] Thor I Fossen et al. *Guidance and control of ocean vehicles*, volume 199. Wiley New York, 1994.
- [8] Thor I Fossen and Ola-Erik Fjellstad. Nonlinear modelling of marine vehicles in 6 degrees of freedom. *Mathematical Modelling of Systems*, 1:17–27, 1995.
- [9] Thor I Fossen, Karl-Petter Lindegaard, and Roger Skjetne. Inertia shaping techniques for marine vessels using acceleration feedback. *IFAC Proceedings Volumes*, 35(1):343–348, 2002.
- [10] Haibo He, Zhen Ni, and Jian Fu. A three-network architecture for online learning and optimization based on adaptive dynamic programming. *Neurocomputing*, 78(1):3–13, 2012.

- [11] Yu Jiang and Zhong-Ping Jiang. Robust adaptive dynamic programming and feedback stabilization of nonlinear systems. *IEEE Transactions on Neural Networks and Learning Systems*, 25(5):882–893, 2014.
- [12] Malvin H Kalos and Paula A Whitlock. *Monte carlo methods*. John Wiley & Sons, 2009.
- [13] Rushikesh Kamalapurkar, Huyen Dinh, Shubhendu Bhasin, and Warren E Dixon. Approximate optimal trajectory tracking for continuous-time nonlinear systems. *Automatica*, 51:40–48, 2015.
- [14] James Nate Knight and Charles Anderson. Stable reinforcement learning with recurrent neural networks. *Journal of Control Theory and Applications*, 9(3):410–420, 2011.
- [15] Jae Young Lee, Jin Bae Park, and Yoon Ho Choi. Integral reinforcement learning for continuous-time input-affine nonlinear systems with simultaneous invariant explorations. *IEEE Transactions on Neural Networks and Learning Systems*, 26(5):916–932, 2014.
- [16] Frank L Lewis and Derong Liu. *Reinforcement learning and approximate dynamic programming for feedback control*, volume 17. John Wiley & Sons, 2013.
- [17] Frank L Lewis, Draguna Vrabie, and Vassilis L Syrmos. *Optimal control*. John Wiley & Sons, 2012.
- [18] Hamidreza Modares, Bahare Kiumarsi, Kyriakos G Vamvoudakis, and Frank L Lewis. Adaptive \mathbb{H}_∞ tracking control of nonlinear systems using reinforcement learning. In *Adaptive Learning Methods for Nonlinear System Modeling*, pages 313–333. Elsevier, 2018.
- [19] Hamidreza Modares and Frank L Lewis. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. *Automatica*, 50(7):1780–1792, 2014.
- [20] Hamidreza Modares, Frank L Lewis, and Zhong-Ping Jiang. \mathbb{H}_∞ tracking control of completely unknown continuous-time systems via off-policy reinforcement learning. *IEEE transactions on neural networks and learning systems*, 26(10):2550–2562, 2015.
- [21] Hamidreza Modares, Frank L Lewis, and Mohammad-Bagher Naghibi-Sistani. Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks. *IEEE transactions on neural networks and learning systems*, 24(10):1513–1525, 2013.

- [22] John J Murray, Chadwick J Cox, George G Lendaris, and Richard Saeks. Adaptive dynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 32(2):140–153, 2002.
- [23] Nguyễn Doãn Phước. *Tối ưu hóa và điều khiển tối ưu*. NXB Bách Khoa Hà Nội, 2010.
- [24] George N Saridis and Chun-Sing G Lee. An approximation theory of optimal control for trainable manipulators. *IEEE Transactions on systems, Man, and Cybernetics*, 9(3):152–159, 1979.
- [25] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [26] Kyriakos G Vamvoudakis and Frank L Lewis. Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica*, 46(5):878–888, 2010.
- [27] Kyriakos G Vamvoudakis, Marcio Fantini Miranda, and João P Hespanha. Asymptotically stable adaptive–optimal control algorithm with saturating actuators and relaxed persistence of excitation. *IEEE transactions on neural networks and learning systems*, 27(11):2386–2398, 2015.
- [28] Kyriakos G Vamvoudakis, Draguna Vrabie, and Frank L Lewis. Online adaptive algorithm for optimal control with integral reinforcement learning. *International Journal of Robust and Nonlinear Control*, 24(17):2686–2710, 2014.
- [29] Draguna Vrabie and Frank Lewis. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Networks*, 22(3):237–246, 2009.
- [30] Paweł Wawrzynski. Autonomous reinforcement learning with experience replay for humanoid gait optimization. *Procedia Computer Science*, 13:205–211, 2012.
- [31] Xin Xu, Lei Zuo, and Zhenhua Huang. Reinforcement learning algorithms with function approximation: Recent advances and applications. *Information Sciences*, 261:1–31, 2014.
- [32] Xiong Yang, Haibo He, Derong Liu, and Yuanheng Zhu. Adaptive dynamic programming for robust neural control of unknown continuous-time non-linear systems. *IET Control Theory & Applications*, 11(14):2307–2316, 2017.

- [33] Yuanheng Zhu, Dongbin Zhao, and Xiangjun Li. Using reinforcement learning techniques to solve continuous-time non-linear optimal tracking problem without system dynamics. *IET Control Theory & Applications*, 10(12):1339–1347, 2016.

Phụ lục 1. Chứng minh ổn định

Để phân tích tính ổn định của hệ thống, xét hàm ứng viên Lyapunov:

$$V_L \triangleq \frac{1}{2}\rho z_\eta^T z_\eta + V^*(X) + V_c(\tilde{W}_c, t) + \frac{1}{2}\tilde{W}_a^T \tilde{W}_a$$

với ρ là 1 hằng số dương. Vì $V^*(X)$ có đạo hàm liên tục và xác định dương nên tồn tại 2 hàm lớp K_{α_1, α_2} thỏa mãn

$$\alpha_1(\|X\|) \leq V^*(X) \leq \alpha_2(\|X\|) \quad (60)$$

Từ (4.53) và (60) ta có:

$$\begin{aligned} & \frac{1}{2}\rho \|z_\eta\|^2 + \alpha_1(\|X\|) + c_1 \|\tilde{W}_c\|^2 + \frac{1}{2}\|\tilde{W}_a\|^2 \leq V_L \\ & \leq \frac{1}{2}\rho \|z_\eta\|^2 + \alpha_2(\|X\|) + c_2 \|\tilde{W}_c\|^2 + \frac{1}{2}\|\tilde{W}_a\|^2 \end{aligned}$$

Lấy đạo hàm V_L ta có:

$$\begin{aligned} \dot{V}_L = & \rho z_\eta^T (J(\eta)z_v - \beta_\eta z_\eta) + \frac{\partial V^*}{\partial X} (F(X) + G(X)\hat{u}) \\ & + \frac{\partial V_c}{\partial t} + \frac{\partial V_c}{\partial \tilde{W}_c} (\Omega_{nom} + \Delta_{per}) - \tilde{W}_a^T \dot{\tilde{W}}_a + \frac{\partial V^*}{\partial X} G(X)\Delta \end{aligned} \quad (61)$$

trong đó

$$\begin{aligned} \Omega_{nom} = & -\eta_c \Gamma \psi \psi^T \tilde{W}_c \\ \Delta_{per} = & \eta_c \Gamma \frac{\omega}{1 + \nu \omega^T \Gamma \omega} \left(\frac{1}{4} \tilde{W}_a^T \left(\frac{\partial \phi}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \phi}{\partial X} \right)^T \tilde{W}_a \right. \\ & \left. - \frac{1}{4} \left(\frac{\partial \epsilon}{\partial X} \right) G R^{-1} G^T \left(\frac{\partial \epsilon}{\partial X} \right)^T - \frac{\partial \epsilon}{\partial X} (F(X) + G(X)u^*) \right) \end{aligned}$$

Từ (2.19) ta có $\frac{\partial V^*}{\partial X} F = -\frac{\partial V^*}{\partial X} G u^* - X^T Q_T X - u^{*T} R u^*$, $\frac{\partial V^*}{\partial X} G = -2u^{*T} R$.

Sử dụng (4.47), (4.53) viết lại (61) thành:

$$\begin{aligned} \dot{V}_L \leq & -\rho \beta_\eta \|z_\eta\|^2 + \rho z_\eta^T J(\eta) z_v - z_v^T Q z_v - u^{*T} R u^* + 2u^{*T} R(u^* - \hat{u}) \\ & - c_3 \|\tilde{W}_c\|^2 + c_4 \|\tilde{W}_c\| \|\Delta_{per}\| + \eta_{a2} \tilde{W}_a^T (\hat{W}_a - \hat{W}_c) - 2u^{*T} R \Delta \\ & + \frac{\eta_{a1}}{\sqrt{1 + \omega^T \omega}} \tilde{W}_a^T \frac{\partial \phi}{\partial X} G(X) R^{-1} G^T(X) \left(\frac{\partial \phi}{\partial X} \right)^T (\hat{W}_a - \hat{W}_c) \delta_{hjb} \end{aligned} \quad (62)$$

Sử dụng bất đẳng thức Young ta có:

$$\begin{aligned} \rho z_\eta^T J(\eta) z_v & \leq \frac{\rho}{2} \|z_\eta\|^2 + \frac{\rho}{2} z_v^T J^T(\eta) J(\eta) z_v \\ & = \frac{\rho}{2} \|z_\eta\|^2 + \frac{\rho}{2} \|z_v\|^2 \end{aligned} \quad (63)$$

Từ (2.19), (4.41) và (4.54) ta có:

$$\begin{aligned}
 2u^{*T}R(u^* - \hat{u}) &= \frac{1}{2}W^T \frac{\partial\phi}{\partial X} GR^{-1}G^T \left(\frac{\partial\epsilon}{\partial X} \right)^T \\
 &\quad + \frac{1}{2}W^T \frac{\partial\phi}{\partial X} GR^{-1}G^T \left(\frac{\partial\phi}{\partial X} \right)^T \tilde{W}_a \\
 &\quad + \frac{1}{2} \frac{\partial\epsilon}{\partial X} GR^{-1}G^T \left(\frac{\partial\phi}{\partial X} \right)^T \tilde{W}_a \\
 &\quad + \frac{1}{2} \frac{\partial\epsilon}{\partial X} GR^{-1}G^T \left(\frac{\partial\epsilon}{\partial X} \right)^T \leq \kappa_4
 \end{aligned} \tag{64}$$

Thành phần Δ_{per} bị chấn bởi:

$$\|\Delta_{per}\| \leq \frac{\eta_c\varphi_0}{2\sqrt{\nu\varphi_1}}\kappa_3 \tag{65}$$

Sử dụng (4.54) và (4.48) ta có

$$\begin{aligned}
 &\frac{\eta_{a1}}{\sqrt{1+\omega^T\omega}}\tilde{W}_a^T \frac{\partial\phi}{\partial X} G(X)R^{-1}G^T(X) \left(\frac{\partial\phi}{\partial X} \right)^T (\hat{W}_a - \hat{W}_c) \delta_{hjb} \\
 &= \frac{\eta_{a1}}{\sqrt{1+\omega^T\omega}}\tilde{W}_a^T \frac{\partial\phi}{\partial X} G(X)R^{-1}G^T(X) \left(\frac{\partial\phi}{\partial X} \right)^T (\tilde{W}_c - \tilde{W}_a) \\
 &\quad \times \left(-\tilde{W}_c^T\omega + \frac{1}{4}\tilde{W}_a^T \left(\frac{\partial\phi}{\partial X} \right) GR^{-1}G^T \left(\frac{\partial\phi}{\partial X} \right)^T \tilde{W}_a \right. \\
 &\quad \left. - \frac{1}{4} \left(\frac{\partial\epsilon}{\partial X} \right) GR^{-1}G^T \left(\frac{\partial\epsilon}{\partial X} \right)^T - \frac{\partial\epsilon}{\partial X}(F(X) + G(X)u^*) \right) \\
 &\leq \eta_{a1}\kappa_1\kappa_2\|\tilde{W}_c\|^2 + \eta_{a1}\kappa_1^2\kappa_2\|\tilde{W}_c\| + \eta_{a1}\kappa_1\kappa_2\kappa_3\|\tilde{W}_c\| + \eta_{a1}\kappa_1^2\kappa_2\kappa_3
 \end{aligned} \tag{66}$$

Sử dụng biến đổi sau:

$$\begin{aligned}
 \eta_{a2}\tilde{W}_a^T(\hat{W}_a - \hat{W}_c) &= \eta_{a2}\tilde{W}_a^T(\tilde{W}_c - \tilde{W}_a) \\
 &\leq \eta_{a2}\kappa_1\|\tilde{W}_c\| - \eta_{a2}\|\tilde{W}_a\|^2
 \end{aligned} \tag{67}$$

Lại có điều kiện:

$$-u^{*T}Ru^* - 2u^{*T}R\Delta \leq \Delta^T R \Delta \leq \lambda_{max}(R)\bar{\Delta}^2 \tag{68}$$

Thay (63) và (67) vào (62) ta có:

$$\begin{aligned}
 \dot{V}_L &\leq -\rho \left(\beta_\eta - \frac{1}{2} \right) \|z_\eta\|^2 - z_v^T(Q - 0.5\rho I)z_v - (c_3 - \eta_{a1}\kappa_1\kappa_2)\|\tilde{W}_c\|^2 \\
 &\quad - \eta_{a2}\|\tilde{W}_a\|^2 + \eta_{a1}\kappa_1^2\kappa_2\kappa_3 + \kappa_4 + \left(\frac{c_4\eta_c\varphi_0}{2\sqrt{\nu\varphi_1}}\kappa_3 + \eta_{a1}\kappa_1\kappa_2\kappa_3 \right. \\
 &\quad \left. + \eta_{a1}\kappa_1^2\kappa_2 + \eta_{a2}\kappa_1 \right) \|\tilde{W}_c\| + \lambda_{max}(R)\bar{\Delta}^2
 \end{aligned} \tag{69}$$

Sử dụng biến đổi $ab \leq \gamma a^2 + \frac{1}{4\gamma}b^2$ ta có:

$$\begin{aligned}\dot{V}_L &\leq -\rho \left(\beta_\eta - \frac{1}{2} \right) \|z_\eta\|^2 - z_v^T (Q - 0.5\rho I) z_v + \lambda_{max}(R) \bar{\Delta}^2 \\ &\quad - (1-\theta) (c_3 - \eta_{a1}\kappa_1\kappa_2) \|\tilde{W}_c\|^2 - \eta_{a2} \|\tilde{W}_a\|^2 + \eta_{a1}\kappa_1^2\kappa_2\kappa_3 + \kappa_4 \\ &\quad + \frac{1}{4\theta(c_3 - \eta_{a1}\kappa_1\kappa_2)} \left(\frac{c_4\eta_c\varphi_0}{2\sqrt{\nu\varphi_1}} \kappa_3 + \eta_{a1}\kappa_1\kappa_2\kappa_3 + \eta_{a1}\kappa_1^2\kappa_2 + \eta_{a2}\kappa_1 \right)^2\end{aligned}\quad (70)$$

với các tham số thỏa mãn $\beta_\eta > \frac{1}{2}$, $\rho < 2\lambda_{min}(Q)$, $0 < \theta < 1$, $c_3 > \eta_{a1}\kappa_1\kappa_2$. Định nghĩa $z = [z_\eta^T, z_v^T, \tilde{W}_c^T, \tilde{W}_a^T]^T$. Tồn tại hàm lớp K α_3, α_4 thỏa mãn

$$\begin{aligned}\alpha_3(\|z\|) &\leq \rho \left(\beta_\eta - \frac{1}{2} \right) \|z_\eta\|^2 + z_v^T (Q - 0.5\rho I) z_v \\ &\quad + (1-\theta) (c_3 - \eta_{a1}\kappa_1\kappa_2) \|\tilde{W}_c\|^2 + \eta_{a2} \|\tilde{W}_a\|^2 \leq \alpha_4(\|z\|)\end{aligned}\quad (71)$$

Viết lại (70) như sau:

$$\begin{aligned}\dot{V}_L &\leq -\alpha_3(\|z\|) + \eta_{a1}\kappa_1^2\kappa_2\kappa_3 + \kappa_4 + \lambda_{max}(R) \bar{\Delta}^2 \\ &\quad + \frac{1}{4\theta(c_3 - \eta_{a1}\kappa_1\kappa_2)} \left(\frac{c_4\eta_c\varphi_0}{2\sqrt{\nu\varphi_1}} \kappa_3 + \eta_{a1}\kappa_1\kappa_2\kappa_3 + \eta_{a1}\kappa_1^2\kappa_2 + \eta_{a2}\kappa_1 \right)^2\end{aligned}\quad (72)$$

Vậy $\|z\|$ là UUB với miền hấp dẫn

$$\Omega_z \triangleq \left\{ z : \|z\| \leq \alpha_3^{-1} \left(\frac{1}{4\theta(c_3 - \eta_{a1}\kappa_1\kappa_2)} \left(\frac{c_4\eta_c\varphi_0}{2\sqrt{\nu\varphi_1}} \kappa_3 + \eta_{a1}\kappa_1\kappa_2\kappa_3 + \eta_{a1}\kappa_1^2\kappa_2 + \eta_{a2}\kappa_1 \right)^2 + \eta_{a1}\kappa_1^2\kappa_2\kappa_3 + \kappa_4 + \lambda_{max}(R) \bar{\Delta}^2 \right) \right\}$$

Phụ lục 2. Chứng minh nghiệm tương đương

Với $u^{(i)} \in \Psi(\Omega)$, $V^{u^{(i)}} \in C^1(\Omega)$ được định nghĩa bởi $V^{u^{(i)}} = \int_t^\infty r(x(\tau), u^{(i)}(x(\tau))) d\tau$, là một hàm Lyapunov cho hệ thống $\dot{x}(t) = f(x(t), u^{(i)}(x(t)))$. $V^{u^{(i)}} \in C^1(\Omega)$ thỏa mãn:

$$(\nabla V_x^{u^{(i)}})^T (f(x(t), u^{(i)}(x(t)))) = -r(x(t), u^{(i)}(x(t))) \quad (73)$$

với $r(x(t), u^{(i)}(x(t))) > 0; x(t) \neq 0$. Tích phân (73) trên khoảng thời gian $[t, t+T]$, ta thu được:

$$V^{u^{(i)}}(x(t)) = \int_t^{t+T} r(x(\tau), u^{(i)}(x(\tau))) d\tau + V^{u^{(i)}}(x(t+T)). \quad (74)$$

Điều này nghĩa là nghiệm duy nhất của (2.21), $V^{u^{(i)}}$ cũng thỏa mãn (74). Để hoàn thiện chứng minh, chúng ta phải chỉ ra rằng phương trình (74) có một nghiệm duy nhất.

Do đó, giả sử tồn tại một hàm chi phí khác $V \in C^1(\Omega)$ thoả mãn (3.3) với điều kiện $V(0) = 0$. Hàm chi phí này cũng thỏa mãn $\dot{V}(x(t)) = -r(x(t), u^{(i)}(x(t)))$. Thay vào (74) ta thu được:

$$\begin{aligned} & \left(\frac{d[V(x(t)) - V^{u^{(i)}}(x(t))]^T}{dx} \right) \dot{x} \\ &= \left(\frac{d[V(x(t)) - V^{u^{(i)}}(x(t))]^T}{dx} \right) \dot{x} \times (f(x(t), u^{(i)}(x(t)))) = 0 \end{aligned} \quad (75)$$

điều này đúng với mọi quỹ đạo trạng thái x được tạo ra của hệ thống với luật điều khiển ổn định $u^{(i)}$. Do đó, $V(x(t)) = V^{u^{(i)}}(x(t)) + c$. Quan hệ này vẫn đúng với $x(t) = 0$ do đó $V(0) = V^{u^{(i)}}(0) + c \rightarrow 0 = c$ và do đó $V(x(t)) = V^{u^{(i)}}(x(t))$. Vậy phương trình (2.21) có một nghiệm duy nhất thì nghiệm này trùng với nghiệm duy nhất của phương trình (3.16). Chứng minh hoàn thành.