

TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI



ĐỒ ÁN TỐT NGHIỆP

THIẾT KẾ BỘ ĐIỀU KHIỂN TỐI ƯU CHO THIẾT BỊ PHẢN ỨNG CSTR CHỊU TÁC DỤNG CỦA NHIỀU VÀ BẤT ĐỊNH

NGUYỄN TRUNG NGHĨA

nghia.nt162913@sis.hust.edu.vn

Ngành KT Điều khiển & 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, 07/2021

**NHIỆM VỤ
ĐỒ ÁN TỐT NGHIỆP**

Họ và tên sinh viên:

Khóa.....Viện: Điện

Ngành: CN ĐK & TĐH

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)

Tóm tắt nội dung đồ án

Nội dung chính của đồ án tốt nghiệp do sinh viên thực hiện tập trung vào thiết kế thuật toán để tính toán bộ điều khiển phản hồi trạng thái tối ưu cho hệ thống bình khuấy trộn phản ứng liên tục (CSTR) trong công nghiệp trong trường hợp hệ chịu tác động của bất định và nhiễu. Mô hình được sử dụng trong đồ án là mô hình phi tuyến, có ràng buộc đầu vào. Bài toán trở thành điều khiển tối ưu phi tuyến có ràng buộc, chịu tác động của nhiễu và bất định. Như ta đã biết, để điều khiển tối ưu một hệ thống ta cần giải các phương trình HJB, Riccati là những phương trình phi tuyến và phụ thuộc vào động học của hệ thống nên rất khó để giải tường minh trong thực tế. Đồ án này đưa ra thuật toán dựa trên nguyên lý quy hoạch động trong thuật toán học tăng cường để giải quyết các phương trình này tìm ra bộ điều khiển tối ưu. Cộng với đó các thuật toán được đề xuất chỉ sử dụng các thông tin từ đầu vào và đầu ra của hệ thống thay vì sử dụng toàn bộ các thông tin về biến trạng thái. Phương án thiết kế được đề xuất cùng với được mô phỏng kiểm chứng trong phần mềm Matlab/Simulink. Kết quả đồ án sẽ chỉ ra phương pháp thiết kế điều khiển phù hợp thiết kế điều khiển không chỉ hệ bình phản ứng hóa học trong công nghiệp mà còn có thể mở rộng điều khiển nhiều hệ thống phi tuyến khác trong các ứng dụng thực tế. Sau khi hoàn thành quá trình triển khai thực hiện đồ án, sinh viên thu được kinh nghiệm áp dụng kiến thức lý thuyết vào thiết kế điều khiển một hệ thống trong thực tế cùng với đó xây dựng được phương pháp tư duy khoa học khi phát triển một thuật toán thiết kế bộ điều khiển mới.

Sinh viên thực hiện
Ký và ghi rõ họ tên

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU	5
1.1. TỔNG QUAN VỀ LÝ THUYẾT ĐIỀU KHIỂN.....	5
1.2. TỔNG QUAN VỀ MÔ HÌNH BÌNH PHẢN ỨNG KHUẤY TRỘN LIÊN TỤC.....	5
CHƯƠNG 2: KIẾN THỨC SỬ DỤNG TRONG ĐỒ ÁN	7
2.1. MỘT SỐ BẤT ĐẲNG THỨC TOÁN HỌC CẦN THIẾT	7
2.2. LÝ THUYẾT ĐIỀU KHIỂN TUYẾN TÍNH.....	7
2.3. LÝ THUYẾT ĐIỀU KHIỂN PHI TUYẾN.....	8
2.4. MỘT SỐ THUẬT TOÁN TỐI ƯU HÓA TRONG ĐIỀU KHIỂN.....	8
2.5. TỔNG QUAN VỀ NGUYÊN LÝ THUẬT TOÁN ADP.....	11
CHƯƠNG 3: PHƯƠNG PHÁP ĐỀ XUẤT THIẾT KẾ BỘ ĐIỀU KHIỂN TỐI ƯU CHO HỆ THỐNG PHI TUYẾN CÓ NHIỀU ĐẦU VÀO	14
3.1. MÔ HÌNH HỆ THỐNG:.....	14
3.2. THIẾT KẾ BỘ QUAN SÁT NHIỀU ĐẦU VÀO CHO HỆ THỐNG:	16
3.3. NHẬN DẠNG MÔ HÌNH HỆ THỐNG:.....	19
3.4. THIẾT KẾ BỘ ĐIỀU KHIỂN TỐI ƯU CHO HỆ THỐNG:.....	24
CHƯƠNG 4: MÔ PHỎNG THUẬT TOÁN ĐỀ XUẤT VỚI QUÁ TRÌNH HOÁ HỌC TRONG THIẾT BỊ CSTR.....	30
4.1. MÔ HÌNH TOÁN HỌC VÀ THAM SỐ CỦA ĐỐI TƯỢNG.....	30
4.2. THIẾT KẾ BỘ QUAN SÁT NHIỀU CHO HỆ THỐNG	32
4.3. NHẬN DẠNG MÔ HÌNH CHO HỆ THỐNG	34
4.4. BỘ ĐIỀU KHIỂN TỐI ƯU CHO HỆ THỐNG	37
CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN ĐỀ TÀI.....	43
TÀI LIỆU THAM KHẢO	44

DANH MỤC HÌNH VẼ

Hình 1. Sơ đồ cấu trúc tổng quát của toàn bộ thuật toán	29
Hình 2. Đồ thị nhiễu đầu vào mô phỏng	32
Hình 3. So sánh nhiễu đầu ra bộ quan sát với bậc quan sát khác nhau	33
Hình 4. Sai lệch của nhiễu xấp xỉ so với thực tế	34
Hình 5. Sự hội tụ của ma trận W_f	35
Hình 6. Sự hội tụ của ma trận V_f	36
Hình 7. Ước lượng mô hình trạng thái so với thực tế	36
Hình 8. Sai lệch khi nhận dạng mô hình	37
Hình 9. Sự hội tụ của trạng thái mô hình	39
Hình 10. Sự hội tụ của trọng số thành phần critic	39
Hình 11. Sự hội tụ của trọng số thành phần actor	40
Hình 12. Tín hiệu điều khiển thực tế của thuật toán	40
Hình 13. Đồ thị nồng độ của sinh khối	41
Hình 14. Đồ thị nồng độ của cơ chất	41
Hình 15. Đồ thị hàm mục tiêu theo thời gian trước và sau cập nhật	42

DANH MỤC BẢNG BIỂU

Bảng 1. Các thông số của đối tượng điều khiển	30
Bảng 2. Các tham số của bộ quan sát nhiễu	32

DANH MỤC TỪ VIẾT TẮT

Từ viết tắt	Thuật ngữ tiếng anh	Giải thích
ADP	Adaptive Dynamic Programming	Quy hoạch động thích nghi, một phương pháp học củng cố để xấp xỉ luật điều khiển tối ưu online (offline)
CSTR	Continous Stirred Tank Reactor	Một thiết bị chuyên dùng trong công nghệ hóa học, thường dùng để trộn lẫn các chất hoặc thực hiện các phản ứng hóa học
HJB	Hamilton – Jacobi – Bellman	Phương trình toán học kinh điển liên quan đến bài toán quy hoạch động trong điều khiển
PE	Persistence of Excitation	Kích thích hệ thống bằng cách thêm nhiễu vào vector tín hiệu vào ra. PE là điều kiện để tham số hội tụ trong nhận dạng và điều khiển thích nghi
UUB	Uniform Ultimate Bounded	Bị chặn tới hạn đều. Thể hiện rằng một tín hiệu sẽ luôn nằm trong một miền xác định khi thời gian đủ lớn

CHƯƠNG 1. GIỚI THIỆU

1.1. TỔNG QUAN VỀ LÝ THUYẾT ĐIỀU KHIỂN

Điều khiển học là một trong những ngành khoa học cơ bản và đóng vai trò quan trọng trong cuộc sống chúng ta. Khi nói về đối tượng điều khiển, chúng ta có rất nhiều cách để chia, tuy nhiên có hai cách chia phổ biến: một là hệ liên tục và hệ rời rạc, hai là hệ tuyến tính và hệ phi tuyến. Đối với hệ tuyến tính, các phương pháp điều khiển cổ điển tỏ ra mạnh mẽ. Đó là phương pháp PID, gain điểm cực, cân bằng tối thiểu,... Những thông số quan trọng đặc biệt ảnh hưởng tới chất lượng hệ thống, đó là sai lệch tĩnh, độ vọt, độ lợi, thời gian quá độ,...

Những đánh giá về các con số này được thể hiện rất rõ trong cuốn điều khiển tuyến tính của thầy Nguyễn Doãn Phước. Đối với hệ phi tuyến, các thông số nói đến ở trên khó đánh giá hơn và chúng ta thường tập trung vào tính ổn định của hệ thống. Chính vì thế mà xuất hiện nhiều phương pháp điều khiển hiện đại và thông minh cho những hệ này. Tuy nhiên, ta có thể thấy có những mảng điều khiển chính như: điều khiển mờ và neural, điều khiển thích nghi, điều khiển bền vững, và điều khiển tối ưu.

1.2. TỔNG QUAN VỀ MÔ HÌNH BÌNH PHẢN ỨNG KHUẤY TRỘN LIÊN TỤC

Bình phản ứng khuấy trộn liên tục (CSTR) là một thiết bị không thể thiếu trong các quá trình công nghệ kỹ thuật và hóa học. Thiết bị khuấy trộn liên tục (CSTR) được biết đến như một thùng chứa lớn hoặc một lò phản ứng, là một dạng thiết bị phản ứng phổ biến trong kỹ thuật hóa học. Một thiết bị CSTR thường dùng để nói đến một mô hình được sử dụng để đánh giá sự thay đổi của các thành phần hợp chất trong quá trình phản ứng, với việc sử dụng một thùng (bể) chứa có thiết bị khuấy hoạt động liên tục, nhằm cho sản phẩm đầu ra theo yêu cầu. Mô hình này làm việc với hầu hết các lưu chất như : chất lỏng, khí đốt, bùn than, xỉ măng...

Trong một lò phản ứng khuấy trộn hoàn hảo, hợp chất ở đầu ra là đồng nhất được cấu thành từ các hợp chất hóa học ở đầu vào, hoạt động với thời gian và tốc độ phản ứng ổn định. Nếu thời gian trộn ổn định là từ 5 đến 10 chu kỳ thì ta coi như đã đạt được yêu cầu kỹ thuật. Các thiết bị CSTR khi được sử dụng thường được đơn giản hóa các công thức tính toán kỹ thuật và có thể được sử dụng để mô tả các nghiên cứu về phản ứng.

Trong thực tế, các quá trình công nghiệp thường được điều khiển bằng phương pháp điều khiển tuyến tính như PID, LQR. Mô hình phi tuyến thường được tuyến tính hóa quanh điểm điểm cân bằng.

Trong thiết bị CSTR trong công nghiệp thường điều khiển 2 đại lượng chính là nhiệt độ và nồng độ. 2 đại lượng này đều thay đổi chậm, quán tính lớn. Ngoài ra, khi điều khiển nhiệt độ hoặc nồng độ thay đổi một lượng lớn thì mô hình sẽ không còn là tuyến tính. Vì vậy nên đề án của em đề xuất một phương pháp khác điều khiển thiết bị CSTR bằng mô hình phi tuyến ban đầu của hệ thống

Trong đề án này xét một phản ứng lên men đơn giản trong công nghiệp. Gồm 2 biến đầu ra và 1 biến đầu vào. Điều khiển lưu lượng dòng chảy của chất phản ứng để nồng độ sản phẩm đạt giá trị yêu cầu

CHƯƠNG 2: KIẾN THỨC SỬ DỤNG TRONG ĐỒ ÁN

2.1. MỘT SỐ BẤT ĐẲNG THỨC TOÁN HỌC CẦN THIẾT

- Bất đẳng thức Young: Cho 2 số thực a và b không âm, 2 số thực p và q thỏa mãn $\frac{1}{p} + \frac{1}{q} = 1$. Khi đó ta có

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad (2.1)$$

Áp dụng đối với 2 vecto cột A, B trong trường hợp $p=q=2$. Ta có

$$A^T B \leq \frac{1}{2} \|A\|^2 + \frac{1}{2} \|B\|^2 \quad (2.2)$$

Trong đó $\|A\|$ được định nghĩa là chuẩn Euclid của ma trận A

- Bất đẳng thức Cauchy Schwarz: Cho 2 ma trận A và B , tích AB tồn tại. Khi đó ta có

$$\|AB\| \leq \|A\| \|B\| \quad (2.3)$$

Các bất đẳng thức toán học về chuẩn của ma trận thường được sử dụng để chứng minh tính ổn định của thuật toán hoặc ma trận

2.2. LÝ THUYẾT ĐIỀU KHIỂN TUYẾN TÍNH

- Tính ổn định của hệ thống tuyến tính

Cho hệ thống tuyến tính có dạng

$$\dot{x} = Ax + Bw \quad (2.4)$$

Trong đó, A là ma trận vuông, w là tín hiệu không xác định, nhưng luôn bị chặn trên.

Khi đó, hệ thống (2.4) ổn định khi và chỉ khi ma trận A là ma trận bền, có các trị riêng đều nằm trên nửa mặt phẳng bên trái trục ảo. Ổn định này không phải là ổn định tiệm cận, nhưng trạng thái luôn nằm trong một miền xác định D lân cận gốc tọa độ.

Điều kiện này được sử dụng để lựa chọn tham số thích hợp cho quá trình. Và ma trận A có thể được chọn một cách phù hợp

2.3. LÝ THUYẾT ĐIỀU KHIỂN PHI TUYẾN

Một số định nghĩa quan trọng trong lý thuyết điều khiển cần dùng trong đồ án

- Khái niệm ổn định tiệm cận

Cho hệ thống phi tuyến tổng quát

$$\dot{x} = f(x, t) \quad \text{với } f(0, t) = 0, \quad \forall t \geq 0 \quad (2.5)$$

Hệ (2.5) được gọi là ổn định tiệm cận toàn cục nếu tồn tại hàm $\beta \in KL$ mà với mọi tín hiệu khởi tạo x_0 và tồn tại mốc thời gian t_0 thỏa mãn

$$|x(t)| \leq \beta(|x_0|, t - t_0), \quad \forall t \geq t_0 \quad \text{và } x(t_0) = x_0 \in R^n \quad (2.6)$$

Hơn nữa, hệ được gọi là ổn định tiệm cận theo hàm mũ nếu hệ ổn định tiệm cận và thêm điều kiện $\beta(|x_0|, t - t_0) = a|x_0|e^{-b(t-t_0)}$ với a và b là 2 hằng số dương

- Khái niệm ổn định UUB

Cho hệ thống phi tuyến (2.5) với trạng thái $x(t) \in R^n$. Điểm cân bằng x_c được gọi là điểm ổn định UUB nếu tồn tại một tập đóng $\Omega_x \subset R^n$, sao cho với mọi $x \in \Omega_x$, luôn tồn tại chặn trên B và thời gian $T_B(B, x)$ để điều kiện $\|x - x_c\| \leq B$ luôn thỏa mãn với mọi $t \geq t_0 + T_B$

- Điều kiện PE

Một vector tín hiệu bị chặn $\delta(t)$ được gọi là thỏa mãn điều kiện PE trong khoảng thời gian $[t, t + T_p]$ với $T_p > 0$ nếu tồn tại 2 số dương β_1 và β_2 sao cho với mọi giá trị t có

$$\beta_1 I \leq \int_t^{t+T_p} \delta(\tau) \delta(\tau)^T d\tau \leq \beta_2 I \quad (2.7)$$

Các định nghĩa trên được sử dụng để chứng minh các biến, tín hiệu trong thuật toán hội tụ về giá trị xác định. Điều kiện PE là điều kiện cần để thuật toán có thể cập nhật trọng số về giá trị tối ưu

2.4. MỘT SỐ THUẬT TOÁN TỐI ƯU HÓA TRONG ĐIỀU KHIỂN

Do bài toán là điều khiển tối ưu, nên chắc chắn cần dùng đến các thuật toán tối ưu hóa. Hai thuật toán tối ưu hóa được sử dụng trong đồ án là: bình phương tối thiểu, gradient descent.

- Phương pháp bình phương tối thiểu

Cho 2 bộ ma trận đầu vào X và đầu ra Y , ta cần tìm một hàm số $f(X)$ sao cho

$$Y = f(X) + \varepsilon \quad (2.8)$$

Chuyển hàm số $f(X)$ về dạng $W^T \sigma(X)$. Trong đó $\sigma(X)$ là một hàm cơ sở, W là một trọng số thực. Khi đó (2.8) trở thành

$$Y = \sigma(X)W + \varepsilon \quad (2.9)$$

Ta đánh giá nghiệm của phương pháp bình phương tối thiểu là trọng số W^* sao cho bình phương sai lệch ε là nhỏ nhất

$$W^* = \arg \min_W (\varepsilon^T \varepsilon) = \arg \min_{W^*} \left((Y - \sigma(X)W)^T (Y - \sigma(X)W) \right) \quad (2.10)$$

Phương trình (2.10) có nghiệm là

$$W^* = (K^T K)^{-1} K^T Y \quad \text{với } K = \sigma(X) \quad (2.11)$$

Nghiệm (2.11) chỉ phù hợp trong trường hợp bộ dữ liệu không thay đổi, và ma trận $K^T K$ phải là ma trận không suy biến.

Nếu kích thước ma trận K luôn tăng lên theo thời gian, thì (2.11) không có là phương pháp phù hợp do W^* cũng thay đổi theo thời gian. Vậy nên ta cần biến đổi theo một phương pháp khác với K , Y đều phụ thuộc theo thời gian

$$W^*(t) = (K(t)^T K(t))^{-1} K(t)^T Y(t) \quad (2.12)$$

Đặt $P(t) = (K(t)^T K(t))^{-1}$. Ta có

$$W^*(t) = P(t) \sum_{i=1}^t k(i) y(i) = P(t) \left[\sum_{i=1}^{t-1} k(i) y(i) + k(t) y(t) \right] \quad (2.13)$$

Mặt khác, ta lại có

$$\begin{aligned} W^*(t-1) &= P(t-1) \sum_{i=1}^{t-1} k(i) y(i) \\ \sum_{i=1}^{t-1} k(i) y(i) &= P^{-1}(t-1) W^*(t-1) \end{aligned} \quad (2.14)$$

Quan hệ giữa $P(t)$ và $P(t-1)$ có thể biểu diễn như sau

$$\begin{aligned}
P^{-1}(t) &= K(t)^T K(t) = \sum_{i=1}^t k(i)k^T(i) = P^{-1}(t-1) + k(t)k^T(t) \\
P^{-1}(t-1) &= P^{-1}(t) - k(t)k^T(t)
\end{aligned} \tag{2.15}$$

Thay (2.15) vào (2.14), ta rút ra được

$$\sum_{i=1}^{t-1} k(i)y(i) = [P^{-1}(t) - k(t)k^T(t)]W^*(t-1) \tag{2.16}$$

Từ (2.16) và (2.13), luật cập nhật $W^*(t)$ theo $W^*(t-1)$ là

$$\begin{aligned}
W^*(t) &= P(t) \{ [P^{-1}(t) - k(t)k^T(t)]W^*(t-1) + k(t)y(t) \} \\
W^*(t) &= W^*(t-1) + P(t)k(t) [y(t) - k^T(t)W^*(t-1)]
\end{aligned} \tag{2.17}$$

Theo định lý về ma trận nghịch đảo.

$$(A^{-1} + BCD)^{-1} = A - AB(C^{-1} + DAB)^{-1}DA \tag{2.18}$$

Mặt khác từ (2.15), ta lại có

$$P(t) = [P(t-1) + k(t)k^T(t)]^{-1} \tag{2.19}$$

Áp dụng (2.18) với $A = P^{-1}(t)$, $B = k(t)$, $C = I$, $D = k^T(t)$, (2.19) tương đương với

$$P(t) = P(t-1) - P(t-1)k(t)(I + k^T(t)P(t-1)k^T(t))^{-1}k^T(t)P(t-1) \tag{2.20}$$

$$\begin{aligned}
P(t)k(t) &= P(t-1)k(t) - P(t-1)k(t)(I + k^T(t)P(t-1)k^T(t))^{-1}k^T(t)P(t-1)k(t) \\
P(t)k(t) &= P(t-1)k(t) \left[I - (I + k^T(t)P(t-1)k^T(t))^{-1}k^T(t)P(t-1)k(t) \right] \\
P(t)k(t) &= P(t-1)k(t)(I + k^T(t)P(t-1)k^T(t))^{-1} [I + k^T(t)P(t-1)k^T(t) - k^T(t)P(t-1)k(t)]
\end{aligned} \tag{2.21}$$

Thay (2.21) vào (2.17), ta có luật cập nhật trọng số W^* là

$$W^*(t) = W^*(t-1) + P(t-1)k(t)(I + k^T(t)P(t-1)k^T(t))^{-1} [y(t) - k^T(t)W^*(t-1)] \tag{2.22}$$

Thuật toán bình phương tối thiểu gồm 2 công thức (2.22) và (2.20) sử dụng đệ quy, phù hợp để tối ưu các mô hình có kích thước thay đổi chỉ bằng cách cập nhật trọng số. Ngoài

ra vấn đề ma trận suy biến cũng được giải quyết do đã loại trừ được phép tính ma trận nghịch đảo trong (2.10).

- Phương pháp gradient descent

Phương pháp này tối ưu một hàm cho trước bằng cách cập nhật trọng số tương tự như phương pháp bình phương tối thiểu. Tuy nhiên, phương pháp này có thể áp dụng được cho cả các hàm phụ thuộc phi tuyến theo trọng số W , trong khi phương pháp bình phương tối thiểu chỉ có thể áp dụng cho các hàm phụ thuộc tuyến tính theo W .

Cho hàm $E(X, W)$. Điểm tối ưu W^* thỏa mãn

$$\frac{\partial E}{\partial W^*} = 0 \quad (2.23)$$

Phương pháp cập nhật trọng số cho W để W tiến dần về giá trị tối ưu

$$\dot{W} = -\eta \frac{\partial E}{\partial W} \quad (2.24)$$

Trong đó, η là một tham số gọi là tốc độ học của phương pháp. Nếu η quá nhỏ sẽ dẫn đến quá trình hội tụ của phương pháp xảy ra rất chậm chạp. Nếu η quá lớn có thể gây mất ổn định, làm W đi xa khỏi W^* .

2.5. TỔNG QUAN VỀ NGUYÊN LÝ THUẬT TOÁN ADP

Để thuận tiện trình bày, ta xét hệ động học có mô hình trạng thái liên tục, dạng affine

$$\dot{x} = f(x) + g(x)u \quad (2.25)$$

Với $x \in R^n$ là biến trạng thái và tín hiệu điều khiển $u \in R^m$.

Luật điều khiển là một hàm $\mu(.): R^n \rightarrow R^m$ có dạng

$$u = \mu(x) \quad (2.26)$$

Hàm mục tiêu thể hiện sự tối ưu cần đạt được của hệ thống là một hàm tích phân liên tục

$$J(t) = \int_t^\infty e^{-\gamma(\tau-t)} r(x, u) d\tau \quad (2.27)$$

Trong đó $\gamma \geq 0$ là hệ số suy giảm, thể hiện ảnh hưởng của hàm chi phí $r(x, u)$ khi thời gian càng dài

Giả sử hệ thống có thể ổn định trong tập xác định $\Omega \in R^n$ thì tồn tại luật điều khiển $u = \mu(x)$ giúp hệ điều khiển vòng kín (2.25) ổn định tiệm cận trong Ω . Tín hiệu điều khiển $u = \mu(x)$ được gọi là tín hiệu chấp nhận được nếu nó là tín hiệu giúp hệ ổn định. Ngoài ra, hàm $J(x)$ do nó gây ra phải là hàm hữu hạn

Mục tiêu của điều khiển tối ưu là phải tìm ra được một hàm $u = \mu^*(x)$ là tín hiệu chấp nhận được, để có thể tối ưu được hàm mục tiêu $J(x)$

$$u^*(x(t)) = \arg \min_{u \in \Omega} \int_t^{\infty} e^{-\gamma(\tau-t)} r(x, u) d\tau \quad (2.28)$$

Đạo hàm hàm chi phí (2.27) theo thời gian, ta có

$$\dot{J}(t) = \int_t^{\infty} \frac{\partial}{\partial t} e^{-\gamma(\tau-t)} r(x, u) d\tau - r(x, u) = \gamma J(t) - r(x, u) \quad (2.29)$$

Hàm Hamilton liên tục được định nghĩa như sau

$$H(x, u, J) = r(x, u) - \gamma J(x) + \frac{\partial J}{\partial x} (f(x) + g(x)u) \quad (2.30)$$

Hàm Hamilton tính toán năng lượng dọc theo diễn biến của hệ thống cũng như phản chiếu sự tối ưu của hệ thống. Phương trình HJB đòi hỏi hàm Hamilton bằng 0

Bản chất của các thuật toán ADP, là thực hiện giải các phương trình HJB phi tuyến nhưng không dựa trên mô hình đối tượng. Đối với các mô hình hệ thống có dạng affine (2.25) thì thuật toán ADP không cần biết bất cứ thông tin gì về hàm $f(x)$. Thuật toán chỉ sử dụng dữ liệu đo được đầu vào và đầu ra của hệ thống, phụ thuộc vào ma trận $g(x)$.

Thuật toán ADP cơ bản chia thành 2 loại cơ bản.

- ADP off policy. Thuật toán thực hiện bằng cách thu một loạt dữ liệu từ việc chạy thử hệ thống. Sau đó tìm ra bộ trọng số và bộ điều khiển tối ưu cho hệ thống. Mô hình dựa trên nhiều dữ liệu trên khoảng thời gian T.

Do số dữ liệu là không thay đổi. Nên việc cập nhật dữ liệu thường được tiến hành offline bằng công thức (2.11). Chúng cập nhật dựa trên chênh lệch giữa đầu vào và đầu ra tại các thời điểm liên tiếp.

Ưu điểm của phương pháp này là thuật toán đơn giản, thu được kết quả ngay lúc ổn định. Nhưng không phù hợp với các mô hình có bất định do mô hình thay đổi liên tục

- ADP on policy. Thuật toán thực hiện bằng cách cập nhật liên tục các ma trận trọng số và bộ điều khiển một cách online

Luật cập nhật phụ thuộc vào đầu vào đầu ra của bước trước đó. Quá trình học online được thực hiện một cách liên tục, nên phù hợp với các mô hình có bất định hoặc tham số thay đổi theo thời gian.

Mô hình khảo sát ở đây là chịu tác dụng của bất định và nhiễu. Nên trong đề án này sẽ sử dụng phương pháp ADP on policy để xác định bộ điều khiển tối ưu.

CHƯƠNG 3: PHƯƠNG PHÁP ĐỀ XUẤT THIẾT KẾ BỘ ĐIỀU KHIỂN TỐI ƯU CHO HỆ THỐNG PHI TUYẾN CÓ NHIỀU ĐẦU VÀO

3.1. MÔ HÌNH HỆ THỐNG:

Xét hệ thống phi tuyến liên tục theo thời gian:

$$\dot{x} = f(x) + g(x)(u + d) \quad (3.1)$$

Đây là mô hình phi tuyến affine, có nhiều đầu vào d không đo được. Trong đó $f(x) \in R^n$, $g(x) \in R^{n \times m}$, $x(t) \in R^n$, $u(t) \in R^m$ là đầu vào bị chặn, $|u_i| \leq \lambda$. $F(x, u) = f(x) + g(x)(u + d)$ là hàm Lipschitz liên tục. $f(x)=0$ khi $x=0$. Nghĩa là hệ phi tuyến này có điểm cân bằng tại $x=0$.

Trong mô hình hệ thống, ma trận $f(x)$ là không biết trước do bất định, $g(x)$ là xác định

Ta khai báo hàm mục tiêu điều khiển cho hệ này:

$$J(x(t)) = \int_t^{\infty} r(x(\tau), u(\tau)) d\tau \quad (3.2)$$

Trong đó:

$$r(x(t), u(t)) = x^T Q x + U(u(t)) \quad (3.3)$$

$$U(u) = 2 \int_0^u \lambda \Phi^{-1}\left(\frac{v}{\lambda}\right) R dv \quad (3.4)$$

Các ma trận Q và R là ma trận đối xứng xác định dương. Để có thể đánh giá hàm mục tiêu với đầu vào bị chặn, ta cần chọn hàm $U(u(t))$ sao cho giá trị của nó là một hàm xác định dương, tập giá trị trong khoảng $[0, \infty)$.

Do đó để đơn giản trong việc tính toán, biến đổi. Chọn hàm Φ^{-1} sao cho nó là một ánh xạ từ $(-\lambda, \lambda)$ đến R

Với hàm $\Phi^{-1}\left(\frac{v}{\lambda}\right) = \tanh^{-1}\left(\frac{v}{\lambda}\right)$ thỏa mãn điều kiện trên. Khi đó, (3.4) trở thành:

$$U(u) = 2 \int_0^u \lambda \tanh^{-1}\left(\frac{v}{\lambda}\right) R dv \quad (3.5)$$

Dễ thấy với hàm r được chọn như trên, hàm mục tiêu J là luôn dương. $r=0$ khi và chỉ khi $x=0, u=0$.

Trước khi thiết kế điều khiển cho hệ phi tuyến trên, ta cần đưa ra các giả thiết cần thiết.

Giả thiết 1: Nhiễu đầu vào d tác dụng lên hệ là nhiễu bị chặn, và luôn nhỏ hơn giá trị cực đại λ của đầu vào u .

$$\|d\|_{\infty} \leq \bar{d} < \lambda \quad (3.6)$$

Giả thiết này nhằm giới hạn nhiễu d là không quá lớn so với đầu vào u để thiết kế bù nhiễu trong bộ điều khiển.

Giả thiết 2: Các ma trận $f(x)$ và $g(x)$ là khả vi đến cấp 2

Giả thiết 3: Ma trận $g(x)$ bị chặn và biết trước

$$0 \leq \|g(x)\|_{\infty} \leq \bar{g} \quad (3.7)$$

Giả thiết 4: Cho hàm liên tục $Y: S \rightarrow R^n$. S là một tập đóng. Luôn tồn tại 2 bộ số W và V thỏa mãn.

$$Y(x) = W^T \sigma(V^T x) + \varepsilon(x) \quad (3.8)$$

Trong đó, σ là các hàm cơ sở của mạng nơ-ron, độc lập tuyến tính với nhau. Ngoài ra, khi số nút N mạng tăng lên càng lớn, sai lệch do xấp xỉ mạng nơ-ron càng nhỏ, và $\varepsilon \rightarrow 0$ khi $N \rightarrow \infty$

Giả thiết 5: Giả thiết các ma trận W, V trong (3.8) là ma trận có chuẩn Euclid bị chặn trên. Tức là

$$\|W\| \leq \bar{W}, \|V\| \leq \bar{V} \quad (3.9)$$

Giả thiết 6: Sai lệch $\varepsilon(x)$ và đạo hàm $\varepsilon'(x)$ đều bị chặn trên.

$$\|\varepsilon(x)\| \leq \bar{\varepsilon}, \|\varepsilon'(x)\| \leq \bar{\varepsilon}' \quad (3.10)$$

Để điều khiển hệ phi tuyến trên, ta chia bài toán thành 3 mục tiêu chính:

- Nhận dạng xấp xỉ mô hình hệ thống
- Ước lượng nhiễu d không đo được bằng giá trị \hat{d} để bù nhiễu trong bộ điều khiển
- Xác định bộ điều khiển u khi không có nhiễu đầu vào để tối ưu hàm mục tiêu (3.2)

Do hàm $J(x(t))$ trong (3.2) luôn dương, $r(x(t), u(t))$ dễ thấy là một hàm lồi, nên bài toán trở thành bài toán tối ưu lồi dưới tác động của bất định và nhiễu

3.2. THIẾT KẾ BỘ QUAN SÁT NHIỀU ĐẦU VÀO CHO HỆ THỐNG:

Ta đặt $d^* = g(x)d$. Khi đó (3.1) trở thành:

$$\dot{x} = f(x) + g(x)u + d^* \quad (3.11)$$

Bộ quan sát được thiết kế ở đây là quan sát nhiễu bậc cao. Được thiết kế để quan sát nhiễu cho đến đạo hàm cấp n của d .

Trước khi thiết kế, ta cần đưa ra giả thiết cho d^* để đảm bảo bộ quan sát làm hệ ổn định:

Giả thiết 8: d^* là hàm số khả vi đến cấp n theo t , và các đạo hàm này là bị chặn trên:

$$\left\| \frac{d^{(n)}(d^*)}{dt^n} \right\| \leq \mu \quad (3.12)$$

Giả sử $f(x)$ là xác định. Bộ quan sát nhiễu bậc n cho hệ (3.11) được thiết kế như sau:

$$\hat{d}_{(i-1)}^* = p_i + L_i x \quad (3.13)$$

$$\dot{p}_i = -L_i \left(f(x) + \hat{d}_{(i)}^* + g(x)u \right) + \hat{d}_{(i)}^* \quad i=1,2,\dots,n-1 \quad (3.14)$$

$$\hat{d}_{(n-1)}^* = p_n + L_n x \quad (3.15)$$

$$\dot{p}_n = -L_n \left(f(x) + \hat{d}_{(n)}^* + g(x)u \right) \quad (3.16)$$

Trong đó, L_i là các ma trận vuông có số chiều thích hợp. Các thành phần $\hat{d}_{(i-1)}^*$ là giá trị ước lượng của $\frac{d^{(i-1)}(d^*)}{dt^{i-1}}$.

Khi $i=1$, ta có $\hat{d}_{(0)}^* = d^*$. Các biến p_i là các biến trung gian dùng để ước lượng nhiễu.

Tiếp theo, ta đánh giá sai lệch của bộ quan sát nhiễu so với nhiễu thực tế.

Định nghĩa sai lệch quan sát là:

$$e_{d^*} = \left[\tilde{d}_{(0)}^*, \tilde{d}_{(1)}^*, \dots, \tilde{d}_{(n-1)}^* \right]^T \quad (3.17)$$

Trong đó:

$$\tilde{d}_{(i-1)}^* = \frac{d^{(i-1)}(d^*)}{dt^{i-1}} - \hat{d}_{(i-1)}^* \quad i=1,2,\dots,n \quad (3.18)$$

Đạo hàm 2 vế (3.18) theo thời gian ta có:

$$\dot{\tilde{d}}_{(i-1)}^* = \frac{d^{(i)}(d^*)}{dt^i} - \dot{\hat{d}}_{(i-1)}^* \quad (3.19)$$

Từ (3.11), (3.13), (3.14). Ta có:

$$\begin{aligned} \dot{\tilde{d}}_{(i-1)}^* &= \dot{p}_i + L_i \dot{x} \\ &= -L_i(f(x) + \hat{d}^* + g(x)u) + \dot{\hat{d}}_{(i)}^* + L_i(f(x) + g(x)u + d^*) \\ &= L_i(d^* - \hat{d}^*) + \dot{\hat{d}}_{(i)}^* \\ &= L_i \tilde{d}^* + \dot{\hat{d}}_{(i)}^* \end{aligned} \quad (3.20)$$

Đặt $\frac{d^{(i-1)}(d^*)}{dt^{i-1}} = d_{(i-1)}^*$. Ta có:

$$\dot{d}_{(i-1)}^* = d_{(i)}^* \quad (3.21)$$

Trừ theo vế (3.21) cho (3.20)

$$\dot{\tilde{d}}_{(i-1)}^* = -L_i \tilde{d}^* + \dot{\hat{d}}_{(i)}^* \quad (3.22)$$

Từ (3.11), (3.15), (3.16). Ta có:

$$\begin{aligned} \dot{\hat{d}}_{(n-1)}^* &= \dot{p}_n + L_n \dot{x} \\ &= -L_n(f(x) + \hat{d}^* + g(x)u) + L_n(f(x) + g(x)u + d^*) \\ &= L_n(d^* - \hat{d}^*) \\ &= L_n \tilde{d}^* \end{aligned} \quad (3.23)$$

Trừ cả 2 vế (3.23) cho $\dot{d}_{(n)}^*$ ta có:

$$\dot{\tilde{d}}_{(n-1)}^* = -L_n \tilde{d}^* + \dot{d}_{(n)}^* \quad (3.24)$$

Kết hợp các công thức (3.22) và (3.24) dưới dạng ma trận:

$$\begin{bmatrix} \dot{\tilde{d}}_{(0)}^* \\ \dot{\tilde{d}}_{(1)}^* \\ \vdots \\ \dot{\tilde{d}}_{(n-1)}^* \end{bmatrix} = \begin{bmatrix} -L_1 & I & 0 & \cdots & 0 \\ -L_2 & 0 & I & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -L_{n-1} & 0 & 0 & \ddots & I \\ -L_n & 0 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} \tilde{d}_{(0)}^* \\ \tilde{d}_{(1)}^* \\ \vdots \\ \tilde{d}_{(n-1)}^* \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ I \end{bmatrix} d_{(n)}^* \quad (3.25)$$

I là các ma trận đơn vị có số chiều phù hợp.

Ta thấy hệ (3.25) có dạng mô hình tuyến tính viết dạng thu gọn như sau:

$$\dot{e}_{d^*} = D_n e_{d^*} + E w \quad (3.26)$$

Điều kiện để (3.26) ổn định là ma trận D_n ở trên phải là ma trận bền. Các điểm cực của D_n nằm bên trái trục ảo. Theo định lý Lyapunov với D_n là ma trận bền, tồn tại một ma trận P_1 đối xứng xác định dương thỏa mãn

$$D_n^T P_1 + P_1 D_n = -Q_1 \quad (3.27)$$

Q_1 là ma trận xác định dương.

Để đánh giá tính ổn định của bộ quan sát trên ta có định lý sau:

Định lý 1: Cho mô hình phi tuyến (3.11) chịu tác động của nhiễu d^* , có $f(x)$ và $g(x)$ xác định. Sử dụng bộ quan sát bậc cao trong các công thức (3.13) – (3.16), thì sai lệch giữa nhiễu đầu vào d^* và nhiễu quan sát được \hat{d}^* trong (3.17) sẽ là bị chặn sau thời gian đủ lớn (UUB)

Chứng minh:

Xét hàm Lyapunov sau:

$$V_1 = e_{d^*}^T P_1 e_{d^*} \quad (3.28)$$

P_1 là nghiệm của phương trình (3.27).

Đạo hàm 2 vế (3.28) theo thời gian, ta có:

$$\begin{aligned}
\dot{V}_1 &= \dot{e}_{d^*}^T P_1 e_{d^*} + e_{d^*}^T P_1 \dot{e}_{d^*} \\
&= (w^T E^T + e_{d^*}^T D_n^T) P_1 e_{d^*} + e_{d^*}^T P_1 (D_n e_{d^*} + Ew) \\
&= e_{d^*}^T (D_n^T P_1 + P_1 D_n) e_{d^*} + 2e_{d^*}^T P_1 Ew \\
&= -e_{d^*}^T Q_1 e_{d^*} + 2e_{d^*}^T P_1 Ew \\
&\leq -e_{d^*}^T Q_1 e_{d^*} + 2\|P_1 E\| \|e_{d^*}\| \|w\| \\
&\leq -\lambda_m \|e_{d^*}\|^2 + 2\|P_1 E\| \|e_{d^*}\| \|w\| \\
&= -\|e_{d^*}\| (\lambda_m \|e_{d^*}\| - 2\|P_1 E\| \|w\|)
\end{aligned} \tag{3.29}$$

Ta có $\dot{V}_1 \leq 0$ khi điều kiện sau thỏa mãn:

$$\|e_{d^*}\| \geq \frac{2\|P_1 E\| \|w\|}{\lambda_m} \tag{3.30}$$

Từ **Giả thiết 8**, sau khoảng thời gian đủ dài, sai lệch quan sát sẽ bị chặn trong miền xác định

$$\|e_{d^*}\| \leq \frac{2\|P_1 E\| \|w\|}{\lambda_m} \leq \frac{2\|P_1 E\| \mu}{\lambda_m} \tag{3.31}$$

Từ (3.31), ta chứng minh được e_{d^*} là UUB. \square

$$\|e_{d^*}\| = \sqrt{e_{d^*}^T e_{d^*}} = \sum_{i=0}^{n-1} [\tilde{d}_{(i)}^*]^T [\tilde{d}_{(i)}^*] \tag{3.32}$$

Nhận xét 1: Với cùng một giá trị của $\|e_{d^*}\|$ trong công thức (3.32), ta thấy khi bậc n của bộ quan sát càng lớn, giá trị trung bình của từng phần tử $[\tilde{d}_{(i)}^*]^T [\tilde{d}_{(i)}^*]$ sẽ càng nhỏ. Khi đó sai lệch $\tilde{d}_{(0)}^* = d^* - \hat{d}^*$ càng nhỏ, chất lượng của bộ quan sát càng được cải thiện. Tuy nhiên, độ phức tạp tính toán cũng tăng lên. Vì thế, cần chọn bậc phù hợp cho bộ quan sát để vừa đảm bảo chất lượng quan sát, vừa đơn giản cho hệ thống tính toán.

3.3. NHẬN DẠNG MÔ HÌNH HỆ THỐNG:

Mô hình hệ thống (3.1) được xét là bất định toàn phần do $f(x)$ là không biết trước nên ta cần xấp xỉ hàm $f(x)$ thành hàm $\hat{f}(x)$, nhằm dùng để thiết kế bộ điều khiển sau này.

Do hàm $f(x)$ là hàm khả vi đến cấp 2, từ **Giả thiết 4** ta có thể biểu diễn hàm $f(x)$ như sau

$$f(x) = W_f^T \sigma(V_f^T x) + \varepsilon_f(x) \quad (3.33)$$

$$\dot{x} = W_f^T \sigma(V_f^T x) + \varepsilon_f(x) + g(x)(u + d) \quad (3.34)$$

Trong đó các ma trận $W_f \in R^{(L_f+1) \times n}$, $V_f \in R^{n \times L_f}$ là các ma trận trọng số lý tưởng nhưng chưa biết trước, σ là các hàm cơ sở của mạng nơ-ron, độc lập tuyến tính với nhau.

Hàm kích hoạt của mạng nơ-ron là:

$$\sigma_f = \sigma(V_f^T x) \in R^{L_f+1} \quad (3.35)$$

Để nhận dạng gần đúng mô hình hệ thống ta thiết kế như sau:

$$\dot{\hat{x}} = \hat{W}_f^T \hat{\sigma}_f + g(x)u + \hat{d}^* + \psi \quad (3.36)$$

Trong đó \hat{d}^* là đầu ra của bộ quan sát nhiễu

$$\psi = k\tilde{x}(t) - k\tilde{x}(0) + v \quad (3.37)$$

Với $\tilde{x}(t) = x(t) - \hat{x}(t)$ là sai lệch trạng thái, k là hằng số thiết kế.

$$\dot{v} = (k\alpha + \gamma)\tilde{x} + \beta_1 \text{sgn}(\tilde{x}) \quad , v(0) = 0 \quad (3.38)$$

Trừ theo về (3.34) cho (3.36), ta có:

$$\dot{\tilde{x}} = W_f^T \sigma_f - \hat{W}_f^T \hat{\sigma}_f + \varepsilon_f(x) + \tilde{d}^* - \psi \quad (3.39)$$

Để khảo sát tính ổn định của hệ thống, ta đặt:

$$\varepsilon_f = \dot{\tilde{x}} + \alpha \tilde{x} \quad (3.40)$$

Đạo hàm 2 về của (3.40) theo thời gian ta thu được:

$$\begin{aligned} \dot{\varepsilon}_f = \ddot{\tilde{x}} + \alpha \dot{\tilde{x}} = & W_f^T \sigma_f V_f^T \dot{x} - \dot{W}_f^T \hat{\sigma}_f - \hat{W}_f^T \hat{\sigma}_f' \dot{V}_f^T \hat{x} - \hat{W}_f^T \hat{\sigma}_f' V_f^T \dot{\hat{x}} \\ & + \dot{\varepsilon}_f(x) + \dot{\tilde{d}}^* - k\dot{\tilde{x}} - k\alpha \tilde{x} - \gamma \tilde{x} - \beta_1 \text{sgn}(\tilde{x}) + \alpha \dot{\tilde{x}} \end{aligned} \quad (3.41)$$

Luật cập nhật các ma trận trọng số W_f và V_f được cho như sau:

$$\begin{aligned} \dot{W}_f &= \text{proj}(\Gamma_{W_f} \hat{\sigma}_f' V_f^T \dot{\tilde{x}} \tilde{x}^T) \\ \dot{V}_f &= \text{proj}(\Gamma_{V_f} \dot{\tilde{x}} \tilde{x}^T \hat{W}_f^T \hat{\sigma}_f') \end{aligned} \quad (3.42)$$

Toán tử proj ở trên là phép toán nhằm để giới hạn giá trị của $\dot{\hat{W}}_f$ và $\dot{\hat{V}}_f$ để đảm bảo tính ổn định cho toàn hệ sẽ được chứng minh ở sau. Định nghĩa toán tử proj như sau:

$$\hat{\theta} = proj(\sigma_i) = \begin{cases} 0 & \text{nếu } \hat{\theta}_i = \theta_{i \max} \text{ và } \sigma_i > 0 \\ 0 & \text{nếu } \hat{\theta}_i = \theta_{i \min} \text{ và } \sigma_i < 0 \\ \sigma_i & \text{trường hợp khác} \end{cases}$$

Điều này sẽ giúp chặn giá trị của \hat{W}_f và \hat{V}_f trong giới hạn cho phép. Trong quá trình cập nhật \hat{W}_f và \hat{V}_f mà không đạt các giá trị giới hạn trong toán tử proj, có thể coi:

$$\begin{aligned} \dot{\hat{W}}_f &= \Gamma_{w_f} \hat{\sigma}_f' \hat{V}_f^T \hat{x} \hat{x}^T \\ \dot{\hat{V}}_f &= \Gamma_{v_f} \hat{x} \hat{x}^T \hat{W}_f^T \hat{\sigma}_f' \end{aligned} \quad (3.43)$$

Ta đưa ra định lý về tính ổn định của hệ thống nhận dạng và quan sát nhiễu như sau:

Định lý 2: Cho mô hình phi tuyến (3.1), có d là nhiễu đầu vào, d^* là đầu ra của bộ quan sát nhiễu, $f(x)$ là mô hình bất định toàn phần. Sử dụng mạng nơ-ron xấp xỉ mô hình trong các công thức (3.36) – (3.38), với luật cập nhật trọng số trong công thức (3.42). Khi đó, các tín hiệu sai lệch tương ứng là \tilde{x} và \tilde{z} , sẽ bị chặn sau thời gian đủ lớn (UUB), đồng thời các ma trận \hat{W}_f và \hat{V}_f sẽ tiến về 1 lân cận ma trận trọng số tối ưu W_f và V_f

Chứng minh:

Gọi các biến sai lệch trọng số $\tilde{W}_f = W_f - \hat{W}_f$ và $\tilde{V}_f = V_f - \hat{V}_f$

Xét hàm Lyapunov sau:

$$V_2 = \frac{1}{2} e_f^T e_f + \frac{1}{2} \gamma \tilde{x}^T \tilde{x} + \frac{1}{4} \alpha \left[trace(\tilde{W}_f^T \Gamma_{w_f}^{-1} \tilde{W}_f) + trace(\tilde{V}_f^T \Gamma_{v_f}^{-1} \tilde{V}_f) \right] \quad (3.44)$$

Lấy đạo hàm 2 vế của (3.44) theo thời gian ta có:

$$\dot{V}_2 = e_f^T \dot{e}_f + \gamma \tilde{x}^T \dot{\tilde{x}} + \frac{1}{2} \alpha \left[trace(\tilde{W}_f^T \Gamma_{w_f}^{-1} \dot{\tilde{W}}_f) + trace(\tilde{V}_f^T \Gamma_{v_f}^{-1} \dot{\tilde{V}}_f) \right] \quad (3.45)$$

Biểu thức (3.41) có thể viết lại như sau:

$$\begin{aligned}
\dot{e}_f &= \mathbf{W}_f^T \sigma_f V_f^T \dot{x} - \dot{\hat{\mathbf{W}}}_f^T \hat{\sigma}_f - \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \dot{V}_f^T \hat{x} + \frac{1}{2} (\tilde{\mathbf{W}}_f^T - \mathbf{W}_f^T) \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' (\tilde{V}_f^T - V_f^T) \dot{x} \\
&\quad + \dot{e}_f(x) + \dot{\tilde{d}}^* - k\dot{x} - k\alpha\tilde{x} - \gamma\tilde{x} - \beta_1 \text{sgn}(\tilde{x}) + \alpha\dot{x} \\
&= \mathbf{W}_f^T \sigma_f V_f^T \dot{x} - \dot{\hat{\mathbf{W}}}_f^T \hat{\sigma}_f - \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \dot{V}_f^T \hat{x} + \frac{1}{2} \tilde{\mathbf{W}}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \tilde{V}_f^T \dot{x} - \frac{1}{2} \mathbf{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} \\
&\quad - \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' V_f^T \dot{x} + \frac{1}{2} \mathbf{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' V_f^T \dot{x} + \dot{e}_f(x) + \dot{\tilde{d}}^* - k\dot{e}_f - \gamma\tilde{x} \\
&\quad - \beta_1 \text{sgn}(\tilde{x}) + \alpha\dot{x}
\end{aligned} \tag{3.46}$$

Mặt khác ta có tính chất vết của ma trận: $\text{trace}(\mathbf{y}\mathbf{x}^T) = \mathbf{x}^T\mathbf{y}$, với \mathbf{x} và \mathbf{y} là 2 vecto cột.

Khai báo các biến trung gian như sau:

$$\begin{aligned}
\tilde{N} &= -\dot{\hat{\mathbf{W}}}_f^T \hat{\sigma}_f - \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \dot{V}_f^T \hat{x} + \frac{1}{2} \mathbf{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' V_f^T \dot{x} + \alpha\dot{x} \\
N_{B1} &= \mathbf{W}_f^T \sigma_f V_f^T \dot{x} - \frac{1}{2} \mathbf{W}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} - \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' V_f^T \dot{x} + \dot{\tilde{d}}^* + \dot{e}_f(x) \\
\hat{N}_{B2} &= \frac{1}{2} \tilde{\mathbf{W}}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \tilde{V}_f^T \dot{x} \\
N_{B2} &= \frac{1}{2} \tilde{\mathbf{W}}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \tilde{V}_f^T \dot{x} \\
\tilde{N}_{B2} &= \frac{1}{2} \tilde{\mathbf{W}}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \frac{1}{2} \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \tilde{V}_f^T \dot{x}
\end{aligned} \tag{3.47}$$

Rút gọn công thức (3.46) từ các công thức (3.47)

$$\dot{e}_f = \tilde{N} + N_{B1} + \hat{N}_{B2} - k\dot{e}_f - \gamma\tilde{x} - \beta_1 \text{sgn}(\tilde{x}) \tag{3.48}$$

Thay các công thức (3.48), (3.43), (3.40) vào công thức (3.45), ta có

$$\begin{aligned}
\dot{V}_2 &= e_f^T \left(\tilde{N} + N_{B1} + \hat{N}_{B2} - k\dot{e}_f - \gamma\tilde{x} - \beta_1 \text{sgn}(\tilde{x}) \right) + \gamma\tilde{x}^T (e_f - \alpha\tilde{x}) \\
&\quad - \frac{1}{2} \alpha \left[\text{trace} \left(\tilde{\mathbf{W}}_f^T \Gamma_{\mathbf{W}_f}^{-1} \Gamma_{\mathbf{W}_f} \hat{\sigma}_f' \hat{V}_f^T \dot{x} \tilde{x}^T \right) + \text{trace} \left(\hat{V}_f^T \Gamma_{\mathbf{V}_f}^{-1} \Gamma_{\mathbf{V}_f} \dot{x} \tilde{x}^T \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \right) \right] \\
&= e_f^T \left(\tilde{N} + N_{B1} + \hat{N}_{B2} \right) - k \|e_f\|^2 - e_f^T \beta_1 \text{sgn}(\tilde{x}) - \alpha \gamma \|\tilde{x}\|^2 \\
&\quad - \frac{1}{2} \alpha \left(\tilde{x}^T \tilde{\mathbf{W}}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} + \tilde{x}^T \hat{\mathbf{W}}_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{x} \right)
\end{aligned} \tag{3.49}$$

Theo công thức \hat{N}_{B2} trong công thức (3.47), (3.49) tương đương với

$$\begin{aligned}
\dot{V}_2 &= e_f^T (\tilde{N} + N_{B1}) - k \|e_f\|^2 - e_f^T \beta_1 \operatorname{sgn}(\tilde{x}) - \alpha \gamma \|\tilde{x}\|^2 + e_f^T \hat{N}_{B2} - \alpha \tilde{x}^T \hat{N}_{B2} \\
&= e_f^T (\tilde{N} + N_{B1}) - k \|e_f\|^2 - e_f^T \beta_1 \operatorname{sgn}(\tilde{x}) - \alpha \gamma \|\tilde{x}\|^2 + \dot{\tilde{x}}^T \hat{N}_{B2} \\
&= e_f^T (\tilde{N} + N_{B1}) - k \|e_f\|^2 - \beta_1 \alpha \tilde{x} \operatorname{sgn}(\tilde{x}) - \alpha \gamma \|\tilde{x}\|^2 + \dot{\tilde{x}}^T (N_{B2} + \beta_1 \operatorname{sgn}(\tilde{x})) - \dot{\tilde{x}}^T \tilde{N}_{B2}
\end{aligned} \tag{3.50}$$

Thay \hat{W}_f và \hat{V}_f trong công thức (3.43) vào \tilde{N} , ta có:

$$\tilde{N} = -\tilde{x} \dot{\tilde{x}}^T \hat{\sigma}_f^T \Gamma_{W_f} \hat{\sigma}_f - \hat{W}_f^T \hat{\sigma}_f' \hat{\sigma}_f^T \hat{W}_f \tilde{x} \dot{\tilde{x}}^T \Gamma_{V_f} \hat{x} + \frac{1}{2} W_f^T \hat{\sigma}_f' \hat{V}_f^T \dot{\tilde{x}} + \frac{1}{2} \hat{W}_f^T \hat{\sigma}_f' V_f^T \dot{\tilde{x}} + \alpha \dot{\tilde{x}} \tag{3.51}$$

Từ các **Giả thiết 3 – 6**, tính bị chặn của \hat{W}_f và \hat{V}_f qua toán tử proj, các ma trận $W_f, V_f, \hat{W}_f, \hat{V}_f, \hat{\sigma}_f, \hat{\sigma}_f', \hat{x}, \dot{\hat{x}}, \hat{e}_f, \hat{d}^*$ là bị chặn trên.

Từ (3.51), kết hợp với (3.40), ta đánh giá được:

$$\|\tilde{N}\| \leq \xi_1 \|\tilde{x}\| + \xi_2 \|e_f\| \tag{3.52}$$

Tương tự ta có:

$$\|N_{B1}\| \leq \xi_3, \|N_{B2}\| \leq \xi_4 \tag{3.53}$$

$$\|\tilde{N}_{B2}\| \leq \xi_5 \|\tilde{x}\| + \xi_6 \|e_f\| \tag{3.54}$$

Trong đó $\xi_1, \xi_2, \xi_3, \xi_4, \xi_5, \xi_6$ là các hằng số dương. Theo bất đẳng thức Cauchy – Schwarz và bất đẳng thức Young, ta có.

$$\begin{aligned}
\dot{V}_2 &\leq \|e_f\| \left(\xi_1 \|\tilde{x}\| + \xi_2 \|e_f\| + \xi_3 \right) - k \|e_f\|^2 - \beta_1 \alpha \sum_i |\tilde{x}_i| - \alpha \gamma \|\tilde{x}\|^2 + \|\tilde{x}\| \left(\xi_4 + \beta_1 + \xi_5 \|\tilde{x}\| + \xi_6 \|e_f\| \right) \\
&= -\beta_1 \alpha \sum_i |\tilde{x}_i| + (\xi_5 - \alpha \gamma) \|\tilde{x}\|^2 + (\xi_2 - k) \|e_f\|^2 + \xi_3 \|e_f\| + (\xi_4 + \beta_1) \|\tilde{x}\| + (\xi_1 + \xi_6) \|\tilde{x}\| \|e_f\| \\
&\leq -\beta_1 \alpha \sum_i |\tilde{x}_i| + (\xi_5 - \alpha \gamma) \|\tilde{x}\|^2 + (\xi_2 - k) \|e_f\|^2 + \xi_3 \|e_f\| + (\xi_4 + \beta_1) \|\tilde{x}\| + \frac{\xi_1 + \xi_6}{2} \|\tilde{x}\|^2 + \frac{\xi_1 + \xi_6}{2} \|e_f\|^2
\end{aligned} \tag{3.55}$$

Chọn các tham số thiết kế k, α, γ sao cho $\alpha \gamma > \xi_5 + \frac{\xi_1 + \xi_6}{2}$ và $k > \xi_2 + \frac{\xi_1 + \xi_6}{2}$. Đặt $\chi_1^2 = \alpha \gamma - \xi_5 - \frac{\xi_1 + \xi_6}{2} > 0, \chi_2^2 = k - \xi_2 - \frac{\xi_1 + \xi_6}{2} > 0$. Ta có

$$\begin{aligned}
\dot{V}_2 &\leq -\beta_1 \alpha \sum_i |\tilde{x}_i| - \chi_1^2 \|\tilde{x}\|^2 + (\xi_4 + \beta_1) \|\tilde{x}\| - \chi_2^2 \|e_f\| + \xi_3 \|e_f\| \\
&\leq -\beta_1 \alpha \sum_i |\tilde{x}_i| - \left(\chi_1 \|\tilde{x}\| - \frac{\xi_4 + \beta_1}{2\chi_1} \right)^2 - \left(\chi_2 \|e_f\| - \frac{\xi_3}{2\chi_2} \right)^2 + \left(\frac{\xi_4 + \beta_1}{2\chi_1} \right)^2 + \left(\frac{\xi_3}{2\chi_2} \right)^2 \quad (3.56)
\end{aligned}$$

Xét miền D chứa gốc tọa độ

$$D = \left\{ z = \begin{bmatrix} \tilde{x}^T & e_f^T \end{bmatrix} \left| \beta_1 \alpha \sum_i |\tilde{x}_i| + \left(\chi_1 \|\tilde{x}\| - \frac{\xi_4 + \beta_1}{2\chi_1} \right)^2 + \left(\chi_2 \|e_f\| - \frac{\xi_3}{2\chi_2} \right)^2 \leq \left(\frac{\xi_4 + \beta_1}{2\chi_1} \right)^2 + \left(\frac{\xi_3}{2\chi_2} \right)^2 \right\} \quad (3.57)$$

Dễ thấy \dot{V}_2 bán xác định dương với mọi điểm nằm trong miền D, sau một khoảng thời gian đủ lớn, hàm V_2 sẽ bị chặn, do đó các tín hiệu \tilde{x} , e_f , \tilde{W}_f , \tilde{V}_f cũng bị chặn trong một lân cận quanh gốc tọa độ. Do \tilde{x} , e_f cùng bị chặn, nên $\dot{\tilde{x}} = e_f - \alpha \tilde{x}$ cũng bị chặn.

Vậy định lý 2 đã được chứng minh. \square

3.4. THIẾT KẾ BỘ ĐIỀU KHIỂN TỐI ƯU CHO HỆ THỐNG:

Ta thiết kế bộ điều khiển như sau:

$$u = u_1 - g^+(x) \hat{d}^* \quad (3.58)$$

Với u_1 là thành phần phản hồi của bộ điều khiển, $g^+(x) \hat{d}^*$ là thành phần bù nhiễu của hệ thống.

Trong đó, $g^+(x)$ là ma trận giả nghịch đảo của $g(x)$. Thay (3.58) và (3.1), ta có

$$\begin{aligned}
\dot{x} &= f(x) + g(x) \left(u_1 - g^+(x) \hat{d}^* + d \right) \\
&= f(x) + g(x) u_1 - \hat{d}^* + d^* \\
&= f(x) + g(x) u_1 + \tilde{d}^* \quad (3.59)
\end{aligned}$$

Để đơn giản trong việc thiết kế, ta coi $\tilde{d}^* \approx 0$. Khi đó bài toán điều khiển hệ thống có nhiễu (3.1) trở thành bài toán điều khiển hệ thống (3.59) không có nhiễu.

Hàm mục tiêu điều khiển là các biểu thức từ (3.2) – (3.5), nhưng thay thế u thành u_1 trong các công thức trên.

$$J(x(t)) = \int_t^\infty r(x(\tau), u_1(\tau)) d\tau \quad (3.60)$$

Ta thiết kế bộ điều khiển u_1 bị chặn, tối ưu cho hệ trên thỏa mãn hàm mục tiêu.

Đạo hàm 2 vế của (3.60) theo thời gian ta có:

$$\dot{J} = -x^T Qx - 2 \int_0^{u_1} \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv = \frac{\partial J}{\partial x} \dot{x} = \nabla J_x \dot{x} = \nabla J_x (f(x) + g(x)u_1) \quad (3.61)$$

Hàm Hamilton của hệ thống (3.59) là

$$H(x, u_1, \nabla J_x) = x^T Qx + 2 \int_0^{u_1} \lambda \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv + \nabla J_x (f(x) + g(x)u_1) \quad (3.62)$$

u_1^* là bộ điều khiển tối ưu của hệ thống, J^* là hàm mục tiêu tương ứng với $u_1 = u_1^*$. Điều kiện để u_1^* là bộ điều khiển tối ưu của hệ thống (3.59) là

$$\begin{aligned} \frac{\partial H}{\partial u_1^*} &= 2\lambda \tanh^{-1} \left(\frac{u_1^*}{\lambda} \right) R + g^T(x) \nabla J_x^* = 0 \\ u_1^* &= -\lambda \tanh \left(\frac{1}{2\lambda} R^{-1} g^T(x) \nabla J_x^* \right) \end{aligned} \quad (3.63)$$

Mặt khác, u_1^* và J^* phải thỏa mãn phương trình HJB sau

$$H(x, u_1^*, \nabla J_x^*) = x^T Qx + 2 \int_0^{u_1^*} \lambda \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv + \nabla J_x^* \dot{x} = 0 \quad (3.64)$$

Thay các giá trị tối ưu trong (3.64) thành các giá trị xấp xỉ tương ứng, phương trình HJB trở thành

$$\hat{H}(x, \hat{x}, \hat{u}_1, \nabla \hat{J}_x) = x^T Qx + 2 \int_0^{\hat{u}_1} \lambda \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv + \nabla \hat{J}_x \dot{\hat{x}} \quad (3.65)$$

Để đánh giá và thiết kế bộ điều khiển tối ưu, ta gọi biến δ_{hjb} là sai lệch Bellman

$$\delta_{hjb} = \hat{H}(x, \hat{x}, \hat{u}_1, \nabla \hat{J}_x) - H(x, u_1^*, \nabla J_x^*) = \hat{H}(x, \hat{x}, \hat{u}_1, \nabla \hat{J}_x) \quad (3.66)$$

Nghiệm của phương trình HJB (3.64) rất khó để xác định chính xác bằng các phương pháp giải tích thông thường, ngay cả khi các hàm $f(x)$ và $g(x)$ là hoàn toàn xác định. Ở đây hàm $f(x)$ chưa được xác định một cách tường minh, nên ta cần một thuật toán xác định u_1^* và J^* một cách gần đúng. Các hàm u_1^* và J^* đều là hàm phi tuyến, để có thể xác định một cách gần đúng, ta đặt

$$\begin{aligned}
J^*(x) &= W^T \phi(x) + \varepsilon_v(x) \\
u_1^*(x) &= -\lambda \tanh \left[\frac{1}{2\lambda} R^{-1} g^T(x) (\phi'(x)^T W + \varepsilon_v'(x)^T) \right]
\end{aligned} \tag{3.67}$$

Trong đó $\phi(x) = [\phi_1(x) \ \phi_2(x) \ \dots \ \phi_N(x)]^T \in R^N$ là vecto các hàm cơ sở của mạng nơ-ron, $\phi'(x) = \frac{\partial \phi}{\partial x} \in R^{N \times n}$ là đạo hàm của $\phi(x)$ theo x . Các hàm $\phi_i(x)$ phải thỏa mãn $\phi_i(0) = 0$ và $\phi_i'(0) = 0$ để đảm bảo tại $x=0$, u_1^* và J^* đều triệt tiêu. Các hàm $\phi_i(x)$ phải độc lập tuyến tính với nhau. Chọn số hàm càng lớn, hàm nhận dạng càng gần với kết quả. $\varepsilon_v \rightarrow 0$ khi $N \rightarrow \infty$.

Bộ điều khiển tối ưu được thiết kế dựa trên cấu trúc actor – critic. Giá trị xấp xỉ của thành phần actor và critic là.

$$\begin{aligned}
\hat{J}(x) &= \hat{W}_c^T \phi(x) \\
\hat{u}_1(x) &= -\lambda \tanh \left[\frac{1}{2\lambda} R^{-1} g^T(x) \phi'(x)^T \hat{W}_a \right]
\end{aligned} \tag{3.68}$$

Các ma trận trọng số $\hat{W}_a \in R^N$ và $\hat{W}_c \in R^N$ là ước lượng của các ma trận actor, critic lý tưởng W_a và W_c .

Sai lệch giữa trọng số lý tưởng và ước lượng là $\tilde{W}_c = W_c - \hat{W}_c \in R^N$ và $\tilde{W}_a = W_a - \hat{W}_a \in R^N$

Thay thế (3.68) vào công thức (3.65). Ta có

$$\delta_{hjb} = x^T Qx + 2 \int_0^{u_1} \lambda \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv + \hat{W}_c \omega \tag{3.69}$$

Trong đó $\omega = \phi'(x)\hat{x} \in R^N$ là biến trung gian được dùng cho thiết kế điều khiển sau. Thuật toán được sử dụng là ADP onpolicy, cấu trúc actor – critic, với các ma trận \hat{W}_a và \hat{W}_c được cập nhật theo nguyên lý tối ưu hàm Bellman.

Gọi hàm tích phân bình phương sai lệch Bellman để tối ưu thành phần critic.

$$E_c = \int_0^t \delta_{hjb}^2(\tau) d\tau \tag{3.70}$$

Cần xác định trọng số \hat{W}_c trong (3.69) sao cho E_c trong (3.70) đạt cực tiểu. Dùng phương pháp bình phương tối thiểu, đạo hàm E_c theo \hat{W}_c có

$$\frac{\partial E_c}{\partial \hat{W}_c} = 2 \int_0^t \delta_{hjb}(\tau) \frac{\partial \delta_{hjb}}{\partial \hat{W}_c} d\tau = 2 \int_0^t \left[x^T Q x + 2 \int_0^{\hat{u}_1} \lambda \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv + \hat{W}_c \omega \right] \omega^T d\tau = 0 \quad (3.71)$$

Giải phương trình (3.71), nghiệm \hat{W}_c thu được là

$$\hat{W}_c = - \left(\int_0^t \omega(\tau) \omega^T(\tau) d\tau \right)^{-1} \left(\int_0^t r(x, \hat{u}_1) \omega(\tau) d\tau \right) \quad (3.72)$$

Nghiệm giải được từ (3.72) có 2 nhược điểm. Nghiệm chỉ tồn tại khi ma trận $\int_0^t \omega(\tau) \omega^T(\tau) d\tau$ là khả nghịch. Ngoài ra, việc lấy tích phân trong quá trình tính toán sẽ gây khó khăn cho việc tính toán. Vì thế ta cần dùng một thuật toán khác xác định \hat{W}_c bằng phương pháp đệ quy như sau

$$\dot{\hat{W}}_c = -\eta_c \Gamma \frac{\omega}{1 + v \omega^T \Gamma \omega} \delta_{hjb} \quad (3.73)$$

Trong đó các hằng số dương η_c và v là các tham số thiết kế. Ma trận $\Gamma(t) = \left(\int_0^t \omega(\tau) \omega^T(\tau) d\tau \right)^{-1} \in R^{N \times N}$ là ma trận đối xứng xác định dương được ước lượng như sau.

$$\dot{\Gamma} = -\eta_c \Gamma \frac{\omega \omega^T}{1 + v \omega^T \Gamma \omega} \Gamma, \quad \Gamma(0) = \varphi_0 I \quad (3.74)$$

Nhận xét 2:

Trong quá trình cập nhật $\Gamma(t)$, có thể do cách chọn η_c hoặc v không phù hợp có thể khiến ma trận $\Gamma(t)$ không xác định dương, vì thế, ta cần thực hiện giới hạn để bảo đảm $\Gamma(t)$ luôn xác định dương bằng giới hạn

$$0 < \varphi_1 I \leq \Gamma(t) \leq \varphi_0 I \quad (3.75)$$

Khi $\Gamma(t)$ vượt ra khỏi giới hạn này sẽ được reset về giá trị $\Gamma(0)$. Từ (3.74), dễ thấy rằng $\Gamma \leq 0$. Do đó $\Gamma(t)$ luôn giảm, đến thời điểm $\Gamma(t)$ gần xấp xỉ bằng 0 thì luật cập nhật (3.73) sẽ diễn ra rất chậm. Vì vậy nên việc reset giá trị của $\Gamma(t)$ khi $\Gamma(t)$ quá nhỏ là cần thiết

Để cập nhật trọng số cho \hat{W}_a , ta vẫn dựa trên nguyên lý tối ưu hàm sai lệch Bellman δ_{hjb} . Tuy nhiên phương pháp tối ưu bằng bình phương tối thiểu không phù hợp để xác định \hat{W}_a vì δ_{hjb} là hàm phi tuyến theo \hat{W}_a .

Gọi hàm bình phương sai lệch Bellman để tối ưu thành phần actor.

$$E_a = \delta_{hjb}^2(t) \quad (3.76)$$

Đạo hàm (3.76) theo \hat{W}_a

$$\frac{\partial E_a}{\partial \hat{W}_a} = 2 \frac{\partial \delta_{hjb}}{\partial \hat{W}_a} \delta_{hjb} = 2 \left(\hat{W}_c^T \phi' \frac{d\hat{x}}{d\hat{u}_1} \frac{d\hat{u}_1}{d\hat{W}_a} + 2\lambda \tanh^{-1} \left(\frac{\hat{u}_1}{\lambda} \right)^T R \frac{d\hat{u}_1}{d\hat{W}_a} \right) \delta_{hjb} \quad (3.77)$$

Do tính phi tuyến của E_a theo \hat{W}_a , để tìm nghiệm \hat{W}_a tối thiểu E_a , ta dùng phương pháp gradient descent để cập nhật. Luật cập nhật \hat{W}_a như sau

$$\dot{\hat{W}}_a = \text{proj} \left\{ \frac{-2\eta_{a1}}{\sqrt{1+\omega^T \omega}} \left[\left(\hat{W}_c^T \phi' \frac{d\hat{x}}{d\hat{u}_1} \frac{d\hat{u}_1}{d\hat{W}_a} \right)^T \delta_{hjb} + 2\lambda \left(\frac{d\hat{u}_1}{d\hat{W}_a} \right)^T R \tanh^{-1} \left(\frac{\hat{u}_1}{\lambda} \right) \delta_{hjb} \right] - \eta_{a2} (\hat{W}_a - \hat{W}_c) \right\} \quad (3.78)$$

Toán tử proj được dùng để giới hạn \hat{W}_a . Thành phần thứ nhất trong (3.78) là luật cập nhật dựa trên thuật toán gradient descent.

Từ (3.68), lấy đạo hàm \hat{u}_1 theo \hat{W}_a . Ta có.

$$\begin{aligned} \frac{\partial \hat{u}_1}{\partial \hat{W}_a} &= -\frac{1}{2} B \left[\tanh \left(\frac{1}{2\lambda} R^{-1} g^T(x) \phi'(x)^T \hat{W}_a \right) \right] R^{-1} g^T(x) \phi'(x)^T \\ &= -\frac{1}{2} B \left(\frac{-\hat{u}_1}{\lambda} \right) R^{-1} g^T(x) \phi'(x)^T \end{aligned} \quad (3.79)$$

Trong đó, phép ánh xạ $B(\theta) \in R^{m \times m}$ được định nghĩa là

$$B(\theta) = \begin{bmatrix} 1-\theta_1^2 & 0 & 0 & 0 \\ 0 & 1-\theta_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \dots & 1-\theta_m^2 \end{bmatrix} \quad (3.80)$$

Qua các biến đổi toán học, ta có

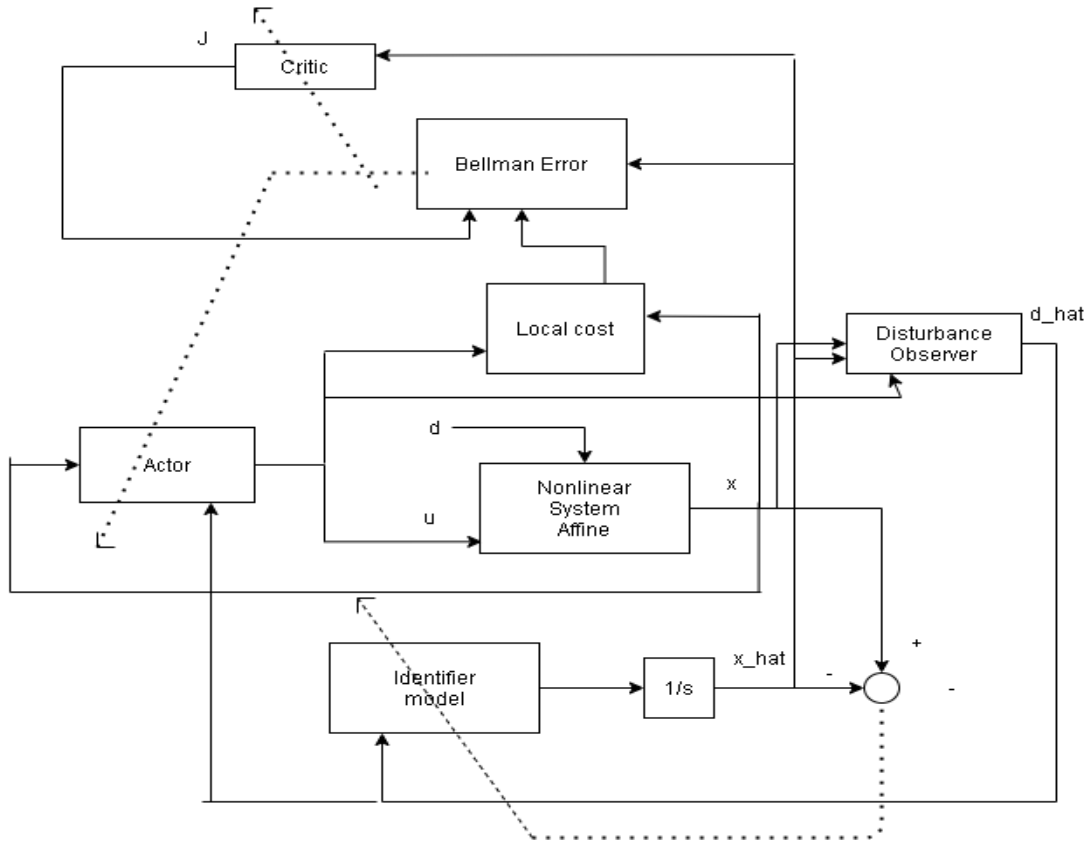
$$\begin{aligned} \left(\hat{W}_c^T \phi'(x) \frac{d\hat{x}}{d\hat{u}_1} \frac{d\hat{u}_1}{d\hat{W}_a} \right)^T \delta_{hjb} &= \left(-\frac{1}{2} \hat{W}_c \phi'(x) g(x) B \left(\frac{-\hat{u}_1}{\lambda} \right) R^{-1} g^T(x) \phi'(x)^T \right)^T \delta_{hjb} \\ &= -\frac{1}{2} \phi'(x) g(x) R^{-1} B \left(\frac{-\hat{u}_1}{\lambda} \right) g^T(x) \phi'(x)^T \hat{W}_c \delta_{hjb} \end{aligned} \quad (3.81)$$

$$\begin{aligned}
& 2\lambda \left(\frac{d\hat{u}_1}{d\hat{W}_a} \right)^T R \tanh^{-1} \left(\frac{\hat{u}_1}{\lambda} \right) \delta_{hjb} \\
& = 2\lambda \frac{-1}{2} \phi'(x) g(x) R^{-1} B \left(\frac{-\hat{u}_1}{\lambda} \right) R \frac{-1}{2\lambda} R^{-1} g^T(x) \phi'(x)^T \hat{W}_a \delta_{hjb} \\
& = \frac{1}{2} \phi'(x) g(x) R^{-1} B \left(\frac{-\hat{u}_1}{\lambda} \right) g^T(x) \phi'(x)^T \hat{W}_a \delta_{hjb}
\end{aligned} \tag{3.82}$$

Lần lượt thay (3.81) và (3.82) vào công thức (3.78), ta có

$$\dot{\hat{W}}_a = \text{proj} \left\{ \frac{-\eta_{a1}}{\sqrt{1+\omega^T \omega}} \phi'(x) g(x) R^{-1} B \left(\frac{-\hat{u}_1}{\lambda} \right) g^T(x) \phi'(x)^T (\hat{W}_a - \hat{W}_c) \delta_{hjb} - \eta_{a2} (\hat{W}_a - \hat{W}_c) \right\} \tag{3.83}$$

Sau khi tổng hợp các thành phần của bộ điều khiển, ta có sơ đồ tổng quát của thuật toán như sau



Hình 1. Sơ đồ cấu trúc tổng quát của toàn bộ thuật toán

CHƯƠNG 4: MÔ PHỎNG THUẬT TOÁN ĐỀ XUẤT VỚI QUÁ TRÌNH HOÁ HỌC TRONG THIẾT BỊ CSTR

4.1. MÔ HÌNH TOÁN HỌC VÀ THAM SỐ CỦA ĐỐI TƯỢNG

Giả thiết rằng phản ứng lên men trong thiết bị xảy ra là liên tục. Bảng sau mô tả các biến và tham số của đối tượng

Biến	Ý nghĩa	Giá trị	Đơn vị
X	Nồng độ của sinh khối		g/l
S	Nồng độ của cơ chất		g/l
F	Lưu lượng dòng chảy cấp vào		l/h
F _d	Nhiều tác động lên dòng chảy do cơ cấu chấp hành		l/h
V	Thể tích phản ứng	4	l
S _F	Nồng độ đầu vào của cơ chất	10	g/l
Y	Hiệu suất phản ứng	0.5	
μ _{max}	Tốc độ lên men cực đại	1	1/h
K ₁	Thông số bão hòa	0.03	g/l
K ₂	Thông số của chất ức chế	0.5	l/g

Bảng 1. Các thông số của đối tượng điều khiển

Dựa vào tài liệu tham khảo [3], ta có mô hình phi tuyến của quá trình lên men như sau:

$$\begin{bmatrix} \dot{X} \\ \dot{S} \end{bmatrix} = \begin{bmatrix} \frac{\mu_{\max} SX}{K_1 + S + K_2 S^2} \\ -\frac{\mu_{\max} SX}{(K_1 + S + K_2 S^2)Y} \end{bmatrix} + \begin{bmatrix} \frac{-X}{V} \\ \frac{S_F - X}{V} \end{bmatrix} (F + F_d) \quad (4.1)$$

Giả sử S₀, X₀, F₀ là điểm cân bằng của đối tượng trên.

Đặt các biến đầu ra và đầu vào: $X - X_0 = x_1$, $S - S_0 = x_2$, $F - F_0 = u$, $F_d = d$. Ta có mô hình

$$\begin{aligned}
\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} &= \begin{bmatrix} \frac{\mu_{\max}(x_1 + X_0)(x_2 + S_0)}{K_1 + (x_2 + S_0) + K_2(x_2 + S_0)^2} \\ \frac{-\mu_{\max}(x_1 + X_0)(x_2 + S_0)}{K_1 + (x_2 + S_0) + K_2(x_2 + S_0)^2} \frac{1}{Y} \end{bmatrix} + \begin{bmatrix} \frac{-(X_0 + x_1)}{V} \\ \frac{S_F - (X_0 + x_1)}{V} \end{bmatrix} (u + F_0 + d) \\
&= \begin{bmatrix} \frac{\mu_{\max}(x_1 + X_0)(x_2 + S_0)}{K_1 + (x_2 + S_0) + K_2(x_2 + S_0)^2} - \frac{F_0(X_0 + x_1)}{V} \\ \frac{-\mu_{\max}(x_1 + X_0)(x_2 + S_0)}{K_1 + (x_2 + S_0) + K_2(x_2 + S_0)^2} \frac{1}{Y} + \frac{(S_F - (X_0 + x_1))F_0}{V} \end{bmatrix} + \begin{bmatrix} \frac{-(X_0 + x_1)}{V} \\ \frac{S_F - (X_0 + x_1)}{V} \end{bmatrix} (u + d) \quad (4.2) \\
\dot{x} &= f(x) + g(x)(u + d)
\end{aligned}$$

Dễ thấy mô hình (4.2) có điểm cân bằng tại gốc tọa độ, các ma trận $f(x)$ và $g(x)$ thỏa mãn các giả thiết đã cho.

Để khảo sát hoạt động của thuật toán trên vào mô hình (4.2), ta cần chọn các giá trị đặt cho nồng độ của hai chất phản ứng, đưa bài toán về tám giá trị đặt

Điều kiện cân bằng của mô hình (4.2) là

$$\begin{bmatrix} \frac{\mu_{\max} X_0 S_0}{K_1 + S_0 + K_2 S_0^2} \\ \frac{-\mu_{\max} X_0 S_0}{K_1 + S_0 + K_2 S_0^2} \frac{1}{Y} \end{bmatrix} + \begin{bmatrix} \frac{-X_0}{V} \\ \frac{S_F - X_0}{V} \end{bmatrix} F_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (4.3)$$

Giải hệ phương trình (4.3), phương trình có nghiệm

$$S_0 = \frac{1}{2} \frac{-2K_1 + 2\sqrt{K_1^2 + S_F^2 K_1 K_2 + S_F K_1}}{S_F K_2 + 1} \quad (4.4)$$

$$X_0 = (S_F - S_0)Y \quad (4.5)$$

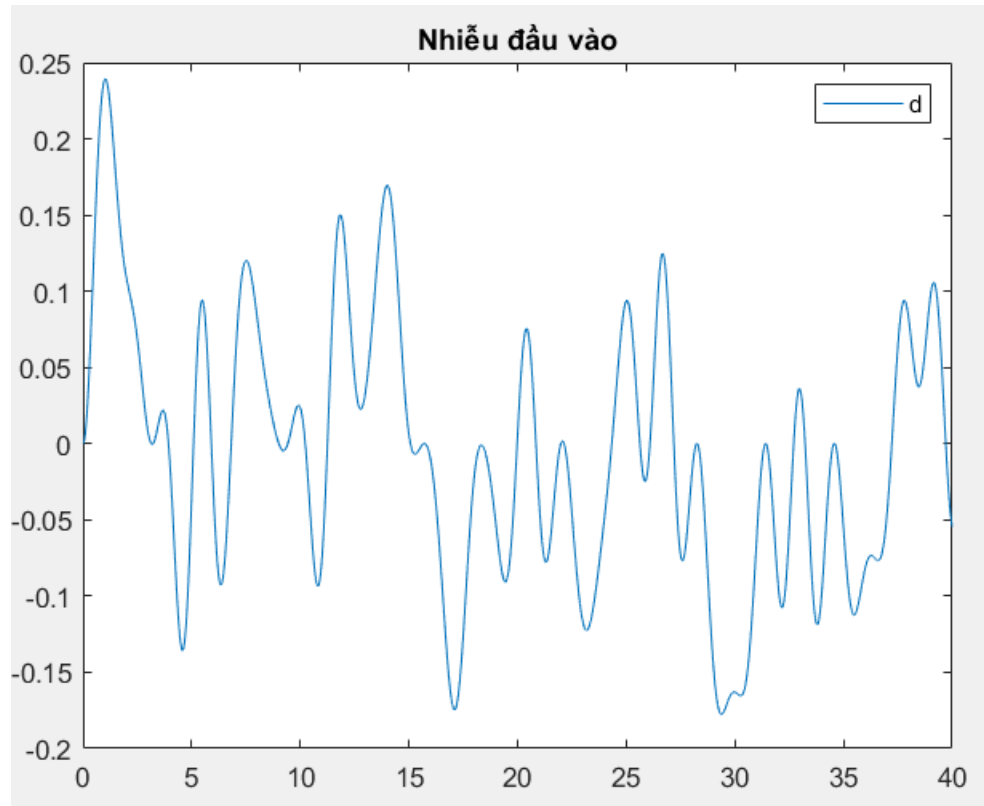
$$F_0 = \frac{\mu_{\max} S_0}{K_2 S_0^2 + S_0 + K_1} \quad (4.6)$$

Thay số vào (4.4) – (4.6), ta có $S_0 = 0.2187(\text{g/l})$, $X_0 = 4.8907(\text{g/l})$, $F_0 = 3.2029(\text{g/l})$

Trong mô hình (4.2), có $\lambda = 2$ là giới hạn cực đại của đầu vào u . Nhiễu d là nhiễu biến thiên phụ thuộc vào thời gian là

$$d(t) = 0.1(\sin^2(t)\cos(t) + \sin^2(2t)\cos(0.1t) + \sin^2(-1.2t)\cos(0.5t) + \sin^5(t)) \quad (4.7)$$

Đồ thị của nhiễu d theo thời gian như sau:



Hình 2. Đồ thị nhiễu đầu vào mô phỏng

Từ đồ thị d , đảm bảo được điều kiện ở **Giả thiết 1**, $0.25 < \bar{d} < \lambda = 2$

4.2. THIẾT KẾ BỘ QUAN SÁT NHIỄU CHO HỆ THỐNG

Theo kết quả ở **Nhận xét 1**, khi số bậc của bộ quan sát càng tăng thì sai lệch nhiễu đo được càng nhỏ. Nhiễu xấp xỉ càng chính xác.

Ta sẽ kiểm chứng bằng ba bộ quan sát nhiễu có bậc từ 1 đến 3. Các tham số bộ quan sát được cho trong bảng sau.

	Bậc 1	Bậc 2	Bậc 3
L_1	1.97	18.76	14.2
L_2	x	54.93	217.1
L_3	x	x	716.3

Bảng 2. Các tham số của bộ quan sát nhiễu

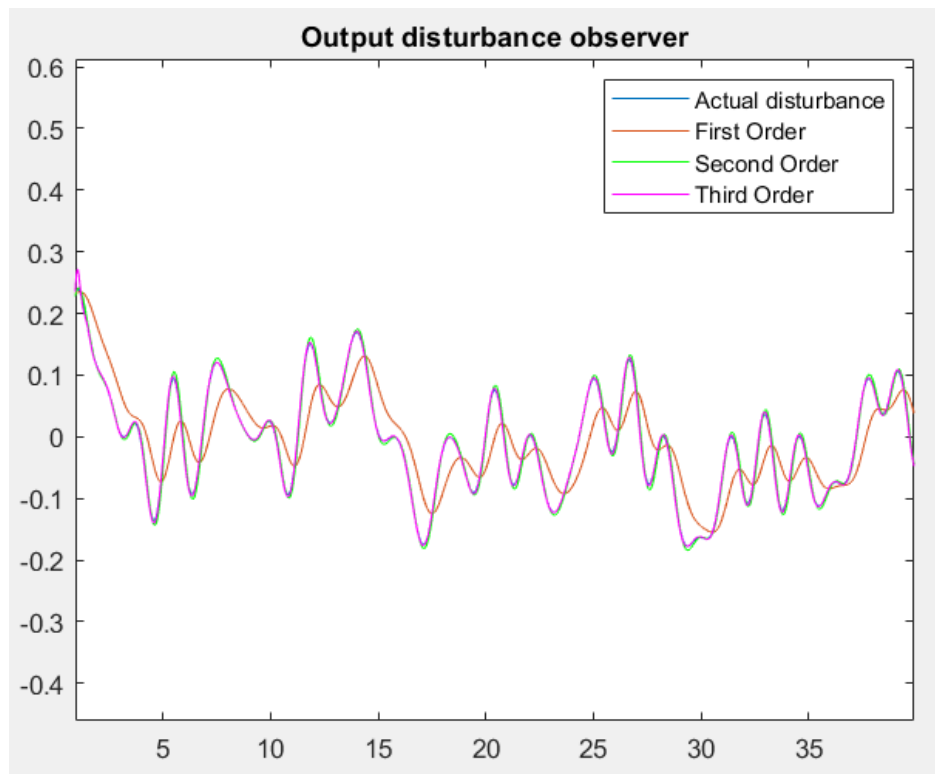
Các ma trận D_n là

$$D_1 = [-1.97], D_2 = \begin{bmatrix} -18.76 & 1 \\ -54.93 & 0 \end{bmatrix}, D_3 = \begin{bmatrix} -14.2 & 1 & 0 \\ -217.1 & 0 & 1 \\ -716.3 & 0 & 0 \end{bmatrix}$$

đều là ma trận bền, thỏa mãn **Định lý 1**

Để khảo sát đặc tính của bộ quan sát nhiễu, ta đặt vào hệ thống một đầu vào bất kỳ làm hệ thống ổn định.

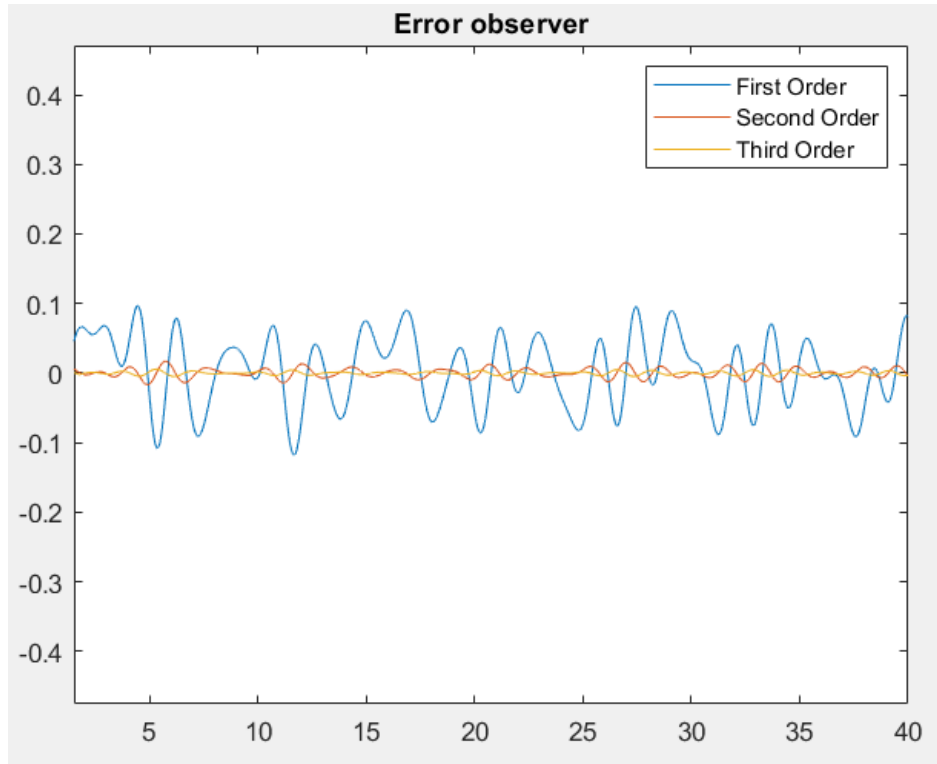
Chọn $u = [0.6734 \quad -0.7676]x$. Đặc tính nhiễu thực tế và các nhiễu quan sát như dưới đây



Hình 3. So sánh nhiễu đầu ra bộ quan sát với bậc quan sát khác nhau

Có thể nhận thấy từ hình vẽ, nhiễu của bộ quan sát bậc 1 sai lệch khá lớn so với nhiễu thực tế, đặc biệt là khi tần số nhiễu lớn. Bộ quan sát bậc 2 và bậc 3 bám khá tốt so với nhiễu thực tế.

Sai lệch quan sát giữa nhiễu của bộ quan sát và nhiễu thực tế như sau



Hình 4. Sai lệch của nhiều xấp xỉ so với thực tế

Sai lệch nhiều của bộ quan sát bậc 1 rất lớn nên chất lượng quan sát không thể dùng để điều khiển hệ thực

Sai lệch nhiều của bộ quan sát bậc 3 là nhỏ nhất, tuy nhiên chất lượng quan sát cải thiện không đáng kể so với bộ quan sát bậc 2, còn làm tăng độ phức tạp tính toán.

Vậy nên khi thiết kế bộ quan sát nhiều theo phương pháp đề xuất, cần chọn bậc và tham số quan sát một cách phù hợp.

4.3. NHẬN DẠNG MÔ HÌNH CHO HỆ THỐNG

Các phản ứng hóa sinh trong quá trình lên men trên có các thông số như thể tích, nồng độ, lưu lượng ... có thể đo được bằng cảm biến. Tuy nhiên các hệ số phản ứng, hiệu suất ... chắc chắn không thể xác định chính xác bằng cảm biến thông thường. Nên mô hình (4.2) là không thể xác định chính xác hoàn toàn, luôn chứa thành phần bất định. Do đó, cần thiết kế nhận dạng cho hệ thống.

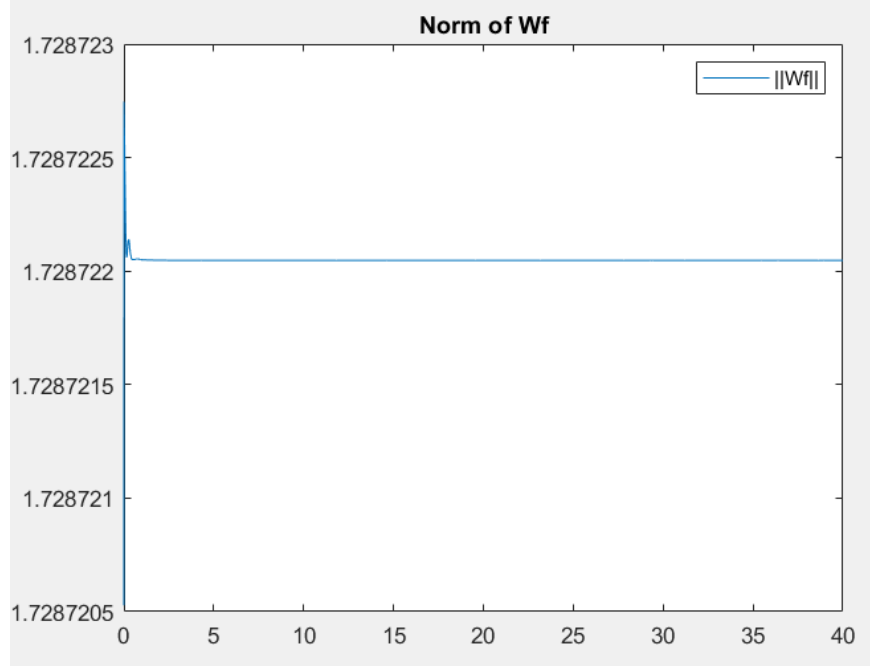
Ta thiết kế mạng nơ-ron nhận dạng hệ thống như các công thức (3.36) – (3.38).

Trong đó, 2 ma trận $W_f \in R^{(L_f+1) \times n}$, $V_f \in R^{n \times L_f}$ có $L_f = 5$. Các hàm $\sigma_i(x)$ là các hàm sigmoid, $\sigma_i(x) = \frac{1}{1+e^{x_i}}$. Tính chất của hàm sigmoid là

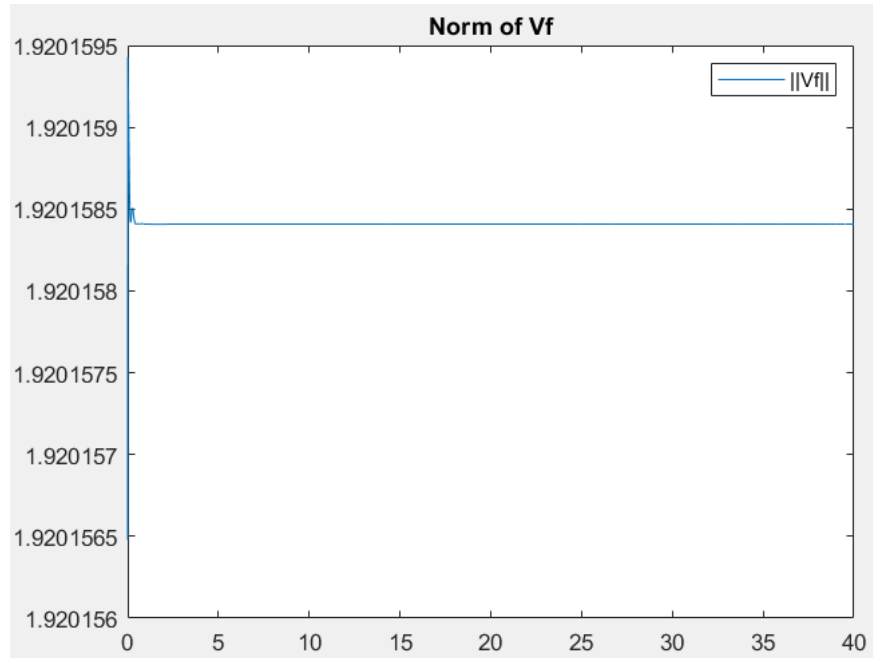
$$\sigma'_i(x_i) = \sigma_i(x_i)(1 - \sigma_i(x_i)) \quad (4.8)$$

Các tham số thiết kế là: $k = 800$, $\alpha = 300$, $\gamma = 5$, $\beta_1 = 0.2$. Các ma trận $\Gamma_{Wf} = 0.1I_{L_f+1}$, $\Gamma_{Vf} = 0.1I_2$.

Bộ điều khiển áp dụng vẫn là $u = [0.6734 \quad -0.7676]x$. Các ma trận \hat{W}_f và \hat{V}_f được khởi tạo ngẫu nhiên trong khoảng từ 0 đến 1. Các đồ thị $\|\hat{W}_f\|$ và $\|\hat{V}_f\|$ như sau



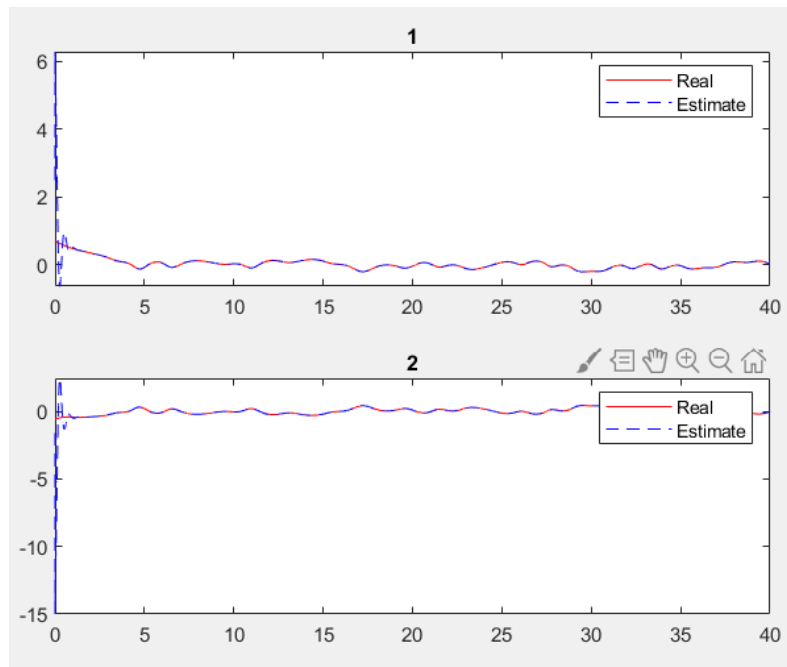
Hình 5. Sự hội tụ của ma trận W_f



Hình 6. Sự hội tụ của ma trận V_f

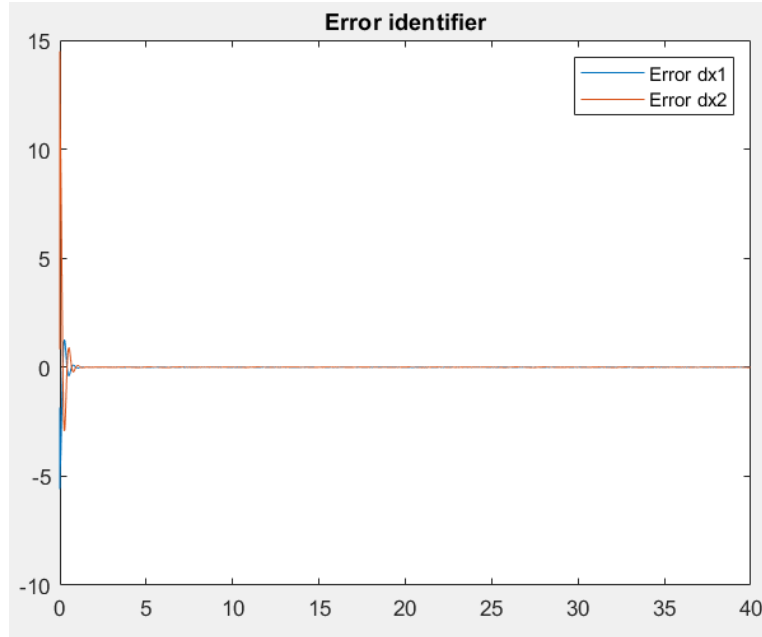
Nhận thấy các ma trận \widehat{W}_f và \widehat{V}_f đều hội tụ về giá trị xác định sau khoảng thời gian đủ lớn, nên $\widehat{W}_f \rightarrow W_f$ và $\widehat{V}_f \rightarrow V_f$.

Đồ thị của giá trị \hat{x} và \dot{x} như sau



Hình 7. Ước lượng mô hình trạng thái so với thực tế

Sai lệch của phép nhận dạng là \tilde{x} là



Hình 8. Sai lệch khi nhận dạng mô hình

Từ 2 đồ thị trên, sai lệch nhận dạng \tilde{x} cũng tiến về 0, trong thời gian đủ lớn. Thuật toán nhận dạng xấp xỉ một cách gần đúng mô hình bất định của đối tượng. **Định lý 2** được nghiệm đúng bằng mô phỏng.

4.4. BỘ ĐIỀU KHIỂN TỐI ƯU CHO HỆ THỐNG

Hệ thống phản ứng lên men trên là quá trình điều khiển nồng độ. Nồng độ là đại lượng có quán tính lớn, hơn nữa nếu phản ứng xảy ra một cách chậm chạp thì việc tối ưu bộ điều khiển là việc cần thiết.

Điều khiển tối ưu hệ (4.2) không thể giải bằng phương trình HJB do mô hình không xác định. Hơn nữa việc giải phương trình HJB rất khó, không có nghiệm tường minh.

Dùng phương pháp được đề xuất có thể giải gần đúng phương trình HJB, tìm trọng số tối ưu cho hàm mục tiêu

Bộ điều khiển tối ưu thiết kế sử dụng công thức (3.67). Luật cập nhật ma trận trọng số cho thành phần critic (3.73) – (3.75). Luật cập nhật ma trận trọng số cho thành phần actor (3.83).

Trong các công thức, thành phần δ_{hjb} còn chứa thành phần tích phân. Nên cần xác định lại công thức δ_{hjb} để loại bỏ thành phần tích phân trên

Ta có

$$\delta_{hjb} = x^T Qx + 2 \int_0^{\hat{u}_1} \lambda \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv + \hat{W}_c \omega \quad (4.9)$$

Dùng tích phân từng phần cho thành phần thứ 2 của (4.9)

$$\delta_{hjb} = x^T Qx + 2\lambda \tanh^{-1} \left(\frac{\hat{u}_1}{\lambda} \right)^T Ru + \lambda^2 \bar{R} \ln \left(1 - \left(\frac{\hat{u}_1}{\lambda} \right)^2 \right) + \hat{W}_c \omega \quad (4.10)$$

Trong đó, ma trận \bar{R} là ma trận hàng, ghép bởi tất cả các phần tử trên đường chéo của R , $\bar{R} = \text{diag}(R)$.

Hàm cơ sở cho mạng nơ-ron actor – critic được chọn là

$$\phi(x) = \begin{bmatrix} x_1^2 & x_1 x_2 & x_2^2 \end{bmatrix}^T \quad (4.11)$$

Hàm mục tiêu có ma trận mục tiêu là $Q = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}$, $R = 1$.

Các tham số thiết kế là: $\nu = 0.1$, $\eta_c = 20$, $\eta_{a1} = 10$, $\eta_{a2} = 50$. Tại thời điểm $t=0$, chọn giá trị $x(0) = \begin{bmatrix} 1 \\ -0.5 \end{bmatrix}$, các ma trận trọng số ban đầu được khởi tạo là giá trị ngẫu nhiên trong khoảng từ -5 đến 5.

Để đạt điều kiện PE, trong thành phần của u_1 được thêm một nhiễu nhỏ nhằm bảo đảm hệ thống hội tụ về giá trị ổn định. Nhiễu này được thêm vào hàm bên trong hàm tanh của bộ điều khiển

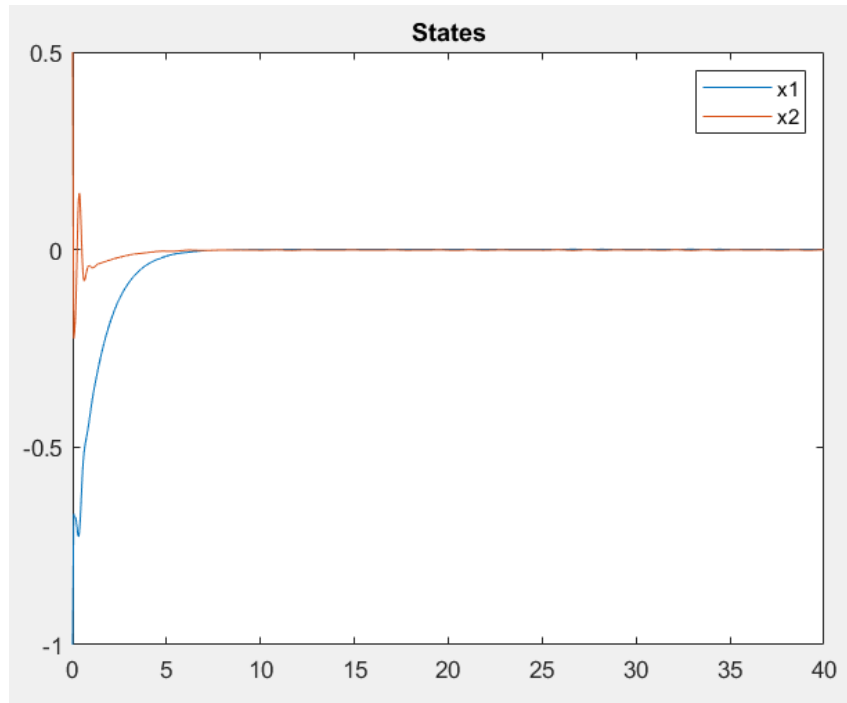
$$e = \sum_{i=1}^{100} 0.00001 \sin(\omega_i t) \quad (4.12)$$

Trong đó, các tần số ω_i được chọn ngẫu nhiên trong khoảng từ -500 đến 500. Nhiễu này được tác dụng trong 16 giờ đầu tiên của quá trình

Vậy tín hiệu u tác dụng lên hệ là

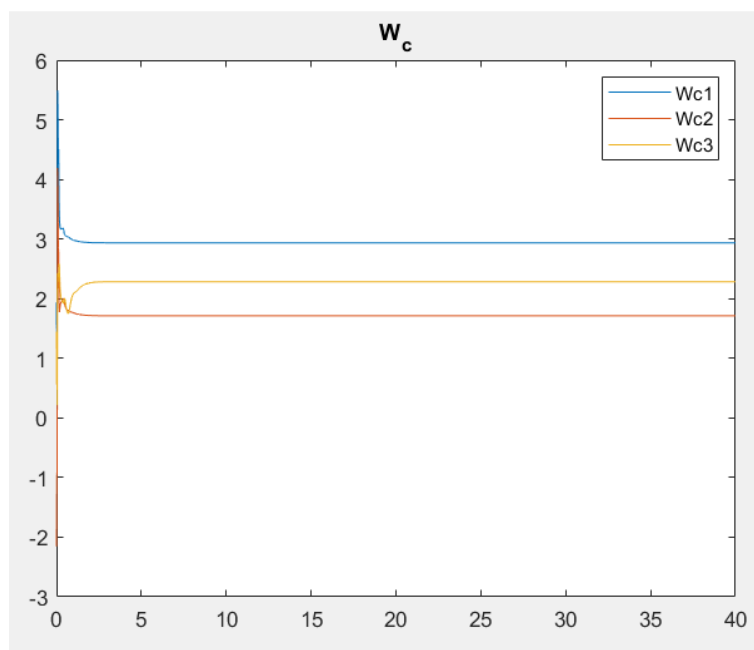
$$u = -\lambda \tanh \left[\frac{1}{2\lambda} R^{-1} g^T(x) \phi'(x)^T \hat{W}_a + e \right] - g^+(x) \hat{d}^* \quad (4.13)$$

Các đồ thị của toàn bộ thuật toán như sau

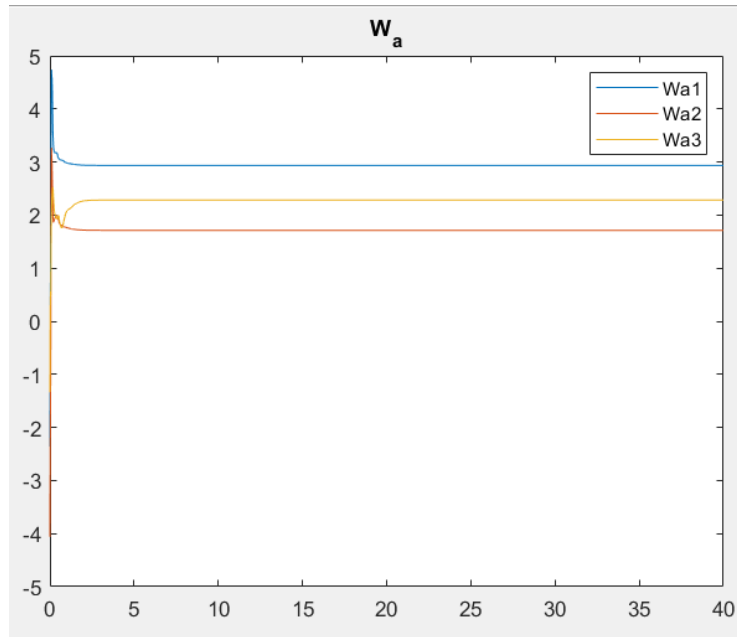


Hình 9. Sự hội tụ của trạng thái mô hình

Các trạng thái của hệ thống đều tiến về điểm cân bằng, nên hệ thống là ổn định. Tuy nhiên có thể nhận thấy trạng thái hệ thống không ổn định tiệm cận về gốc tọa độ do bộ quan sát không thể triệt tiêu hoàn toàn nhiễu đầu vào mà chỉ làm giảm ảnh hưởng của nhiễu

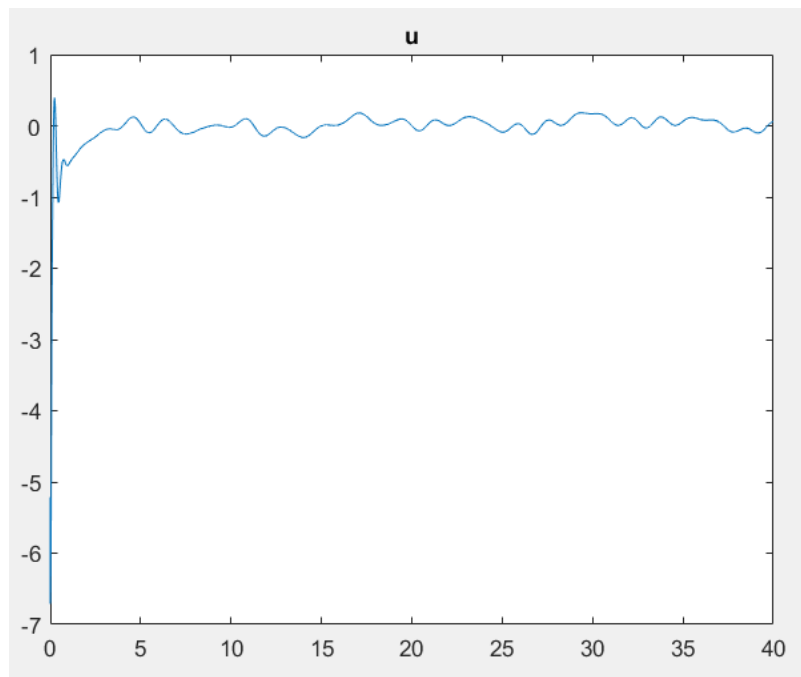


Hình 10. Sự hội tụ của trọng số thành phần critic



Hình 11. Sự hội tụ của trọng số thành phần actor

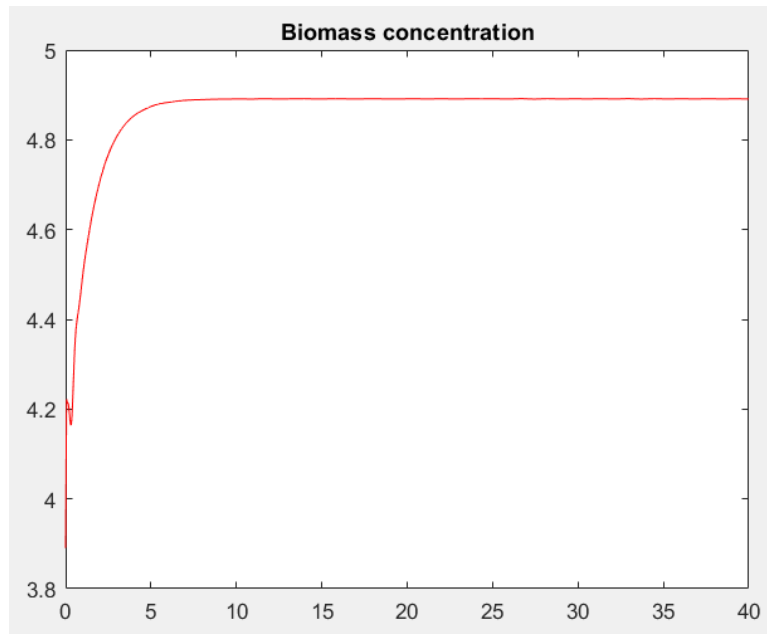
Khi ở trạng thái ổn định, các ma trận trọng số \widehat{W}_c và \widehat{W}_a cùng hội tụ khi t đủ lớn. Ngoài ra chúng cùng hội tụ về cùng 1 giá trị, do bản chất của chúng đều là giá trị xấp xỉ của ma trận W tối ưu.



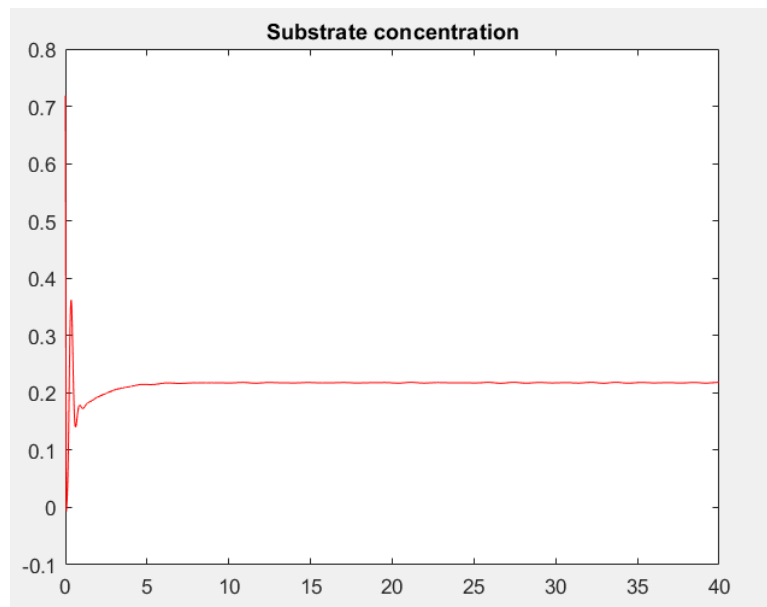
Hình 12. Tín hiệu điều khiển thực tế của thuật toán

Tín hiệu điều khiển luôn dao động xung quanh điểm cân bằng mà không thể ổn định do trong bộ điều khiển có thành phần bù nhiễu.

Chuyển mô hình (4.2) về dạng (4.1). Ta sẽ thu được đồ thị nồng độ của chất phản ứng và sản phẩm như sau



Hình 13. Đồ thị nồng độ của sinh khối



Hình 14. Đồ thị nồng độ của cơ chất

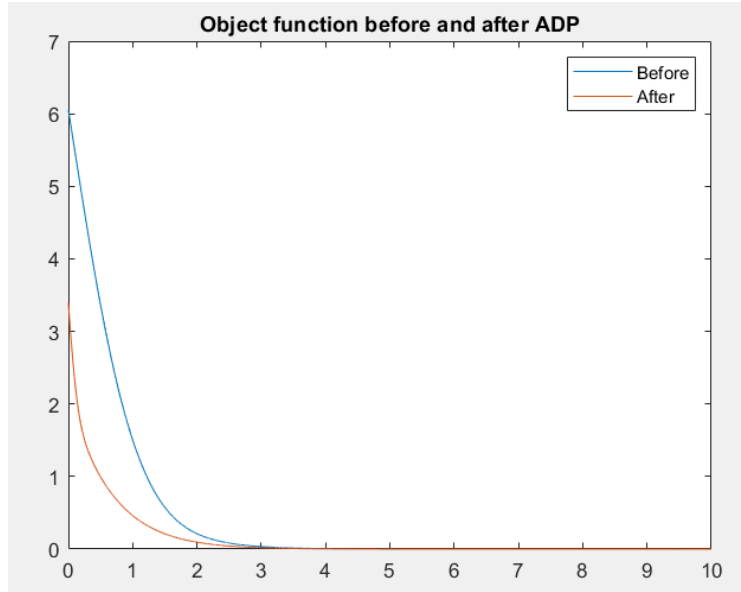
Cuối cùng, ta sẽ kiểm chứng tính tối ưu của thuật toán bằng cách so sánh hàm mục tiêu $J(t)$ khi trọng số bộ điều khiển là $\hat{W}_a(0)$ và $\hat{W}_a(T)$

Coi như mô hình không chịu ảnh hưởng của nhiễu. Hàm mục tiêu $J(t)$ là

$$J(t) = \int_0^t \left(x^T Q x + 2 \int_0^{\hat{v}} \lambda \tanh^{-1} \left(\frac{v}{\lambda} \right) R dv \right) d\tau \quad (4.14)$$

$$u = -\lambda \tanh \left[\frac{1}{2\lambda} R^{-1} g^T(x) \phi'(x)^T \hat{W}_a \right] \quad (4.15)$$

Với $\hat{W}_a(0) = [2.3642 \quad 2.2522 \quad 1.3138]^T$, $\hat{W}_a(T) = [2.936 \quad 1.7126 \quad 2.2839]^T$. Đồ thị hàm $J(t)$ trong 2 trường hợp sẽ như sau



Hình 15. Đồ thị hàm mục tiêu theo thời gian trước và sau cập nhật

Quá trình mô phỏng được thực hiện trên MATLAB. Tất cả các file đều được viết dưới dạng m-file.

CHƯƠNG 5. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN ĐỀ TÀI

Đồ án đã trình bày đầy đủ về phương pháp thiết kế bộ điều khiển tối ưu cho mô hình phi tuyến đơn giản về một phản ứng hóa học trong thiết bị CSTR. Các chỉ tiêu đầu ra về nhiễu, bất định, tín hiệu vào, tín hiệu ra và hàm mục tiêu đều đạt chất lượng tốt. Thuật toán được đưa ra đã được chứng minh sự hội tụ và tối ưu trong cả lý thuyết và chỉ ra bởi các mô phỏng. Ta thấy thuật toán không chỉ áp dụng với thiết bị CSTR dạng đơn giản, mà còn có thể áp dụng được với các hệ thống có mô hình phi tuyến dạng affine khác như con lắc ngược, xe cân bằng, lò nhiệt

Code mô phỏng kèm theo được viết dưới dạng rời rạc hóa nên cũng có khả năng áp dụng được cho hệ thống thực tế. Tuy nhiên thuật toán vẫn còn nhiều hạn chế, cần cải thiện và phát triển trong tương lai. Dưới đây là một số hướng phát triển đề xuất.

- *Hướng phát triển 1:* Mô hình được mô phỏng trong đồ án là có 2 biến. Nên mở rộng thuật toán cho các mô hình có bậc cao hơn. Vì hầu hết các hệ điện cơ hoặc robot thường có số bậc tự do lớn hơn 2.
- *Hướng phát triển 2:* Mô hình đang xét là mô hình có dạng affine. Nên cần mở rộng thuật toán cho các mô hình có dạng không thể chuyển về dạng affine. Hoặc các mô hình affine, nhưng bất định toàn phần. Không biết cả $f(x)$ và $g(x)$
- *Hướng phát triển 3:* Phương pháp gradient descent trong luật cập nhật có nhược điểm là nghiệm tối ưu có thể chỉ là nghiệm cục bộ. Ngoài ra tính hội tụ còn phụ thuộc vào các tham số được lựa chọn. Nên có thể chọn một thuật toán tối ưu khác tốt hơn để cải thiện tính hội tụ và đảm bảo hệ ổn định với cả những hệ phức tạp hơn.

TÀI LIỆU THAM KHẢO

- [1] N. D. Phước, Lý thuyết điều khiển tuyến tính, Hà Nội: NXB khoa học và kỹ thuật kỹ thuật, 2009.
- [2] N. D. Phước, Tối ưu hóa trong điều khiển và điều khiển tối ưu: NXB khoa học và kỹ thuật kỹ thuật, 2016.
- [3] H. Nhâm, Hóa học vô cơ cơ bản, lý thuyết đại cương về hóa học: NXB giáo dục Việt Nam, 2014.
- [4] S. Bhasin, R. Kamalapurkar, M. Johnson, K. G. Vamvoudakis, F. L. Lewis, and W. E. Dixon, “A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems,” *Automatica*, vol. 49, no. 1, pp. 82–92, Jan. 2013.
- [5] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Springer, 2017.