ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

THÁNG SINH VIÊN NGHIÊN CỨU KHOA HỌC và SÁNG TẠO 2025

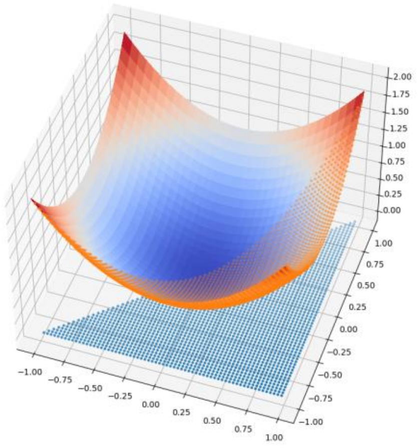
ĐẠI HỌC
BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY
OF SCIENCE AND TECHNOLOGY

THÁNG SINH VIÊN NGHIÊN CỨU KHOA HỌC và SÁNG TẠO 2025

Empowering Robotic Control with Artificial Intelligence: Optimizing Performance through Reinforcement Learning

(Talented Program in Intelligent Control and Automation)
Minh Hiep Phung, Khanh Tien Nguyen
Supervised by Assoc. Prof. Phuong Nam Dao

REINFORCEMENT LEARNING BASED OPTIMAL CONTROL



Optimal control

Optimal control theory is a branch of control theory that deals with finding a controller for a dynamical system over a period of time that an objective function is optimized

$$J(\bar{x}_0, \Delta u_k) = \sum_{k=0}^{\infty} \left(\bar{x}_k^T \bar{Q} \bar{x}_k + \Delta u_k^T R \Delta u_k \right)$$

$$J(X(t_0), u_{RL}) := \int_{t_0}^{\infty} r(X(\rho), u_{RL}(\rho)) d\rho$$

$$P - \bar{A}^T P \bar{A} + \bar{A}^T P \bar{B} \left(R + \bar{B}^T P \bar{B} \right)^{-1} \bar{B}^T P \bar{A} - \bar{Q} = 0$$

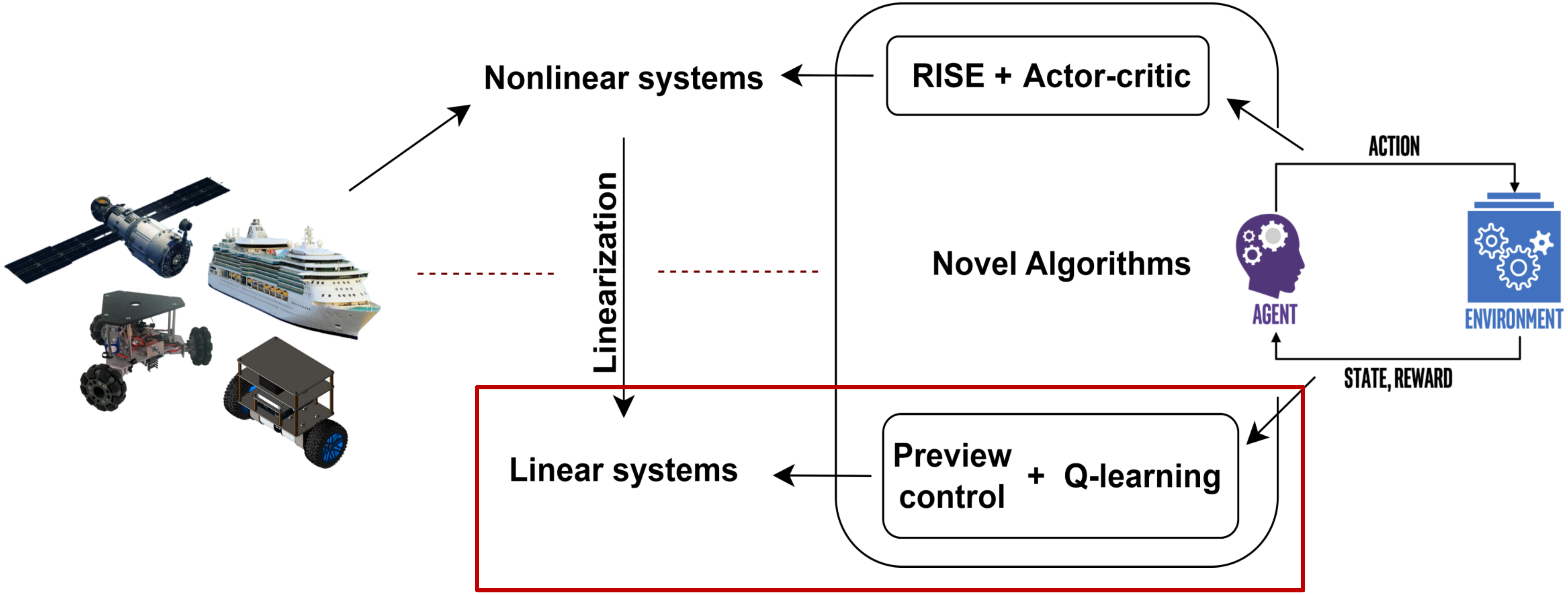
$$0 = Q + \nabla_X V^{*T} F - \frac{1}{4} \nabla_X V^{*T} G R^{-1} G^T \nabla_X V^*$$

Reinforcement learning

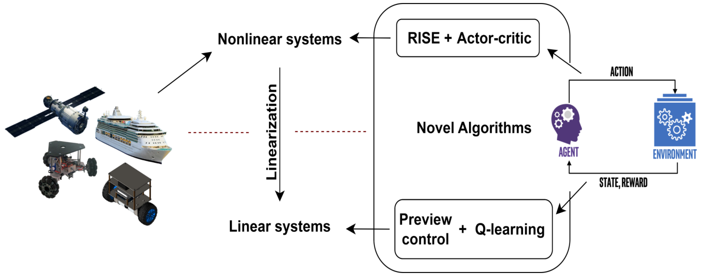
Reinforcement learning is the process where an agent learns to act by interacting with an environment to maximize long-term reward:

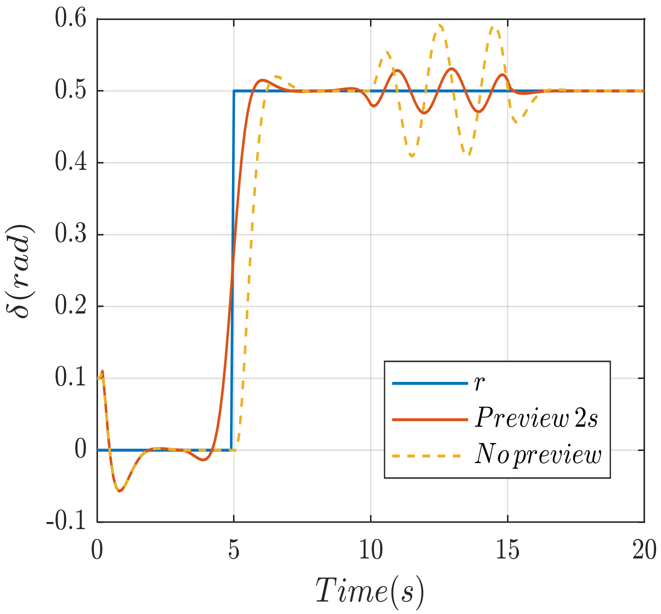
$$\pi^* = \arg\max_{\pi} \mathbb{E} \left[\sum_{t=0}^{\infty} \gamma^t r_t \right]$$

BACKGROUND OF OUR RESEARCH



ROBUST OPTIMAL PROBLEM FORMULATION





Preview control

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + Dd_k \\ y_k = Cx_k \end{cases}$$

$$\boxed{\tilde{x}_k = \begin{bmatrix} e_k^T & \Delta x_k^T \end{bmatrix}^T}$$

$$\boxed{x_k^r = [\Delta r_k^T, \Delta r_{k+1}^T, \dots, \Delta r_{k+M_r}^T]^T}$$

$$\boxed{x_k^d = [\Delta d_k^T, \Delta d_{k+1}^T, \dots, \Delta d_{k+M_d}^T]^T}$$

Preview Information

$$\boxed{\bar{x}_k = [\tilde{x}_k^T, (x_k^r)^T, (x_k^d)^T]^T}$$

$$\bar{x}_{k+1} = \underbrace{\begin{bmatrix} \tilde{A}_d & \tilde{G}^r & \tilde{G}^d \\ 0 & A^r & 0 \\ 0 & 0 & A^d \end{bmatrix}}_{\bar{A}} \bar{x}_k + \underbrace{\begin{bmatrix} \tilde{B}_d \\ 0 \\ 0 \end{bmatrix}}_{\bar{B}} \Delta u_k$$

Optimal Problem

The objective is to find the optimal policy to minimize the following infinite horizon cost:

$$J(\bar{x}_0, \Delta u_k) = \sum_{k=0}^{\infty} \left(\bar{x}_k^T \bar{Q} \bar{x}_k + \Delta u_k^T R \Delta u_k \right)$$

In order to find the optimal control solution, we need to the following equation for the value function:

$$P - \bar{A}^T P \bar{A} + \bar{A}^T P \bar{B} \left(R + \bar{B}^T P \bar{B} \right)^{-1} \bar{B}^T P \bar{A} - \bar{Q} = 0$$

After finding the optimal value function, it is used to deduce the optimal control function as follows:

$$K_d = -\left(R + \bar{B}^T P \bar{B} \right)^{-1} \bar{B}^T P \bar{A} = \begin{bmatrix} K^e & K^x & K^r & K^d \end{bmatrix}$$

$$\Delta u_k = -K^e e_k - K^x \Delta x_k - \sum_{i=0}^{M_r} k_i^r \Delta r_{k+i} - \sum_{i=0}^{M_d} k_i^d \Delta d_{k+i}$$

NON-AUTONOMOUS SYSTEMS

Discrete-time linear periodic systems

$$\mathbf{x}_{k+1} = \mathbf{A}_k\mathbf{x}_k + \mathbf{B}_k\mathbf{u}_k + \mathbf{D}_k\mathbf{d}_k$$

$$\mathbf{J}^* = \lim_{p\to\infty}\left(min\frac{1}{2}\mathbf{x}_p^T\mathbf{Q}_p\mathbf{x}_p + \frac{1}{2}\sum_{k=0}^{p-1}\mathbf{x}_k^T\mathbf{Q}_k\mathbf{x}_k + \mathbf{u}_k^T\mathbf{R}_k\mathbf{x}_k \right)$$

$$\mathbf{P}_k = \mathbf{Q}_k + \mathbf{A}_k^T\mathbf{P}_{k+1}\mathbf{A}_k - \mathbf{A}_k^T\mathbf{P}_{k+1}\mathbf{B}_k(\mathbf{R}_k + \mathbf{B}_k^T\mathbf{P}_{k+1}\mathbf{B}_k)^{-1}\mathbf{B}_k^T\mathbf{P}_{k+1}\mathbf{A}_k$$

Preview control for DTLP systems

$$\mathbf{X}_{k+1} = \mathbf{\Psi}_k\mathbf{X}_k + \tilde{\mathbf{B}}_k\Delta\mathbf{u}_k$$

$$\begin{cases} \mathbf{X}_{Nk+1} &= \mathbf{\Psi}_0\mathbf{X}_{Nk} + \mathbf{\Lambda}_0\mathbf{U}_{Nk}, \\ \mathbf{X}_{Nk+2} &= \mathbf{\Psi}_1\mathbf{X}_{Nk+1} + \mathbf{\Lambda}_1\mathbf{U}_{Nk+1}, \\ \dots & \\ \mathbf{X}_{Nk+N} &= \mathbf{\Psi}_{N-1}\mathbf{X}_{Nk+N-1} + \mathbf{\Lambda}_{N-1}\mathbf{U}_{Nk+N-1}, \end{cases}$$

$$\mathbf{J} = \sum_{K=0}^{\infty}[\bar{\mathbf{X}}_K^T\bar{\mathbf{Q}}^n\bar{\mathbf{X}}_K + \bar{\mathbf{U}}_K^T\bar{\mathbf{R}}^n\bar{\mathbf{U}}_K]$$

$$\Delta\mathbf{u}_k = -\tilde{\mathbf{K}}\mathbf{X}_k = -\mathbf{K}_k^x\Delta\mathbf{x}_k - \sum_{i=0}^{a+N}\mathbf{K}_k^{d,i}\Delta\mathbf{d}_{k+i}$$

System Lifting

$$\bar{\mathbf{X}}_K := \begin{bmatrix} \mathbf{X}_{Nk+1} \\ \mathbf{X}_{Nk+2} \\ \vdots \\ \mathbf{X}_{Nk+N} \end{bmatrix} = \left[\begin{array}{ccc|c} \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Psi}_0 \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Psi}_1\mathbf{\Psi}_0 \\ \vdots & \vdots & \vdots & \vdots \\ \hline \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Psi}_{N-1}\dots\mathbf{\Psi}_1\mathbf{\Psi}_0 \end{array}\right]\begin{bmatrix} \mathbf{X}_{N(k-1)+1} \\ \mathbf{X}_{N(k-1)+2} \\ \vdots \\ \mathbf{X}_{N(k-1)+N} \end{bmatrix}$$

$$+\left[\begin{array}{cccc} \mathbf{\Lambda}_0 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{\Psi}_1\mathbf{\Lambda}_0 & \mathbf{\Lambda}_1 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots \\ \hline \mathbf{\Psi}_{N-1}\dots\mathbf{\Psi}_1\mathbf{\Lambda}_0 & \mathbf{\Psi}_{N-1}\dots\mathbf{\Psi}_2\mathbf{\Lambda}_1 & \dots & \mathbf{\Lambda}_{N-1} \end{array}\right]\begin{bmatrix} \mathbf{U}_{Nk} \\ \mathbf{U}_{Nk+1} \\ \vdots \\ \mathbf{U}_{Nk+N-1} \end{bmatrix}$$

$$:= \left[\begin{array}{c|c} \mathbf{0} & \bar{\mathbf{\Psi}}_1 \\ \hline \mathbf{0} & \bar{\mathbf{\Psi}}_2 \end{array}\right]\bar{\mathbf{X}}_K + \left[\begin{array}{c} \bar{\mathbf{\Lambda}}_1 \\ \hline \bar{\mathbf{\Lambda}}_2 \end{array}\right]\bar{\mathbf{U}}_K = \bar{\mathbf{\Psi}}\bar{\mathbf{X}}_K + \bar{\mathbf{\Lambda}}\bar{\mathbf{U}}_K$$

Modified Riccati Equation

$$\bar{\mathbf{\Psi}}_2^T\bar{\mathbf{P}}_{22}^n\bar{\mathbf{\Psi}}_2 - \bar{\mathbf{P}}_{22}^n - (\bar{\mathbf{\Psi}}_2^T\bar{\mathbf{P}}_{22}^n\bar{\mathbf{\Lambda}}_2 + \bar{\mathbf{\Psi}}_1^T\bar{\mathbf{Q}}_1^n\bar{\mathbf{\Lambda}}_1)(\bar{\mathbf{\Lambda}}_2^T\bar{\mathbf{P}}_{22}^n\bar{\mathbf{\Lambda}}_2 + \bar{\mathbf{R}}^n + \bar{\mathbf{\Lambda}}_1^T\bar{\mathbf{Q}}_1^n\bar{\mathbf{\Lambda}}_1)^{-1}$$
$$(\bar{\mathbf{\Lambda}}_2^T\bar{\mathbf{P}}_{22}^n\bar{\mathbf{\Psi}}_2 + \bar{\mathbf{\Lambda}}_1^T\bar{\mathbf{Q}}_1^n\bar{\mathbf{\Psi}}_1) + \bar{\mathbf{Q}}_2^n + \bar{\mathbf{\Psi}}_1^T\bar{\mathbf{Q}}_1^n\bar{\mathbf{\Psi}}_1 = \mathbf{0}$$

Theorem of Equivalent Model

$$\begin{cases} \mathbf{X}_{k+1} = \mathbf{\Psi}_k\mathbf{X}_k + \mathbf{\Lambda}_k\mathbf{U}_k \\ \mathbf{y}_k = \mathbf{C}\mathbf{X}_k \end{cases}$$

$$\begin{cases} \begin{bmatrix} \Delta\mathbf{x}_{k+1} \\ \mathbf{X}_{k+1}^d \end{bmatrix} = \begin{bmatrix} \mathbf{A}_k & \mathbf{G}_k \\ \mathbf{0} & \mathbf{A}^d \end{bmatrix}\begin{bmatrix} \Delta\mathbf{x}_k \\ \mathbf{X}_k^d \end{bmatrix} + \begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \varepsilon \end{bmatrix}\begin{bmatrix} \Delta\mathbf{u}_k \\ \mathbf{u}_k^i \end{bmatrix}, \\ \mathbf{y}_k = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \Delta\mathbf{x}_k \\ \mathbf{X}_k^d \end{bmatrix} \end{cases}$$

Proof:

Output-Based Approximation Model

$$\mathbf{y}_k = \Delta\mathbf{x}_k^{new} = \mathbf{A}_{k-1}\Delta\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{X}_{k-1}^d + \mathbf{B}_{k-1}\Delta\mathbf{u}_{k-1}$$
$$= \mathbf{A}_{k-1}\left(\mathbf{A}_{k-2}\Delta\mathbf{x}_{k-2} + \mathbf{G}_{k-2}\mathbf{X}_{k-2}^d + \mathbf{B}_{k-1}\Delta\mathbf{u}_{k-1}\right) + \mathbf{G}_{k-1}\left(\mathbf{X}_{k-2}^d + \varepsilon\mathbf{u}_{k-2}^{im}\right) + \mathbf{B}_{k-1}\Delta\mathbf{u}_{k-1}$$
$$= \Delta\mathbf{x}_k^{old} + \left(\mathbf{G}_{k-1}\varepsilon\sum_{i=1}^{k-2}\mathbf{u}_i^{im}\right) + \prod_{i-1}^{k-2}\left(\left(\sum_{z=i}^{k-1}\mathbf{A}_{k-1+i-z}\right)\mathbf{G}_i\varepsilon\prod_{j=1}^{i}\mathbf{u}_j^{im}\right) = \Delta\mathbf{x}_k^{old} + \boldsymbol{\kappa}(\varepsilon)$$

Riccati Equation for the Approximation Model

$$\mathbf{Q}_k^n + \mathbf{\Psi}_k^T\mathbf{P}_{k+1}^n\mathbf{\Psi}_k - \mathbf{\Psi}_k^T\mathbf{P}_{k+1}^n\mathbf{\Lambda}_k\left(\mathbf{R}_k^n + \mathbf{\Lambda}_k^T\mathbf{P}_{k+1}^n\mathbf{\Lambda}_k\right)^{-1}\mathbf{\Lambda}_k^T\mathbf{P}_{k+1}^n\mathbf{\Psi}_k - \mathbf{P}_k^n$$
$$\approx \mathbf{Q}_k^n + \mathbf{\Psi}_k^T\mathbf{P}_{k+1}^n\mathbf{\Psi}_k - \mathbf{\Psi}_k^T\mathbf{P}_{k+1}^n\begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\left(\mathbf{R}_k^n + \begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^T\mathbf{P}_{k+1}^n\begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right)^{-1}$$
$$\times\begin{bmatrix} \mathbf{B}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}^T\mathbf{P}_{k+1}^n\mathbf{\Psi}_k - \mathbf{P}_k^n$$

$$\begin{cases} \bar{\mathbf{U}}_K^{j'}(\bar{\mathbf{X}}_K) \\ \bar{\mathbf{X}}_K = \bar{\mathbf{\Psi}}\bar{\mathbf{X}}_K + \bar{\mathbf{\Lambda}}(\bar{\mathbf{U}}_K + \epsilon_K^U) \end{cases}$$

$$\begin{cases} \Delta\bar{\mathbf{u}}_K^{j'}(\bar{\mathbf{X}}_K) \\ \bar{\mathbf{X}}_K = \bar{\mathbf{\Psi}}\bar{\mathbf{X}}_K + \bar{\bar{\mathbf{B}}}(\Delta\bar{\mathbf{u}}_K^j + \epsilon_K^{\Delta u}) + \begin{bmatrix} \mathbf{0} \\ \bar{\varepsilon} \end{bmatrix}(\bar{\mathbf{u}}_K^i + \epsilon_K^{ui}) \end{cases}$$

$$\begin{cases} \Delta\bar{\mathbf{u}}_K^{j'}(\bar{\mathbf{X}}_K) \\ \bar{\mathbf{X}}_K = \bar{\mathbf{\Psi}}\bar{\mathbf{X}}_K + \bar{\bar{\mathbf{B}}}(\Delta\bar{\mathbf{u}}_K + \epsilon_K^{\Delta u}) + \begin{bmatrix} \mathbf{0} \\ \epsilon_{X^d} \end{bmatrix} \end{cases}$$

$$\begin{cases} \Delta\bar{\mathbf{u}}_K^{j'}(\bar{\mathbf{X}}_K) \\ \bar{\mathbf{X}}_K = \bar{\mathbf{\Psi}}\bar{\mathbf{X}}_K + \bar{\bar{\mathbf{B}}}(\Delta\bar{\mathbf{u}}_K + \epsilon_K^{\Delta u}) + \epsilon_K^X \end{cases}$$

REINFORCEMENT LEARNING FOR LINEAR SYSTEMS

THÁNG SINH VIÊN NGHIÊN CỨU KHOA HỌC và SÁNG TẠO 2025

Q-function: $Q(\bar{x}_k, \Delta u_k) = r(\bar{x}_k, \Delta u_k) + V(\bar{x}_{k+1}) = \bar{x}_k^T \bar{Q} \bar{x}_k + \phi_k^T H \phi_k$ where $\phi_k = \begin{bmatrix} \bar{x}_k^T & \Delta u_k^T \end{bmatrix}^T$

Algorithm 1 On-policy Q learning

1: **Initialization:** Start with a stabilizing policy $\Delta u_k = -K^{(0)} \bar{x}_k$ and apply it to the augmented state $\bar{x}_k$. Add an exploration noise $\epsilon_k^{\Delta u}$ to the control input and a noise ϵ_k^x to the state variable.

2: **Policy Evaluation:** Use the collected data to construct the linear equation:
$Z \cdot \mathrm{vecs}(H) = Y$

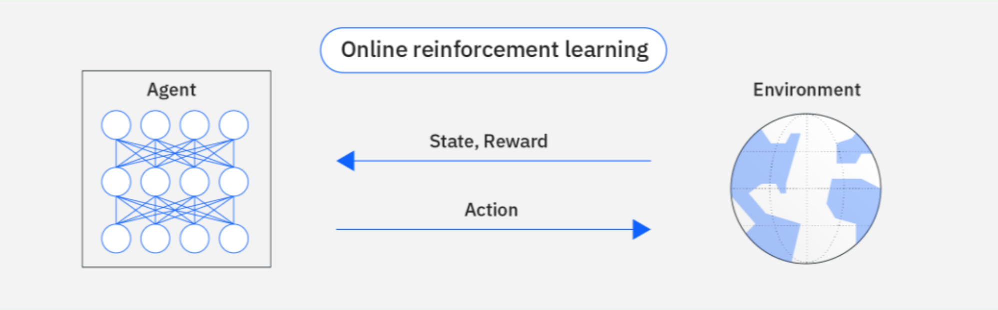
3: **Policy Update:** Update the policy as:
$K^{(j+1)} = (H_{22})^{-1} H_{21}$

4: **Checking:** If $\|K^{(j+1)} - K^{(j)}\| \leq \sigma$, then stop the iteration; otherwise, set $j \leftarrow j+1$ and repeat the process.

$$Z = \begin{bmatrix} vecv(\phi_k)^T - vecv(\phi_{k+1})^T \\ vecv(\phi_{k+1})^T - vecv(\phi_{k+2})^T \\ \vdots \\ vecv(\phi_{k+s-1})^T - vecv(\phi_{k+s})^T \end{bmatrix}, Y = \begin{bmatrix} \rho_k \\ \rho_{k+1} \\ \vdots \\ \rho_{k+s-1} \end{bmatrix}$$

$$\rho_k = r(\bar{x}_k, \Delta u_k) + (\bar{x}_{k+1})^T \bar{Q} \bar{x}_{k+1} - (\bar{x}_k)^T \bar{Q} \bar{x}_k$$

$$\boxed{\begin{cases} \bar{x}_{k+1} = \begin{bmatrix} \tilde{A} & \tilde{G}^r & \tilde{G}^d \\ 0 & A^r & 0 \\ 0 & 0 & A^d \end{bmatrix} \bar{x}_k + \begin{bmatrix} \tilde{B} & 0 \\ 0 & \varepsilon I \\ 0 & \varepsilon I \end{bmatrix} \begin{bmatrix} \Delta u_k \\ u_k^i \end{bmatrix} \\ \bar{y}_k = \begin{bmatrix} I & 0 \end{bmatrix} \bar{x}_k. \end{cases}}$$



Algorithm 2 Off-policy Q learning

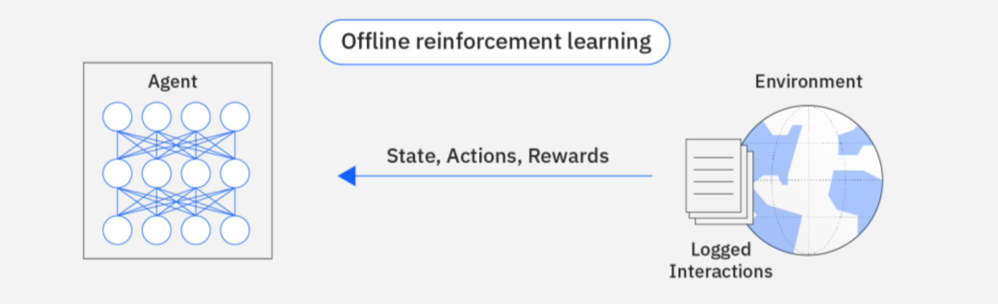
1: **Data collection:** Collect $s \geq (p+n+p(M_r+1)+q(M_d+1)+m) \times (p+n+p(M_r+1)+q(M_d+1)+m+1)/2$ data sets of system data $\bar{x}$ and store them in the sample sets Z^j and Y^j by using a stabilizing behavior control policy $\Delta u = -K\bar{x}_k + \epsilon_k^{\Delta u}$ in the system $\bar{x}_{k+1} = \bar{A}\bar{x}_k + \bar{B}\Delta u_k + \epsilon_k^x$
Set j = 0.

2: **Implementing Q-learning:** Solve the equation (26) by least-squares method using the data in Step 1 and update the target policy gain in term of $\bar{K}_{12}^{j+1} = (H_{22}^{j+1})^{-1}(H_{12}^{j+1})^T$.

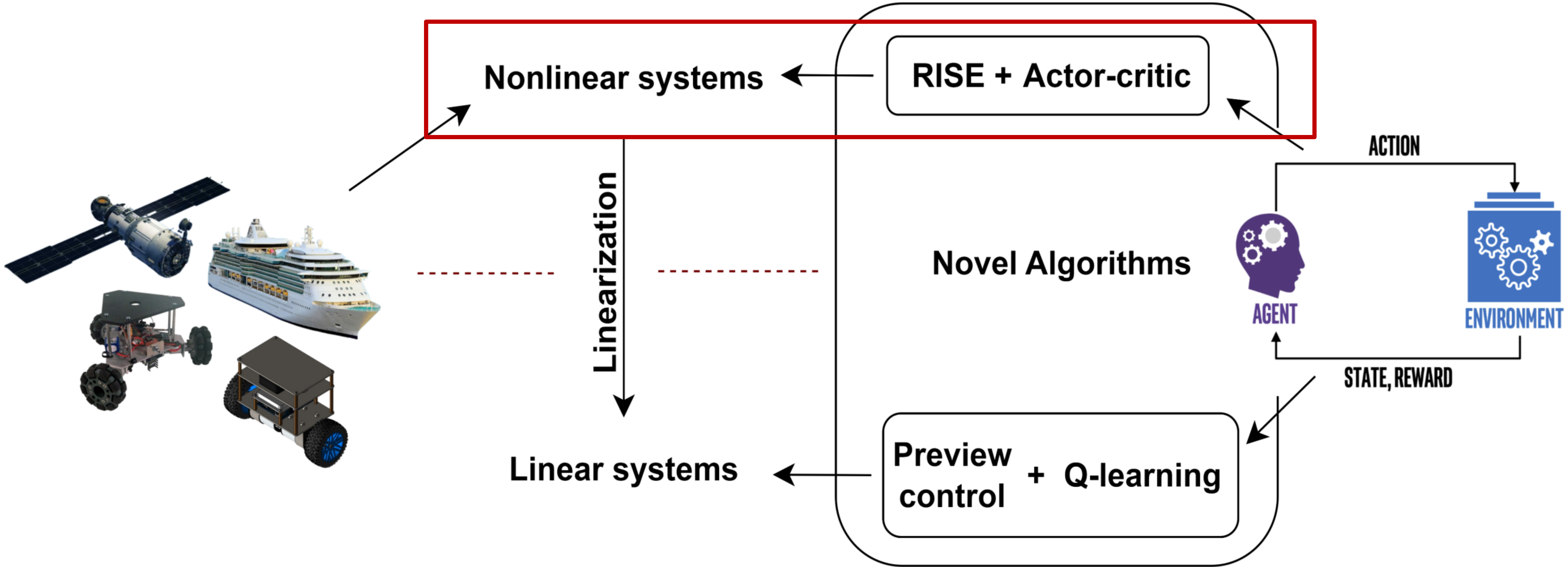
$$Z^j \begin{bmatrix} vecs(H_{11}^{j+1}) \\ vec(H_{12}^{j+1}) \\ vecs(H_{22}^{j+1}) \end{bmatrix} = Y^j \qquad (26)$$

3: **Checking:** If $\|K^{(j+1)} - K^{(j)}\| \leq \sigma$, then stop the iteration; otherwise, set $j \leftarrow j+1$ and repeat the process.

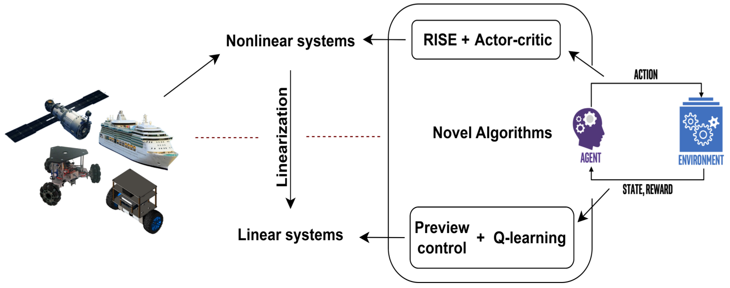
$$\Gamma_{(11)K}^j = vecv(\bar{x}_k^2)^T - vecv(\bar{x}_{k+1}^2)^T$$
$$\Gamma_{(12)K}^j = 2\Delta u_k^T \otimes (\bar{x}_{k+1}^2)^T + 2(\bar{K}_{12}^j \bar{x}_{k+1}^2)^T \otimes \bar{x}_{k+1}^2$$
$$\Gamma_{(22)K}^j = vecv(\Delta u_k)^T - vecv(\bar{K}_{12}^j \bar{x}_{k+1}^2)^T$$
$$\Gamma_k^j = \begin{bmatrix} (\Gamma_{(11)K}^j & \Gamma_{(12)K}^j & \Gamma_{(22)K}^j \end{bmatrix}$$
$$\rho_k^j = (\bar{x}_k^2)^T \bar{Q} \bar{x}_k^2 + \Delta u_k^T R \Delta u_k + (\bar{x}_{k+1}^1)^T \bar{Q}_1^+ \bar{x}_{k+1}^1,$$
$$Z^j = \begin{bmatrix} (\Gamma_k^j)^T & (\Gamma_{K+1}^j)^T & \dots & (\Gamma_{K+s-1}^j)^T \end{bmatrix}^T$$
$$Y^j = \begin{bmatrix} \rho_k^j & \rho_{K+1}^j & \dots & \rho_{K+s-1}^j \end{bmatrix},$$

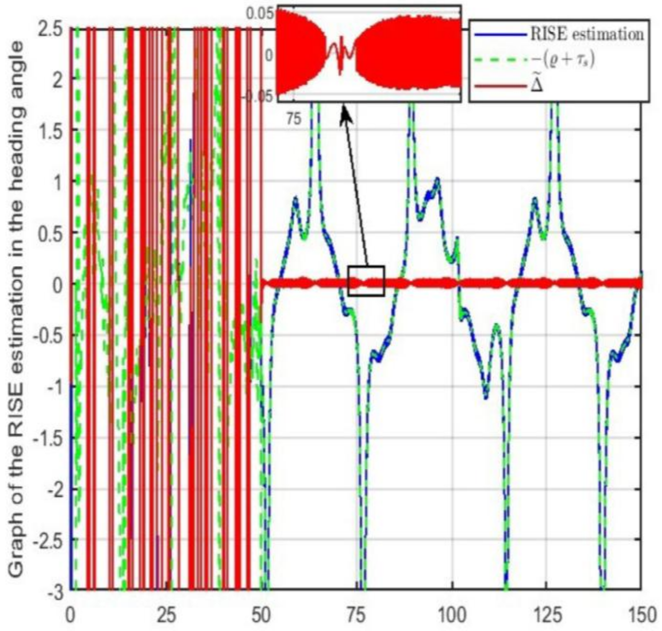


BACKGROUND OF OUR RESEARCH



ROBUST OPTIMAL PROBLEM FORMULATION





Robust Integral Sign Error

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

$$\boxed{e_1 := x - x_d}$$

$$\boxed{e_2 := \dot{e}_1 + \alpha_1 e_1}$$

$$\boxed{X := [e_1^T, e_2^T, x_d^T]^T}$$

Known model for learning

RISE for handling unknown factors

$$\dot{x} = \boxed{F(X) + G(X)u_{RL}} + \boxed{\widetilde{\Delta}(X, u_{RISE}, d(t))}$$

$$u_{RISE} := \lambda e_2 + \int_{t_0}^{t} \left(\lambda\alpha_2 e_2(\tau) + \beta_1 \mathrm{sgn}(e_2(\tau))\right) d\tau$$

Optimal Problem

The objective is to find the optimal policy to minimize the following infinite horizon cost:

$$J(X(t_0), u_{RL}) := \int_{t_0}^{\infty} r(X(\rho), u_{RL}(\rho)) d\rho$$

$$\text{subject to: } \dot{x} = F(X) + G(X)u_{RL},$$

In order to find the optimal control solution, we need to the following equation for the value function:

$$0 = Q + \nabla_X V^{*T} F - \frac{1}{4} \nabla_X V^{*T} G R^{-1} G^T \nabla_X V^*$$

After finding the optimal value function, it is used to deduce the optimal control function as follows:

$$u_{RL}^* := -\tfrac{1}{2} R^{-1} G^T(X) \nabla_X V^*(X)$$

ĐẠI HỌC BÁCH KHOA HÀ NỘI

REINFORCEMENT LEARNING FOR NONLINEAR SYSTEMS

The optimal value function and optimal control policy are approximated as follows:

$$\widehat{V}(X, W_c) = W_c^T \phi(X),$$
$$\widehat{u}_{RL}(X, W_a) = -\frac{1}{2} R^{-1} G^T(X) \nabla_X \phi^T(X) W_a,$$

HJB equation

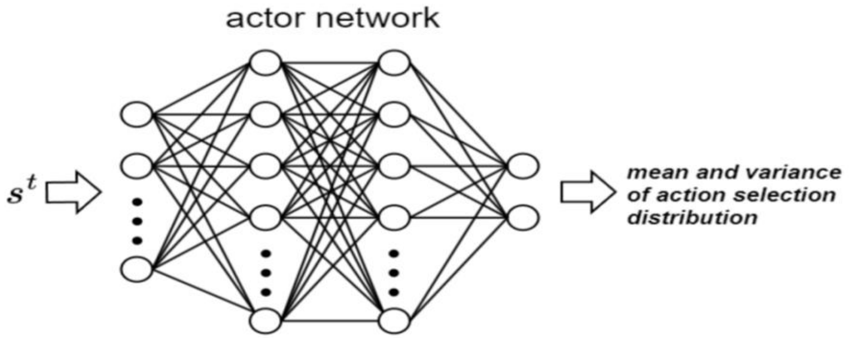
$$\delta_{HJB}(X, W_a, W_c) = W_c^T \sigma(X, W_a) + r(X, \widehat{u}_{RL}(X, W_a))$$
$$\sigma(X, W_a) := \nabla_X \phi(X)(F(X) + G(X)\widehat{u}_{RL}(X, W_a))$$

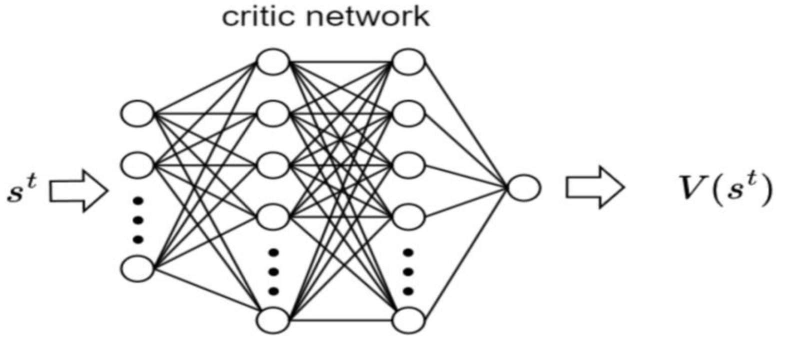
The learning law of Critic weights to minimize $E_c := \int_{t_0}^{t} \delta_{HJB}^2 d\tau$ **leveraging least-square method:**

$$\dot{W}_c := -\eta_c \Gamma \frac{\sigma}{1 + \nu \sigma^T \Gamma \sigma} \delta_{HJB}, \quad \dot{\Gamma} := -\eta_c \left(-\beta_2 \Gamma + \Gamma \frac{\sigma \sigma^T}{1 + \nu \sigma^T \Gamma \sigma} \right),$$

A gradient update law is developed for Actor to minimize the square Bellman error $E_a := \delta_{HJB}^2$ **as :**

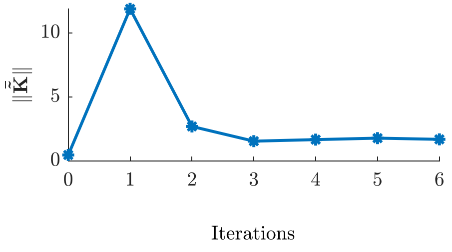
$$\dot{W}_a := -\eta_{a1} \frac{1}{\sqrt{1 + \sigma^T \sigma}} \nabla_X \phi G R^{-1} G^T \nabla_X \phi^T (W_a - W_c) \times \delta_{HJB} - \eta_2 (W_a - W_c),$$



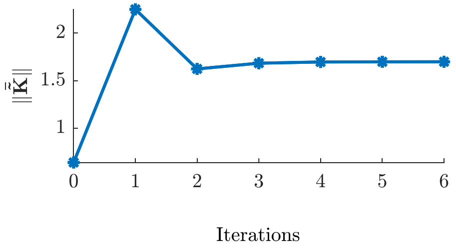


SIMPLE SYSTEM

Convergence of K

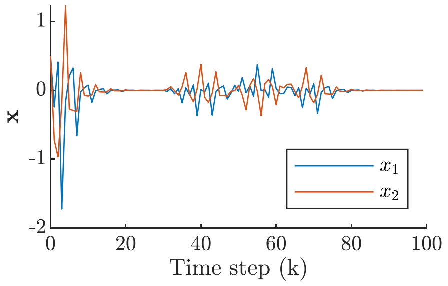


(a) On-policy

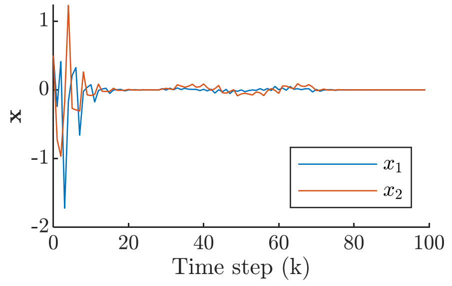


(b) Off-policy

Classical Q-learning vs. Improved Q-learning

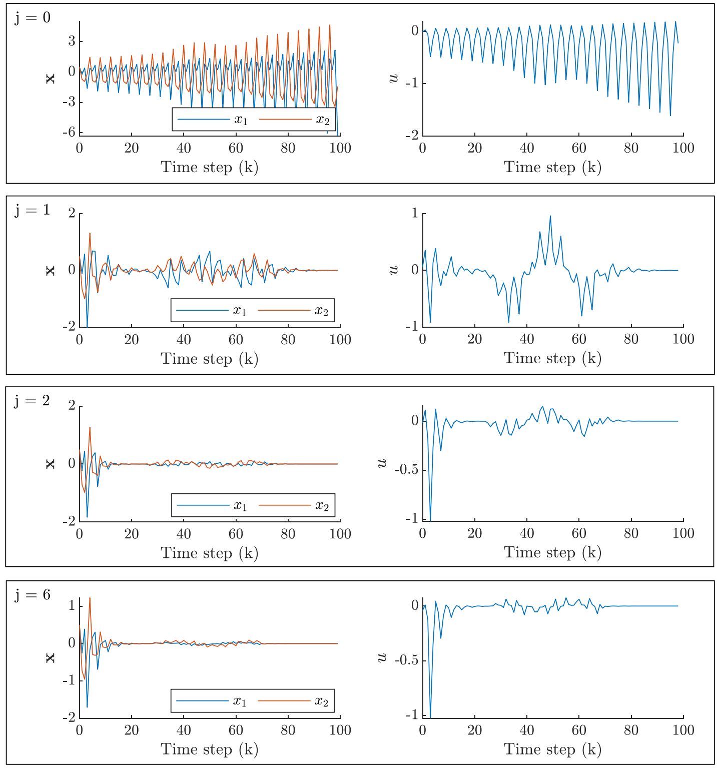


(a)

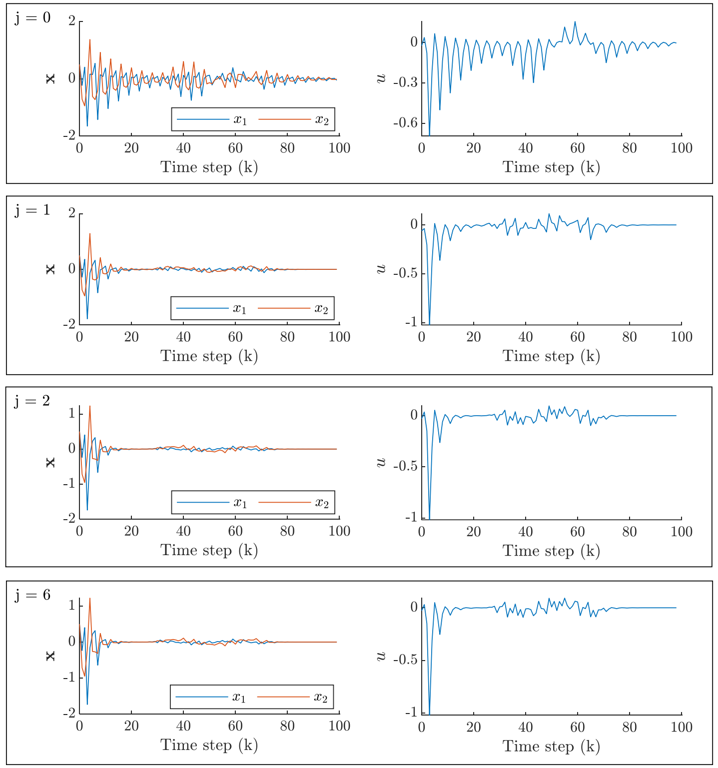


(b)

On-policy Q-learning

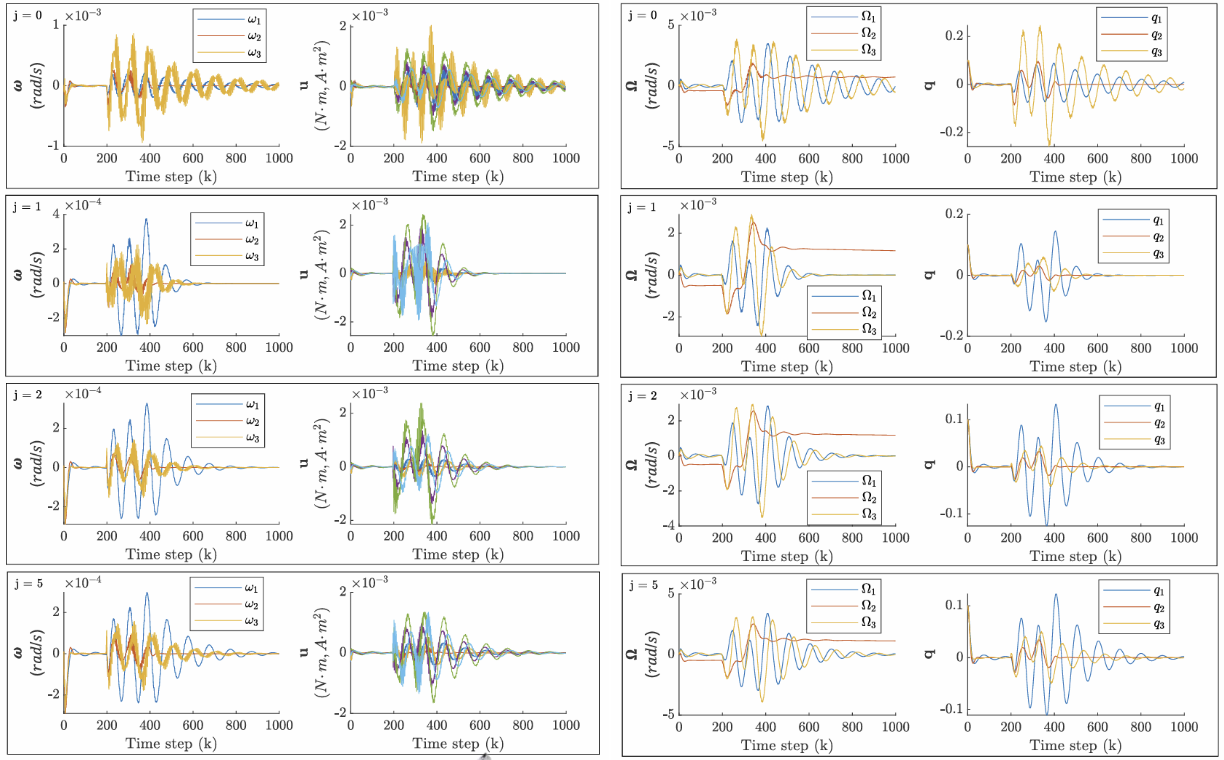


Off-policy Q-learning

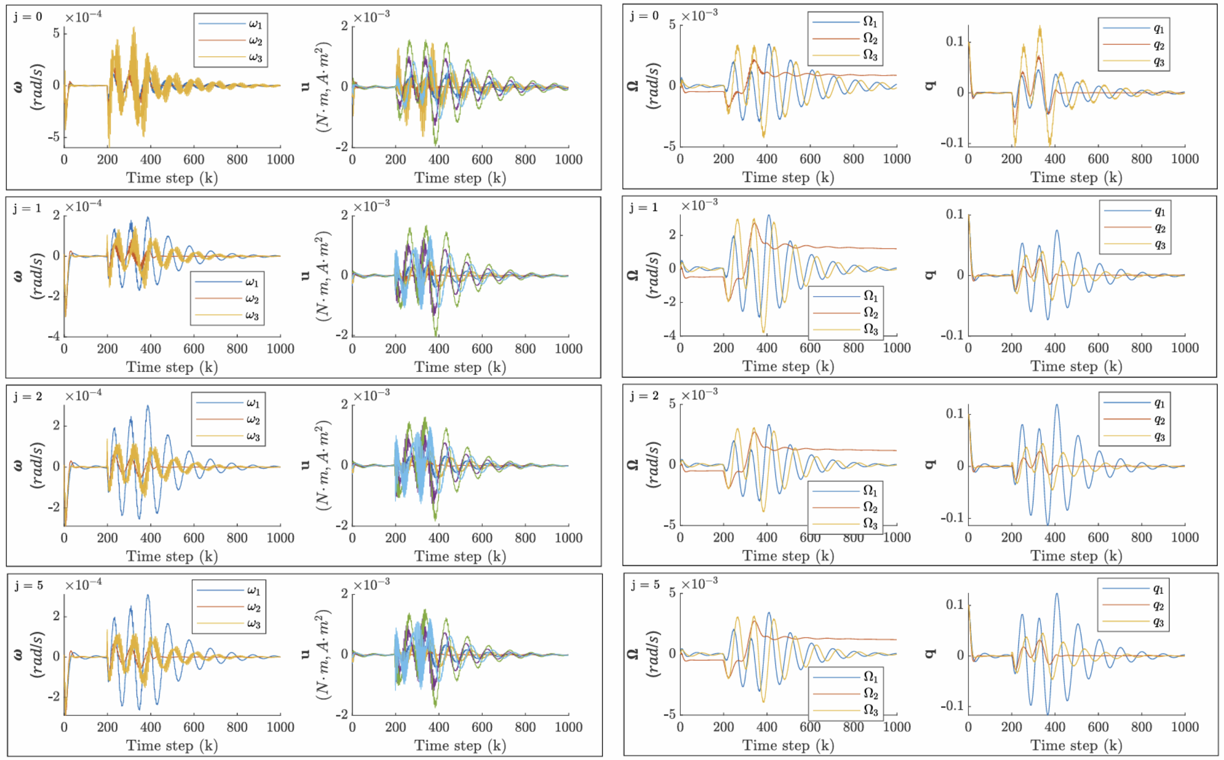


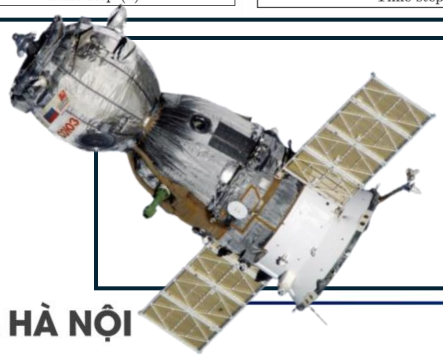
SPACECRAFT (NON-AUTONOMOUS SYSTEM)

On-policy Q-learning

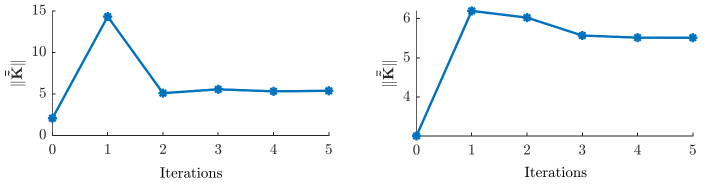


Off-policy Q-learning





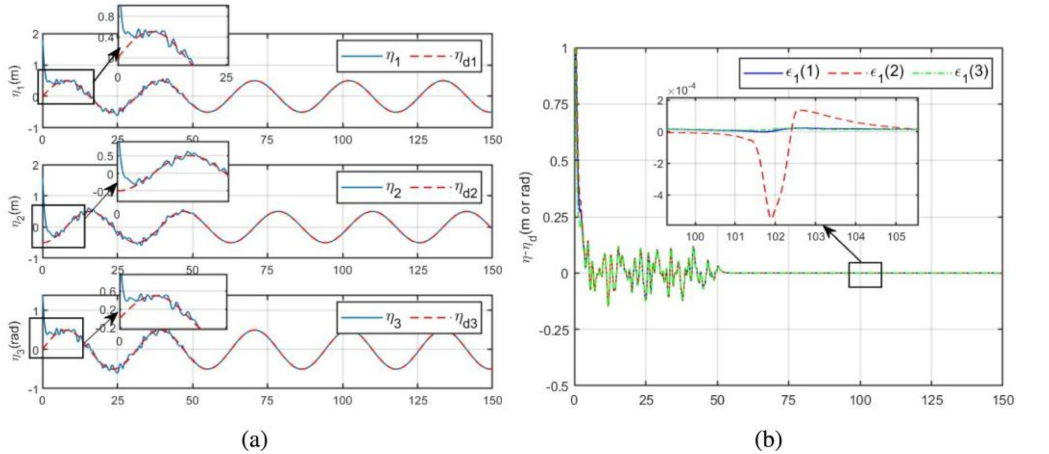
Convergence of K



(a) On-policy (b) Off-policy

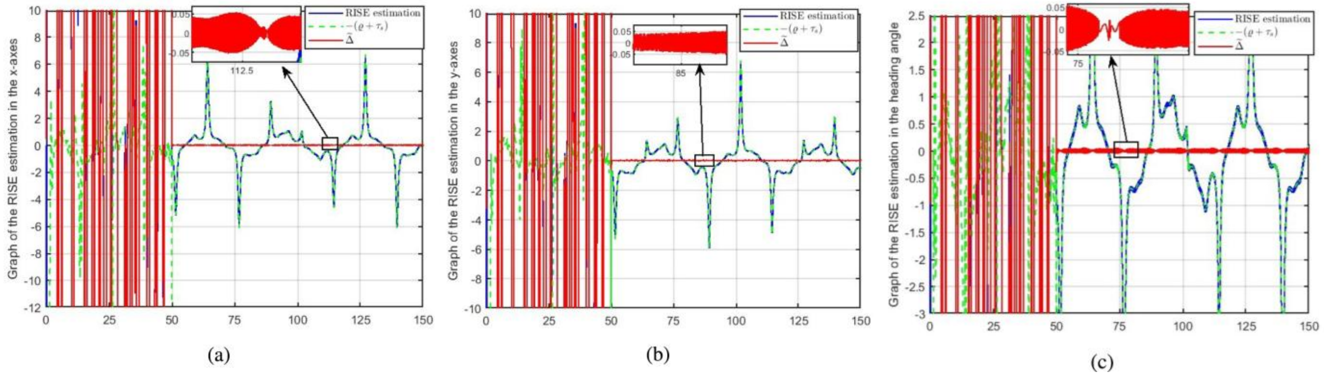
THREE-WHEELED MOBILE ROBOT WITH MECANUM WHEELS

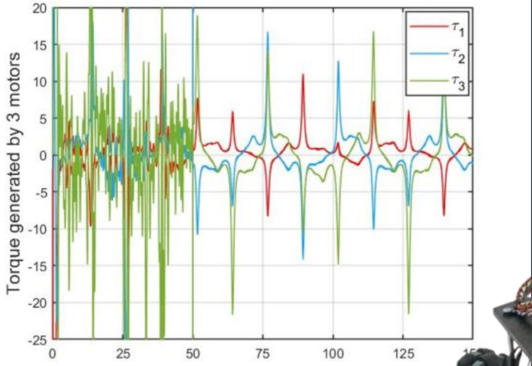
THÁNG SINH VIÊN NGHIÊN CỨU và KHOA HỌC SÁNG TẠO 2025

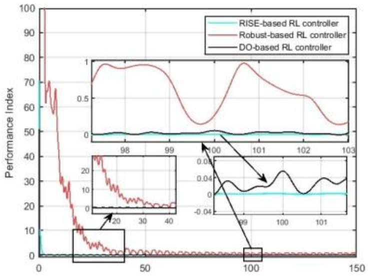


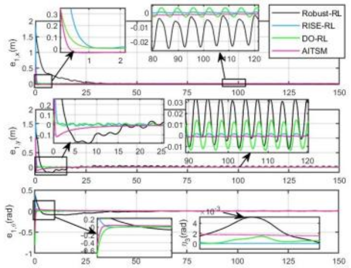
Tracking results and errors

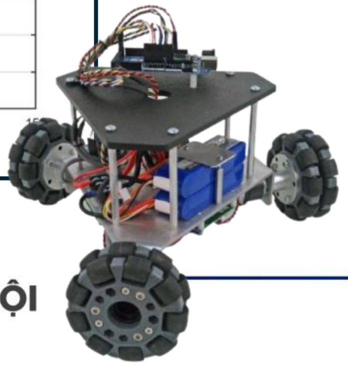
Disturbance under the compensation of the RISE controller

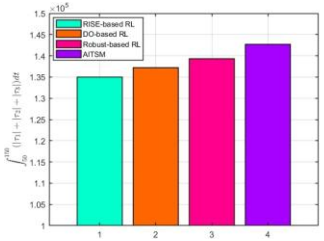


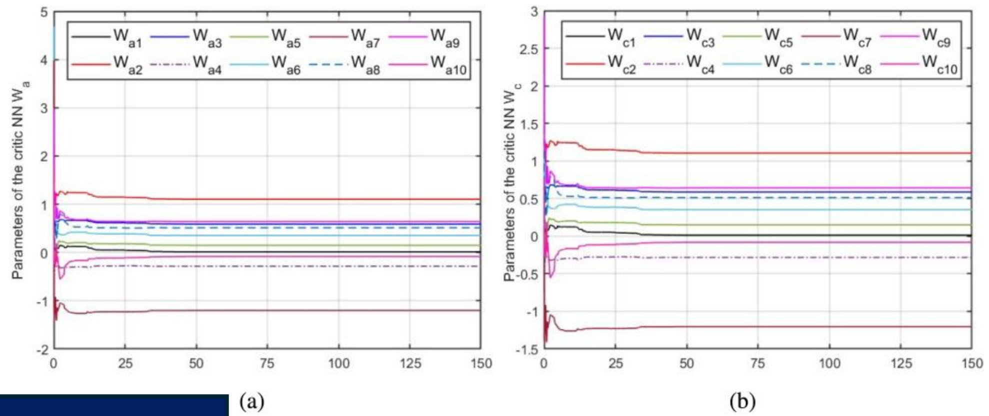












Convergence of W


ĐẠI HỌC BÁCH KHOA HÀ NỘI
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY
THÁNG SINH VIÊN
NGHIÊN CỨU
KHOA HỌC
và
SÁNG TẠO
2025

THANK YOU