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

School of Electrical Engineering

Robust Optimal Control for Nonlinear Systems Based on Adaptive Reinforcement Learning

Le Cong Nhat Anh
Nguyen Xuan Khai
Supervisor: Dao Phuong Nam, Ph.D.

July 17, 2021

ONE LOVE. ONE FUTURE.

Contents

Tracking control

- Nonlinear systems with disturbances and uncertainties
- Robust adaptive approach
- Traditional optimal approach

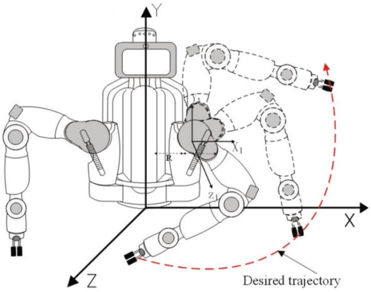


Figure 1: Tracking control

Tracking control

- Nonlinear systems with disturbances and uncertainties
- Robust adaptive approach
- Traditional optimal approach

Reinforcement learning (RL)

- Multi-purpose
- Multi-component

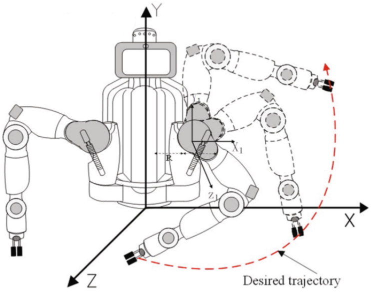


Figure 1: Tracking control

Tracking control

- Nonlinear systems with disturbances and uncertainties
- Robust adaptive approach
- Traditional optimal approach

Reinforcement learning (RL)

- Multi-purpose
- Multi-component

Objectives

- Deal with disturbances/uncertainties
- Improve tracking performance
- Adaptive optimal solution

⇒ **Propose two control structures**

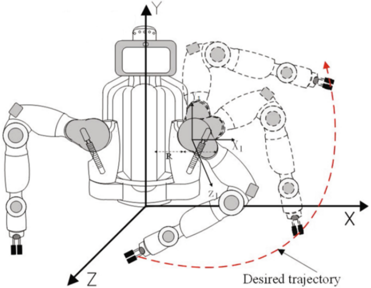


Figure 1: Tracking control

Time-Varying RISE and ARL Control Structure for Nonlinear Systems

- Combination of time-varying RISE[1] and ARL control structure for
- Robot manipulator: second-order CT nonlinear MIMO systems
- Sliding variable to achieve reduced-order system
- Transformed to autonomous affine system
- Comparison with RISE+ARL structure in [1]:
  - Clear exploration signal and initial conditions
  - Time-varying RISE implementation leads to better weight convergences

[1] Robust integral of the sign of the error

Disturbance Observer and ARL Control Structure for Nonlinear Systems

- Combination of disturbance observer, ARL-based control,
- Kinematic and feed-forward control structure for
- 3-DOF Surface vessel system: second-order CT nonlinear MIMO systems
- Transformed to autonomous affine system
- Comparison with previous works:
  - Simpler computation over [2] and [3]
  - Better tracking performance over [4]

Model of Robot Manipulator

The dynamic model of an n-link arm [1]:

$$M(\eta)\ddot{\eta} + C(\eta,\dot{\eta})\dot{\eta} + G(\eta) + F(\dot{\eta}) + d(t) = \tau \quad (1)$$

$\eta(t)$	joint variables
$M(\eta)$	inertia matrix
$C(\eta,\dot{\eta})$	Coriolis/centripetal matrix
$G(\eta)$, $F(\dot{\eta})$	gravity forces and friction
$d(t)$	disturbance vector
τ	control input

with several properties and necessary assumptions.

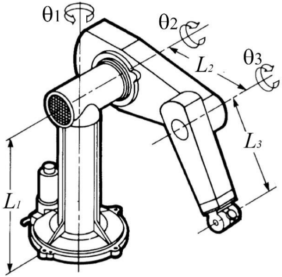


Figure 2: Illustration of a manipulator

Control Objective

Design a control structure for nonlinear system (1) in order to

- Track reference trajectory η_{ref} under disturbances d
- Minimize a predefined cost function $J(\cdot)$

Control Objective

Design a control structure for nonlinear system (1) in order to

- Track reference trajectory η_{ref} under disturbances d $\Rightarrow$ Time-varying RISE
- Minimize a predefined cost function $J(\cdot)$ $\Rightarrow$ On-policy AC

Firstly, with the joint error $e_1(t) = \eta_{ref} - \eta, \alpha_1 \in \mathbb{R}^{n\times n} > 0$, define a sliding variable

$$s(t) = \dot{e}_1 + \alpha_1 e_1 \tag{2}$$

Substitute into (1), the dynamics of $s(t)$ is obtained

$$M\dot{s} = -Cs - \tau + f + d \tag{3}$$

where $f = M(\ddot{\eta}_{ref} + \alpha_1 \dot{e}_1) + C(\dot{\eta}_{ref} + \alpha_1 e_1) + G + F$

With the system described in (3), design the control input

$$\tau = f + d - u = \varepsilon - u \tag{4}$$

ε – time-varying RISE estimation	$u(X)$ – On-policy AC-based control

With the system described in (3), design the control input

$$\tau = f + d - u = \varepsilon - u \tag{4}$$

ε – time-varying RISE estimation $u(X)$ – On-policy AC-based control

The dynamics of $x = [e_1^T, s^T]^T$ is

$$\dot{x} = \begin{bmatrix} -\alpha_1 e_1 + s \\ -M(\eta_{ref} - e_1)^{-1} C(\eta_{ref} - e_1, \dot{\eta}_{ref} + \alpha_1 e_1 - s)s \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{n \times n} \\ M^{-1} \end{bmatrix} u \tag{5}$$

To obtain **autonomous affine system**, assume that $\dot{\eta}_{ref}(t) = f_{ref}(\eta_{ref})$, then

$$\dot{X} = A(X) + B(X)u \tag{6}$$

with $X = [x^T, \eta_{ref}^T, \dot{\eta}_{ref}^T]^T$ and the matrices:

$$A(X) = \begin{bmatrix} -\alpha_1 e_1 + s \\ -M(\cdot)^{-1} C(\cdot) s \\ f_{ref}(\eta_{ref}) \\ \dot{f}_{ref}(\eta_{ref}) \end{bmatrix}, \quad B(X) = \begin{bmatrix} \mathbf{0} \\ M^{-1} \\ \mathbf{0} \end{bmatrix} \tag{7}$$

The infinite horizon cost function to be minimized:

$$J = \int_0^\infty \left(\frac{1}{2} X^T Q_T X + \frac{1}{2} u^T R u\right) dt \tag{8}$$

On-Policy Actor-Critic Architecture

The infinite horizon cost function to be minimized:

$$J = \int_0^\infty (\frac{1}{2}X^T Q_T X + \frac{1}{2}u^T R u)dt \tag{8}$$

Hard-to-solve HJB equation with the optimal value function $V^*(X)$ and the optimal feedback controller $u^*(X) \rightarrow$ **approximated using a NN**

$$V^*(X) = W^T \psi(X) + \varepsilon_v(X), \tag{9}$$

$$u^*(X) = -\frac{1}{2}R^{-1}B^T(X)\Big((\frac{\partial \psi}{\partial X})^T W + (\frac{\partial \varepsilon_v(X)}{\partial X})^T\Big) \tag{10}$$

On-Policy Actor-Critic Architecture

Using two separate NNs for easier weight update and stability analysis [5]

$$\hat{V}(X) = \hat{W}_c^T \psi(X), \quad \hat{u}(X) = -\frac{1}{2} R^{-1} B^T(X) \left(\frac{\partial \psi}{\partial X}\right)^T \hat{W}_a \tag{11}$$

Actor and critic approximators are tuned simultaneously to minimize Bellman error δ_{hjb}. The least-squares update of critic weights is

$$\dot{\hat{W}}_c = -k_c \lambda \frac{\sigma}{1 + \upsilon \sigma^T \lambda \sigma} \delta_{hjb}, \quad \dot{\lambda} = -k_c \lambda \frac{\lambda \sigma^T}{1 + \upsilon \sigma^T \psi \sigma} \lambda \tag{12}$$

The actor adaptation law is based on gradient descent method

$$\dot{\hat{W}}_a = -\frac{k_{a1}}{\sqrt{1 + \sigma^T \sigma}} \frac{\partial \psi}{\partial X} B R^{-1} B^T \frac{\partial \psi^T}{\partial X} (\hat{W}_a - \hat{W}_c) \delta_{hjb} - k_{a2} (\hat{W}_a - \hat{W}_c) \tag{13}$$

ε is the estimation of $f + d$, based on the time-varying RISE [6].

$$\varepsilon(t) = (K_s(\cdot) + 1)s(t) - (K_s(t_0) + 1)s(0) + \rho(t) \tag{14}$$

$$\dot{\rho} = (k_{s0} + 1)\alpha(\cdot)s(t) + \beta\,\mathrm{sgn}(s(t)) \tag{15}$$

with $K_s(\cdot)$ and $\alpha(\cdot)$ are two nonlinear feedback functions designed as

$$K_s(\cdot) \equiv K_s(s, \gamma_1, \delta_1) = \begin{cases} k_{s0}|s|^{\gamma_1 - 1}, |s| > \delta_1 \\ k_{s0}{\delta_1}^{\gamma_1 - 1}, |s| \leq \delta_1 \end{cases} \tag{16}$$

$$\alpha(\cdot) \equiv \alpha(s, \gamma_2, \delta_2) = \begin{cases} \alpha_0|\int s|^{\gamma_2 - 1}, |\int s| > \delta_2 \\ \alpha_0{\delta_2}^{\gamma_2 - 1}, |\int s| \leq \delta_2 \end{cases} \tag{17}$$

where $k_{s0}, \alpha_0, \gamma_1, \delta_1, \gamma_2, \delta_2$ are positive parameters which need designing carefully.

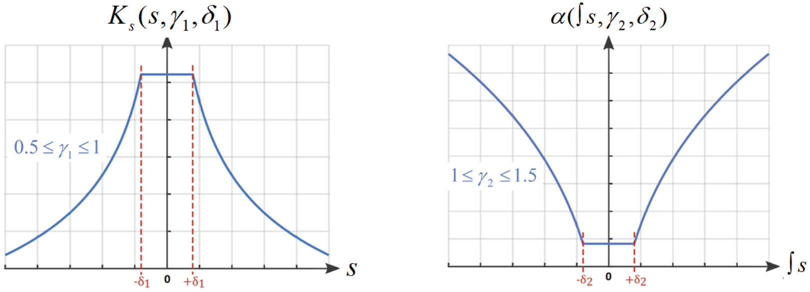
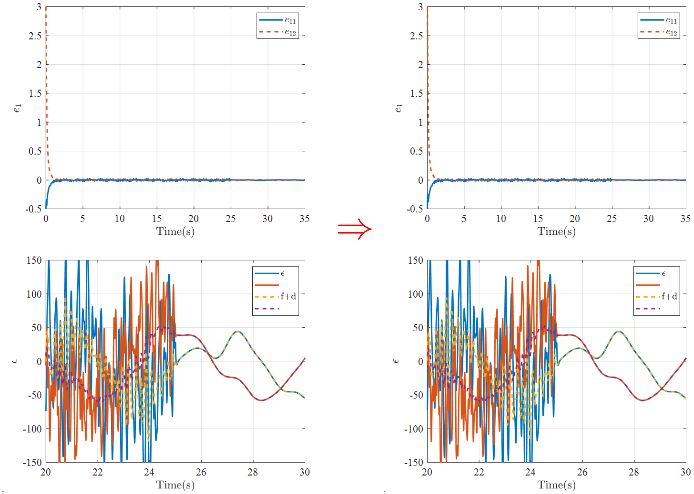


Figure 3: Functions of the control gains with respect to their arguments



- RMSE: 0.2073 and 0.2072
- Control reduces 0.52%
- Similar estimation

Figure 4: Tracking errors and estimation of the two controllers

Simulation Results

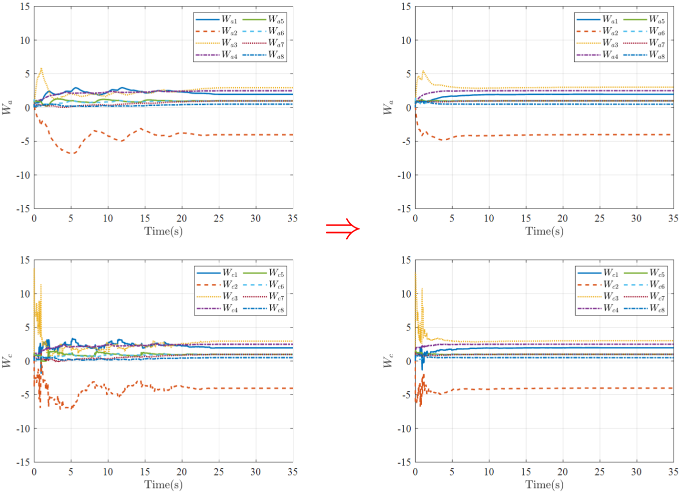


Figure 5: Weight convergences

- Both show great convergences
- Convergence error reduces 41.63%
- Faster and smoother

$\eta = [x, y, \psi]^T$ denotes the 3-DOF position
$\nu = [u, v, r]^T$ denotes the corresponding velocities
The model of SV system [7]:

$$\begin{cases} \dot{\eta} = R(\eta)\nu(t) \\ M\dot{\nu} = \tau + \Delta(\eta, \nu) - f(\eta, \nu) \end{cases} \quad (18)$$

with dynamics $f(\eta, \nu)$ is modeled by

$$f(\eta, \nu) = C(\nu)\nu + D(\nu)\nu + g(\eta, \nu) \quad (19)$$

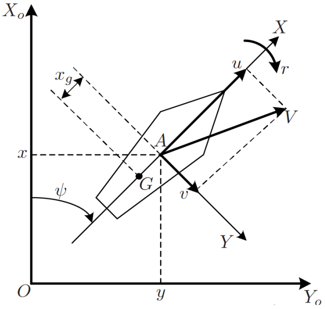


Figure 6: Earth-fixed and body-fixed coordinate frames of an SV

Control Objective

Design a control structure for nonlinear system (18) in order to

- Track reference trajectory η_d
- Robust with disturbances Δ
- Minimize a predefined cost function $J(\cdot)$

Control Objective

Design a control structure for nonlinear system (18) in order to

- Track reference trajectory $\eta_d \Rightarrow$ Kinematic and feed-forward control
- Robust with disturbances $\Delta \Rightarrow$ Disturbance observer
- Minimize a predefined cost function $J(\cdot) \Rightarrow$ On-policy AC

Kinematic and Feed-Forward Control Structure

With the trajectory tracking error $e_\eta = \eta - \eta_d$, design the kinematic control law

$$\nu_d = R^{-1}(\eta)(\dot{\eta}_d - \beta_\eta e_\eta) \tag{20}$$

Kinematic and Feed-Forward Control Structure

With the trajectory tracking error $e_\eta = \eta - \eta_d$, design the kinematic control law

$$\nu_d = R^{-1}(\eta)(\dot{\eta}_d - \beta_\eta e_\eta) \tag{20}$$

With the body-fixed velocity error: $e_\nu = \nu - \nu_d$, a feed-forward term τ_{ff} is added

$$\tau = u + \tau_{ff}, \quad \tau_{ff} = M\dot{\nu}_d + f(\eta_d, \nu_d) \tag{21}$$

With the trajectory tracking error $e_\eta = \eta - \eta_d$, design the kinematic control law

$$\nu_d = R^{-1}(\eta)(\dot{\eta}_d - \beta_\eta e_\eta) \tag{20}$$

With the body-fixed velocity error: $e_\nu = \nu - \nu_d$, a feed-forward term τ_{ff} is added

$$\tau = u + \tau_{ff}, \quad \tau_{ff} = M\dot{\nu}_d + f(\eta_d, \nu_d) \tag{21}$$

Let $X = [e_\nu^T \quad e_\eta^T \quad \eta_d^T]^T$ is an augmented state for the dynamic subsystem of SV. The **autonomous system** can be represented concisely as

$$\dot{X} = F(X) + G_u(X)u + G_d(X)\Delta \tag{22}$$

Kinematic and Feed-Forward Control Structure

The control input is designed as

$$u = d(X)\hat{\Delta} + u_r(X) \tag{23}$$

$d(X)\hat{\Delta}$ – disturbance compensator $u_r(X)$ – RL-based control for nominal system

The control input is designed as

$$u = d(X)\hat{\Delta} + u_r(X) \tag{23}$$

$d(X)\hat{\Delta}$ – disturbance compensator $u_r(X)$ – RL-based control for nominal system

According to [8], with $\dot{\Delta} \simeq 0$, a disturbance observer is constructed

$$\begin{cases} \hat{\Delta} = y + P(X) \\ \dot{y} = -\frac{\partial P(X)}{\partial X}(G_u(X)u + G_d(X)y + G_d(X)P(X) + F(X)) \end{cases} \tag{24}$$

If $\frac{\partial P(X)}{\partial X} G_d(X)$ is positive definite, $\tilde{\Delta}$ is exponentially stable, as $t \to \infty$.

The disturbance compensation gain is

$$d(X) = -G_u^{+} G_d \tag{25}$$

Disturbance Observer and ARL Control Structure

Simulation Results

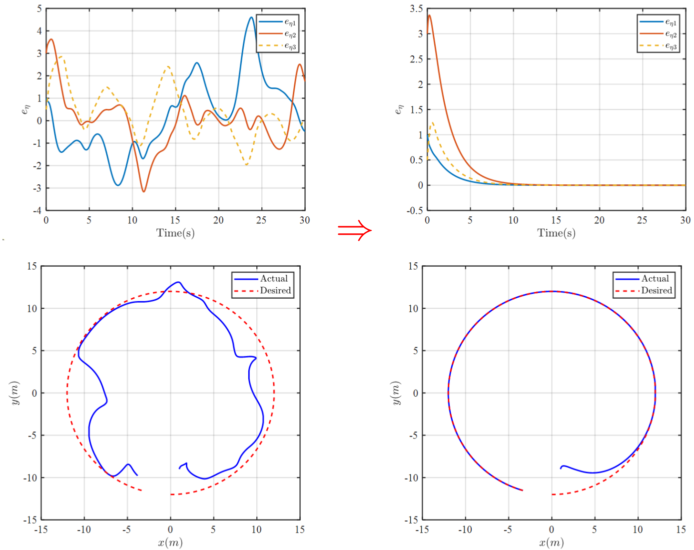


Figure 7: Tracking errors and trajectories of the two controllers

- RMSE reduces 66%

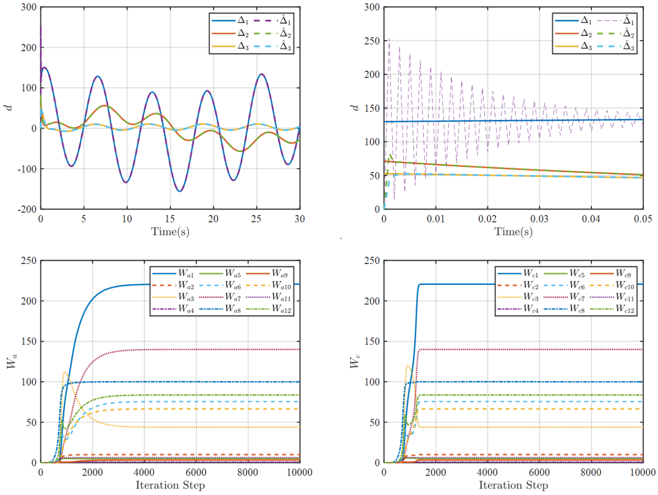


Figure 8: Estimation error and weight convergence

- Fast and precise estimation
- AC convergence

Conclusion

Two problems are accomplished satisfying:

- ✓ Great trajectory tracking performance
- ✓ Robustness to disturbance
- ✓ Minimization of predefined cost function

Conclusion

Two problems are accomplished satisfying:

- ✓ Great trajectory tracking performance
- ✓ Robustness to disturbance
- ✓ Minimization of predefined cost function

Successful implementation of frameworks:

- ✓ System modeling
- ✓ On-policy actor-critic architecture (ARL/ADP)
- ✓ Sliding variable, kinematic and feed-forward control
- ✓ Time-varying RISE, and disturbance observer...

Conclusion

Two problems are accomplished satisfying:

- ✓ Great trajectory tracking performance
- ✓ Robustness to disturbance
- ✓ Minimization of predefined cost function

Successful implementation of frameworks:

- ✓ System modeling
- ✓ On-policy actor-critic architecture (ARL/ADP)
- ✓ Sliding variable, kinematic and feed-forward control
- ✓ Time-varying RISE, and disturbance observer...

Limitations:

- Requirement of drift matrix and input gain matrix
- Lack of some explicit proof of stability and optimality

Future Work

- A number of intriguing unresolved challenges
- Partially/completely model-free RL approaches
- Explicit optimality and stability demonstration
- Differential games and multi-agent systems

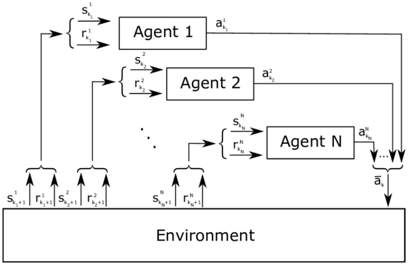


Figure 9: RL of multi-agent system

- Phuong Nam Dao, Pham Thanh Loc, and Trang Quang Huy. "Sliding Variable-based Online Adaptive Reinforcement Learning of Uncertain/Disturbed Nonlinear Mechanical Systems". In: *Journal of Control, Automation and Electrical Systems* 32 (2 2021), pp. 281–290.
- Zhao Yin et al. "Control Design of a Marine Vessel System Using Reinforcement Learning". In: *Neurocomputing* 311 (2018), pp. 353–362.
- Guoxing Wen et al. "Adaptive Tracking Control of Surface Vessel Using Optimized Backstepping Technique". In: *IEEE Transactions on Cybernetics* 49.9 (2019), pp. 3420–3431.
- Van Tu Vu et al. "Online Actor-Critic Reinforcement Learning Control for Uncertain Surface Vessel Systems with External Disturbances". In: *International Journal of Control, Automation and Systems* (). Forthcoming.
- S. Bhasin et al. "A Novel Actor-Critic-Identifier Architecture for Approximate Optimal Control of Uncertain Nonlinear Systems". In: *Automatica* 49.1 (2013), 82–92.
- Hussein Saied et al. "A New Time-Varying Feedback RISE Control for 2nd-Order Nonlinear MIMO Systems: Theory and Experiments". In: *International Journal of Control* (Dec. 2019).

Ning Wang et al. "Fast and Accurate Trajectory Tracking Control of an Autonomous Surface Vehicle With Unmodeled Dynamics and Disturbances". In: *IEEE Transactions on Intelligent Vehicles* 1.3 (2016), pp. 230–243.

Wen-Hua Chen. "Disturbance observer based control for nonlinear systems". In: *IEEE/ASME Transactions on Mechatronics* 9.4 (2004), pp. 706–710.

Thank you!