SCIENCE CHINA
Information Sciences

• LETTER •

June 2026, Vol. 69, Iss. 6, 169202:1–169202:2
https://doi.org/10.1007/s11432-025-4700-4

Adaptive dynamic programming control based on dual-critic networks of a flexible two-link manipulator with elastic vibration

Hejia GAO[1,2,4], Zele YU[1,2,4], Jiangxu LIU[1,2,4] & Changyin SUN[1,2,3,4*]

[1]*School of Artificial Intelligence, Anhui University, Hefei 230601, China*
[2]*Engineering Research Center of Autonomous Unmanned System Technology, Ministry of Education, Hefei 230601, China*
[3]*School of Automation, Southeast University, Nanjing 214135, China*
[4]*Anhui Provincial Key Laboratory of Security Artificial Intelligence, Anhui University, Hefei 230601, China*

Received 7 April 2025/Revised 6 September 2025/Accepted 16 October 2025/Published online 15 January 2026

Citation Gao H J, Yu Z L, Liu J X, et al. Adaptive dynamic programming control based on dual-critic networks of a flexible two-link manipulator with elastic vibration. Sci China Inf Sci, 2026, 69(6): 169202, https://doi.org/10.1007/s11432-025-4700-4

As automation advances, flexible robotic manipulators are increasingly vital in high-precision applications like aerospace assembly and minimally invasive surgery. However, their inherent structural flexibility leads to challenging nonlinear dynamics and vibrations, which can severely degrade performance [1, 2]. Traditional model-based controllers struggle with the unavoidable uncertainties in these systems, while conventional reinforcement learning (RL) approaches often require extensive, model-specific prior knowledge, limiting their adaptability. To overcome these limitations, this study proposes an innovative adaptive dynamic programming (ADP) method [3]. Our approach utilizes a novel dual-critic network architecture to achieve superior vibration suppression and trajectory tracking for a flexible two-link manipulator (FTLM) without requiring a precise system model. The main contributions are twofold: first, the introduction of a dual-critic structure that significantly reduces reliance on predefined models by generating adaptive internal goals; and second, the development of an adaptive control scheme that robustly handles system uncertainties, leading to enhanced performance.

Methodology. The dynamics of the FTLM are derived using the assumed modes method, which discretizes the infinite-dimensional PDE model into a finite-dimensional ordinary differential equation (ODE) representation. The system's motion is governed by the Lagrangian dynamic equation

$$A(x)\ddot{x} + O(x,\dot{x})\dot{x} + H(x) = \tau(t), \tag{1}$$

where $x = [\theta, p]^{\mathrm{T}}$ is the state vector containing joint angles θ and flexible coordinates p. $A(x)$ is the inertia matrix, $O(x,\dot{x})$ represents Coriolis and centrifugal effects, $H(x)$ is the stiffness matrix, and $\tau(t)$ is the control torque vector. The design of our controller is based on the following fundamental assumptions.

Assumption 1. The system states are available for feedback.

Assumption 2. The uncertainties in the system dynamics, including unknown parameters and external disturbances, are bounded.

To control this complex system without precise knowledge of its dynamics, we designed a novel ADP controller based on a dual-critic network architecture. This structure synergistically integrates three collaborative radial basis function neural networks (RBFNNs) to achieve online learning and optimal control. The objective is to find a control policy $\tau(t)$ that minimizes a long-term cost function $J(t)$, which is an integral of an instantaneous cost $\kappa(t) = e_1^{\mathrm{T}} D e_1 + \tau^{\mathrm{T}} S \tau$, where e_1 is the tracking error. The three networks collaborate as follows.

- **Reference network.** This network is central to our dual-critic design. It adaptively generates a continuous internal reinforcement signal $R(t) = \hat{W}_r^{\mathrm{T}} S_r(U_r)$, acting as a dynamic, state-dependent goal to guide the learning process based on the system state vector U_r.
- **Critic network.** The critic approximates the cost function with $\hat{J}(t) = \hat{W}_c^{\mathrm{T}} S_c(U_c)$. It evaluates the system's performance by comparing the actual state transitions against the adaptive goals $R(t)$ from the reference network, providing crucial feedback for policy refinement.
- **Actor network.** The actor formulates the optimal control policy. Guided by the critic's evaluation, it generates the control torque using the following law:

$$\tau(t) = -e_1 - K_2 e_2 + \hat{W}_a^{\mathrm{T}} S_a(U_a), \tag{2}$$

where e_1 and e_2 are tracking error terms, K_2 is a gain matrix, and $\hat{W}_a^{\mathrm{T}} S_a(U_a)$ is the RBFNN output that compensates for system uncertainties.

The controller's learning process is driven by a set of precisely defined error signals. The primary and secondary tracking errors, based on a backstepping design, are defined as

$$e_1 = x - x_r, \tag{3}$$

$$e_2 = \dot{e}_1 + K_1 e_1, \tag{4}$$

where x_r is the reference trajectory and K_1 is a positive definite gain matrix. The critic network's learning is guided by the temporal difference (TD) error, which represents the discrepancy in the

* Corresponding author (email: cysun@ahu.edu.cn)

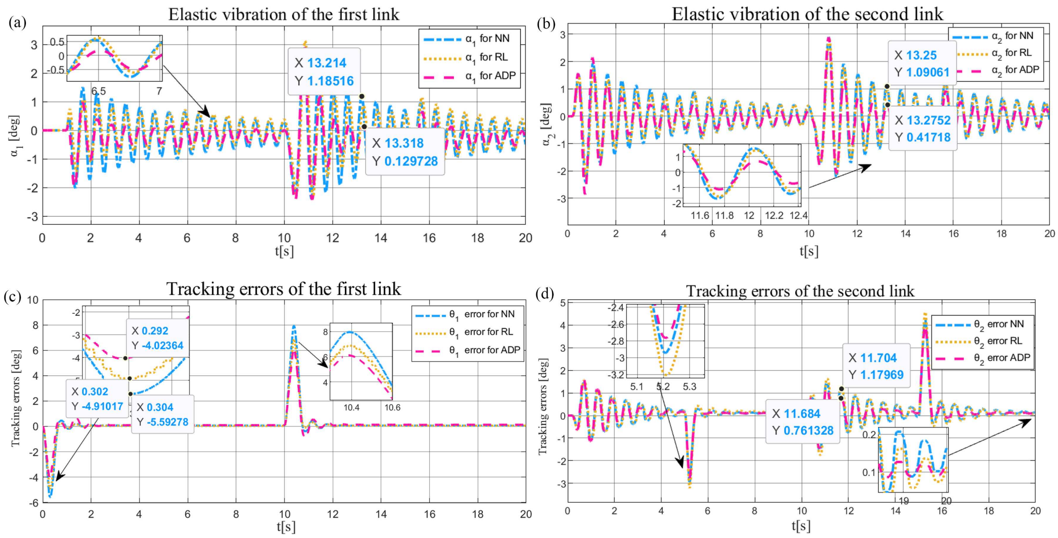

Figure 1 (Color online) Experimental results comparing elastic vibration and tracking errors. (a) and (b) show the vibration suppression for Link 1 and Link 2, respectively. (c) and (d) show the corresponding tracking errors. The results highlight the superior performance of the proposed ADP controller.

Bellman equation:

$$e_c(t) = \kappa(t) + \dot{\hat{J}}(t). \tag{5}$$

The weights of the three RBFNNs $(\hat{W}_a, \hat{W}_c, \hat{W}_r)$ are tuned online using gradient descent-based update laws designed to minimize their respective error functions. These laws generally take the form $\dot{\hat{W}} = -\sigma e S$, where σ is a positive learning rate, e represents an error signal like e_c, and S is the vector of RBF activations.

The stability of the entire closed-loop system is rigorously proven, as formally stated in Theorem 1. The proof is based on the Lyapunov direct method, utilizing the following candidate function:

$$V = \frac{1}{2}e_1^{\mathrm{T}}e_1 + \frac{1}{2}e_2^{\mathrm{T}}Ae_2 + \frac{1}{2}\tilde{W}_r^{\mathrm{T}}\tilde{W}_r + \frac{1}{2}\tilde{W}_a^{\mathrm{T}}\tilde{W}_a + \frac{1}{2}\tilde{W}_c^{\mathrm{T}}\tilde{W}_c. \tag{6}$$

Theorem 1. Consider the FTLM system described by (1) under Assumptions 1 and 2, with the proposed dual-critic ADP control scheme. For bounded initial states, all signals in the closed-loop system, including the tracking errors e_1, e_2 and the NN weight estimation errors $\tilde{W}_r, \tilde{W}_a, \tilde{W}_c$, are semi-globally uniformly ultimately bounded (SGUUB).

The proof, based on the Lyapunov direct method, is detailed in the supplementary document. It confirms that the system remains stable and that tracking errors converge to a small residual set.

Results. The proposed dual-critic ADP controller was experimentally validated on a Quanser FTLM platform. Its performance was benchmarked against a standard neural network (NN) controller [4] and a state-of-the-art reinforcement learning (RL) controller [5], which notably utilizes a single-critic architecture with an adaptive law. All controllers were tested using identical square wave reference trajectories to challenge their vibration suppression capabilities.

The experimental results, depicted in Figure 1, demonstrate the clear superiority of our ADP method. As shown in Figures 1(a) and (b), our method achieves significantly more effective vibration damping for both links compared to the NN and RL controllers, with the vibration amplitude being noticeably smaller. Similarly, Figures 1(c) and (d) highlight the enhanced tracking performance, where the ADP controller results in smaller peak errors and faster settling times. Quantitatively, for the first link, it reduced the maximum tracking error by up to 22.8% compared to the NN controller and by 11.6% compared to the RL controller. The performance improvement was even more pronounced in vibration suppression, which is the primary challenge for flexible manipulators. Our method successfully attenuated the steady-state vibration of the first link by a remarkable 49.1% compared to the NN controller and by 46.2% compared to the RL controller. Furthermore, the ADP controller achieved these results while maintaining a smooth and stable control input profile, indicating high control efficiency and enhanced overall system stability. These quantitative improvements, supported by the visual evidence in Figure 1, strongly validate the effectiveness of the proposed dual-critic architecture in navigating the complex dynamics of flexible systems.

Conclusion. This work introduced and experimentally validated a novel dual-critic ADP framework that significantly improves both vibration damping and trajectory accuracy in FTLM systems. By generating adaptive internal reinforcement signals, the controller overcomes the limitations of traditional model-based and RL methods, demonstrating superior performance in a real-world experimental setting. Future work will explore the extension of this framework to more complex spatial manipulators, its application in human-robot interaction, and a rigorous analysis of its robustness against practical challenges like impulsive sensor noise.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 62573003, 62303010) and University Synergy Innovation Program of Anhui Province (Grant No. GXXT-2023-039).

References

1 Li B, Li X, Gao H, et al. Advances in flexible robotic manipulator systems—part I: overview and dynamics modeling methods. IEEE ASME Trans Mechatron, 2024, 29: 1100–1110
2 Li B, Li X, Gao H, et al. Advances in flexible robotic manipulator systems—part II: planning, control, applications, and perspectives. IEEE ASME Trans Mechatron, 2024, 29: 1680–1689
3 Yao X Q, Xu X M, He W, et al. Asymptotical event-based input-output constrained boundary control of flexible manipulator agents under a signed digraph. Sci China Inf Sci, 2025, 68: 142201
4 Gao H, He W, Zhou C, et al. Neural network control of a two-link flexible robotic manipulator using assumed mode method. IEEE Trans Ind Inf, 2019, 15: 755–765
5 He W, Gao H, Zhou C, et al. Reinforcement learning control of a flexible two-link manipulator: an experimental investigation. IEEE Trans Syst Man Cybern Syst, 2021, 51: 7326–7336