Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Reinforcement Learning-based Nonlinear Systems Optimal Control of Distributed Drive Electric Vehicles: A High-Order Fully Actuated System Approach

Yuchen Wang[1†], Wenzhuang Wang[2], Jiayi Fang[3], Jizhe Wang[4] and Yahui Zhang[5]

[1] Heibei Light Structure Design and Preparation Technology lnnovation Center and
School of Mechanical Engineering, Yanshan Universitly, Qinhuangdao 066004, China
(zhangyahui@ysu.edu.cn)
[3]The School of Mechanical Engineering, Beijing Institute of Technology, Beijing 100081, China

Abstract: Distributed drive electric vehicles (DDEVs) feature redundant actuators, offering enhanced control flexibility. To address the trajectory tracking problem in complicated nonlinear system with strong coupling of actuators, this paper proposes a model-free reinforcement learning-based tracking strategy within High- order Fully Actuated (HOFA) system framework. The proposed method integrates feedforward and feedback control. Specifically, a direct parametric design of HOFA system as the feedforward control law, while the feedback control law is obtained by approximating the value function using a neural network. The neural network weights are updated based on the Bellman error and least squares method, ensuring system stability while approximating the optimal control strategy. Simulation results demonstrate that the proposed control method enables smooth and accurate trajectory tracking, validating its effectiveness in improving vehicle stability and path-following performance.

Keywords: Actor-critic structure, distributed drive electric vehicles (DDEVs), high-order fully actuated system approach (HOFA), path tracking, reinforcement learning control.

1. INTRODUCTION

In nonlinear system control, the classical optimal control theory typically relies on solving the Hamilton-Jacobi-Bellman (HJB) equation [1]. However, distributed drive electric vehicles (DDEVs) are characterized by multiple actuators, strong coupling, and complex nonlinear dynamics, making model predictive approach [2] [3] the predominant approach among researchers for optimal control. Recently, neural network-based approximation methods have gained significant attention, enabling the adaptive numerical solution of feedback control problems [4] [5]. This advancement has also contributed to the growing application of reinforcement learning (RL) techniques, which have emerged as a promising trend in the field.

These studies focus on developing advanced control strategies to improve vehicle performance and safety. [6] proposes a hierarchical control framework that prioritizes handling stability while minimizing energy loss, optimizing the trade-off between stability and efficiency in DDEVs control. Similarly, [7] designs a slip rate control system to enhance traction and prevent slipping across various road conditions. [8] introduces a control channel recombination method, transforming an overactuated system into a standard square system to improve compatibility and control efficiency. [9] investigates collision avoidance strategies for DDEVs with controlled longitudinal speed, emphasizing accident prevention and safety enhancement. Furthermore, [10] presents a method to reduce tire slip energy while ensuring vehicle stability, employing the C/GMRES algorithm for efficient and rapid control allocation. [11] explores the integration of electromechanical systems in DDEVs traction control, improving overall system performance and efficiency.

In 2020, G.R Duan elaborated on robust control, adaptive control, and optimal control methods for high-Order fully actuated (HOFA) system [12] [13] [14]. It is well understood that designing additional control inputs within a fully actuated system can mitigate uncertainties, disturbances, and faults [15]. However, in complex DDEVs systems, strong nonlinearities and significant uncertainties pose major challenges for control design. To address these issues, researchers have developed sophisticated algorithms to achieve coordinated control of vehicle stability and path tracking. However, such complexity also increases the risk of control system bugs.

The main contributions of this paper are as follows:
1. A reinforcement learning-based approach is introduced to handle uncertainties in fully actuated systems, eliminating the need for explicit system dynamics modeling through iterative learning.
2. The actor-critic framework is employed for partial dynamics approximation.
3. A fully actuated system transformation is applied to DDEVs, incorporating path tracking as the primary driving objective.

2. DIRECT PARAMETERIZATION METHOD BASED FEEDFORWARD TRACKING CONTROLLER

2.1. Vehicle HOFA model

The physical characteristics of the HOFA require that the derivative term has directly related properties. Consider establishing a vehicle motion equation into the

† Yuchen Wang is the presenter of this paper.

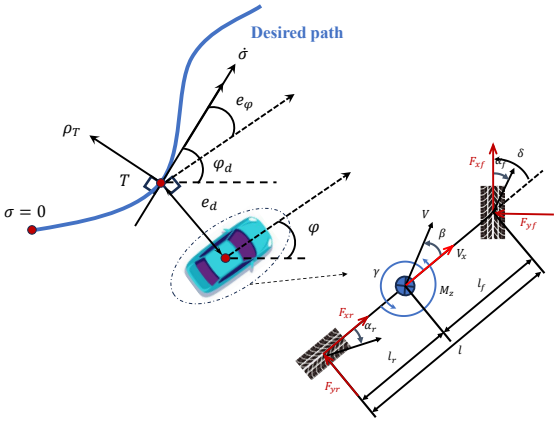

Fig. 1: Vehicle dynamical model

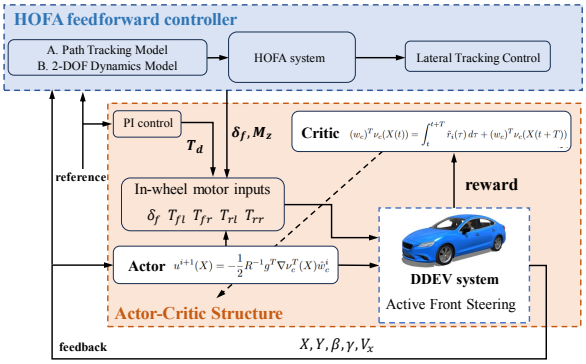

Fig. 2: control scheme

HOFA model, which includes a path tracking model and a vehicle dynamics model.

As shown in Fig.1, the geometric relationship of vehicle motion is presented, where e_y and e_ε represent the lateral error and heading angle error of the projection point, respectively. β and γ respectively represent the lateral deviation angle and yaw rate of the center of mass, v_x is the longitudinal vehicle speed, and ρ is the expected curvature of the projection point.

$$\begin{cases} \dot{e}_y = v_x e_\varepsilon + v_x \beta \\ \dot{e_\varepsilon} = \gamma - \rho_T(\sigma) v_x \end{cases} \tag{1}$$

Correspondingly, a two degree of freedom vehicle dynamics model is constructed, in which the lateral force can be represented by the tire lateral stiffness and lateral angle, and a linear steady system is considered under small turning angles

$$\begin{bmatrix} \dot{\beta} \\ \dot{\gamma} \end{bmatrix} = A \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + B \begin{bmatrix} \delta \\ M_z \end{bmatrix} \tag{2}$$

where

$$A = \begin{bmatrix} \frac{C_f + C_r}{m v_x} & \frac{l_f C_f - l_r C_r}{m v_x^2} - 1 \\ \frac{l_f C_f - l_r C_r}{I_z} & \frac{{l_f}^2 C_f + {l_r}^2 C_r}{I_z v_x} \end{bmatrix} \quad B = \begin{bmatrix} -\frac{C_f}{m v_x} & 0 \\ -\frac{l_f C_f}{I_z} & \frac{1}{I_z} \end{bmatrix}$$

l_f and l_r represent distances of front and rear axles from the center of mass, respectively. δ_f and M_z are control inputs for front steering angle and direct yaw moment.

Consider constructing the following HOFA system:

$$x^{(n)} = f + \Delta f + gu \tag{3}$$

Combining (1) and (2), the vehicle HOFA model can be described as:

$$\ddot{x} = H^T(x, \zeta, \xi, t)\theta + q(x, \zeta, \xi, t) + gu \tag{4}$$

where $x = [e_d, \varepsilon]^T$, $\zeta = [\beta, \gamma]^T$, ξ is other time-varying term, θ is an arbitrary constant.

$$H_{11} = \left(\frac{\zeta_1}{m} + \frac{\zeta_2 l_f}{m v_x}\right) C_f, H_{12} = \left(\frac{\zeta_1}{m} - \frac{\zeta_2 l_r}{m v_x}\right) C_r,$$
$$H_{21} = \left(\frac{l_f \zeta_1}{I_z} + \frac{\zeta_2 {l_f}^2}{I_z v_x}\right) C_f, H_{22} = \left(\frac{\zeta_2 {l_r}^2}{I_z v_x} - \frac{l_r \zeta_1}{I_z}\right) C_r$$
$$q_1 = \dot{v_x}(e_\phi + \zeta_1) - \rho_T(\sigma) {v_x}^2$$
$$q_2 = -\dot{\rho_T}(\sigma) v_x - \rho_T(\sigma) \dot{v_x}$$
$$g_{11} = -\frac{C_f}{m}, g_{12} = 0, g_{21} = -\frac{l_f C_f}{I_z}, g_{22} = \frac{1}{I_z}$$

2.2. Direct parameterization method

Consider using the following controller for (5):

$$\begin{cases} u_c = -G^{-1}(u_1 + u_2) \\ u_1 = H^T(x, \zeta, \xi, t)\theta + q(x, \zeta, \xi, t) \\ u_2 = [A_0 \quad A_1][x \quad \dot{x}]^T + Gv \end{cases} \tag{5}$$

where v is an external input signal. with $det(V) \neq 0$, the multi parameter range of Z and F provides reliable trajectory tracking performance for the gain A_i.

Based on HOFA system (9) and (10), consider the following closed-loop system:

$$\dot{z} = \Phi(A_{0-1})z + Bv \tag{6}$$

where $z = [x, \dot{x}]^T$, $B = [0_{2\times 1}, g]^T$.

Therefore, the following feedforward trajectory is defined

$$\dot{z_a} = \Phi(A_{0-1})z_a \tag{7}$$

this trajectory corresponds to input u_c. the feedback control designed next has the same effect as v.

3. MODEL-FREE REINFORCEMENT LEARNING ALGORITHM BASED HOFA MODEL FEEDBACK CONTROLLER

In this section, a reinforcement learning-based control strategy is proposed to compensate for system uncertainties as much as possible, as illustrated in Fig.2.

3.1. Policy Iteration Algorithm Based on Fully Actuated System

we propose a new model-free reinforcement learning that utilizes the HOFA feedforward trajectory.

For system (12), the actual z can be decoupled from two different trajectories, one is the direct parameterization trajectory and the other is $e = z - z_a$. By using low-level expression of (9) and (12), the following error dynamics can be obtained

$$\dot{e} = F(z) - \Phi(A_{0-1})z_a + Gv \tag{8}$$

where

$$F(z) = \begin{bmatrix} 0_{2\times 1} \\ H^T(x,\zeta,\xi,t)\theta + q(x,\zeta,\xi,t) \end{bmatrix}, G = \begin{bmatrix} 0_{2\times 1} \\ g \end{bmatrix}$$

Define the augmented state as $X = col(z_a, e)$ Then, the augmented system can be written as follows:

$$\dot{X} = F(X) + \hat{G}v \tag{9}$$

where

$$F(X) = \begin{bmatrix} \Phi(A_{0-1})z_a \\ F(e+z_a) - \Phi(A_{0-1})z_a \end{bmatrix}, \hat{G} = \begin{bmatrix} 0_{4\times 1} \\ G \end{bmatrix}$$

3.2. Actor-Critic Neural Network Structure Approximation of Value Function

This section will introduce specific algorithm explanations for reinforcement learning.

At policy evaluation step, integrating (17) over $[t, t+T]$ along system (14) with $\tilde{\mu}_i(X) = \mu_{i-1}(X)$ yields

$$\tilde{V}_i(X(t+T)) - \tilde{V}_i(X(t)) + \int_t^{t+T} \tilde{r}_i(\tau)\, d\tau = 0 \tag{10}$$

At policy improvement step, an updated strategy is as follows

$$\tilde{\mu}_{i+1}(X) = -\frac{1}{2}R^{-1}G^{\top}\frac{\partial \tilde{V}_i}{\partial X} \tag{11}$$

The final strategy $\tilde{\mu}_i(X)$ can converge to μ^*, resulting in

$$J = V^*(X) = \min_{u\in A}\int_0^{\infty} r(\tau)\, d\tau$$

As shown in Fig.2, by considering the following neural network approximations for policy evaluation and policy improvement, they represent actor and critic respectively.

The corresponding strategy evaluation can be approximated by the following neural network.

$$\tilde{V}(X) = \sum_{j=1}^{L}\omega_j\phi_j(X) = (w_c)^T\nu_c(X) \tag{12}$$

where ν_c denotes the activation function, w_c denotes the weight vector. Thus, by combining (18) and (20), it can be expressed as follows

$$\begin{aligned}(w_c)^T\nu_c(X(t)) = &\int_t^{t+T}\tilde{r}_i(\tau)\, d\tau \\ &+ (w_c)^T\nu_c(X(t+T))\end{aligned} \tag{13}$$

Due to the approximation of the value function through neural networks, the following Bellman error can be constructed

$$\begin{aligned}\delta(X(t),T) = &\int_t^{t+T} r(\tau)d\tau + (\hat{w}_c^i)^T \\ &\times [\nu_c(X(t+T)) - \nu_c(X(t))]\end{aligned} \tag{14}$$

By using the Bellman error, the weights of the neural network can be learned, and for (19), the following iterative strategy can be obtained

$$u^{i+1}(X) = -\frac{1}{2}R^{-1}g^T\nabla\nu_c^T(X)\hat{w}_c^i \tag{15}$$

From the above, it can be found that the neural network weights of actors and critics are consistent, which enables the value function of Bellman error update to solve the optimal control strategy

The following applies the recursive least square method to update the weights of the critic neural network.

$$\begin{aligned}&\hat{w}_c = -K\delta \\ &K = P\Phi \\ &\dot{P} = -\frac{P\Phi\Phi^T P}{1+\Phi^T P\Phi}\end{aligned} \tag{16}$$

where

$$\Phi = \nu_c(X(t+T)) - \nu_c(X(t))$$

and P represents matrix gain.

Based on the above analysis, we can address the following optimal regulation problem.

$$\begin{aligned}&Minimize: J = \int_0^{\infty} x^T Qx + u^T Ru\, d\tau \\ &Subject \quad to: \dot{X} = F(X) + Gv\end{aligned} \tag{17}$$

The torque distribution for maintaining vehicle speed and additional yaw moment under longitudinal speed requirements is as follows

$$T_d = K_p(v_{xref} - v_x) + K_i\int_0^{\infty}(v_{xref} - v_x)dt \tag{18}$$

$$\begin{aligned}&T_{fl} = \frac{1}{4}T_d - \frac{M_z R_w}{2l_w}, T_{fr} = \frac{1}{4}T_d + \frac{M_z R_w}{2l_w} \\ &T_{rl} = \frac{1}{4}T_d - \frac{M_z R_w}{2l_w}, T_{rr} = \frac{1}{4}T_d + \frac{M_z R_w}{2l_w}\end{aligned} \tag{19}$$

4. SIMULATION AND ANALYSIS

To validate the effectiveness of the proposed reinforcement learning-based control strategy and ensure the desired path tracking performance of the DDEVs, simulations were conducted under specific road conditions. The road surface adhesion coefficient was set to 0.85, with a target driving speed of 72 km/h following a single-lane change trajectory. A feedforward-feedback control approach was adopted, where the feedforward component was designed with appropriate tracking parameters. Through repeated iterations, the tracking accuracy was continuously improved by adjusting the lateral position and yaw angle.

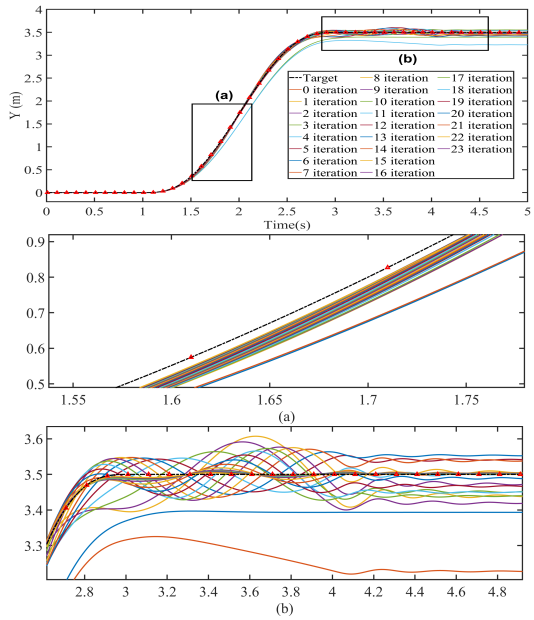

Fig. 3: Lateral trajectory

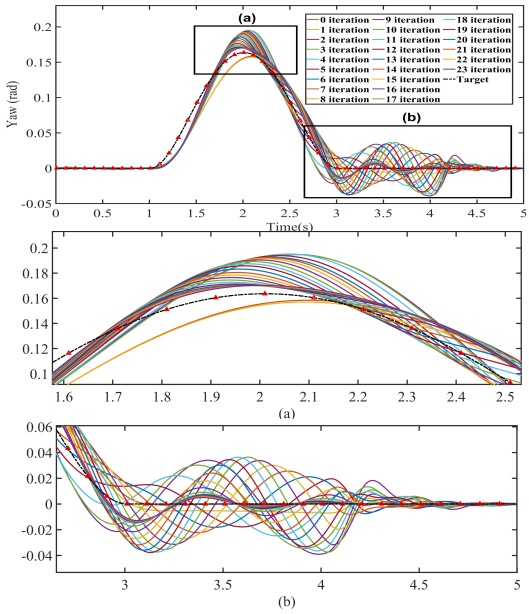

Fig. 4: Yaw trajectory

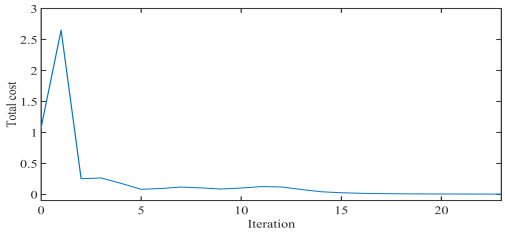

Fig. 5: Total cost

5. SIMULATIONS AND ANALYSIS

The learning performance is illustrated in Fig. 3 and Fig. 4. It can be observed that the fully actuated feedforward control, as the initial 0th iteration, maintains a rough trajectory-tracking effect. As the number of iterations increases, significant trajectory oscillations appear between the first and fifth iterations. This is due to the neural network initially exploring within a wide weight range. However, since the control update follows the gradient direction of the value function, the vehicle trajectory gradually converges to the reference path around the 15th iteration. By the 23rd iteration, marked by the star-shaped line, the trajectory successfully tracks the desired route. In the early phase before 2.5s, the vehicle maintains a stable left-lane change. Fig. 5 further confirms that the total cost function progressively decreases, demonstrating the effectiveness of the proposed reinforcement learning-based control strategy.

REFERENCES

[1] J. J. Murray, C. J. Cox, G. G. Lendaris, and R. Saeks, "Adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 32, no. 2, pp. 140–153, 2002.

[2] N. Guo, X. Zhang, Y. Zou, B. Lenzo, G. Du, and T. Zhang, "A supervisory control strategy of distributed drive electric vehicles for coordinating handling, lateral stability, and energy efficiency," *IEEE Trans. Transp. Electr.*, vol. 7, no. 4, pp. 2488–2504, 2021.

[3] Z. Wang, X. Ding, and L. Zhang, "Chassis coordinated control for full x-by-wire four-wheel-independent-drive electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 72, no. 4, pp. 4394–4410, 2022.

[4] T. Cheng, F. L. Lewis, and M. Abu-Khalaf, "Fixed-final-time-constrained optimal control of nonlinear systems using neural network hjb approach," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1725–1737, 2007.

[5] R. W. Beard, *Improving the closed-loop performance of nonlinear systems*. Rensselaer Polytechnic Institute, 1995.

[6] X. Li, G. Yin, Y. Ren, F. Wang, R. Fang, and A. Li, "Hierarchical control for distributed drive electric vehicles considering handling stability and energy efficiency," in *2023 IEEE International Automated Vehicle Validation Conference (IAVVC)*, pp. 1–6, IEEE, 2023.

[7] B. Leng, L. Xiong, Z. Yu, K. Sun, and M. Liu, "Robust variable structure anti-slip control method of a distributed drive electric vehicle," *IEEE Access*, vol. 8, pp. 162196–162208, 2020.

[8] X. Hou, J. Zhang, Y. Ji, W. Liu, and C. He, "Autonomous drift controller for distributed drive electric vehicle with input coupling and uncertain disturbance," *ISA transactions*, vol. 120, pp. 1–17, 2022.

[9] R.-Y. Zhang, B. Zhang, P.-C. Shi, Y. Mei, Y.-F. Du, and Y.-L. Feng, "Research on the high-speed collision avoidance method of distributed drive electric vehicles," *IEEE Sensors Journal*, vol. 23, no. 14, pp. 15813–15830, 2023.

[10] N. Guo, X. Zhang, Y. Zou, B. Lenzo, T. Zhang, and D. Göhlich, "A fast model predictive control allocation of distributed drive electric vehicles for tire slip energy saving with stability constraints," *Control Engineering Practice*, vol. 102, p. 104554, 2020.

[11] X. Gao and C. Lin, "Electromechanical coupling approach for traction control system of distributed drive electric vehicles," in *E3S Web of Conferences*, vol. 236, p. 01007, EDP Sciences, 2021.

[12] G. Duan, "High-order fully actuated system approaches: Part i. models and basic procedure," *Int. J. Syst. Sci.*, vol. 52, no. 2, pp. 422–435, 2021.

[13] G. Duan, "High-order fully actuated system approaches: Part ii. generalized strict-feedback systems," *Int. J. Syst. Sci.*, vol. 52, no. 3, pp. 437–454, 2021.

[14] G. Duan, "High-order fully actuated system approaches: Part iv. adaptive control and high-order backstepping," *Int. J. Syst. Sci.*, vol. 52, no. 5, pp. 972–989, 2021.

[15] R. Dong, C. Hua, K. Li, and R. Meng, "Adaptive fault-tolerant control for high-order fully actuated system with full-state constraints," *Journal of the Franklin Institute*, vol. 360, no. 12, pp. 8062–8074, 2023.