Convergence-Triggered Reinforcement Adaptive Relearning Control for Virtually-coupled Trains

Wenxiao Si[a], Shigen Gao[a,*], Krishna Busawon[b], Xuewu Dai[c], Tao Wen[a], Hongwei Wang[a]

[a]School of Automation and Intelligence, Beijing Jiaotong University, Beijing 100044, China
[b]Department of Mechanical and Construction Engineering, Northumbria University, Newcastle NE1 8ST, UK
[c]Department of Mathematics, Physics and Electrical Engineering, Northumbria University, Newcastle NE1 8ST, UK

Abstract

This paper proposes a convergence-triggered reinforcement adaptive relearning control (CTRAR) framework for virtual coupling (VC) train formation systems, integrating a novel dual-criteria performance monitor mechanism to ensure optimal and safe operation. Distinct from prior approaches, the proposed mechanism enforces boundary constraints and minimum dwell-time conditions, triggering either convergence to the optimal mode with suspended actor-critic (AC) weight updates or nonconvergence to the nonoptimal mode which indicates a restart of the reinforcement learning (RL) procedure upon violation, termed CTRAR. To enhance learning efficiency visualization, a coefficient-enhanced Gaussian learning rate is developed for AC weight updates, ensuring persistent excitation of tracking error signals without compromising control performance. By leveraging train in-transit data from Beijing South Railway Station (BSRS) to Jinan West Railway Station (JWRS), the simulations successfully demonstrate the execution of reinforcement relearning under convergence-triggered conditions, as well as rigorous validation of the CTRAR algorithm through both error-convergence data illustration and computational resource savings quantification.

Keywords: reinforcement relearning, virtual coupling, convergence-triggered, adaptive control, boundary constraint

1. Introduction

The ongoing evolution of railway transportation systems toward intelligent and high-efficiency paradigms has progressively revealed the intrinsic constraints of conventional train formation methodologies. Fixed-formation train configurations exhibit substantial limitations in addressing temporal passenger flow variations and optimizing capacity utilization, thereby necessitating innovative technological interventions. Since the relative braking distance concept was primarily introduced in [25], it has enabled more efficient braking strategies by safely allowing trailing trains to exceed the positions of leading trains during deceleration, with mathematically guaranteed collision avoidance. The VC concept [4, 3] revolutionizes train connections by replacing physical couplers with controlled spatial gaps and utilizing advanced train-to-train communications to permit dynamic formation reconfiguration. This paradigm-shifting technology presents a fundamental redesign of train operation frameworks, on the one hand, by addressing systemic capacity limitations arising from escalating passenger demand and heterogeneous spatiotemporal distribution patterns, it simultaneously enhances operational flexibility and service quality metrics [28]. On the other hand, significant reduction in energy consumption through optimized formation strategies, which aligns with sustainable rail development objectives [34]. Hence, the scholarly investigation of virtual coupling (VC) technologies constitutes a pivotal response to contemporary transportation challenges, epitomizing the synergistic integration of technological innovation with evolving societal demands.

The realization of VC technology relies on an advanced train operation control system, the core of which lies in breaking through the fixed-block limitations of traditional train control systems to achieve dynamic collaborative operation between trains. In recent years, research on train formation operation control has focused on the following key directions, such as modeling and scenario construction [28, 30, 29, 26, 14], feedback control mechanisms [9, 18, 19, 28], manifold optimal control methods [23, 7, 20, 2, 22], safety operation control [44, 29, 33, 12, 8, 31], communication and perception of cooperative operation [46], energy efficiency and resource conservation [34, 15, 32],

*Corresponding author. *Email address:* gaoshigen@bjtu.edu.cn (Shigen Gao)

computational intelligence control [21, 36, 42, 41, 10]. Meanwhile, in [21], the authors employed a multi-agent RL algorithm to solve the VC train set co-optimization model, designing a prediction-based action management module for non-compliant decision adjustment and a data supervision mechanism for effective value-critical data learning. In [41], the authors employed boundary constraints and pre-evaluated reinforcement learning to enhance train energy efficiency and operational quality, significantly reducing the exploration space while ensuring correct reward signals for the agents. RL, as a powerful machine learning framework for decision-making in complex environments, enables continuous policy optimization to maximize cumulative rewards through agent-environment interactions. Grounded in an iterative learning process, the autonomous agent refines its decision-making policy by RL through continuous interaction with its operational environment. By leveraging reward-based feedback, the agent progressively enhances its actions to achieve long-term objective maximization without relying on explicit system modeling. This model-free approach renders RL particularly advantageous for complex dynamical systems that lack precise mathematical representations or exhibit time-varying parameters, it effectively overcomes the "curse of dimensionality" inherent in traditional dynamic programming methods, owing to these strengths, RL has been successfully deployed across diverse domains, including robotics, transportation (including aerospace and vehicle control) and power systems [24, 13, 43, 39, 40, 27].

The remarkable progress in RL have significantly propelled the development of optimal control systems, particularly, the innovative dual-network architecture of AC has demonstrated remarkable potential by concurrently learning policy and value functions to achieve efficient controller optimization, characterized by its concurrent learning of policy and value functions, has demonstrated exceptional efficacy across diverse control applications. In [45], the authors introduced a neural network-enhanced dynamic programming approach that directly optimizes the Hamilton-Jacobi-Bellman equation in temporal progression, effectively circumventing traditional policy iteration requirements while mitigating persistent excitation constraints through innovative historical data integration. In [6], the authors developed a modified continuous-time AC algorithm incorporating extended policy iteration principles, which relaxed initial stabilizing policies by leveraging inherent system input-affine properties. However, existing AC methodologies fail to achieve automatic adaptive switching by the corresponding triggered cases in real-time operation when faced with varying environmental demands, making them susceptible to performance degradation when handling unknown or time-varying parameters. This limitation becomes particularly critical in VC train operation control systems, where the architecture of AC online learning capability, characterized by its adaptive feedback-variation-reaction mechanism, must continuously address uncertainties arising from high-dimensional state spaces in multi-train coordination and complex aerodynamic coupling effects.

The implementation efficiency of these advanced control methods critically depends on the optimal allocation of computational resources, particularly when handling the high-dimensional state space problems arising from multi-train cooperative control, where computational resource conservation becomes a key limiting factor for practical deployment. With the increasing complexity of VC systems, the computational resource consumption of operational control algorithms has become increasingly prominent. The multi-train cooperative control requires real-time processing of high-dimensional state information and millisecond-level decision-making, posing stringent challenges to onboard computing units. From the perspective of resource conservation, optimizing the computational efficiency of control algorithms is of critical importance. In [34], authors systematically formulated four distinct operational modes for virtually coupled (VC) train formations, with particular emphasis on energy efficiency and ride comfort. Their investigation revealed statistically significant correlations between energy consumption patterns, control modalities, and safety performance metrics across these operational modes. In [5], authors proposed an event-triggered control approach to address the persistent control challenges arising from the complex physical structure of high-speed trains. Their study demonstrated that this method effectively reduces unnecessary control actions through intelligent threshold decision mechanisms, while significantly mitigating mechanical wear caused by frequent control updates. In [10], authors proposed an integrated backstepping control and neural network-based AC framework that online approximates solutions to the Hamilton-Jacobi-Bellman (HJB) equations for coordinated virtual position adjustment and actual speed regulation. Their design achieved stable formation separation control among multiple trains while incorporating a hysteresis performance monitor to identify critical transitions between knowledge-sufficient and knowledge-deficient states in dynamic environments. However, few research has yet focused on providing an optimized control considering resource conservation for VC operation under a RL framework that can be stalled and restarted of uncertain external environmental changes.

Synthesizing the preceding analysis, the principal contributions of this work can be summarized as follows:

1. A CTRAR has been developed for VC formation system. Distinct from prior approaches [6, 35], a novel performance monitor mechanism incorporating dual detection criteria—boundary constraints and minimum dwell-

time constraints—is proposed. This mechanism rigorously evaluates the satisfaction of both conditions, triggering convergence to the optimal operational mode when fulfilled (with suspended RL weight updates), while initiating RL procedure restart upon violation of either condition (i.e., reinforcement relearning).

2. Compared with [37, 38], this work develops a coefficient-enhanced, time-varying Gaussian learning-rate–based adaptive law for updating the AC weights in RL method. The proposed design preserves control performance while ensuring the persistent excitation (PE) condition for any admissible values of train operation error signals. Moreover, the learning rate exhibits a decaying characteristics that facilitates an analytical stability proof, thereby inducing convergence-triggering dynamics for the AC weights in both optimal and non-optimal operational regions.

3. The implementation of CTRAR ensures computational resource efficiency in RL. Leveraging simulated position and velocity data from high-speed trains operating between BSRS and JWRS, the framework considers safe operation under multiple operational scenarios. Comparative simulation results and achieved computational quantified savings further validate the effectiveness of the proposed methodology.

Note: For illustration purposes, some mathematical symbols used in this paper are explained. $\mathbb{R}^n$ means n-dimensional real number space, $\mathbb{R}^{n\times m}$ denotes $n \times m$-dimensional real number space, $\mathbb{Z}_+$ is the set of positive integers. Let the set $[N]$ be defined as $[N] = \{1, 2, \cdots, n\}$, where n represents the number of trains in VC formation.

2. Model formulation and some preliminaries

2.1. Model description

Given the nonlinear dynamical system defined by

$$\dot{x}(t) = \mathcal{F}(x(t)) + \mathcal{G}(x(t))u(x)) \tag{1}$$

where $x(t) \in \mathbb{R}^n$ refers to the system state, $\mathcal{F}(x(t)) \in \mathbb{R}^n$ represents continuous nonlinear functions with $\mathcal{F}_i(0) = 0$, $i = 1, 2, \cdots, n$. $\mathcal{G}(x(t)) \in \mathbb{R}^{n\times m}$ is the control gain function, $u \in \mathbb{R}^m$ denotes control input, $m,\ n \in \mathbb{Z}_+$.

The performance index $\mathcal{J}(x)$ is defined by the complete characterization:

$$\mathcal{J}(x) = \int_0^\infty \mathcal{C}(x(s), u(x(s)))\mathrm{d}s, \tag{2}$$

where $\mathcal{C}(x, u) = x^\top \mathcal{P}_1 x + u^\top \mathcal{P}_2 u$ is cost function, $\mathcal{P}_1$ and $\mathcal{P}_2$ are positive definite symmetric matrices, and $\mathcal{P}_1 \in \mathbb{R}^{n\times n}$ and $\mathcal{P}_2 \in \mathbb{R}^{m\times m}$, x and u are defined in (1). According to [37], let $\Lambda(\Omega)$ denote the set of admissible control policies $u(x)$ with $u(x) \in \Omega$. If $u(x)$ is continuous and satisfies $u_j(0) = 0$ for $j = 1, 2, \ldots, m$, then the performance index $\mathcal{J}(x)$ defined in (2) remains finite over $\Lambda(\Omega)$.

The optimal performance index $\mathcal{J}^*(x)$, in combination with the optimal control $u^*(x)$, is given as

$$\mathcal{J}^*(x) = \int_t^\infty \mathcal{C}(x(s), u^*(x))\mathrm{d}s = \min_{u\in\Lambda(\Omega)} \int_t^\infty \mathcal{C}(x(s), u(x))\mathrm{d}s, \tag{3}$$

Then, differentiating and reformulating both sides of (3) leads to the following HJB equation:

$$\mathcal{H}\left(x, u^*, \Theta(x)\right) = \Theta(x)\left(\mathcal{F}(x) + \mathcal{G}(x)u^*(x)\right)\right) + x^\top \mathcal{P}_1 x + u^{*\top} \mathcal{P}_2 u^* = 0 \tag{4}$$

where $\Theta(x) = \partial\mathcal{J}^*(x)/\partial x$. Solving $\partial\mathcal{H}\left(x, u^*, \Theta(x)\right)/\partial u^* = 0$ yields that

$$u^* = -\frac{1}{2}\mathcal{P}_2^{-1}\mathcal{G}^\top(x)\Theta(x). \tag{5}$$

Then, by substituting (5) into (4), we get

$$\mathcal{H}\left(x, u^*, \Theta(x)\right) = x^\top \mathcal{P}_1 x + \frac{\mathrm{d}\mathcal{J}^*(x)}{\mathrm{d}x^\top}\mathcal{F}(x) - \frac{1}{4}\frac{\mathrm{d}\mathcal{J}^*(x)}{\mathrm{d}x^\top}\mathcal{G}(x)\mathcal{P}_2^{-1}\mathcal{G}^\top(x)\frac{\mathrm{d}\mathcal{J}^*(x)}{\mathrm{d}x} \tag{6}$$

2.2. Train model description

Based on (1), the Newtonian mechanics model considering the rear train in the VC formation is presented as follows

$$\dot{p}_i = v_i,\ \dot{v}_i = u_i + \lceil \mathscr{F}_i(v_i, p_i) \rfloor_{\Omega_i},\ i \in [N]. \tag{7}$$

where p_i, v_i and u_i are position, speed and control input of train i during the VC operation, where $i \in [N] = \{1, 2, \cdots, n\}$ with $n \in \mathbb{Z}^+$ being the total number of trains in the formation. The resistance force per unit mass is given by: $\lceil \mathscr{F}_i(v_i, p_i) \rfloor_{\Omega_i} = f_{b,i}(v_i) + f_{a,i}(p_i)$, $f_{b,i}(v_i)$ represents basic resistance generally described by Davis equation $f_{b,i}(v_i) = -a_{co,i} - b_{co,i}v_i - c_{co,i}v_i^2$, where $a_{co,i}$, $b_{co,i}$ and $c_{co,i}$ are the resistance coefficients under environmental variation conditions shown by *Condition* 1 $\sim$ *Condition Q* in Fig. 1. For convenience, $f_{b,i}(v_i)$ is abbreviated as $= a_i + b_i v_i + c_i v_i^2$, with coefficients a_i, b_i and $c_i > 0$. $f_{a,i}(p_i)$ means additional resistance, $f_{a,i}(p_i) = g\sin\left(\theta_i^{par-g}(p_i)\right)$, where $\theta_i^{par-g}(p_i)$ is partitioned gradient profile. As described above, denote the target-pair $[p_{d,i}, v_{d,i}]^\top$ as the desired position and speed of the measured pair $[p_i, v_i]^\top$ in the VC formation, $i \in \mathbb{Z}_+$. The target pair of $[p_1, v_1]^\top$ of the leading train is preset as $[p_{d,1}, v_{d,1}]^\top$; and the target pair for $[p_i, v_i]^\top$ of the following train is set as $[p_{d,i}, v_{d,i}]^\top = [p_i - \mathscr{L}_{f,i-1}, v_{i-1}]^\top$; where $\mathscr{L}_{f,i-1}$ is the preset inter-train coupling distance for the formation operation between train i and train $i-1$, $\mathscr{L}_{f,0} = 0$.

The given $\lceil \mathscr{F}_i(v_i, p_i) \rfloor_{\Omega_i}$ in (7) represents the variations in (7) induced by uncertain piecewise environmental fluctuations during train formation operation. Partitioned gradient profile $\theta_i^{par-g}(p_i)$ is composed of bounded, unknown and finite number of value ranges $\left\{\theta_i^{par-g,1}(p_i), \theta_i^{par-g,2}(p_i), \cdots, \theta_i^{par-g,k}(p_i)\right\}$ with $k \in \mathbb{Z}_+$. For any partitioned gradient interval, $p_{i,j}^+$ ($p_{i,j}^-$) is the largest (smallest) actual distance in the interval $\mathscr{D}_i^{par-g} := \left[\left(p_{i,j}, \theta_i^{par-g,j}\right), \left(p_{i,j+1}, \theta_i^{par-g,j+1}\right)\right)$. For ceiling speed $v_i^{j,\max}$, i.e., the maximum allowed running speed for ith HST, the theoretical minimum running time $t_i^{j,\min}$ consumed in the railway interval $\mathscr{D}_i^{par-g}$ is thus $t_i^{j,\min} := \frac{p_{i,j}^+ - p_{i,j}^-}{v_i^{j,\max}}$. It is easy to reach that the actual duration time $t_i^{j,act}$ satisfies $t_i^{j,act} \geq t_i^{j,\min} = \frac{p_{i,j}^+ - p_{i,j}^-}{v_i^{j,\max}}$ without doubt. As a consequence, for a stretch of railway line composed of N_s numbers of distance intervals, $[N_s] = 1, 2, \cdots, N_s$, there exists a positive constant $\tau^\# \in \mathbb{Z}_+$ satisfying $\tau^\# \geq \min\left\{t_i^{1,\min} = \frac{p_{i,1}^+ - p_{i,1}^-}{v_i^{1,\max}}, \cdots, t_i^{N_s,\min} = \frac{p_{i,N_s}^+ - p_{i,N_s}^-}{v_i^{N_s,\max}}\right\}$, $t_i^{j,\min} = \frac{p_{i,j}^+ - p_{i,j}^-}{v_i^{j,\max}}$ is known as the minimum dwell time on corresponding distance interval. Together with the "non-rollback" running principle for HSTs, one knows that $\tau^\#$ is the theoretically minimum dwell time for a controlled train running on the interval $\left[\left(p_{i,j}, \theta_i^{par-g,j}\right), \left(p_{i,j+1}, \theta_i^{par-g,j+1}\right)\right)$, $\forall\, j \in [N_s]$, with homologous time interval $\left[t_i^{\text{ini}}|\theta_i^{par-g,j}, t_i^{\text{fin}}|\theta_i^{par-g,j+1}\right)$. Without loss of generality, for a specific rail line, the following three situations form all potential running time periods subsets caused by partitioned gradient profiles:

$$\Omega_i : \begin{cases} \text{S1} : t \in \left[t_i^{\text{ini}}|\theta_i^{par-g,1}, t_i^{\text{fin}}|\theta_i^{par-g,2}\right) \cup \cdots \cup \left[t_i^{\text{ini}}|\theta_i^{par-g,\mathcal{S}^1-1}, t_i^{\text{fin}}|\theta_i^{par-g,\mathcal{S}^1}\right],\ t_i^{\text{fin}}|\theta_i^{par-g,\mathcal{S}^1} \in \mathbb{Z}_+ \\ \text{S2} : t \in \left[t_i^{\text{ini}}|\theta_i^{\text{cons}}, t_i^{\text{fin}}|\theta_i^{\text{cons}}\right),\ t_i^{\text{fin}}|\theta_i^{\text{cons}} \in \mathbb{Z}_+ \\ \text{S3} : t \in \left[t_i^{\text{ini}}|\theta_i^{par-g,1}, t_i^{\text{fin}}|\theta_i^{par-g,2}\right) \cup \cdots \cup \left[t_i^{\text{ini}}|\theta_i^{par-g,\mathcal{S}^2-1}, t_i^{\text{fin}}|\theta_i^{par-g,\mathcal{S}^2}\right),\ t_i^{\text{fin}}|\theta_i^{par-g,\mathcal{S}^2} \in \mathbb{Z}_+ \end{cases} \tag{8}$$

herein, $t_i^{\text{fin}}|\theta_i^{par-g,\mathcal{S}^1}$ is the terminal relative time moment for some specific short-term running in S1, and $t_i^{\text{fin}}|\theta_i^{\text{cons}}$ and $t_i^{\text{fin}}|\theta_i^{par-g,\mathcal{S}^2}$ are the terminal relative time moments for some long-terms running in S2 and S3. An illustrative explanation of such three cases is given in Fig. 1 using pictorial demonstrations.

Assumption 1. *The target function pair $[p_{d,i}, v_{d,i}]^\top$ is differentiable, continuous and bounded curves.*

Remark 1. In the aforementioned model description of virtual coupling operation, due to the uncertainties in the formation running environment, let S-point denote the starting point of the train operation and E-point denote the endpoint. Two typical scenarios may occur during operation: 1) Uncertain variations in gradient profiles. As illustrated in Fig. 1 (cases S1–S3), S1: The *distance-gradient* profile exhibits continuous variations throughout the entire segment. S2: The *distance-gradient* profile remains constant across the whole segment. S3: A hybrid scenario combining features of S1 and S2. 2) Variations in fundamental resistance. As indicated by the background color identification in Fig. 1 (Condition 1 to Condition Q), in practical VC operation scenarios, the parameters $a_{co,i}$, $b_{co,i}$ and $c_{co,i}$ are assumed to remain invariant within each condition. However, these parameters may exhibit condition-dependent variations, i.e., their values can differ across distinct operational conditions. Furthermore, Condition 1 to Condition Q represent cases where the basic resistance changes asynchronously with the gradient profile. Together with the simplified representation in Fig. 1, these scenarios form the basis for the modeling considerations in (7) of this study.

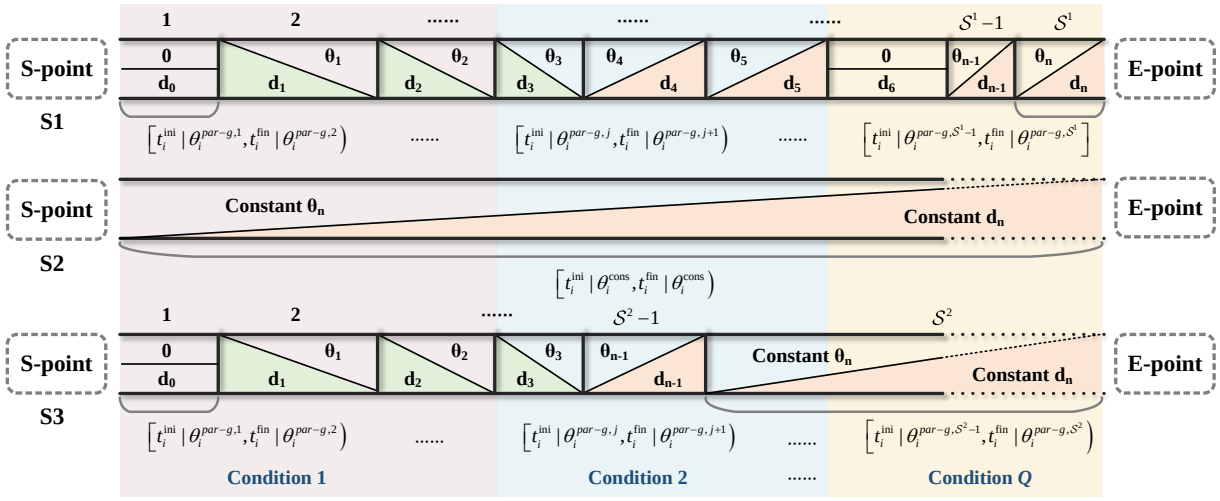


Figure 1: Illustrative explanation of differentially partitioned basic resistance and gradient profile.

2.3. NN estimation

NNs are widely adopted as universal function approximators, capable of reproducing arbitrary continuous nonlinear functions with arbitrary precision. Theoretical studies (as evidenced in references [11, 37, 17, 10]) have demonstrated that any continuous function can be uniformly approximated by a linear combination of Gaussian basis functions. On the basis of Radial basis function neural network (RBF NN), the training architecture utilizes a fixed nonlinear transformation in its hidden layer—devoid of adjustable parameters—to map the input space onto an intermediate space. Subsequently, the output layer performs a linear combination of the outputs from this intermediate layer to produce the network's final output. Subsequently, consider a continuous mapping f_{NN}, the neural network approximation within a compact region Ω_x is mathematically represented as follows:

$$f_{NN}(x) = \mathcal{W}^{\top}\xi(x), \tag{9}$$

where $\mathcal{W} \in \mathbb{R}^q$ is the NN weight vector with q representing the neuron number, and $\xi(x) = [\xi_1(x), \xi_2(x), \cdots, \xi_q(x)]^{\top} \in \mathbb{R}^{q\times 1}$ means basis function vector written as $\xi_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i}\exp\left(-\frac{\|x-c_i\|^2}{2\sigma_i^2}\right) \in \mathbb{R}$, where $\xi_i(x)$ stands for the output of the i-th RBF neuron, x signifies the input feature vector, c_i is the center (mean vector) of the i-th neuron, σ_i refers to the standard deviation (bandwidth) controlling the width of the Gaussian. $\|x - c_i\|$ is the Euclidean distance between x and c_i. For the following derivation, denote $f^*_{NN}(x)$ represents an ideal NN, there exists an ideal weight $\mathcal{W}^* \in \mathbb{R}^{q\times 1}$ which is interpreted as $\mathcal{W}^* = \arg\min_{\mathcal{W}\in\mathbb{R}^q}\{\sup_{x\in\Lambda_x}|f^*_{NN}(x) - \mathcal{W}^{\top}\xi(x)|\}$ such that

$$f^*_{NN}(x) = \mathcal{W}^{*\top}\xi(x) + \epsilon(x). \tag{10}$$

where $|\epsilon(x)| < \bar{\epsilon}$ is a bounded approximation error.

Lemma 1. *Consider the Gaussian RBF neural networks presented in (9), define* $\mathcal{T} := \frac{1}{2}\min_{i\neq j}\left\|c_i - c_j\right\|$. *Invoking that n is the dimension of input vector x, and* σ_k *is the width coefficient of Gaussian kernel function. One can conclude that* $\|\Phi(x)\|$ *is bounded as* $\|\Phi(x)\| \leq \sum_{k=0}^{+\infty} 3n\,(k+2)^{n-1}\exp\left(\frac{-2\mathcal{T}^2k^2}{\sigma_k^2}\right) := \Phi^*$ *[16]. According to ratio test theorem [1], one knows that the infinite series* $\left\{3n\,(k+2)^{n-1}\exp\left(\frac{-2\mathcal{T}^2k^2}{\sigma_k^2}\right)\right\}$, $\forall k = 0, 1, \cdots, +\infty$ *is convergent, meanwhile, it is known that* $\Phi^* \in \mathbb{R}$ *is independent of input vector dimension n and number of implemented neurons q.*

3. Main results

3.1. RL for VC train formation operation

In this section, the optimized formation control strategy is presented based on RL. At first, the tracking error of the system is presented as follows.

$$e_{p,i}(t) = p_i(t) - p_{d,i}(t), \quad e_{v,i}(t) = v_i(t) - v_{d,i}(t) - \alpha_i(t), \ i \in [N] \tag{11}$$

where $\alpha_i(t)$ is intermediate virtual control signal. By (11), $\dot{e}_{p,i}(t) = v_i(t) - v_{d,i}(t) = e_{v,i}(t) + \alpha_i(t)$. The controller α_i and u_i will be further elaborated on for the subsequent content.

Part 1: By (2), denote performance function of $e_{p,i}$ as

$$\mathcal{J}_{i1}(e_{p,i}) = \int_t^\infty \mathcal{C}(e_{p,i}, \alpha_i(e_{p,i}))\mathrm{d}s, \tag{12}$$

where $\mathcal{C}(e_{p,i}, \alpha_i(e_{p,i})) = \mathcal{P}_1 e_{p,i}^2 + \alpha_i^2(e_{p,i})$, $\mathcal{P}_1$ is a positive constant for adjusting the weight of $e_{p,i}$ and α_i. The optimal performance index $\mathcal{J}_{i1}^*(e_{p,i}) = \int_t^\infty \mathcal{C}(e_{p,i}, \alpha_i^*(e_{p,i}))\mathrm{d}s = \min_{u\in\Lambda(\Omega)} \int_t^\infty \mathcal{C}(e_{p,i}, \alpha_i(e_{p,i}))\mathrm{d}s$. By (4), differentiating the both side of $\mathcal{J}_{i1}^*(e_{p,i})$ yields the HJB equation as

$$\mathcal{H}_{i1}\left(\mathcal{J}_{i1}^*, \alpha_i^*, e_{p,i}^*\right) = \mathcal{C}(e_{p,i}, \alpha_i^*(e_{p,i})) + \frac{\mathrm{d}\mathcal{J}_{i1}^*(e_{p,i})}{\mathrm{d}e_{p,i}}\dot{e}_{p,i} = \mathcal{P}_1 e_{p,i}^2 + \alpha_i^2(e_{p,i}) + \frac{\mathrm{d}\mathcal{J}_{i1}^*(e_{p,i})}{\mathrm{d}e_{p,i}}(\alpha_i^* + e_{v,i}) \tag{13}$$

By solving $\partial\mathcal{H}_{i1}/\partial\alpha_i^* = 0$, one can obtain

$$\alpha_i^* = -\frac{1}{2}\frac{\mathrm{d}\mathcal{J}_{i1}^*(e_{p,i})}{\mathrm{d}e_{p,i}} \tag{14}$$

$\mathcal{J}_{i1}^0(e_{p,i})$ in (14) can also be written as

$$\frac{\mathrm{d}\mathcal{J}_{i1}^*(e_{p,i})}{\mathrm{d}e_{p,i}} = 2\kappa_{i1}e_{p,i} + \mathcal{J}_{i1}^0(e_{p,i}) \tag{15}$$

where $\mathcal{J}_{i1}^0(e_{p,i}) = -2\kappa_{i1}e_{p,i} + \mathrm{d}\mathcal{J}_{i1}^*(e_{p,i})/\mathrm{d}e_{p,i}$, adjustable parameters $\kappa_{i1} > 0$. Substituting (15) into (14), one arrives at

$$\alpha_i^* = -\kappa_{i1}e_{p,i} - \frac{1}{2}\mathcal{J}_{i1}^0(e_{p,i}). \tag{16}$$

By means of the estimation principle of NNs in Section 3.1, $\mathcal{J}_{i1}^0(e_{p,i})$ can be restructured as

$$\mathcal{J}_{i1}^0(e_{p,i}) = \mathcal{W}_{i1}^{*\top}\xi_{i1}(e_{p,i}) + \epsilon_{i1}(e_{p,i}) \tag{17}$$

where $\mathcal{W}_{i1}^* \in \mathbb{R}^{q_{i1}}$, q_{i1} is the number of neurons. Substituting (17) into (15) and (16) yields that

$$\frac{\mathrm{d}\mathcal{J}_{i1}^*(e_{p,i})}{\mathrm{d}e_{p,i}} = 2\kappa_{i1}e_{p,i} + \mathcal{W}_{i1}^{*\top}\xi_{i1}(e_{p,i}) + \epsilon_{i1}(e_{p,i}), \tag{18}$$

$$\alpha_i^* = -\kappa_{i1}e_{p,i} - \frac{1}{2}\left(\mathcal{W}_{i1}^{*\top}\xi_{i1}(e_{p,i}) + \epsilon_{i1}(e_{p,i})\right). \tag{19}$$

where $\mathcal{W}_{i1}^*$ is the ideal yet unknown weight. Then, by [37], the critic NN and actor NN can be utilized for the estimation of $\mathrm{d}\mathcal{J}_{i1}^*(e_{p,i})/\mathrm{d}e_{p,i}$ and $\hat{\alpha}_i^*$, therefore, the estimation of the optimal value function and optimal virtual control are given as

$$\frac{\mathrm{d}\hat{\mathcal{J}}_{i1}^*(e_{p,i})}{\mathrm{d}e_{p,i}} = 2\kappa_{i1}e_{p,i} + \hat{\mathcal{W}}_{c,i1}^\top\xi_{i1}(e_{p,i}) \tag{20}$$

$$\hat{\alpha}_i^* = -\kappa_{i1}e_{p,i} - \frac{1}{2}\hat{\mathcal{W}}_{a,i1}^\top\xi_{i1}(e_{p,i}). \tag{21}$$

where $\mathrm{d}\hat{\mathcal{G}}_{i1}^{*}(e_{p,i})/\mathrm{d}e_{p,i}$ and $\hat{\alpha}_{i}^{*}$ are the estimation of $\mathrm{d}\mathcal{G}_{i1}^{*}(e_{p,i})/\mathrm{d}e_{p,i}$ and α_{i}^{*}, the critic weight $\hat{\mathcal{W}}_{c,i1}$ and actor weight $\hat{\mathcal{W}}_{a,i1}$ are provided as

$$\dot{\hat{\mathcal{W}}}_{c,i1} = -\iota_{c,i1}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{c,i1}(t) \tag{22a}$$

$$\dot{\hat{\mathcal{W}}}_{a,i1} = -\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\left(\iota_{a,i1}\hat{\mathcal{W}}_{a,i1} - \iota_{c,i1}\hat{\mathcal{W}}_{c,i1}\right) \tag{22b}$$

where $\iota_{c,i1}$ and $\iota_{a,i1} > 0$ are critic gain parameter and actor gain parameter; $\sigma_{i1} > 0$ is modification parameters, and I_{q_1} is the $q_1 \times q_1$ unit matrix, q_1 represents the neuron number. $\hat{\mathcal{W}}_{c,i1}$ and $\hat{\mathcal{W}}_{a,i1}$ are respectively the estimation of $\mathcal{W}_{i1}^{*}$ for the critic and actor part.

Proposition 1. *The parameters* $\iota_{a,i1} > \frac{1}{2}$, $\iota_{a,i1} > \iota_{c,i1} > 0$, $\kappa_{i1} > \frac{3}{4}$.

Proof. Define weight estimation of actor and critic $\tilde{\mathcal{W}}_{a,i1} = \hat{\mathcal{W}}_{a,i1} - \mathcal{W}_{i1}^{*}$, $\tilde{\mathcal{W}}_{c,i1} = \hat{\mathcal{W}}_{c,i1} - \mathcal{W}_{i1}^{*}$, let Lyapunov function $V_{i1} = \frac{1}{2}e_{p,i}^{2} + \frac{1}{2}\tilde{\mathcal{W}}_{c,i1}^{\top}\tilde{\mathcal{W}}_{c,i1} + \frac{1}{2}\tilde{\mathcal{W}}_{a,i1}^{\top}\tilde{\mathcal{W}}_{a,i1}$, through differentiating V_{i1}, by (21) and Young's inequality, we obtain

$$\begin{aligned}
\dot{V}_{i1} =& e_{p,i}\dot{e}_{p,i} + \tilde{\mathcal{W}}_{c,i1}^{\top}\dot{\hat{\mathcal{W}}}_{c,i1} + \tilde{\mathcal{W}}_{a,i1}^{\top}\dot{\hat{\mathcal{W}}}_{a,i1} \\
=& e_{p,i}(e_{v,i} + \alpha_i) + \tilde{\mathcal{W}}_{c,i1}^{\top}\Big[-\iota_{c,i1}(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1})\hat{\mathcal{W}}_{c,i1}(t)\Big] + \tilde{\mathcal{W}}_{a,i1}^{\top}\Big[-\Big(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) \\
&+\sigma_{i1}I_{q_1}\Big)\times\Big(\iota_{a,i1}\hat{\mathcal{W}}_{a,i1} - \iota_{c,i1}\hat{\mathcal{W}}_{c,i1}\Big)\Big] \\
=& e_{p,i}e_{v,i} + e_{p,i}\left(-\kappa_{i1}e_{p,i} - \frac{1}{2}\hat{\mathcal{W}}_{a,i1}^{\top}\xi_{i1}(e_{p,i})\right) - \iota_{c,i1}\tilde{\mathcal{W}}_{c,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{c,i1} \\
&- \iota_{a,i1}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} + \iota_{c,i1}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{c,i1} \\
\leq& \frac{1}{2}e_{p,i}^{2} + \frac{1}{2}e_{v,i}^{2} - \kappa_{i1}e_{p,i}^{2} - \frac{1}{2}e_{p,i}\mathcal{W}_{a,i1}^{\top}\xi_{i1}(e_{p,i}) - \iota_{c,i1}\tilde{\mathcal{W}}_{c,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} \\
&- \iota_{a,i1}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} + \iota_{c,i1}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{c,i1}
\end{aligned} \tag{23}$$

By the expression of $\tilde{\mathcal{W}}_{a,i1}$ and $\tilde{\mathcal{W}}_{c,i1}$ above (23), one can obtain

$$\begin{aligned}
&-\iota_{a,i1}\tilde{\mathcal{W}}_{a,i1}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} \\
&\quad\leq -\frac{\iota_{a,i1}}{2}\Big[\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\tilde{\mathcal{W}}_{a,i1} + \hat{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} \\
&\qquad+\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^{*}\Big]
\end{aligned} \tag{24}$$

Accordingly, substituting (24) into (23), it arrives

$$\begin{aligned}
\dot{V}_{i1} \leq& \frac{1}{2}e_{p,i}^{2} + \frac{1}{2}e_{v,i}^{2} - \kappa_{i1}e_{p,i}^{2} + \frac{1}{4}e_{p,i}^{2} + \frac{1}{4}\hat{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} \\
&- \frac{\iota_{a,i1} - \iota_{c,i1}}{2}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\tilde{\mathcal{W}}_{a,i1} - \frac{\iota_{c,i1}}{2}\tilde{\mathcal{W}}_{c,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\tilde{\mathcal{W}}_{c,i1} \\
&- \frac{\iota_{a,i1}}{2}\hat{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} + \frac{\iota_{a,i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^{*} \\
&+ \frac{\iota_{c,i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^{*} \\
\leq& -(\kappa_{i1} - \frac{3}{4})e_{p,i}^{2} + \frac{1}{2}e_{v,i}^{2} - \left(\frac{\iota_{a,i1}}{2} - \frac{1}{4}\right)\times\hat{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} \\
&- \frac{\iota_{a,i1} - \iota_{c,i1}}{2}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\tilde{\mathcal{W}}_{a,i1} - \frac{\iota_{c,i1}}{2}\tilde{\mathcal{W}}_{c,i1}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right) + \mathcal{B}_{i1} \\
\leq& -\mathcal{P}_1 V_{i1} + \frac{1}{2}e_{v,i}^{2} + \mathcal{B}_{i1}
\end{aligned} \tag{25}$$

with the presentation of given V_{i1}, $\mathcal{B}_{i1} = \frac{\iota_{a,i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^{*} + \frac{\iota_{c,i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^{*}$, and $\mathcal{P}_1 = \min\left\{2(\kappa_{i1} - \frac{3}{4}), (\iota_{a,i1} - \iota_{c,i1})\,\sigma_{i1}, \iota_{c,i1}\sigma_{i1}\right\}$. According to Lemma 1, $\mathcal{B}_{i1}$ is bounded. At this point, the given

parameter conditions in the Proposition 1 may potentially ensure the negative definiteness property. Subsequently, Proposition 2 will provide further analysis regarding $e_{v,i}$. □

Part 2: Denote performance function of $e_{v,i}$ as

$$\mathcal{J}_{i2}(e_{v,i}) = \int_t^{\infty} \mathcal{C}(e_{v,i}, u_i(e_{v,i}))\mathrm{d}s, \tag{26}$$

where $\mathcal{C}(e_{v,i}, u_i(e_{v,i})) = \mathcal{P}_2 e_{v,i}^2 + u_i^2(e_{v,i})$, $\mathcal{P}_2$ is a positive constant for adjusting the weight of $e_{v,i}$ and u_i. The optimal performance index $\mathcal{J}_{i2}^*(e_{v,i}) = \int_t^{\infty} \mathcal{C}(e_{v,i}, u_i^*(e_{v,i}))\mathrm{d}s = \min_{u\in\Lambda(\Omega)} \int_t^{\infty} \mathcal{C}(e_{v,i}, u_i(e_{v,i}))\mathrm{d}s$. By virtue of (11), one can get $\dot{e}_{v,i} = \dot{v}_i - \dot{v}_{d,i} - \dot{\hat{\alpha}}_i^* = u_i^* + \mathcal{F}_i - \dot{v}_{d,i} - \dot{\hat{\alpha}}_i^*$.

By differentiating the both side of $\mathcal{J}_{i2}^*(e_{v,i})$, we obtain HJB equation

$$\mathcal{H}_{i2}\left(\mathcal{J}_{i2}^*, u_i^*, e_{v,i}^*\right) = \mathcal{C}(e_{v,i}, u_i^*(e_{v,i})) + \frac{\mathrm{d}\mathcal{J}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}}\dot{e}_{v,i} = \mathcal{P}_2 e_{v,i}^2 + u_i^{*2} + \frac{\mathrm{d}\mathcal{J}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}}(u_i^* + \mathcal{F}_i - \dot{v}_{d,i} - \hat{\alpha}_i^*) = 0. \tag{27}$$

Solving $\partial\mathcal{H}_{i2}/\partial u_i^* = 0$, one can obtain

$$u_i^* = -\frac{1}{2}\frac{\mathrm{d}\mathcal{J}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}} \tag{28}$$

It can also be written as

$$\frac{\mathrm{d}\mathcal{J}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}} = 2\kappa_{i2}e_{v,i} + 2\mathcal{F}_i + \mathcal{J}_{i2}^0(e_{v,i}) \tag{29}$$

where $\mathcal{J}_{i2}^0(e_{v,i}) = -2\kappa_{i2}e_{v,i} - 2\mathcal{F}_i + \frac{\mathrm{d}\mathcal{J}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}}$. Substituting (15) into (14) yields that

$$u_i^* = -\kappa_{i2}e_{v,i} - \mathcal{F}_i - \frac{1}{2}\mathcal{J}_{i2}^0(e_{v,i}). \tag{30}$$

$\mathcal{J}_{i2}^0(e_{v,i})$ can be estimated by

$$\mathcal{J}_{i2}^0(e_{v,i}) = \mathcal{W}_{i2}^{*\top}\xi_{i2}(e_{v,i}) + \epsilon_{i2}(e_{v,i}), \tag{31}$$

where $\mathcal{W}_{i2}^{*\top} \in \mathbb{R}^{q_{i2}\times 1}$, q_{i2} is the number of neurons. Substituting (31) into (29) and (30), one can obtain

$$\frac{\mathrm{d}\mathcal{J}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}} = 2\kappa_{i2}e_{v,i} + 2\mathcal{F}_i + \mathcal{W}_{i2}^{*\top}\xi_{i2}(e_{v,i}) + \epsilon_{i2}(e_{v,i}) \tag{32}$$

$$u_i^* = -\kappa_{i2}e_{v,i} - \mathcal{F}_i - \frac{1}{2}\left[\mathcal{W}_{i2}^{*\top}\xi_{i2}(e_{v,i}) + \epsilon_{i2}(e_{v,i})\right]. \tag{33}$$

Similar to the previous steps, the designed $\frac{\mathrm{d}\hat{\mathcal{J}}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}}$ and $\hat{u}_i^*$ by the critic weight presentation $\hat{\mathcal{W}}_{c,i1}$ and actor weight presentation $\hat{\mathcal{W}}_{a,i1}$ in RL are provided to estimate the $\frac{\mathrm{d}\mathcal{J}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}}$ and u_i^*:

$$\frac{\mathrm{d}\hat{\mathcal{J}}_{i2}^*(e_{v,i})}{\mathrm{d}e_{v,i}} = 2\kappa_{i2}e_{v,i} + 2\mathcal{F}_i + \hat{\mathcal{W}}_{c,i2}^{\top}\xi_{i2}(e_{v,i}) \tag{34}$$

$$\hat{u}_i^* = -\kappa_{i2}e_{v,i} - \mathcal{F}_i - \frac{1}{2}\hat{\mathcal{W}}_{a,i2}^{\top}\xi_{i2}(e_{v,i}). \tag{35}$$

and the critic weight $\hat{\mathcal{W}}_{c,i2}$ and actor weight $\hat{\mathcal{W}}_{a,i2}$ are designed as

$$\dot{\hat{\mathcal{W}}}_{c,i2} = -\iota_{c,i2}\left(\xi_{i2}(e_{v,i})\xi_{i2}^{\top}(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\hat{\mathcal{W}}_{c,i2}(t) \tag{36a}$$

$$\dot{\hat{\mathcal{W}}}_{a,i2} = -\left[\xi_{i2}(e_{v,i})\xi_{i2}^{\top}(e_{v,i}) + \sigma_{i2}I_{q_2}\right]\left(\iota_{a,i2}\hat{\mathcal{W}}_{a,i2} - \iota_{c,i2}\hat{\mathcal{W}}_{c,i2}\right) \tag{36b}$$

where $\iota_{c,i2} > 0$, and $\iota_{a,i2} > 0$ are critic gain parameter and actor gain parameter; $\sigma_{i2} > 0$ is modification parameters, and I_{q_2} is the $q_2 \times q_2$ unit matrix, q_2 represents the neuron number. $\hat{\mathcal{W}}_{c,i2}$ and $\hat{\mathcal{W}}_{a,i2}$ are respectively the estimation of $\mathcal{W}_{i2}^*$ for the critic and actor part.

Proposition 2. *The parameters $\iota_{a,i2} > \frac{1}{2}$, $\iota_{a,i2} > \iota_{c,i2} > 0$, $\kappa_{i2} > 2$.*

Proof. Define positive definite Lyapunov function $V_{i2} = \frac{1}{2}e_{v,i}^2 + \frac{1}{2}\tilde{\mathcal{W}}_{c,i2}^\top\tilde{\mathcal{W}}_{c,i2} + \frac{1}{2}\tilde{\mathcal{W}}_{a,i2}^\top\tilde{\mathcal{W}}_{a,i2}$, through differentiating V_{i2}, by (21) and Young's inequality, we obtain

$$
\begin{aligned}
\dot{V}_{i2} =& e_{v,i}\dot{e}_{v,i} + \tilde{\mathcal{W}}_{c,i2}^\top\dot{\hat{\mathcal{W}}}_{c,i2} + \tilde{\mathcal{W}}_{a,i2}^\top\dot{\hat{\mathcal{W}}}_{a,i2} \\
=& e_{v,i}(u_i^* + \mathcal{F}_i - \dot{v}_{d,i} - \dot{\hat{\alpha}}_i^*) + \tilde{\mathcal{W}}_{c,i2}^\top\Big[- \iota_{c,i2}(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2} \times I_{q_2})\hat{\mathcal{W}}_{c,i2}(t)\Big] \\
&+ \tilde{\mathcal{W}}_{a,i2}^\top\Big[- (\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}) \times \Big(\iota_{a,i2}\hat{\mathcal{W}}_{a,i2} - \iota_{c,i2}\hat{\mathcal{W}}_{c,i2}\Big)\Big] \\
=& e_{v,i}\left[-\kappa_{i2}e_{v,i} - \frac{1}{2}\hat{\mathcal{W}}_{a,i2}^\top\xi_{i2}(e_{v,i}) - \dot{v}_{d,i} - \dot{\hat{\alpha}}_i^*\right] - \iota_{c,i2}\tilde{\mathcal{W}}_{c,i2}^\top(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2})\hat{\mathcal{W}}_{c,i2}(t) \\
&- \iota_{a,i2}\tilde{\mathcal{W}}_{a,i2}^\top(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2})\hat{\mathcal{W}}_{a,i2}(t) + \iota_{c,i2}\tilde{\mathcal{W}}_{a,i2}^\top(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2})\hat{\mathcal{W}}_{c,i2}(t) \\
\leq& - \kappa_{i2}e_{v,i}^2 + \frac{1}{4}\tilde{\mathcal{W}}_{a,i2}^\top\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i})\hat{\mathcal{W}}_{a,i2} - \frac{\iota_{c,i2}}{2}\tilde{\mathcal{W}}_{c,i2}^\top\left(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\tilde{\mathcal{W}}_{c,i2} \\
&- \frac{\iota_{a,i2}}{2}\hat{\mathcal{W}}_{a,i2}^\top\left(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\hat{\mathcal{W}}_{a,i2} + \frac{\iota_{a,i2}}{2}\mathcal{W}_{i2}^{*\top}\left(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\mathcal{W}_{i2}^* \\
&+ \frac{\iota_{c,i2}}{2}\mathcal{W}_{i2}^{*\top}\left(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\mathcal{W}_{i2}^* - \frac{\iota_{a,i2} - \iota_{c,i2}}{2}\tilde{\mathcal{W}}_{a,i2}^\top\left(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\right) \\
&\times \tilde{\mathcal{W}}_{a,i2} + 2e_{v,i}^2 + \frac{1}{2}\dot{\hat{\alpha}}_i^{*2} + \frac{1}{2}\dot{v}_{d,i}^2 \\
\leq& - (\kappa_{i2} - 2)\, e_{v,i}^2 + \left(\frac{1}{4} - \frac{\iota_{a,i2}}{2}\right)\hat{\mathcal{W}}_{a,i2}^\top[\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}]\hat{\mathcal{W}}_{a,i2} \\
&- \frac{\iota_{a,i2} - \iota_{c,i2}}{2}\tilde{\mathcal{W}}_{a,i2}^\top\Big(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\Big)\tilde{\mathcal{W}}_{a,i2} + \mathcal{B}_{i2}. \\
\leq& - \mathcal{P}_{i2}V_{i2} + \mathcal{B}_{i2}
\end{aligned}
\tag{37}
$$

where $\mathcal{P}_{i2} = \min\{2\kappa_{i2} - 4, (\iota_{a,i2} - \iota_{c,i2})\sigma_{i2}, \iota_{c,i2}\sigma_{i2}\}$, $\mathcal{B}_{i2} = \frac{1}{2}\dot{\hat{\alpha}}_i^{*2} + \frac{1}{2}\dot{v}_{d,i}^2 + + \frac{\iota_{a,i2}}{2}\mathcal{W}_{i2}^{*\top}\left(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\mathcal{W}_{i2}^* + \frac{\iota_{c,i2}}{2}\mathcal{W}_{i2}^{*\top}\left(\xi_{i2}(e_{v,i})\xi_{i2}^\top(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\mathcal{W}_{i2}^*$. In virtue of Lemma 1, $\mathcal{B}_{i2}$ is bounded. Therefore, it can be concluded that when the parameter conditions in the Proposition 2 are met, the negative-definite property of (37) is guaranteed with the parameter conditions in Proposition 2. The proof of Proposition 2 completes. □

Theorem 1. *Consider the VC train operation dynamics in (7), if Assumption 1 holds, the RL-based optimized formation controllers $\hat{\alpha}_i^*$ in (21) and $\hat{u}_i^*$ in (35), designed with the prerequisites Proposition 1 and Proposition 2, together with the AC weight update presentation $\dot{\hat{\mathcal{W}}}_{a,ij}$ in (22b) and (36b), and $\dot{\hat{\mathcal{W}}}_{c,ij}$ in (22a) and (36a), $i \in [N]$, $j = 1, 2$, guaranteeing that $\mathcal{I}_{i1}(e_{p,i}) + \mathcal{I}_{i2}(e_{v,i})$ are minimized; $e_{p,i} \leq \sqrt{\mathcal{B}_i/\mathcal{P}_i}$ with $t \to \infty$; and the resulting closed-loop system is semi-globally uniformly ultimately bounded (SGUUB).*

Proof. Denote a positive definite Lyapunov function as $V_i = V_{i1} + V_{i2}$, $i \in [N]$, according to the proof of Proposition 1 and Proposition 2, for *i*th train, one can obtain

$$
\dot{V}_i(t) \leq -\mathcal{P}_i V + \mathcal{B}_i, \tag{38}
$$

According to the characteristic equation method for solving ordinary differential equations $\dot{V}_i(t) = -\mathcal{P}_i V + \mathcal{B}_i$, with $\mathcal{P}_i = \min_{i\in[N]}\{\mathcal{P}_{i1}, \mathcal{P}_{i2}\}$, $\mathcal{B}_i = \sum_{i=1}^n \mathcal{B}_{i1} + \mathcal{B}_{i2}$, the solution of the expression can be divided into the following steps.

Step 1: The general solution of the homogeneous equation part $\dot{V}_i(t) = -\mathcal{P}_i V$ is obviously given as $V_i(t) = C \cdot \exp\{-\mathcal{P}_i t\}$, $C \in \mathbb{R}$. *Step 2*: The particular solution $V_i(t)$ of the differential equation $\dot{V}_i(t) = -\mathcal{P}_i V + \mathcal{B}_i$ is obtained by solving the equilibrium condition: $-\mathcal{P}_i V + \mathcal{B}_i = 0$, yielding the steady-state solution $V_i = \mathcal{B}_i/\mathcal{P}_i$. *Step 3*: Combination of the general solution and the special solution yields the solution of $V_i(t)$, i.e, $V_i(t) = C \cdot \exp\{-\mathcal{P}_i t\} + \mathcal{B}_i/\mathcal{P}_i$, $C \in \mathbb{R}$. It

is obvious that $\lim_{t\to\infty} V_i(t) = \mathcal{B}_i/\mathcal{P}_i$, so its constituent elements $e_{p,i}$, $\tilde{\mathcal{W}}_{c,i1}$,$\tilde{\mathcal{W}}_{a,i1}$, $e_{v,i}$, $\tilde{\mathcal{W}}_{c,i2}$,$\tilde{\mathcal{W}}_{a,i2}$ are bounded in a finite time. And $t \to \infty$, $e_{p,i}$ is kept within the boundary $\sqrt{\mathcal{B}_i/\mathcal{P}_i}$. The proof completes. □

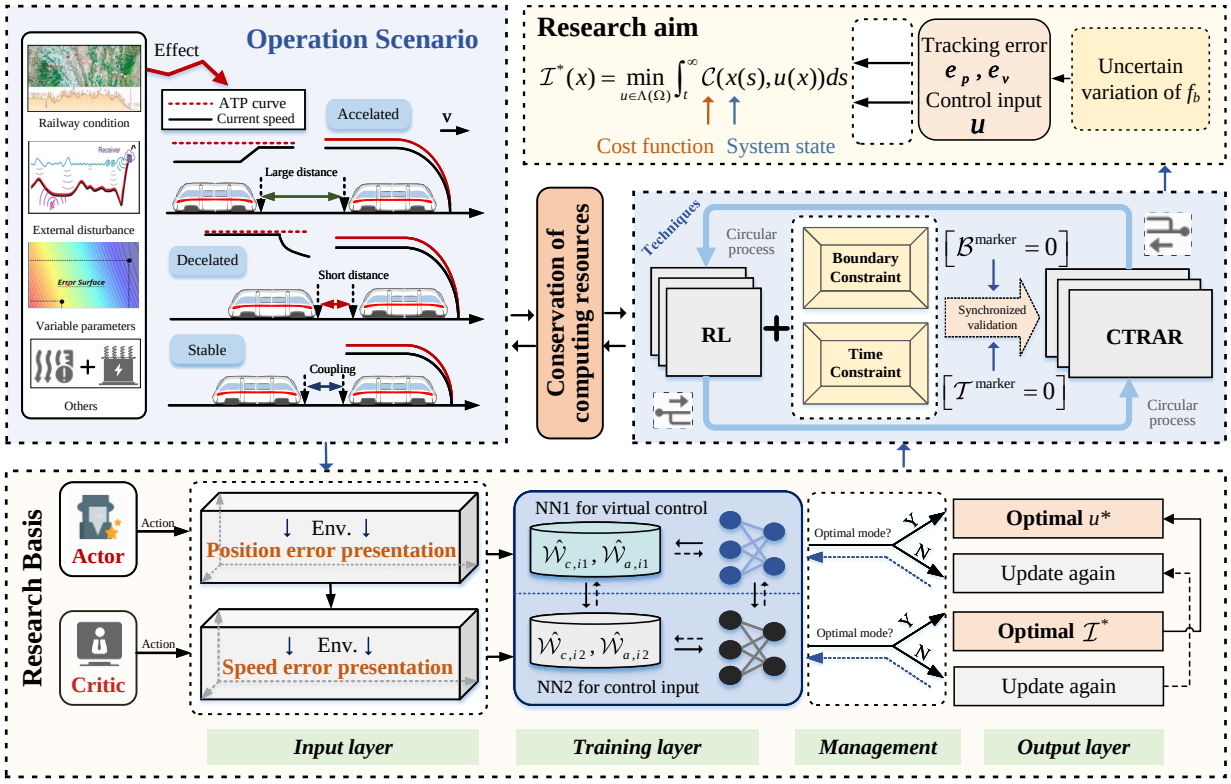


Figure 2: Logic framework.

3.2. *CTRAR for VC train formation operation*

In the previous section, a RL-based NN estimation approach was established, through which the convergence of tracking errors was achieved in multi-train VC systems under the objective of cost function minimization via an online learning mechanism. However, significant uncertainties are exhibited in practical rail transit environments (e.g., load fluctuations, track disturbances, etc.), substantial uncertainties and time-varying disturbances inherent in real-world rail transit environments necessitate continuous online adaptation of pre-trained network weights. While, computational overload is induced by frequent weight retraining, particularly for the limited computing power of onboard edge devices. To resolve this fundamental trade-off between adaptation necessity and computational feasibility, this section proposes a CTRAR strategy featuring: 1) A performance monitoring mechanism with stability-guaranteed thresholds; 2) Switching-time constrained weight updates; 3) The convergence-triggered execution logic is implemented within the proposed framework, with the complete logical architecture presented in Fig. 2 and illustrated through subsequent remark.

Theorem 2. *Consider the system dynamics governed by* (7) *over a partitioned time domain* $\mathcal{T} := \bigcup_{q=0}^{Q}[t_q, t_{q+1})$, $t_0 = 0$, $Q \in \mathbb{N}$, *where each subinterval* $T_q := [t_q, t_{q+1})$ *admits a piecewise continuously differentiable function* $f_b(\cdot)$ *characterizing the system's time-varying behavior. Then, incorporating boundary constraints on system states, dwell-time requirements for mode transitions and adaptive interval partitioning criteria, a performance monitor integrates the following characteristics:*

1) The weight update triggering condition $\mathcal{B}_j^{marker}$ *is defined as:*

$$\mathcal{B}_j^{marker} = \begin{cases} 0, \ \forall\|e_{p,i}\| < B_{e_{p,i}} \wedge \forall\|e_{v,i}\| < B_{e_{v,i}}, \\ \quad \forall\|\hat{\tilde{\mathcal{W}}}_i\| < B_{\hat{\tilde{\mathcal{W}}}_i}, \forall i \in [N]. \\ 1, \ \textit{others}. \end{cases} \tag{39}$$

where $\mathcal{B}_j^{marker} = 0$ *and* $\mathcal{B}_j^{marker} = 1$ *indicate the suspension and activation of weight updates, respectively,* $j \in \mathbb{N}$. $\|\dot{\hat{\mathcal{W}}}_i\| = \sum_{j=1}^{2} \|\dot{\hat{\mathcal{W}}}_{c,ij}\| + \sum_{j=1}^{2} \|\dot{\hat{\mathcal{W}}}_{a,ij}\|$, $\hat{\mathcal{W}}_{c,i1}(\hat{\mathcal{W}}_{a,i1})$ *represents the norm of the weight matrices for the critic (actor) network in the first-stage virtual control law* $\hat{\alpha}_i^*$ *of the ith VC train formation.* $\hat{\mathcal{W}}_{c,i2}$ $(\hat{\mathcal{W}}_{a,i2})$ *represents the norm of the weight matrices for the critic (actor) network in the second-stage controller* $\hat{u}_i^*$ *of the ith train in the VC formation.* $B_{e_{p,i}}$, $B_{e_{v,i}}$, $B_{\hat{\mathcal{W}}_i}$ *are the corresponding preset upper bounds of the estimates of* $\|e_{p,i}\|$, $\|e_{v,i}\|$ *and* $\|\hat{\mathcal{W}}_i\|$.

2) The dwell-time triggering indicator $\mathcal{T}_j^{marker}$ *is defined as:*

$$\mathcal{T}_j^{marker} = \begin{cases} 0, & \Delta t_k > \mathcal{T}_{dwell} \\ 1, & others. \end{cases} \tag{40}$$

where Δt_k *represents the duration of "* $\mathcal{B}_i^{marker} = 0$*",* $\mathcal{T}_{dwell} \in \mathbb{R}$.

3) The mode partition can be presented as:

$$\begin{aligned} &I_{j+1}^{mode}(RL \rightarrow CTRAR \; or \; CTRAR \rightarrow RL) \\ &= \begin{cases} 0, & if \underbrace{\mathcal{B}_j^{marker} = 0 \wedge \mathcal{T}_j^{marker} = 0}_{Optimal\ conditions} \\ 1, & if \underbrace{\mathcal{B}_j^{marker} \neq 0 \vee \mathcal{T}_j^{marker} \neq 0}_{Non\text{-}optimal\ conditions} \end{cases} \end{aligned} \tag{41}$$

where $I_{j+1}^{mode} = 0$ $(I_{j+1}^{mode} = 1)$ *represents the forthcoming optimal mode (non-optimal mode),* $j \in \mathbb{N}$. *Define* $t_{1,0}^{trigger}$ *as the time convergent time trigger point for the pattern from 1 to 0, and* $t_{0,1}^{trigger}$ *as the nonconvergent time trigger point from 0 to 1, and the corresponding counts are denoted as* $\Delta CN_{1,0}$ *and* $\Delta CN_{0,1}$. *The total mode transition count* ΔCN*, derived from the marker conditions in (39)-(41), is given by*

$$\Delta CN = \Delta CN_{1,0} + \Delta CN_{0,1} = \begin{cases} 1, & I_{j+1}^{mode} \neq I_j^{mode}, \\ 0, & I_{j+1}^{mode} = I_j^{mode}. \end{cases} \tag{42}$$

Proof. The finiteness of switching instants is guaranteed by the proposed performance monitoring mechanism, where switching occurrences are defined as transitions between distinct operational modes. As specified in (42), the switching count ΔCN is incremented if and only if a mode transition is detected (i.e., when $I_{j+1}^{\text{mode}} \neq I_j^{\text{mode}}$, $j \in \mathbb{N}$).

The mode definitions in (41) establishes two discrete states: RL and CTRAR, with transitions classified into optimal mode and non-optimal mode categories. For the optimal mode, switching is permitted only when the following conditions are satisfied: 1) The real-time variations of $\|e_{p,i}\|$, $\|e_{v,i}\|$ and $\|\hat{\mathcal{W}}_i\|$ are maintained below their respective thresholds $B_{e_{p,i}}$, $B_{e_{v,i}}$ and $B_{\hat{\mathcal{W}}_i}$. 2) The duration in optimal mode exceeds the prescribed dwell time $\mathcal{T}_{\text{dwell}}$. When these conditions are met, the count defined in (42) is automatically incremented by one unit. Consequently, the total switching count $CN = CN_0 + \Delta CN = \Delta CN$, where ΔCN represents the accumulated transitions. As proven UUB of the system states guaranteed within finite time in Theorem 1, thus ΔCN remains finite for any bounded time interval. This conclusively establishes the finiteness property of the switching mechanism.

Conversely, the existence of infinitely many transitions between optimal and non-optimal states by (41) would violate the minimum dwell-time constraint imposed on mode switching $\mathcal{T}_{\text{dwell}}$ in (40), resulting in a fundamental contradiction. The proposed switching mechanism, governed by threshold-bound conditions ($\forall\|e_{p,i}\| < B_{e_{p,i}} \wedge \forall\|e_{v,i}\| < B_{e_{v,i}}, \forall\|\dot{\hat{\mathcal{W}}}_i\| < B_{\hat{\mathcal{W}}_i}, \forall i \in [N]$.) and mandatory dwell-time adherence $\Delta t_k > \mathcal{T}_{\text{dwell}}$, ensures that only a finite number of mode transitions can occur over any bounded time horizon. This design rigorously precludes Zeno behavior by enforcing a minimum temporal separation between successive switches, thus, the switching logic guarantees both mathematical soundness and physical realizability in practical control execution. □

Remark 2. The logical framework for this study is presented in Fig. 2. As illustrated, the logical framework diagram in Fig. 2 encompasses several categories, including operational scenarios, research basis, research objectives, and research techniques. Building upon the original RL algorithm, CTRAR introduces a convergence-triggered constraint mechanism to ensure the safe and stable operation of VC trains while conserving resources. Specifically, the continuous updating of RL training weights will halt only when both boundary constraints and time constraints are satisfied.

However, if any boundary is violated, the system will resume weight updates. This mechanism offers the following advantages: 1) The adopted mechanism operates under multiple constraints and is underpinned by rigorous stability theory, thereby ensuring the acquisition of an optimal stability mode based on the RL algorithm. 2) RL does not require consecutive weight updates; the intermittent cessation of computational resources, which meets safe operational standards, facilitates its application in scenarios with limited computational resources.

Theorem 3. *Consider the VC train operation dynamics in (7), if Assumption 1 holds, the RL-based optimized formation controllers $\hat{\alpha}_i^*$ in (21) and $\hat{u}_i^*$ in (35), designed with the prerequisites Proposition 1 and Proposition 2, together with the rewritten AC weight update presentation $\dot{\hat{\mathcal{W}}}_{a,ij}$ and $\dot{\hat{\mathcal{W}}}_{c,ij}$, $i \in [N]$, $j = 1, 2$, ensuring that $\mathcal{I}_{i1}(e_{p,i}) + \mathcal{I}_{i2}(e_{v,i})$ are minimized; $e_{p,i} \leq \sqrt{\mathcal{B}_i/\mathcal{P}_i}$ with $t \to \infty$ and the resulting closed-loop system is SGUUB. Meanwhile, the referred $\dot{\hat{\mathcal{W}}}_{a,ij}$ and $\dot{\hat{\mathcal{W}}}_{c,ij}$ is specified as follows:*

$$\dot{\hat{\mathcal{W}}}_{c,i1} = -\iota_{c,i1}\aleph_{i1}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{c,i1}(t) \tag{43a}$$

$$\dot{\hat{\mathcal{W}}}_{a,i1} = -\left[\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right]\left(\iota_{a,i1}\aleph_{i1}\hat{\mathcal{W}}_{a,i1} - \iota_{c,i1}\aleph_{i1}\hat{\mathcal{W}}_{c,i1}\right) \tag{43b}$$

and

$$\dot{\hat{\mathcal{W}}}_{c,i2} = -\iota_{c,i2}\aleph_{i2}\left(\xi_{i2}(e_{v,i})\xi_{i2}^{\top}(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\hat{\mathcal{W}}_{c,i2}(t) \tag{44a}$$

$$\dot{\hat{\mathcal{W}}}_{a,i2} = -\left[\xi_{i2}(e_{v,i})\xi_{i2}^{\top}(e_{v,i}) + \sigma_{i2}I_{q_2}\right]\left(\iota_{a,i2}\aleph_{i2}\hat{\mathcal{W}}_{a,i2} - \iota_{c,i2}\aleph_{i2}\hat{\mathcal{W}}_{c,i2}\right) \tag{44b}$$

where $\iota_{c,ij}$, and $\iota_{a,ij} > 0$ are critic gain parameter and actor gain parameter; $\sigma_{ij} > 0$ is modification parameters, and I_{q_j} is the $q_j \times q_j$ unit matrix. $\hat{\mathcal{W}}_{c,ij}$ and $\hat{\mathcal{W}}_{a,ij}$ are respectively the estimation of $\mathcal{W}_{ij}^*$ for the critic and actor part. The meaning of $\aleph_{ij}$ ($i \in [N]$; $j = 1, 2$) is delineated below:

$$\aleph_{i1} = I_{q_1} + \operatorname{diag}\left\{g_{i1}\exp\left(\frac{-\|e_{p,i}\|}{2\sigma_{i1}^2}\right)\left(1 + \frac{1}{\delta_{i1}t_{effect}}\right)\right\}, \aleph_{i2} = I_{q_2} + \operatorname{diag}\left\{g_{i2}\exp\left(\frac{-\|e_{v,i}\|}{2\sigma_{i2}^2}\right)\left(1 + \frac{1}{\delta_{i2}t_{effect}}\right)\right\}, \tag{45}$$

where $g_{ij}, \delta_{ij} > 0$ are positive constant, $I_{q_j} \in \mathbb{R}^{q_j}$, σ_{ij} is the width of Gauss function, $i \in [N]$; $j = 1, 2$. And t_{effect} is dynamic time points which is given by

$$t_{effect} = \begin{cases} t - t_{1,0}^{trigger}, & I_{j+1}^{mode} \neq I_j^{mode}, \\ t, & I_{j+1}^{mode} = I_j^{mode}. \end{cases} \tag{46}$$

where $t_{1,0}^{trigger}$ is convergent time trigger point for the pattern from 1 to 0, and the denotation of I_j^{mode} is provided in (41).

Proof. By (43), denote Lyapunov function $\bar{V}_{i1} = \frac{1}{2}e_{p,i}^2 + \frac{1}{2}\tilde{\mathcal{W}}_{c,i1}^{\top}\tilde{\mathcal{W}}_{c,i1} + \frac{1}{2}\tilde{\mathcal{W}}_{a,i1}^{\top}\tilde{\mathcal{W}}_{a,i1}$, through differentiating $\bar{V}_{i1}$, the similar process from (23)∼(25) yields that

$$\begin{aligned}
\dot{\bar{V}}_{i1} \leq & \frac{1}{2}e_{p,i}^2 + \frac{1}{2}e_{v,i}^2 - \kappa_{i1}e_{p,i}^2 + \frac{1}{4}e_{p,i}^2 + \frac{1}{4}\hat{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} \\
& - \frac{(\iota_{a,i1} - \iota_{c,i1})\aleph_{i1}}{2}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\tilde{\mathcal{W}}_{a,i1} - \frac{\iota_{c,i1}\aleph_{i1}}{2}\tilde{\mathcal{W}}_{c,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\tilde{\mathcal{W}}_{c,i1} \\
& - \frac{\iota_{a,i1}\aleph_{i1}}{2}\hat{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} + \frac{\iota_{a,i1}\aleph_{i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^* \\
& + \frac{\iota_{c,i1}\aleph_{i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^* \\
\leq & -(\kappa_{i1} - \frac{3}{4})e_{p,i}^2 + \frac{1}{2}e_{v,i}^2 - \left(\frac{\iota_{a,i1}\aleph_{i1}}{2} - \frac{1}{4}\right)\times\hat{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\hat{\mathcal{W}}_{a,i1} \\
& - \frac{(\iota_{a,i1} - \iota_{c,i1})\aleph_{i1}}{2}\tilde{\mathcal{W}}_{a,i1}^{\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\tilde{\mathcal{W}}_{a,i1} - \frac{\iota_{c,i1}\aleph_{i1}}{2}\tilde{\mathcal{W}}_{c,i1}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right) + \mathcal{B}_{i1} \\
\leq & -\mathcal{P}_1 V_{i1} + \frac{1}{2}e_{v,i}^2 + \mathcal{B}_{i1}
\end{aligned} \tag{47}$$

with the presentation of given $\bar{V}_{i1}, \mathcal{B}_{i1} = \frac{\iota_{a,i1}\aleph_{i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^{*} + \frac{\iota_{c,i1}\aleph_{i1}}{2}\mathcal{W}_{i1}^{*\top}\left(\xi_{i1}(e_{p,i})\xi_{i1}^{\top}(e_{p,i}) + \sigma_{i1}I_{q_1}\right)\mathcal{W}_{i1}^{*}$, and $\mathcal{P}_1 = \min\left\{2(\kappa_{i1} - \frac{3}{4}), (\iota_{a,i1} - \iota_{c,i1})\aleph_{i1}\sigma_{i1}, \iota_{c,i1}\aleph_{i1}\sigma_{i1}\right\}$. Therefore, it can be concluded that when the parameter conditions in the Proposition 1 are met, the negative-definite property of the Lyapunov function in the expression holds true, so the condition in Proposition 1 holds. Similarly, it can be proved that the conditions in Proposition 2 hold true, that is, let $\bar{V}_{i2} = \frac{1}{2}e_{v,i}^2 + \frac{1}{2}\tilde{\mathcal{W}}_{c,i2}^{\top}\tilde{\mathcal{W}}_{c,i2} + \frac{1}{2}\tilde{\mathcal{W}}_{a,i2}^{\top}\tilde{\mathcal{W}}_{a,i2}$, as the process of (37), by the given $\hat{\dot{\mathcal{W}}}_{c,i2}$ and $\hat{\dot{\mathcal{W}}}_{a,i2}$ in (44), $\bar{V}_{i2} \leq -\mathcal{P}_{i2}V_{i2} + \mathcal{B}_{i2}$. where $\mathcal{P}_{i2} = \min\{2\kappa_{i2} - 4, (\iota_{a,i2} - \iota_{c,i2})\aleph_{i2}\sigma_{i2}, \iota_{c,i2}\aleph_{i2}\sigma_{i2}\}$, $\mathcal{B}_{i2} = \frac{1}{2}\hat{\dot{\alpha}}_i^{*2} + \frac{1}{2}\dot{v}_{d,i}^2 + + \frac{\iota_{a,i2}\aleph_{i2}}{2}\mathcal{W}_{i2}^{*\top}\left(\xi_{i2}(e_{v,i})\xi_{i2}^{\top}(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\mathcal{W}_{i2}^{*} + \frac{\iota_{c,i2}\aleph_{i2}}{2}\mathcal{W}_{i2}^{*\top}\left(\xi_{i2}(e_{v,i})\xi_{i2}^{\top}(e_{v,i}) + \sigma_{i2}I_{q_2}\right)\mathcal{W}_{i2}^{*}$. Synthetically, by virtue of the aforementioned proof, the conditions in Proposition 1 and Proposition 2 holds, thus the proof of Theorem 1 has been completed. □

Remark 3. Building upon Theorem 1, the coefficient augmentation term $\aleph_{i1}$ and $\aleph_{i2}$ which modulate $\iota_{c,i1}$ and $\iota_{a,i1}$ in (43) and (44) are fundamentally characterized by a Gaussian-modulated time-decay structure. Formally, these terms admit the representation:

$$\psi(t, \mathcal{X}, \mathcal{L}) = \underbrace{\exp\left(-\frac{(\mathcal{X} - c_1)^2}{2\sigma_1^2}\right)}_{\text{Gaussian kernel}} \cdot \underbrace{\frac{1}{1+\beta t}}_{\text{Time decay}}, \quad i \in [N], j \in \{1, 2\}, \mathcal{L} \in \mathbb{R}. \tag{48}$$

where σ_1 presents the sensitivity bandwidth, β indicates the decay rate. From (48), ψ can be divided into the component of *Gaussian kernel* and *Time decay*. Let $\sigma_1 = 1, \beta = 0.3$, $\mathcal{X} = \texttt{linspace}(-3, 3, 50)$, $\mathcal{L} = \texttt{linspace}(-3, 3, 50)$, $[\mathcal{X}, \mathcal{L}] = \texttt{meshgrid}(\mathcal{X}, \mathcal{L})$, $T = [0, 3, 6, 10]$. Subsequently, a three-dimensional plot of the function at several specified points is presented as *Multivariate Gaussian* × *time decay*. As can be observed from the Fig. 3, the function exhibits a decaying trend and approaches zero as time progresses.

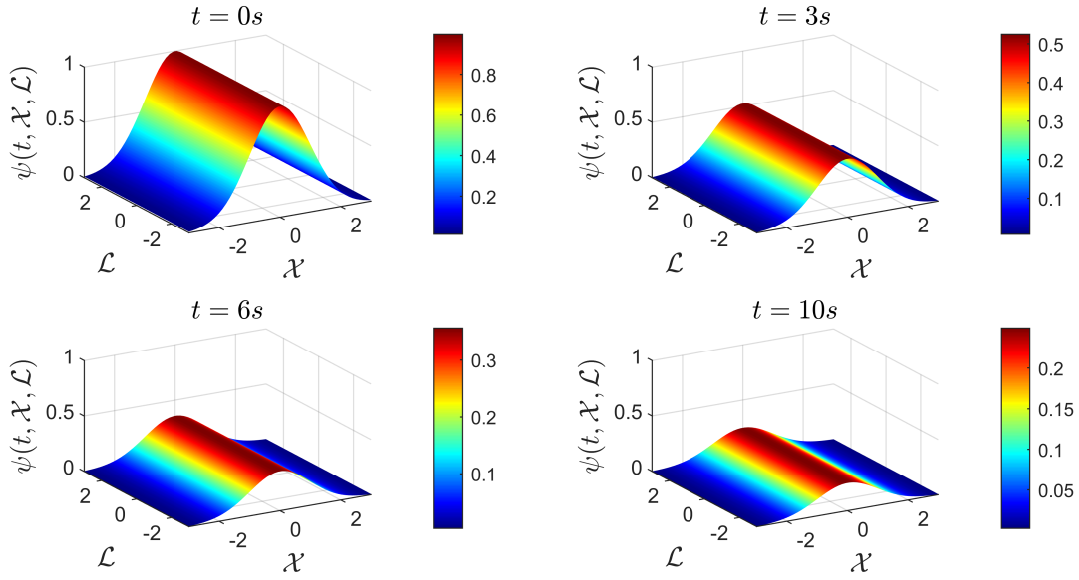


Figure 3: General evolutionary properties of enhanced term-like $\aleph_{i1}$ and $\aleph_{i2}$ in (45) of Theorem 3.

According to (45), as $\aleph_{i1}$ varies with $e_{p,i}$, and $\aleph_{i2}$ varies with $e_{v,i}$, taking $\aleph_{i1}$ as an example, it is known that $\aleph_{i1}$ and $\aleph_{i2}$ are all positive terms with positive constants $g_{ij}, \delta_{ij} > 0$. Moreover, when $t_{\text{effect}} \to \infty$, if $e_{p,i}$ indicates big value, $\aleph_{i1} \to I_{q_1}$; if $e_{p,i} \to 0$, $\aleph_{i1} \to I_{q_1} + C$, C is an arbitrary positive constant. In conclusion, the appearance of $\aleph_{ij}$ does not change the direction of convergence of $\hat{\dot{\mathcal{W}}}_{c,ij}$ and $\hat{\dot{\mathcal{W}}}_{a,ij}$; it only affects the transient response and amplitude of $\hat{\dot{\mathcal{W}}}_{c,ij}$ and $\hat{\dot{\mathcal{W}}}_{a,ij}$ based on the floating moment point of $e_{p,i}$ and $e_{v,i}$, the $\aleph_{i1}$ and $\aleph_{i2}$ The incorporation of coefficients $\aleph_{i1}$ and $\aleph_{i2}$ ensures persistent excitation under small-value position/speed errors of $e_{p,i}/e_{v,i}$, $i \in [N]$, while preserving the baseline dynamics mechanism for large-value errors of $e_{p,i}/e_{v,i}$.

4. Simulation

In this section, simulation experiments are further conducted utilizing data from high-speed train G1 operating on the route from Beijing South Railway Station (BSRS) to Jinan West Railway Station (JWRS). On the basis of the illustration of (7), $a_{co,i}$, $b_{co,i}$ and $c_{co,i}$ appeared in Condition 1 to Condition Q of Fig. 1 may exhibit varying dynamics due to operation environmental fluctuations, which will be specified later during parameter determination in this section. Furthermore, the segmented slope characteristics depicted in the cases of S1 to S3 of Fig. 1 derives from actual measurements, herein, the corresponding elevation profile and gradient data of VC operation process from BSRS to JWRS are presented in Fig. 4.

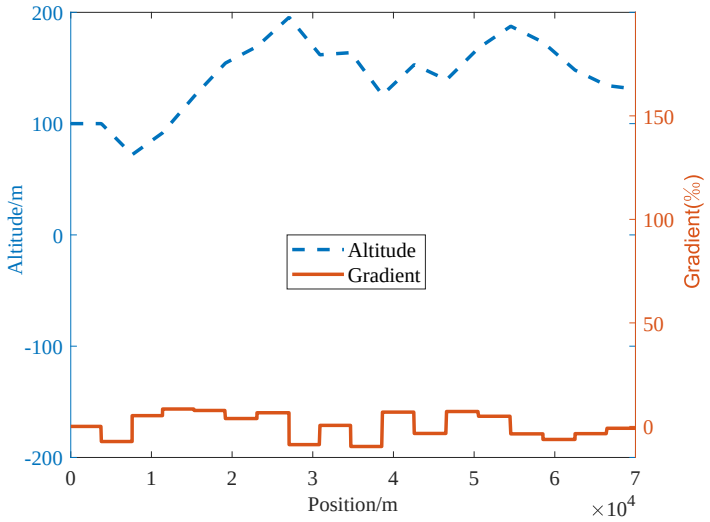


Figure 4: Railway line gradient information.

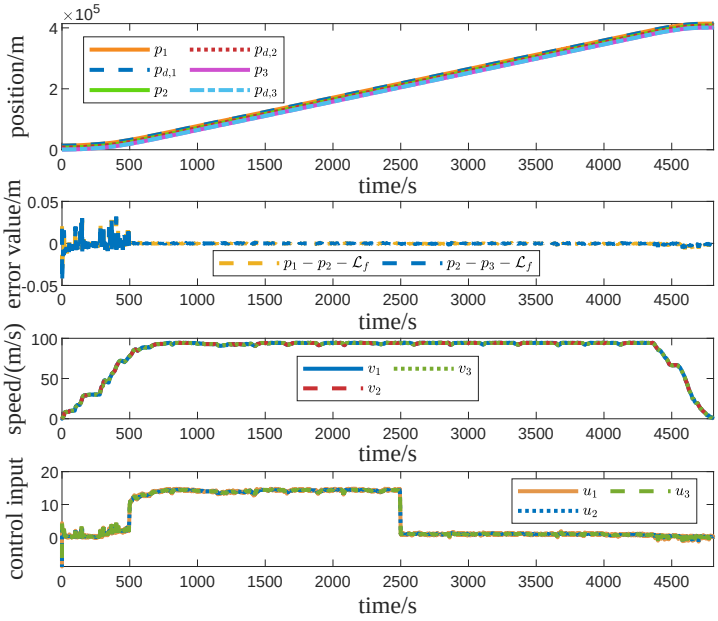


Figure 5: The information of position, safety margin evolution, speed and control input of 3-train VC formation operation.

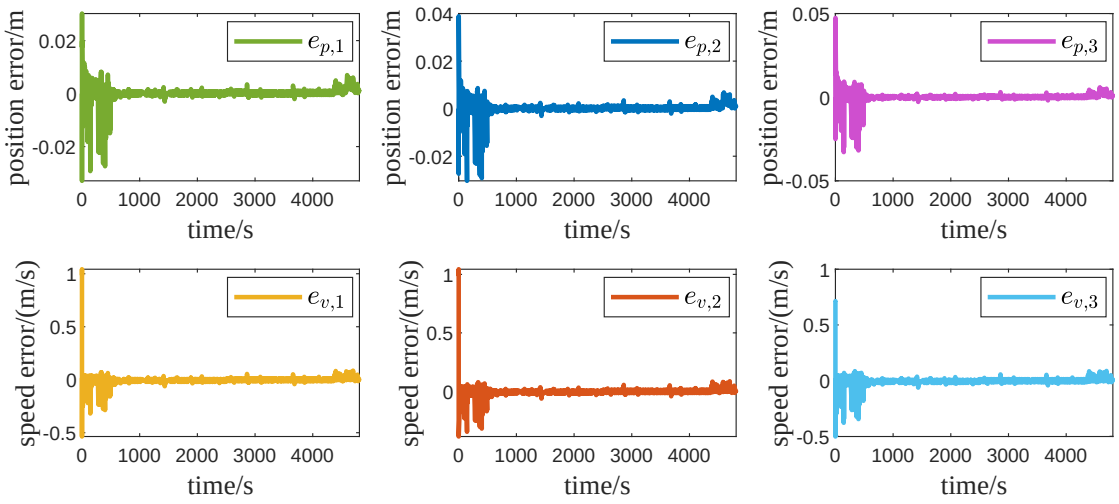


Figure 6: Position and speed errors of train 1 ~ train 3 in 3-train VC formation with variant $f_{b,i}$.

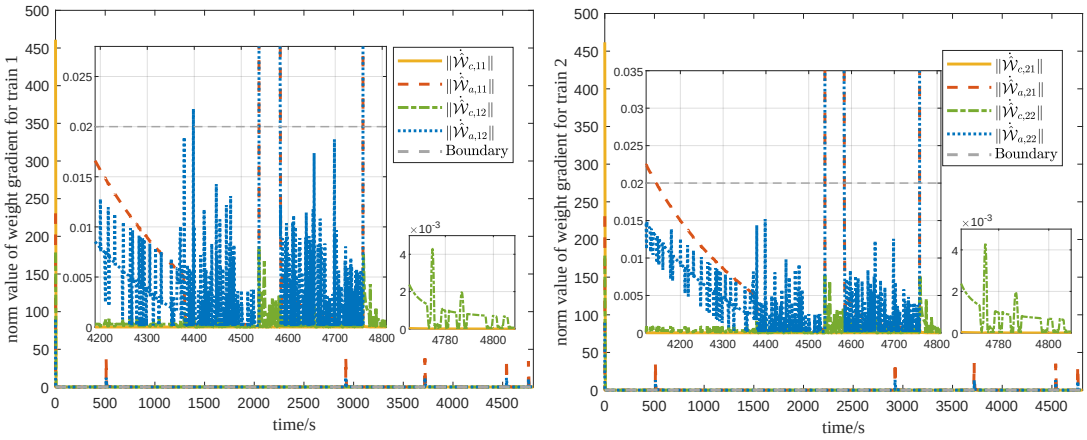
4.1. CTRAR algorithm for 3-train formation

Let p_{G1}, v_{G1} be the actual position and speed operation data of train G1 from BSRS to JWRS. The initial position and speed are set as $p_{d,1}(0) = p_1(0) = p_{G1}(0) + \mathfrak{C}_m + 2\mathcal{L}_{f,i}$, $p_{d,2}(0) = p_2(0) = p_{G1}(0) + \mathfrak{C}_m + \mathcal{L}_{f,i}$. $p_{d,3}(0) = p_2(0) = p_{G1}(0) + \mathfrak{C}_m$. $v_{d,1}(0) = v_1(0) = v_{G1}(0)$, $v_{d,2}(0) = v_2(0) = v_{G1}(0)$, $v_{d,3}(0) = v_3(0) = 0$. where the coupling distance is set as $\mathcal{L}_{f,i} = 6500$m, and the safety margin is defined as $\mathfrak{C}_m = 500$m. To simulate the variation of basic resistance during train operation, the parameters of basic resistance are set in segments here: $a_i = 0.2$, $b_i = 0.00045$, $c_i = 0.00005$ for $t \in [0, 300)$; $a_i = 0.5$, $b_i = 0.0075$, $c_i = 0.00015$ for $t \in [300, 500)$; $a_i = 0.88$, $b_i = 0.00075$, $c_i = 0.0015$ for $t \in [500, 2500)$; $a_i = 0.5$, $b_i = 0.00045$, $c_i = 0.00005$ for $t \in [2500, 4000)$; $a_i = 0.3$, $b_i = 0.00025$, $c_i = 0.00005$ for $t \in [4000, t_{\text{end}}]$; $i = 1, 2, 3$, which leads to the characterization of variant basic resistance $f_{b,i}$. Adjustable parameters for actor NN and critic NN parts are specified as $\iota_{c,11} = \iota_{c,12} = \iota_{c,21} = \iota_{c,22} = 2$, $\iota_{a,11} = \iota_{a,12} = \iota_{a,21} = \iota_{a,22} = 1$, $\iota_{c,31} = \iota_{c,32} = 4; \iota_{a,31} = \iota_{a,32} = 3$. $\sigma = 0.01$, $\kappa_{i1} = 10$, $\kappa_{i2} = 8$, $i = 1, 2, 3$ for $t \in [0, 500) \cup [2500, t_{\text{end}})$; and $\kappa_{11} = 18$, $\kappa_{12} = 20$, $\kappa_{21} = 18$, $\kappa_{22} = 20$, $\kappa_{31} = 20$, $\kappa_{32} = 20$, $i = 1, 2, 3$ for $t \in [500, 2500)$; Additionally, define $E \in \mathbb{R}^{q_{i,j}}$ as all-1 matrix. $\hat{\mathcal{W}}_{c,11}(0) = \hat{\mathcal{W}}_{a,11}(0) = \hat{\mathcal{W}}_{c,21}(0) = \hat{\mathcal{W}}_{a,21}(0) = \hat{\mathcal{W}}_{c,31}(0) = \hat{\mathcal{W}}_{a,31}(0) = 0.8E \in \mathbb{R}^{q_{i,j}}$, $\hat{\mathcal{W}}_{c,12}(0) = \hat{\mathcal{W}}_{a,12}(0) = \hat{\mathcal{W}}_{c,22}(0) = \hat{\mathcal{W}}_{a,22}(0) = \hat{\mathcal{W}}_{c,32}(0) = \hat{\mathcal{W}}_{a,32}(0) = 0.7E \in \mathbb{R}^{q_{i,j}}$, where neuron number $q_{i,j} = 10$, $i, j = 1, 2$. The simulation sampling interval for data acquisition is set to $h = 0.001$s.

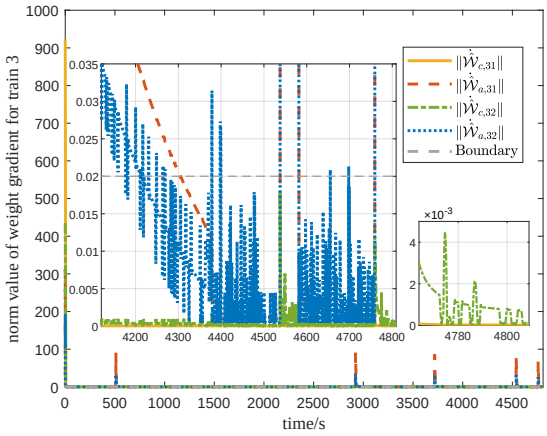
Based on the main findings in Theorem 2 and Theorem 3, Fig. 5 demonstrates the position $p_{d,i}(t)$, safety margin evolution $p_i - p_{i+1} - \mathcal{L}_f$, $i = 1, 2, 3$, speed and control input of 3-train VC formation. It is evident from $p_1 - p_2 - \mathcal{L}_f$ and $p_2 - p_3 - \mathcal{L}_f$ that the intra-formation distance between the front and rear trains in the 3-train formation progressively decreases, and the formation operation is successfully achieved with the specified coupling distance $\mathcal{L}_{f,i}$. The piecewise position and speed errors of the leading train and following trains $e_{p,i}$ and $e_{v,i}$, $i = 1, 2, 3$, are presented in Fig. 6 with the effect of the variant $f_{b,i}$.

As given the proof of first and second stage of $\hat{\alpha}_i^*$ and $\hat{u}_i^*$, denote $\mathcal{W}_{c,ij}$ and $\mathcal{W}_{a,ij}$ as critic weight and actor weight for train i, $i \in [N]$, $j = 1, 2, 3$, Fig. 7 provides norm value of AC-weight gradient for the leading train and following trains in 3-train formation. It is known from Theorem 3 that at the time points where the error changes within the time axis range, the error changes show an attenuating variation as detail figures in Fig. 7(a)~Fig. 7(c), due to the supplemented design of items $\aleph_{i1}$ and $\aleph_{i2}$. Fig. 8 gives cost functions C_{ij} in the first stage of $\hat{\alpha}_i^*(j = 1)$ and second stage of $\hat{u}_i^*(j = 2)$, $i = 1, 2, 3$. The analysis reveals that C_{ij} is influenced by the presence and variation of e_p, e_v, and the resistance $\mathcal{F}_i$ in the control input u_i. Ultimately, when the result of the virtual coupled operation of the three trains is finally accomplished, both C_{ij}, $i, j = 1, 2$, remain close to the zero, indicating that the performance index at this stage has reached the expected value level.

Accordingly, the weight update and mode partition of three trains in 3-train formation are shown in Fig. 9, The composite norm of weight variations $\|\dot{\hat{\mathcal{W}}}_i\| = \sum_{j=1}^{2} \|\dot{\hat{\mathcal{W}}}_{c,ij}\| + \sum_{j=1}^{2} \|\dot{\hat{\mathcal{W}}}_{a,ij}\|$, is defined in accordance with Theorem 2, with prescribed error bounds $B_{e_{p,i}} = 0.05$, $B_{e_{v,i}} = 0.05$ and weight derivative bounds $B_{\dot{\hat{\mathcal{W}}}_i} = 0.02$ being established, $i = 1, 2, 3$. In the 3-train virtual coupling system, the optimality condition for actor-critic network weights is determined by



(a) Gradient norm evolution of AC weights for Train 1 . (b) Gradient norm evolution of AC weights for Train 2 .



(c) Gradient norm evolution of AC weights for Train 3 .

Figure 7: Gradient norm evolution of AC weights for Train 1 ~ Train 3 in 3-train VC formation by the main results in this paper.

both error variations $\|e_{p,i}\|$, $\|e_{v,i}\|$ and weight gradient changes $\|\dot{\hat{\mathcal{W}}}_i\|$. The optimal mode is entered when $\forall\|e_{p,i}\| < B_{e_{p,i}} \wedge \forall\|e_{v,i}\| < B_{e_{v,i}}, \forall\|\dot{\hat{\mathcal{W}}}_i\| < B_{\dot{\hat{\mathcal{W}}}_i}, \Delta t_k > \mathcal{T}_{\text{dwell}} = 50s, \forall i = 1, 2, 3.$ are satisfied; otherwise, the non-optimal mode is activated.

Owing to autonomous switching between RL ($I_{j+1}^{\text{mode}} = 1,\ j \in \mathbb{N}$) and CTRAR ($I_{j+1}^{\text{mode}} = 0,\ j \in \mathbb{N}$), mode is performed by the designed performance monitor through finite transitions verified by the proof of Theorem 2. Furthermore, Computational resources are conserved during weight update suspension, while instantaneous control reactivation is guaranteed when performance thresholds are violated. Both operational efficiency and disturbance rejection capability are thus ensured by this mechanism.

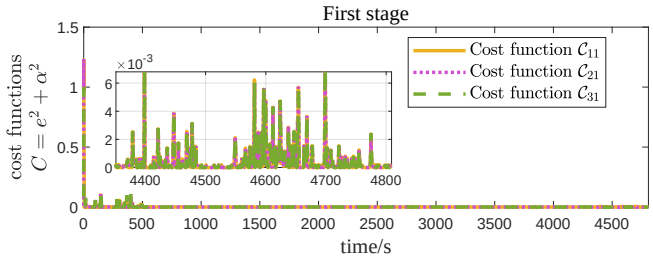
4.2. Analysis of quantitative indicators for the main results

To illustrate the efficacy of the core contributions of this work, the following will explain from two aspects: algorithm comparison and resource conservation.

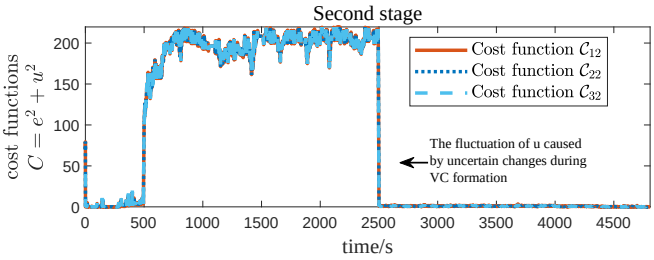
4.2.1. Algorithm comparison

This section conducts a comparative analysis among three control algorithms: Algorithm 1 implements proportional control; Algorithm 2 adopts the deterministic approach from Theorem 1 in [10]; and the algorithm in this paper means the proposed synthetical method in Theorem 2 and Theorem 3. The detailed control strategy and parameter selection are illustrated as follow:

(1) Alg#1: $e_{p,i}(t) = p_i(t) - p_{d,i}(t), e_{v,i}(t) = v_i(t) - v_{d,i}(t) - \alpha_{pc,i}(t), i \in [N], \alpha_{pc,i} = -k_{pc,1}e_{p,i}, u_{pc,i} = -k_{pc,2}e_{v,i}$.



(a) The cost function of first stage related to $\mathcal{C} = e^2 + \alpha^2$.



(b) The cost function of first stage related to $\mathcal{C} = e^2 + u^2$.

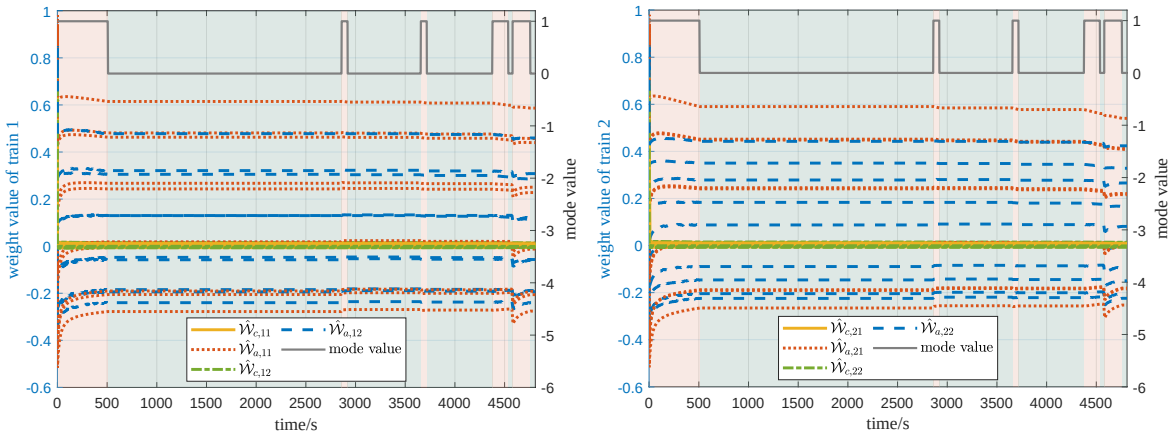
Figure 8: Cost function $\mathcal{C}_{ij}$ of first and second stages, $i = 1, 2, 3$; $j = 1, 2$.

(2) Alg#2: $e_{p,i}(t) = p_i(t) - p_{d,i}(t), e_{v,i}(t) = v_i(t) - v_{d,i}(t) - \alpha_{rc,i}(t), i \in [N]$;
$\hat{\alpha}^*_{rc,i} = -\beta_{i1} e_{p,i} - \frac{1}{2}\hat{W}^{\mathrm{T}}_{ra,i1} S_{i1}\left(e_{p,i}\right)$, $\hat{u}^*_{rc,i} = -\beta_{i2} e_{v,i} - \frac{1}{2}\hat{W}^{\mathrm{T}}_{ra,i2} S_{i2}\left(e_{v,i}\right) - f_i\left(v_i, \varphi_i\right)$;
$\dot{\hat{W}}_{rc,i1} = -\gamma_{c,i1} S_{i1}\left(e_{p,i}\right) S^{\mathrm{T}}_{i1}\left(e_{p,i}\right) \hat{W}_{rc,i1} - \frac{1}{2} S_{i1}\left(e_{p,i}\right) e_{p,i}$, $\dot{\hat{W}}_{ra,i1} = -\gamma_{a,i1} S_{i1}\left(e_{p,i}\right) S^{\mathrm{T}}_{i1}\left(e_{p,i}\right)\left(\hat{W}_{ra,i1} - \hat{W}_{rc,i1}\right)$;
$\dot{\hat{W}}_{rc,i2} = -\gamma_{c,i2} S_{i2}\left(e_{v,i}\right) S^{\mathrm{T}}_{i2}\left(e_{v,i}\right) \hat{W}_{rc,i2} - \frac{1}{2} S_{i2}\left(e_{v,i}\right) e_{v,i}$, $\dot{\hat{W}}_{ra,i2} = -\gamma_{a,i2} S_{i2}\left(e_{v,i}\right) S^{\mathrm{T}}_{i2}\left(e_{v,i}\right)\left(\hat{W}_{ra,i2} - \hat{W}_{rc,i2}\right)$.

(3) Algorithm in this paper: $e_{p,i}(t) = p_i(t) - p_{d,i}(t), e_{v,i}(t) = v_i(t) - v_{d,i}(t) - \alpha_{cc,i}(t), i \in [N]$;
$\hat{\alpha}^*_{cc,i} = -\kappa_{i1} e_{p,i} - \frac{1}{2}\hat{\mathcal{W}}^{\top}_{ca,i1}\xi_{i1}(e_{p,i}), \hat{u}^*_{cc,i} = -\kappa_{i2} e_{v,i} - \mathcal{F}_i - \frac{1}{2}\hat{\mathcal{W}}^{\top}_{ca,i2}\xi_{i2}(e_{v,i})$;
$\dot{\hat{\mathcal{W}}}_{cc,i1} = -\iota_{c,i1}\aleph_{i1}\left(\xi_{i1}(e_{p,i})\xi^{\top}_{i1}(e_{p,i}) + \sigma_{i1} I_{q_1}\right)\hat{\mathcal{W}}_{cc,i1}(t)$,
$\dot{\hat{\mathcal{W}}}_{ca,i1} = -\left[\xi_{i1}(e_{p,i})\xi^{\top}_{i1}(e_{p,i}) + \sigma_{i1} I_{q_1}\right]\left(\iota_{a,i1}\aleph_{i1}\hat{\mathcal{W}}_{ca,i1} - \iota_{c,i1}\aleph_{i1}\hat{\mathcal{W}}_{cc,i1}\right)$
$\dot{\hat{\mathcal{W}}}_{cc,i2} = -\iota_{c,i2}\aleph_{i2}\left(\xi_{i2}(e_{v,i})\xi^{\top}_{i2}(e_{v,i}) + \sigma_{i2} I_{q_2}\right)\hat{\mathcal{W}}_{cc,i2}(t)$
$\dot{\hat{\mathcal{W}}}_{ca,i2} = -\left[\xi_{i2}(e_{v,i})\xi^{\top}_{i2}(e_{v,i}) + \sigma_{i2} I_{q_2}\right]\left(\iota_{a,i2}\aleph_{i2}\hat{\mathcal{W}}_{ca,i2} - \iota_{c,i2}\aleph_{i2}\hat{\mathcal{W}}_{cc,i2}\right)$
with the triggered and stalled presentation of $\dot{\hat{\mathcal{W}}}_{cc,i1}$, $\dot{\hat{\mathcal{W}}}_{cc,i2}$, $\dot{\hat{\mathcal{W}}}_{ca,i1}$ and $\dot{\hat{\mathcal{W}}}_{ca,i2}$ by mode partition I^{mode}_{j+1} in (41), and the illustration of $\aleph_{i1}$ and $\aleph_{i2}$ in (45).

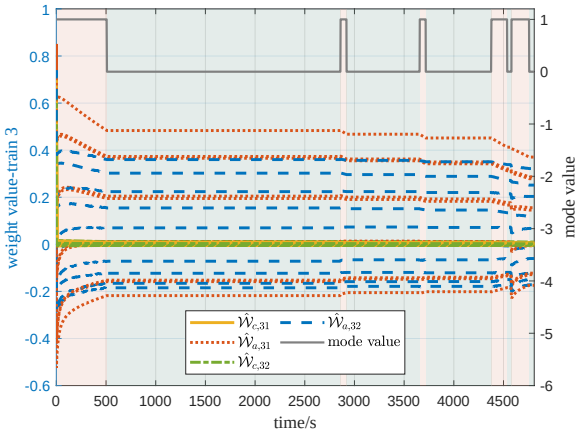
The relevant parameters are specified as follows, with the exception of the specially annotated parameters below, all other parameters in Alg#1 and Alg#2 maintain consistency with those in the algorithm in this paper. For Alg#1, $k_{pc,1} = k_{pc,2} = 3$. For Alg#2, $\gamma_{c,i1} = 4$, $\gamma_{a,i1} = 3$, $\gamma_{c,i2} = 7$, $\gamma_{a,i2} = 5$, $\beta_{i1} = \beta_{i2} = 7$. For the algorithm in this paper, adjustable parameters for actor NN and critic NN parts are specified as $\iota_{c,11} = \iota_{c,12} = \iota_{c,21} = \iota_{c,22} = 2$, $\iota_{a,11} = \iota_{a,12} = \iota_{a,21} = \iota_{a,22} = 1$, $\iota_{c,31} = \iota_{c,32} = 4$; $\iota_{a,31} = \iota_{a,32} = 3$. $\sigma = 0.01$, $\kappa_{i1} = 10$, $\kappa_{i2} = 8$, $i = 1, 2, 3$ for $t \in [0, 500) \cup [2500, t_{\mathrm{end}})$; and $\kappa_{11} = 18$, $\kappa_{12} = 20$, $\kappa_{21} = 18$, $\kappa_{22} = 20$, $\kappa_{31} = 20$, $\kappa_{32} = 20$, $i = 1, 2, 3$ for $t \in [500, 2500)$; Additionally, define $E \in \mathbb{R}^{q_{i,j}}$ as all-1 matrix. $\hat{\mathcal{W}}_{c,11}(0) = \hat{\mathcal{W}}_{a,11}(0) = \hat{\mathcal{W}}_{c,21}(0) = \hat{\mathcal{W}}_{a,21}(0) = \hat{\mathcal{W}}_{c,31}(0) = \hat{\mathcal{W}}_{a,31}(0) = 0.8E \in \mathbb{R}^{q_{i,j}}$, $\hat{\mathcal{W}}_{c,12}(0) = \hat{\mathcal{W}}_{a,12}(0) = \hat{\mathcal{W}}_{c,22}(0) = \hat{\mathcal{W}}_{a,22}(0) = \hat{\mathcal{W}}_{c,32}(0) = \hat{\mathcal{W}}_{a,32}(0) = 0.7E \in \mathbb{R}^{q_{i,j}}$, where neuron number $q_{i,j} = 10$, $i, j = 1, 2$. The simulation sampling interval for data acquisition is set to $h = 0.001$s.

Accordingly, on the one hand, as for the visualization aspect, 2D *time-error* responses by Alg#1, Alg#2 and the algorithm in this paper for $e_{p,i}$ and $e_{v,i}$ of train i are respectively expressed are presented in Fig. 10(a)~Fig. 10(b), Fig. 10(c)~Fig. 10(d), Fig. 10(e)~Fig. 10(f). Furthermore, 3D phase diagrams of *time-position error-speed error* for train i ($i \in [N]$) by Alg#1, Alg#2 and the algorithm in this paper are illustrated in Fig. 11. To provide an intuitive characterization of amplitude errors, data-driven line plots for Alg#1, Alg#2 and the algorithm in this paper, are constructed in Fig. 12 and Fig. 13, using the data set in optimized (convergent) and non-optimized



(a) AC weight values and mode partition of train 1.

(b) AC weight values and mode partition of train 2.

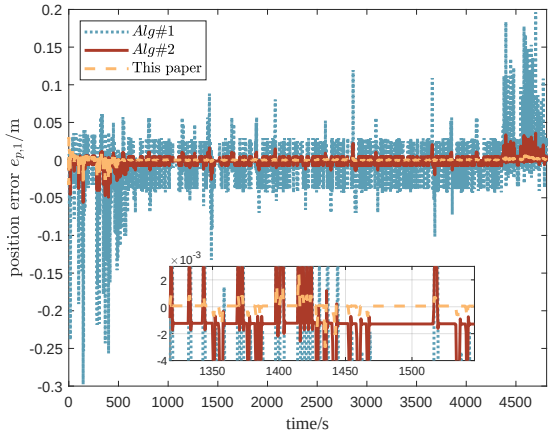


(c) AC weight values and mode partition of train 3.

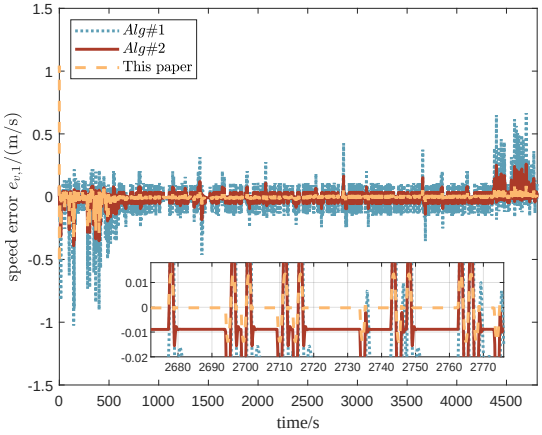
Figure 9: AC weight values and mode partition of train 1~train 3 in 3-train VC formation by CTRAR framework.

(nonconvergent) operational modes, as defined in Theorem 2. The timepoint selection follows the methodology of convergent and nonconvergent modes. The set of all points where the mode transitions exactly from 1 to 0 are selected as critical nodes for comparing algorithm performance.Thus, for convergent modes, 5 critical instants $t_{1,0}^{\text{trigger}} = 509.28s, 2921.085s, 3717.867s, 4537.791s$ and $4759.104s$ are specifically captured where the mode indicator transitions from 1 to 0, supplemented by 2 randomly sampled points $t = 1773.47s$ and $3301.17s$ during sustained $I_{j+1}^{\text{mode}} = 0, j = 0, 1, 2, 3, 4$, collectively forming 7 representative datapoints, as enumerated in point count 1-7 of Table 1. Simultaneously, 3 arbitrary points $t = 144.109s, 2861.53s$ and $4698.48s$ from nonconvergent modes are selected to assist in verifying the error performance under nonconvergent conditions, as enumerated in point count 1-3 of Table 2. The dual dataset strategy ensures comprehensive coverage of both transient switching phenomena, convergent and nonconvergent-state behavior while maintaining statistical relevance. This structured approach yields two distinct datasets - stable sequences (in Table 1) and unstable sequences (in Table 2)-enabling unambiguous visual comparison of error magnitudes across all three algorithmic implementations. As evidenced by the experimental results presented from Fig. 10 to Fig. 13, the proposed algorithm demonstrates superior performance with significantly reduced amplitude in both position and velocity tracking errors. It indicates that our control scheme effectively regulates train operation to achieve the predefined tracking objectives with enhanced precision.

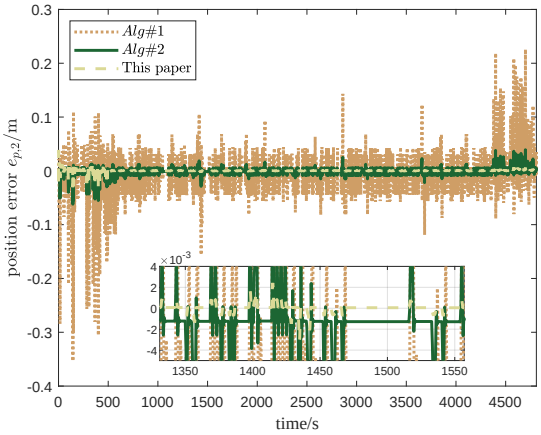
On the other hand, as for the visualization aspect, the corresponding times tamps are indicated in the *time* row (derived from mode transition indices in Fig. 9). At these temporal nodes, each algorithm's positional errors (Ep) and velocity errors (Ev) are systematically retrieved, with comprehensive data presented in Table 1. And the computational methodologies for the "Means" and "Per. (Percentage)" columns are explicitly defined below: 1) The calculation of



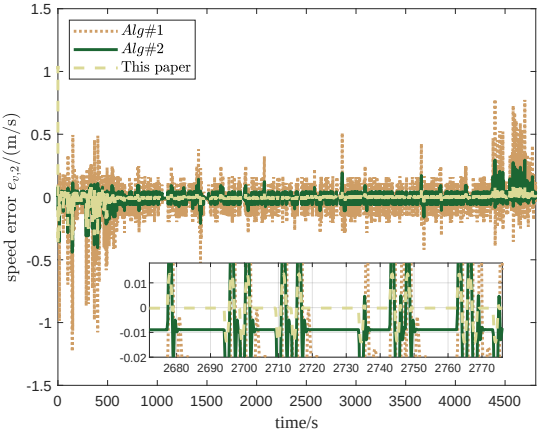
(a) Comparison of $e_{p,1}$ in 3-train VC formation.



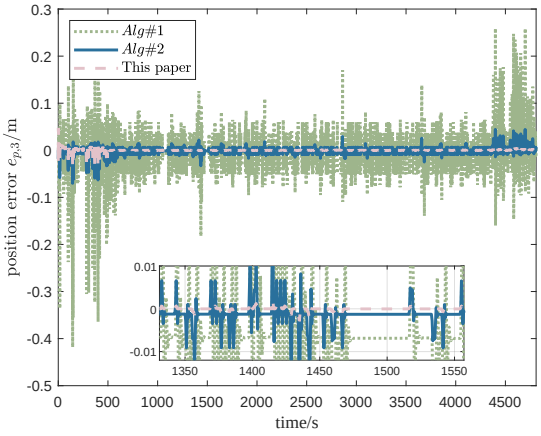
(b) Comparison of $e_{v,1}$ in 3-train VC formation.



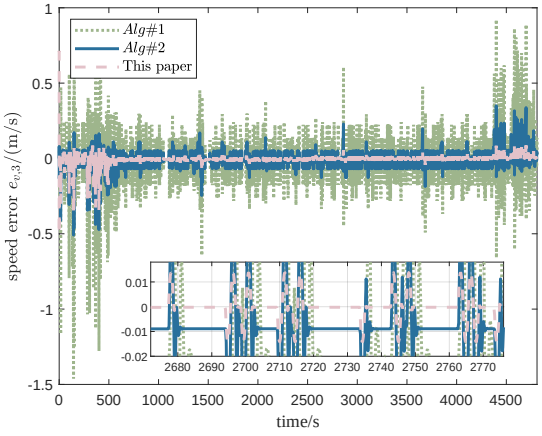
(c) Comparison of $e_{p,2}$ in 3-train VC formation.



(d) Comparison of $e_{v,2}$ in 3-train VC formation.

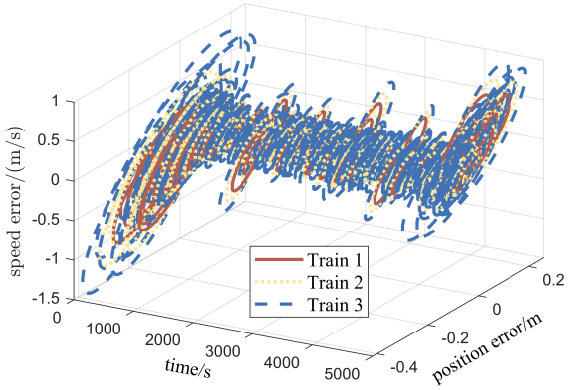


(e) Comparison of $e_{p,3}$ in 3-train VC formation.

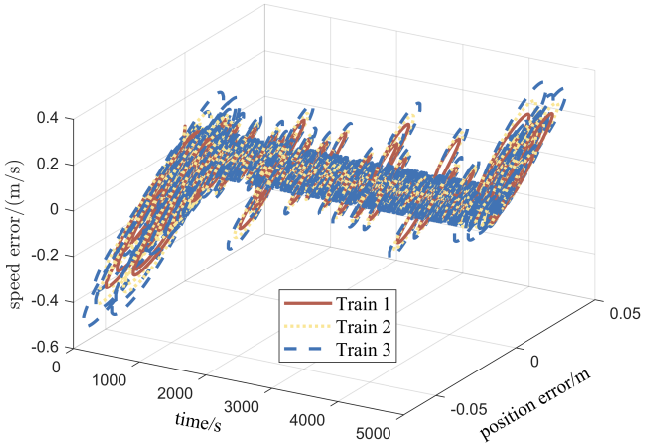


(f) Comparison of $e_{v,3}$ in 3-train VC formation

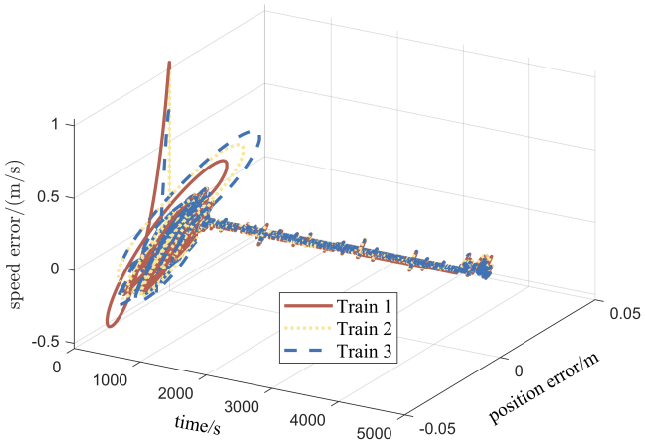
Figure 10: Comparison of $e_{p,i}$ and $e_{v,i}$ by the three algorithms, $i = 1, 2, 3$.



(a) 3-dimensional phase portrait of tracking errors for $Ag\#1$.

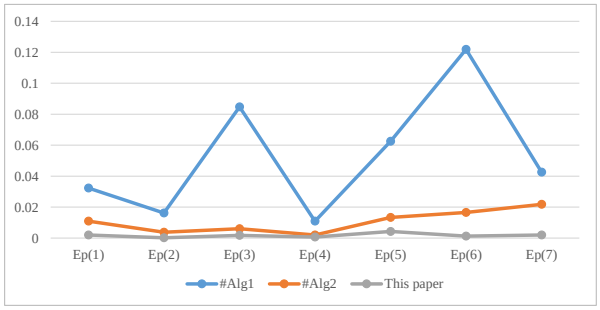


(b) 3-dimensional phase portrait of tracking errors for $Ag\#2$.

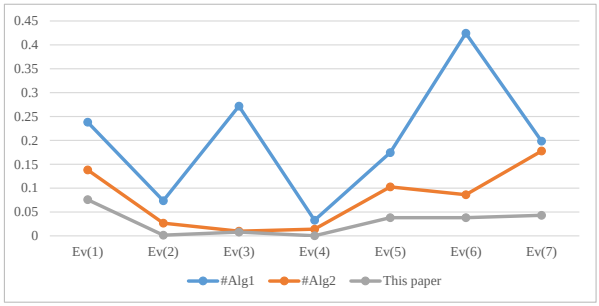


(c) 3-dimensional phase portrait of tracking errors for the algorithm in this paper.

Figure 11: Comparison of 3-dimensional phase portrait of position and speed errors $e_{p,i}$ and $e_{v,i}$ in three algorithms, $i = 1, 2, 3$.



(a) Performance comparison of position errors (Ep).



(b) Performance comparison of velocity errors (Ev).

Figure 12: Performance comparison of position errors (Ep) and velocity errors (Ev) among three algorithmic approaches across seven convergence points in Table 1.

Table 1: Quantitative comparison of three-algorithm performance at convergence-state time points

Convergent point count	1		2		3		4	
Time/s	509.280		2921.085		3717.867		4537.791	
Error / Algorithm	$Ep_{(1)}$	$Ev_{(1)}$	$Ep_{(2)}$	$Ev_{(2)}$	$Ep_{(3)}$	$Ev_{(3)}$	$Ep_{(4)}$	$Ev_{(4)}$
Ag#1	0.0323	0.2379	0.0162	0.0736	0.0847	0.2718	0.0109	0.0327
Ag#2	0.0109	0.1381	0.0038	0.0266	0.0061	0.0098	0.0020	0.0141
This paper	0.0020	0.0757	0.0002	0.0013	0.0017	0.0079	0.0007	0.0002

Convergent point count	5		6		7		Mean		Per./%
Time/s	4759.104		1773.47		3301.17				
Error / Algorithm	$Ep_{(5)}$	$Ev_{(5)}$	$Ep_{(6)}$	$Ev_{(6)}$	$Ep_{(7)}$	$Ev_{(7)}$	$Ep_{(\beta)}$	$Ev_{(\beta)}$	$E_{(\beta)}$
Ag#1	0.0626	0.1743	0.1219	0.4242	0.0426	0.1985	0.0530	0.2018	0.9114
Ag#2	0.0133	0.1024	0.0166	0.0860	0.0218	0.1776	0.0106	0.0792	0.7342
This paper	0.0043	0.0379	0.0013	0.0379	0.0020	0.0431	0.0017	0.0291	0 (Base)

Table 2: Quantitative comparison of three-algorithm performance at non-convergence time points

Transient point count	1		2		3		Mean		Per./%
Time	144.109		2861.53		4698.48				
Error / Algorithm	$Ep_{(1)}$	$Ev_{(1)}$	$Ep_{(2)}$	$Ev_{(2)}$	$Ep_{(3)}$	$Ev_{(3)}$	$Ep_{(\gamma)}$	$Ev_{(\gamma)}$	$E_{(\gamma)}$
Ag#1	0.4316	2.3954	0.4022	1.2793	0.3005	1.4572	0.3781	1.7106	0.8227
Ag#2	0.1815	1.3354	0.0353	0.1173	0.1107	0.8866	0.1092	0.7798	0.5493
This paper	0.0892	1.0307	0.0044	0.0308	0.0168	0.2586	0.0368	0.4400	0(Base)

Table 1 is given below:

$$Mean: Ep_{(\beta),Ag} = \frac{\sum_{m=1}^{\beta} Ep_{(m),Ag}}{\beta}; Mean: Ev_{(\beta),Ag} = \frac{\sum_{m=1}^{\beta} Ev_{(m),Ag}}{\beta} \tag{49}$$

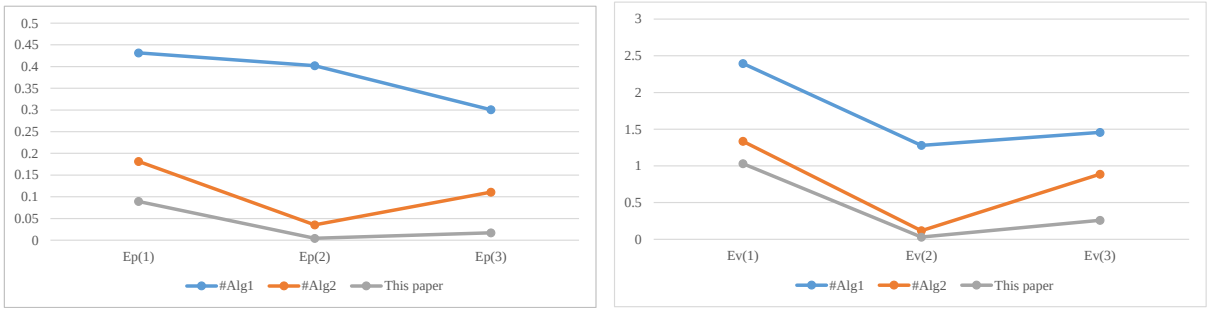
where $Ep_{(\beta),Ag}$ means the value $Ep_{(\beta)}$ under different algorithm, $\beta = 7$ is total record times including I_{j+1}^{mode} turns from 1 to 0, and other two trigger times randomly selected from convergence mode $I_{j+1}^{\text{mode}} = 0$, $j = 0, 1, 2, 3, 4$.

And the percentage reduction presentation in amplitude among the three algorithms on the basis of algorithm in this paper (including Theorem 2 and Theorem 3) is given:

$$Per.: E_{(\beta),Ag} = \frac{1}{2} \cdot \left(\frac{Ep_{(\beta),Ag} - Ep_{(m),Ag}}{Ep_{(\beta),Ag}} + \frac{Ev_{(\beta),Ag} - Ev_{(m),Ag}}{Ev_{(\beta),Ag}} \right), m \in \{1, 2, \cdots, \beta\}. \tag{50}$$

2) Simultaneously, in order to more comprehensively demonstrate the effectiveness of the algorithm, several non-convergent points of the transient state are given:

$$Mean: Ep_{(\gamma),Ag} = \frac{\sum_{n=1}^{\gamma} Ep_{(n),Ag}}{\gamma}; Mean: Ev_{(\gamma),Ag} = \frac{\sum_{n=1}^{\gamma} Ev_{(n),Ag}}{\gamma} \tag{51}$$



(a) Performance comparison of position errors (Ep).

(b) Performance comparison of velocity errors (Ev).

Figure 13: Performance comparison of position errors (Ep) and velocity errors (Ev) among three algorithmic approaches across three non-convergence points in Table 2.

where $Ep_{(\gamma),Ag}$ means the value $Ep_{(\gamma)}$ under different algorithm, r is the sequence number of algorithm, $\gamma = 3$ is trigger times randomly selected from convergence mode $I_{j+1}^{\text{mode}} = 1$, $j = 0, 1, 2, 3, 4$.

And the percentage reduction in amplitude among the three algorithms on the basis of the algorithm in this paper (including Theorem 2 and Theorem 3) in this paper is given:

$$Per.: E_{(\gamma),Ag} = \frac{1}{2} \cdot \left(\frac{Ep_{(\gamma),Ag} - Ep_{(n),Ag}}{Ep_{(\gamma),Ag}} + \frac{Ev_{(\gamma),Ag} - Ev_{(n),Ag}}{Ev_{(\gamma),Ag}} \right), n \in \{1, 2, \cdots, \gamma\}. \tag{52}$$

Since the algorithm in this paper serves as the reference for comparison, by virtue of (51) and (52), it is straightforward to denote "Per.: Ag" in this paper as 0 in Table 1 and Table 2.

4.2.2. Explanation of the results of resource conservation

To demonstrate the resource savings in train operations enabled by the proposed mode partition scheme, a comparison is made between the total number of switches required under this method and that of the original control approach without mode switching.

$$\begin{aligned} C_{\text{CTRAR}} =& C_{\text{non-convergence}} + C_{\text{non-convergence}} = C_{I_{j+1}^{\text{mode}}=1} + C_{I_{j+1}^{\text{mode}}=0} \\ =& T_1 * 1000 + 5 \\ C_{\text{previous}} =& T2 \cdot 1000 \end{aligned} \tag{53}$$

where T_1 is the time duration of $I_{j+1}^{\text{mode}} = 1$ (5 time period), T_2 is the total time duration (4808.953s) in this simulation conditions. Thus, by the results in Fig. 9, it can be known that the starting and ending time orders $C_{\text{non-convergence}}$ of the 5 segments in the non-convergent region ($I_{j+1}^{\text{mode}} = 1$, the pink area in Fig. 9(a)~Fig. 9(c)) are respectively listed as $0 \sim 509.279$, $2860.584 \sim 2921.084$, $3657.539 \sim 3717.866$, $4378.532 \sim 4537.790$, $4580.615 \sim 4759.103$, so $T_1 = 509279 + 60500 + 60327 + 159258 + 178488 = 967852$, it yields that $C_{\text{CTRAR}} = 967857$, $C_{\text{previous}} = 4808953$. Then, define the percentage savings in the number of switches as P_{switch}, then

$$P_{\text{switch}} = \frac{C_{\text{previous}} - C_{\text{CTRAR}}}{C_{\text{previous}}} \times 100\% \tag{54}$$

Therefore, $P_{\text{switch}} = \frac{4808953 - 967857}{4808953} \times 100\% = 79.879\%$, which means that, compared with the previous continuous weight update, CTRAR allows the stagnation of weight updates, and the saving rate in this calculation form is 79.879%.

Remark 4. Based on the defined piecewise function $f_{b,i}$, the control input $\hat{u}_i^*$ exhibits oscillatory behavior, resulting in substantial fluctuations in the cost function. Throughout this process, both the actor and critic networks undergo continuous training. As established in Theorem 2, the optimal mode I_{j+1}^{mode} can only be achieved when both conditions (39) and (40) are satisfied simultaneously.

In Section 4.1, the following settings are implemented:

- Error permission $|e_{p,i}| < B_{e_{p,i}}$ and $|e_{v,i}| < B_{e_{v,i}}$ with variance for $e_{p,i}$ and $e_{v,i}$ (visualized in Fig. 6) with boundary constraint $B_{e_{p,i}} = 0.05$, $B_{e_{v,i}} = 0.05$.
- Weight permission $\|\dot{\hat{\mathcal{W}}}_i\| < B_{\hat{\mathcal{W}}_i}$, where gradient total variations $\|\dot{\hat{\mathcal{W}}}_i\| = \sum_{j=1}^{2} \|\dot{\hat{\mathcal{W}}}_{c,ij}\| + \sum_{j=1}^{2} \|\dot{\hat{\mathcal{W}}}_{a,ij}\|$ with the expression of $\hat{\mathcal{W}}_{c,ij}$ and $\hat{\mathcal{W}}_{a,ij}$, $i = 1, 2, 3$; $j = 1, 2$ (shown in Fig. 7) with boundary constraint $B_{\hat{\mathcal{W}}_i} = 0.02$.
- Dwell-time constraint $\Delta t_k > \mathcal{T}_{\text{dwell}}$, with boundary constraint $\mathcal{T}_{\text{dwell}} = 50s$.

When all conditions are met, $I_{j+1}^{\text{mode}} = 0$, $j \in \mathbb{N}$; otherwise, $I_{j+1}^{\text{mode}} = 1$. This mode switching behavior is clearly depicted by the segmented line in Fig. 9. Through real-time estimation of the basic resistance parameters (a_i, b_i, c_i), the network weights are dynamically adjusted via the performance monitor in Theorem 2, enabling near-optimal operation in $I_{j+1}^{\text{mode}} = 0$.

5. Conclusion

A CTRAR control framework is introduced for VC formation systems, where the proposed approach incorporates a dual-criteria performance monitor that evaluates boundary constraints and minimum dwell-time conditions to regulate learning behavior. Optimal mode activation and weight update suspension occur exclusively when VC tracking error variation, AC network update rate, and dwell time all meet predefined thresholds. To improve learning efficiency, an adaptive Gaussian learning rate is incorporated into AC updates, ensuring persistent excitation while preserving control stability. Theoretical analysis confirms convergence properties of the proposed method, and comparative simulations and quantified indicator using high-speed train operational data from BSRS to JWRS, demonstrate superior computational efficiency over conventional approaches. Future work will further integrate the scenario analysis of train operation, focusing on the operation control strategies including reinforcement relearning identification or fault tolerant control for guaranteeing the security of VC formation operation.

CRediT authorship contribution statement

Wenxiao Si: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation. **Shigen Gao:** Conceptualization, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Krishna Busawon:** Methodology, Formal analysis, Writing – review & editing. **Xuewu Dai:** Methodology, Formal analysis, Writing – review & editing. **Tao Wen:** Methodology, Formal analysis, Writing – review & editing, Funding acquisition. **Hongwei Wang:** Methodology, Formal analysis, Writing – review & editing.

Acknowledgments

This work is supported by the Beijing Natural Science Foundation under Grant L231017.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Owing to the sensitive nature of the data, the datasets underlying this study's findings are available from the corresponding author, SG, upon reasonable request from qualified researchers.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT and DeepSeek to refine the language and grammar. All content was subsequently reviewed and edited by the authors, who take full responsibility for the final version of the publication.

References

[1] Apostol, T.M., 1963. Mathematical analysis. Addison-Wesley Publishing Company Inc.

[2] Bai, W., Zhang, Z., Wu, X., Xiong, Y., Dong, H., 2024. Coordinated operation of virtual coupling high-speed trains under a novel sequential model predictive control framework. Unmanned Systems 12, 359–373.

[3] Bock, U., Bikker, G., 2000. Design and development of a future freight train concept–"Virtually coupled train formations". IFAC Proceedings Volumes 33, 395–400.

[4] Bock, U., Varchmin, J., 1999. Enhancement of the occupancy of railroads using virtually coupled train formations, in: World Congress on Railway Research (WCRR), pp. 19–23.

[5] Dong, H., Zheng, Y., Song, X., Gao, S., 2024. Reinforced safe performance cooperative control with event-triggered implementation for train formation. IEEE Transactions on Industrial Informatics 20, 9588–9598.

[6] Duan, J., Li, J., Ge, Q., Li, S.E., Bujarbaruah, M., Ma, F., Zhang, D., 2023. Relaxed actor-critic with convergence guarantees for continuous-time optimal control of nonlinear systems. IEEE Transactions on Intelligent Vehicles 8, 3299–3311.

[7] Felez, J., Kim, Y., Borrelli, F., 2019. A model predictive control approach for virtual coupling in railways. IEEE Trans Intell Transp Syst 20, 2728–2739.

[8] Gao, S., Hou, Y., Dong, H., Stichel, S., Ning, B., 2019. High-speed trains automatic operation with protection constraints: A resilient nonlinear gain-based feedback control approach. IEEE/CAA Journal of Automatica Sinica 6, 992–999.

[9] Gao, S., Wei, J., Song, H., Zhang, Z., Dong, H., Hu, X., 2020. Fuzzy adaptive automatic train operation control with protection constraints: A residual nonlinearity approximation-based approach. Engineering Applications of Artificial Intelligence 96, 103986.

[10] Gao, S., Xu, C., Zhang, H., Zhao, N., Shen, T., Dong, H., 2024. Deterministic reinforcement learning for optimized formation control of virtually-coupled trains via performance index monitor. Expert Systems with Applications 237, 121421.

[11] Ge, S., Hang, C., Zhang, T., 1999. Adaptive neural network control of nonlinear systems by state and output feedback. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29, 818–828.

[12] Ge, X., Han, Q.L., Zhang, X.M., 2025. Secure virtual coupling control of connected train platoons under cyber attacks. Vehicle System Dynamics 63, 93–120.

[13] Huang, C., Wang, G., Zhou, Z., Zhang, R., Lin, L., 2022. Reward-adaptive reinforcement learning: Dynamic policy gradient optimization for bipedal locomotion. IEEE transactions on pattern analysis and machine intelligence 45, 7686–7695.

[14] Ji, Y., Quaglietta, E., Goverde, R.M., Ou, D., 2025. An artificial potential field approach for virtual coupling train control with complete braking curve supervision. Transportation Research Part C: Emerging Technologies 173, 105050.

[15] Kumar, G.M.S., Cao, S., 2023. Leveraging energy flexibilities for enhancing the cost-effectiveness and grid-responsiveness of net-zero-energy metro railway and station systems. Applied Energy 333, 120632.

[16] Kurdila, A., Narcowich, F.J., Ward, J.D., 1995. Persistency of excitation in identification using radial basis function approximants. SIAM journal on control and optimization 33, 625–642.

[17] Lan, J., Liu, Y.J., Yu, D., Wen, G., Tong, S., Liu, L., 2024. Time-varying optimal formation control for second-order multiagent systems based on neural network observer and reinforcement learning. IEEE Transactions on Neural Networks and Learning Systems 35, 3144–3155.

[18] Li, J., Tian, D., Zhou, J., Duan, X., Sheng, Z., Zhao, D., Cao, D., 2024. Distributed robust model predictive control for virtual coupling under structural and external uncertainty. IEEE Transactions on Intelligent Transportation Systems 25, 8751–8769.

[19] Li, S., Yang, L., Li, K., Gao, Z., 2014. Robust sampled-data cruise control scheduling of high speed train. Transportation Research Part C: Emerging Technologies 46, 274–283.

[20] Lin, J., Li, M., Qiu, X., 2025. A multi-objective optimization approach for the virtual coupling train set driving strategy. Railway Engineering Science 33, 169–191.

[21] Liu, H., Lang, Y., Luo, X., Tang, T., Chai, M., 2025. Coordinated operation strategy design for virtually coupled train set: A multiagent reinforcement learning approach. IEEE Intelligent Transportation Systems Magazine 17, 6–24.

[22] Liu, X., Dabiri, A., Wang, Y., Xun, J., De Schutter, B., 2024. Distributed model predictive control for virtually coupled heterogeneous trains: Comparison and assessment. IEEE Transactions on Intelligent Transportation Systems 25, 20753–20766.

[23] Liu, Y., Zhou, Y., Su, S., Xun, J., Tang, T., 2021. An analytical optimal control approach for virtually coupled high-speed trains with local and string stability. Transp Res Part C Emerg Technol 125, 102886.

[24] Löppenberg, M., Yuwono, S., Diprasetya, M., Schwung, A., 2024. Dynamic robot routing optimization: State–space decomposition for operations research-informed reinforcement learning. Robotics and Computer-Integrated Manufacturing 90, 102812.

[25] Ning, B., 1998. Absolute braking and relative distance braking-train operation control modes in moving block systems. WIT Transactions on The Built Environment 37.

[26] Ning, Z., Ou, D., Xie, C., Zhang, L., Gao, B., He, J., 2023. Optimal convoy composition for virtual coupling trains at junctions: A coalition formation game approach. Transportation Research Part C: Emerging Technologies 154, 104277.

[27] Onile, A., Belikov, J., Levron, Y., Petlenkov, E., 2023. Energy efficient behavior modeling for demand side recommender system in solar microgrid applications using multi-agent reinforcement learning model. Sustainable Cities and Society 90, 104392.

[28] Park, J., Lee, B.H., Eun, Y., 2020. Virtual coupling of railway vehicles: Gap reference for merge and separation, robust control, and position measurement. IEEE Transactions on Intelligent Transportation Systems 23, 1085–1096.

[29] Quaglietta, E., Spartalis, P., Wang, M., Goverde, R.M., van Koningsbruggen, P., 2022. Modelling and analysis of virtual coupling with dynamic safety margin considering risk factors in railway operations. Journal of Rail Transport Planning & Management 22, 100313.

[30] Quaglietta, E., Wang, M., Goverde, R.M., 2020. A multi-state train-following model for the analysis of virtual coupling railway operations. Journal of Rail Transport Planning & Management 15, 100195.

[31] Si, W., Gao, S., 2025. Safety prescribed performance control of virtually-coupled trains with multifold kinematic targets and switching constraints. ISA transactions 160, 19–40.

[32] Song, Y., Wang, H., Ma, Y., Yin, X., Cao, F., 2022. Energetic, economic, environmental investigation of carbon dioxide as the refrigeration alternative in new energy bus/railway vehicles' air conditioning systems. Applied energy 305, 117830.

[33] Su, S., Liu, W., Zhu, Q., Li, R., Tang, T., Lv, J., 2022. A cooperative collision-avoidance control methodology for virtual coupling trains. Accident Analysis & Prevention 173, 106703.

[34] Tian, S., Yan, F., Shang, W.L., Majumdar, A., Chen, H., Chen, M., Zeinab, M., Tian, Y., 2025. Energy consumption analysis of trains based on multi-mode virtual coupling operation control strategies. Applied Energy 377, 124684.

[35] Vamvoudakis, K., Lewis, F., 2010. Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem. Automatica 46, 878–888.

[36] Wang, H., Zhao, Q., Lin, S., Cui, D., Luo, C., Zhu, L., Wang, X., Tang, T., 2021. A reinforcement learning empowered cooperative control approach for IIoT-based virtually coupled train sets. IEEE Transactions on Industrial Informatics 17, 4935–4945.

[37] Wen, G., Xu, L., Li, B., 2023. Optimized backstepping tracking control using reinforcement learning for a class of stochastic nonlinear strict-feedback systems. IEEE Transactions on Neural Networks and Learning Systems 34, 1291–1303.

[38] Wen, G., Zhou, R., Zhao, Y., Niu, B., 2024a. Optimized backstepping combined with dynamic surface technique for single-input–single-output nonlinear strict-feedback system. IEEE Transactions on Systems, Man, and Cybernetics: Systems 54, 4210–4221.

[39] Wen, L., Zhou, W., Liu, J., Ren, G., Zhang, N., 2024b. Real-time optimization of urban rail transit train scheduling via advantage actor–critic deep reinforcement learning. Journal of Transportation Engineering, Part A: Systems 150, 04024046.

[40] Ye, Y., Papadaskalopoulos, D., Yuan, Q., Tang, Y., Strbac, G., 2022. Multi-agent deep reinforcement learning for coordinated energy trading and flexibility services provision in local electricity markets. IEEE Transactions on Smart Grid 14, 1541–1554.

[41] Zhang, H., Xu, K., Huang, D., He, D., Wu, S., Xian, G., 2024. Hybrid decision-making for intelligent high-speed train operation: A boundary constraint and pre-evaluation reinforcement learning approach. IEEE Transactions on Intelligent Transportation Systems 25, 17979–17992.

[42] Zhang, L., U, L.H., Zhou, M., Li, Z., 2022. An intelligent train operation method based on event-driven deep reinforcement learning. IEEE Transactions on Industrial Informatics 18, 6973–6980.

[43] Zhang, M., Pan, C., 2024. Hierarchical optimization scheduling algorithm for logistics transport vehicles based on multi-agent reinforcement learning. IEEE Transactions on Intelligent Transportation Systems 25, 3108–3117.

[44] Zhang, Q., Wang, H., Zhang, Y., Chai, M., 2023. An adaptive safety control approach for virtual coupling system with model parametric uncertainties. Transportation research part C: emerging technologies 154, 104235.

[45] Zhao, J., 2023. Adaptive dynamic programming-based adaptive optimal tracking control of a class of strict-feedback nonlinear system. International Journal of Control, Automation and Systems 21, 1349–1360.

[46] Zhu, L., Li, X., Huang, D., Qin, N., 2025. Cooperative control of virtually coupled trains considering train-to-train communication cycle: a tube-based model predictive control approach. Vehicle System Dynamics 63, 3–25.