Weighted-Sign Based Sequential Synchronization Control via Residual Reinforcement Learning

Xiaodong Zheng, Shuangsi Xue, and Hui Cao

***Abstract*—This paper addresses sequential synchronization for first-order affine systems in the presence of bounded disturbances. We develop a reinforcement learning (RL) based controller that combines an anisotropic weighted power-sign baseline with a projected residual actor-critic. The baseline uses a diagonal weighting to shape the direction field of a transformed error and thereby encode the desired convergence order across state channels. A lightweight projection comprising a norm cap and a half-space alignment with the descent direction filters the residual policy so as to preserve the baseline Lyapunov decrease. We establish fixed-time stability and sequential convergence for the baseline via a composite Lyapunov argument, and prove that the RL residual maintains these properties under the projection. A spacecraft translational tracking case study with bounded sinusoidal disturbances validates the design: the position states converge in the prescribed order, the residual actions remain consistently bounded with respect to the baseline control magnitude and direction, and learning signals stabilize. Quantitatively, axis-wise first-hit times, control energy, and peak control corroborate the theoretical guarantees while illustrating performance benefits of the residual layer.**

***Index Terms*—Sequential synchronization, fixed-time stability, weighted sign function, residual reinforcement learning, actor-critic**

I. Introduction

SEQUENTIAL synchronization and fast convergence are crucial for high-precision spacecraft docking and on-orbit servicing, where different state channels must converge with task-driven priorities under bounded disturbances and modeling errors. Classical finite-/fixed-/predefined-time control offers powerful non-asymptotic convergence guarantees and has recently seen renewed progress on consensus and coordination for nonlinear and networked systems [1–4]. In parallel, safe reinforcement learning (RL) with control-theoretic certificates (control Lyapunov function (CLF), control barrier function (CBF), and control Lyapunov-barrier function (CLBF)) has advanced rapidly, aiming to reconcile performance learning with stability and safety guarantees [5–9]. In particular, composite Lyapunov constructions and direction-field shaping provide a natural pathway to encode convergence priorities with uniform settling-time bounds, and multi-focus fixed-time stability (FTS) results without data-driven residual fusion have been reported [1, 10, 11]. However, integrating these ideas with residual learning policies remains under-explored.

For spacecraft guidance and control, multiple works have demonstrated the promise of RL-based policies for proximity maneuvers and docking, including six degree of freedom (DOF) settings and hierarchical formulations [12–15]. Meanwhile, theoretical efforts on continuous-time policy gradient/actor-critic provide stability or convergence guarantees under appropriate regularity or Lyapunov conditions [16–19]. These trends motivate hybrid approaches that blend certified nonlinear controllers with learning-based residuals to enhance performance while retaining provable properties, as advocated by residual RL in robotics and increasingly in control [20, 21]. Compared with CLF/CBF safety filters that typically rely on real-time quadratic programs (QPs), a projection-based residual layer imposes simple half-space and norm constraints on the learned action, making the architecture lightweight and amenable to continuous-time analysis while still preserving descent directions [22–24].

Existing safe-RL pipelines often focus on constraint satisfaction and asymptotic stability; they do not explicitly encode sequential convergence among state channels. On the other hand, fixed-/predefined-time schemes provide strong non-asymptotic convergence but can be conservative or sensitive when model uncertainty and disturbances are significant. There is a need for an integrated design that (1) preserves non-asymptotic convergence with an order-programmable mechanism, (2) admits adaptive residual learning to compensate uncertainty, and (3) retains Lyapunov-style guarantees end-to-end. Moreover, for spaceborne translation tasks, evidence beyond qualitative plots is desirable: hit-time statistics for each axis, control-energy and peak-control metrics, and parameter trajectories of the value approximation provide a compact and reproducible validation protocol [25].

Motivated by the discussion above, this paper proposes a reinforcement learning based sequential synchronization control strategy for first-order affine spacecraft systems. A

Manuscript received: 9 October 2025; revised: 31 October 2025; accepted: 7 November 2025. *(Corresponding author: Shuangsi Xue.)*

Citation: X. Zheng, S. Xue, and H. Cao, Weighted-sign based sequential synchronization control via residual reinforcement learning, *Int. J. Intell. Control Syst.*, 2025, 30(4), 336–342.

Xiaodong Zheng, Shuangsi Xue, and Hui Cao are with National Key Labratory for High Energy Pulsed Power, Xi'an Jiaotong University, Xi'an 710049, China, and also with School of Electrical Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: zxd_xjtu@stu.xjtu.edu.cn; xssxjtu@xjtu.edu.cn; huicao@mail.xjtu.edu.cn).

Hui Cao is also with The Faculty of Engineering & Advanced Robotics Centre, National University of Singapore, Singapore 117580, Singapore.

Digital Object Identifier 10.62678/IJICS202512.10255

weighted nonlinear sign function is introduced in the baseline controller to achieve sequential convergence of position states. An actor-critic residual reinforcement learning framework is incorporated to compensate for modeling uncertainties and external disturbances. Rigorous Lyapunov analysis demonstrates closed-loop stability. Numerical simulations with spacecraft translational dynamics validate the effectiveness of the proposed approach. The main contributions are:

(1) We introduce an anisotropic weighting matrix within a weighted sign operator to encode the desired convergence order. This yields finite-/fixed-time style contraction in a transformed error, aligning with recent non-asymptotic control trends [1, 10, 11, 26].

(2) A residual actor-critic augments the baseline while being projected to respect descent and magnitude constraints relative to the baseline, echoing safety-layer ideas in CLBF/CBF literature [5–7, 9]. We adopt a half-space alignment and a capped residual norm, forming a lightweight alternative to online QPs [5, 25].

(3) A composite Lyapunov proof showing FTS of the baseline and FTS preservation under the projected residual is provided, together with sequential convergence under bounded disturbances, building on continuous-time actor-critic theory [16, 17, 19].

The remainder of this paper is organized as follows. Section II formalizes the plant class, transformed errors, and assumptions. Section III introduces the anisotropic weighted-sign baseline and establishes fixed-time stability. Section IV proves sequential synchronization and shows that a projected residual actor-critic preserves both the fixed-time and sequence properties. Section V reports the simulation results. Section VI concludes and outlines future directions.

II. Problem Setup and Assumptions

We formalize the plant class, synchronization errors, and objectives. We also state the mild assumptions used by both the fixed-time analysis and the residual RL layer. The notation is aligned with the experimental section (single-spacecraft translation as a special case with $N = 1$).

Consider a multi-agent system (MAS) of N nonlinear affine agents

$$\dot{x}_i = f_i(x_i) + g_i(x_i)u_i,\ i = 1,2,\ldots,N \tag{1}$$

where $x_i \in \mathbb{R}^n$ and $u_i \in \mathbb{R}^m$. We assume that f_i and g_i are locally Lipschitz on a forward-invariant compact set $\mathcal{X} \subset \mathbb{R}^n$, and $g_i(x_i)$ has full column rank on $\mathcal{X}$, ensuring well-posedness and feedback implementability. Let $\mathcal{G} = (\mathcal{V},\mathcal{E},A)$ be the communication graph with weights $A = [a_{ij}]$ and neighbor set $\mathcal{N}_i$. We denote the (out-)Laplacian by $L := D - A$ and assume that $\mathcal{G}$ is weight-balanced and has a spanning tree, which guarantees consensus feasibility for standard diffusive couplings.

The consensus error is

$$e_i = \sum_{j\in\mathcal{N}_i} a_{ij}(x_i - x_j) \tag{2}$$

To encode sequence rules (limited/prescribed order, batching), we introduce a shaped error

$$\tilde{e}_i = T_i e_i + \eta_i \tag{3}$$

where $T_i \in \mathbb{R}^{n\times n}$ is invertible and η_i is an optional feedforward term (e.g., admissible arrival windows). This preserves $\tilde{e}_i = 0 \Leftrightarrow e_i = 0$ while enabling order programmability downstream. In particular, batch-wise common-mode offsets are filtered by T_i, and η_i can encode admissible windows or leader references. We assume that T_i is known and uniformly bounded with $\|T_i\|, \|T_i^{-1}\| \leqslant \bar{T}$.

Define the weighted error $y_i := \varepsilon_i\tilde{e}_i$ with diagonal weights $\varepsilon_i = \mathrm{diag}(\varepsilon_{i,1}, \varepsilon_{i,2}, \ldots, \varepsilon_{i,n}) > 0$, and the Lyapunov candidate

$$V = \frac{1}{2}\sum_{i=1}^{N}\|y_i\|^2 \tag{4}$$

We use $\|\cdot\|$ for the Euclidean norm and $\langle a,b\rangle = a^{\mathrm{T}}b$. For compactness, define $r_i := \|y_i\|$ and $S_2 := \sum_i r_i^2 = 2V$.

Objective: For a prescribed partial order $\prec$ over groups or channels, let $r_i(t) := \|y_i(t)\|$ and $T_i(\delta) := \inf\{t \geqslant 0 : r_i(t) \leqslant \delta\}$.

We seek: (1) fixed-time convergence (FTC) to synchronization; (2) sequential convergence $i \prec j \Rightarrow T_i(\delta) \leqslant T_j(\delta)$; and (3) favorable energy/transients, e.g., $\int_0^\infty u^{\mathrm{T}}Ru\,\mathrm{d}t$ and peak control. The single-spacecraft tracking case uses $N = 1$ and $e_1 = r - r_d$.

A. Assumptions

Assumption 1 Error-channel matching For $y_i = \varepsilon_i\tilde{e}_i$, the weighted error dynamics satisfy

$$\dot{y}_i = -\alpha_1\|y_i\|^{\ell_1-1}y_i - \alpha_2\|y_i\|^{\ell_2-1}y_i + d_i \tag{5}$$

with $\alpha_1, \alpha_2 > 0$, $\ell_1 > 1$, $0 < \ell_2 < 1$, and perturbations d_i obeying $\sum_{i=1}^{N} y_i^{\mathrm{T}}d_i \leqslant c_0\sum_{i=1}^{N}\|y_i\|^{\ell_2+1}$ for some $c_0 \in [0,\alpha_2)$. In particular, Eq. (5) is the specific relation that the weighted error dynamics satisfy, together with the stated growth bound.

Assumption 2 Weighted ratio invariance [27] Before any y_i hits zero, the projection on $\mathrm{sign}_{\varepsilon_i}(\tilde{e}_i) = y_i/\|y_i\|$ cancels common-mode couplings within batches and does not reverse the order induced by $\{\varepsilon_i\}$. Equivalently, along trajectories prior to any zero-crossing, batch-wise projections onto $\mathrm{span}\{\mathrm{sign}_{\varepsilon_k}(\tilde{e}_k)\}_{k\in\mathcal{B}}$ are order-preserving under the partial order defined by $\{\varepsilon_i\}$.

Assumption 3 Adaptive dynamic programming (ADP) regularity (full) Let $y = \varepsilon\tilde{e}$ be the transformed error used by the RL layer.

(1) The critic $V(y)$ and the actor $\Delta u(y)$ are chosen from sufficiently expressive, locally Lipschitz parametric families on a forward-invariant compact set $\mathcal{Y}$; parameter updates are projected onto compact convex sets, so all parameters remain bounded.

(2) The critic and actor regressors admit a mild persistence-of-excitation condition over a finite window, ensuring identifiability of the local value gradient and policy directions on $\mathcal{Y}$.

(3) The critic and actor use diminishing stepsizes $\{\beta_k\}$ and $\{\alpha_k\}$ satisfying Robbins-Monro conditions $\sum\beta_k = \infty$, $\sum\beta_k^2 < \infty$, $\sum\alpha_k = \infty$, $\sum\alpha_k^2 < \infty$, with $\alpha_k/\beta_k \to 0$.

(4) The residual action is the Euclidean projection of the actor's output onto the closed convex set defined by Formula (20); in particular, the projection is nonexpansive and preserves the baseline descent cone.

These standard conditions for continuous-time actor-critic/ ADP imply bounded iterates, well-posed TD errors, and convergence/improvement within the stabilizing policy class [16–19].

III. Weighted-Sign Baseline and Fixed-Time Stability

In this section, we define the anisotropic weighted-sign operator, design a two-exponent baseline, and prove fixed-time stability by a two-term Lyapunov inequality. Fig. 1 compacts the proposed baseline-plus-residual architecture. We explicitly define the weighted sign function and its power variant as follows.

Definition 1 Weighted sign and power-sign [27] For $\varepsilon = \mathrm{diag}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) > 0$, we have

$$\begin{aligned} \mathrm{sign}_\varepsilon(x) &= \frac{\varepsilon x}{\|\varepsilon x\|}, \\ \mathrm{sig}_\varepsilon^\ell(x) &= \|\varepsilon x\|^\ell \, \mathrm{sign}_\varepsilon(x) \end{aligned} \tag{6}$$

A. Baseline Control

For agent i, we design the baseline controller as

$$u_{i,\mathrm{base}} = -\alpha_1 \, \mathrm{sig}_{\varepsilon_i}^{\ell_1}(\tilde{e}_i) - \alpha_2 \, \mathrm{sig}_{\varepsilon_i}^{\ell_2}(\tilde{e}_i) \tag{7}$$

with $\ell_1 > 1$, $0 < \ell_2 < 1$, and $\alpha_1, \alpha_2 > 0$. Larger $\varepsilon_{i,k}$ accelerates the k-th channel, enabling order programming.

Lemma 1 FTS criterion Let V be positive definite and radially unbounded. If

$$\dot{V} \leqslant -(\alpha V^p + \beta V^g)^\chi \tag{8}$$

where $\alpha, \beta > 0$, $p \in (0,1)$, $g > 1$, and $\chi \in (0,1]$, then the origin is FTS with settling-time bound

$$T \leqslant \frac{1}{\alpha^\chi (1 - p\chi)} + \frac{1}{\beta^\chi (g\chi - 1)} \tag{9}$$

Proof 1 For $V \geqslant 1$, $V^{g\chi} \geqslant V^{p\chi}$, thus $\dot{V} \leqslant -\beta^\chi V^{g\chi}$ and

$$\int_{V(0)}^{1} \frac{\mathrm{d}V}{V^{g\chi}} \leqslant -\beta^\chi \int_0^{T_1} \mathrm{d}t \tag{10}$$

Then, we have

$$T_1 \leqslant \frac{1}{\beta^\chi (g\chi - 1)} (1 - V(0)^{1 - g\chi}) \leqslant \frac{1}{\beta^\chi (g\chi - 1)} \tag{11}$$

For $V \in [0,1]$, $\dot{V} \leqslant -\alpha^\chi V^{p\chi}$ yields

$$\int_1^0 \frac{\mathrm{d}V}{V^{p\chi}} \leqslant -\alpha^\chi \int_0^{T_2} \mathrm{d}t \;\Rightarrow\; T_2 \leqslant \frac{1}{\alpha^\chi (1 - p\chi)}.$$

Therefore, $T \leqslant T_1 + T_2$ and the bound is uniform in $V(0)$, proving fixed-time stability. When $\chi = 1$, we recover the standard two-exponent criterion widely used in FTS. ■

Theorem 1 FTS of the baseline Under Assumption 1, the closed-loop system with Eq. (7) satisfies

$$\dot{V} \leqslant -\alpha V^p - \beta V^g,$$

for some $\alpha, \beta > 0$, with $p = (\ell_2 + 1)/2 \in (0,1)$ and $g = (\ell_1 + 1)/2 > 1$. Hence the origin is FTS with bound in Formula (9) (for $\chi = 1$).

Proof 2 From Eq. (5), we have

$$\dot{V} = \sum_{i=1}^{N} y_i^{\mathrm{T}} \dot{y}_i = -\alpha_1 \sum_{i=1}^{N} \|y_i\|^{\ell_1 + 1} - \alpha_2 \sum_{i=1}^{N} \|y_i\|^{\ell_2 + 1} + \sum_{i=1}^{N} y_i^{\mathrm{T}} d_i \tag{12}$$

Assumption 1 gives $\sum_i y_i^{\mathrm{T}} d_i \leqslant c_0 \sum_i \|y_i\|^{\ell_2 + 1}$ with $c_0 < \alpha_2$, hence

$$\dot{V} \leqslant -\alpha_1 \sum_i \|y_i\|^{\ell_1 + 1} - (\alpha_2 - c_0) \sum_i \|y_i\|^{\ell_2 + 1}.$$

Let $r_i := \|y_i\| \geqslant 0$ and $S_2 := \sum_i r_i^2 = 2V$. By norm equivalence (for $s \geqslant 1$), we have

$$\sum_{i=1}^{N} r_i^s = \|r\|_s^s \geqslant N^{1 - \frac{s}{2}} \|r\|_2^s = N^{1 - \frac{s}{2}} (2V)^{\frac{s}{2}}.$$

Setting $s_1 = \ell_1 + 1 > 2$ and $s_2 = \ell_2 + 1 \in (1,2)$ yields explicit constants $k_1 = N^{1 - \frac{s_1}{2}}$ and $k_2 = N^{1 - \frac{s_2}{2}}$ such that

$$\begin{aligned} \sum_i r_i^{\ell_1 + 1} &\geqslant k_1 (2V)^{\frac{\ell_1 + 1}{2}}, \\ \sum_i r_i^{\ell_2 + 1} &\geqslant k_2 (2V)^{\frac{\ell_2 + 1}{2}} \end{aligned} \tag{13}$$

Collect constants to obtain $\dot{V} \leqslant -\alpha V^g - \beta V^p$ with $p = (\ell_2 + 1)/2$ and $g = (\ell_1 + 1)/2$, then apply Lemma 1 (with $\chi = 1$). ■

IV. Sequential Synchronization and Residual RL Preservation

In this section, we show that anisotropic weights induce ordered hitting times (sequential synchronization). We then overlay a projected residual actor-critic and prove that fixed-time and sequence properties are preserved.

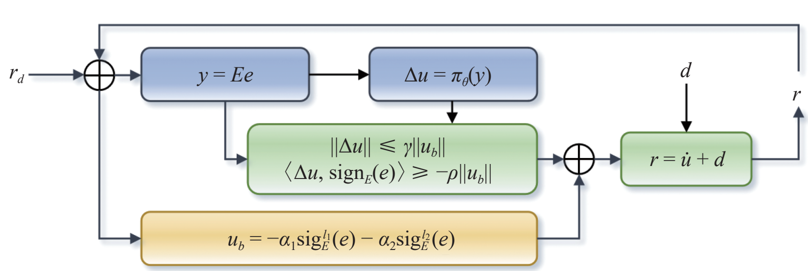

Fig. 1 Compact control architecture.

A. Sequential Synchronization via Weighting

Let $s_i \in \mathbb{N}$ be the order label (smaller s_i earlier). Define $r_i(t) := \|y_i(t)\|$ and $T_i(\delta) := \inf\{t \geqslant 0 : r_i(t) \leqslant \delta\}$.

Theorem 2 Restricted sequence synchronization If Assumptions 1 and 2 hold and

$$\begin{aligned} s_i < s_j &\Rightarrow \varepsilon_i \geqslant \varepsilon_j, \\ s_i = s_j &\Rightarrow \varepsilon_i = \varepsilon_j \end{aligned} \tag{14}$$

with initial consistency

$$\begin{aligned} s_i < s_j &\Rightarrow r_i(0) \leqslant r_j(0), \\ s_i = s_j &\Rightarrow r_i(0) = r_j(0) \end{aligned} \tag{15}$$

then for any small $\delta > 0$, we have

$$\begin{aligned} s_i < s_j &\Rightarrow T_i(\delta) \leqslant T_j(\delta), \\ s_i = s_j &\Rightarrow T_i(\delta) = T_j(\delta) \end{aligned} \tag{16}$$

Proof 3 From Eq. (5) and $r_i = \|y_i\|$, we have

$$\dot{r}_i = \frac{y_i^{\mathrm{T}} \dot{y}_i}{\|y_i\|} = -\alpha_1 r_i^{\ell_1} - \alpha_2 r_i^{\ell_2} + \frac{y_i^{\mathrm{T}} d_i}{r_i}.$$

Assumption 1 implies $|y_i^{\mathrm{T}} d_i|/r_i \leqslant c_0 r_i^{\ell_2}$, hence

$$-\alpha_1 r_i^{\ell_1} - (\alpha_2 + c_0) r_i^{\ell_2} \leqslant \dot{r}_i \leqslant -\alpha_1 r_i^{\ell_1} - (\alpha_2 - c_0) r_i^{\ell_2} \tag{17}$$

For $h_{ij}(t) := r_i(t) - r_j(t)$, monotonicity of $z \mapsto z^{\ell}$ yields

$$\begin{aligned} \dot{h}_{ij} &\leqslant -\alpha_1 (r_i^{\ell_1} - r_j^{\ell_1}) - (\alpha_2 - c_0)(r_i^{\ell_2} - r_j^{\ell_2}), \\ \dot{h}_{ij} &\geqslant -\alpha_1 (r_i^{\ell_1} - r_j^{\ell_1}) - (\alpha_2 + c_0)(r_i^{\ell_2} - r_j^{\ell_2}), \end{aligned}$$

so h_{ij} cannot change sign unless $r_i = r_j$. With Formula (15), $h_{ij}(0) \leqslant 0$ for $s_i \leqslant s_j$ and remains so until thresholds. Within the same batch, Assumption 2 implies $r_i(t) \equiv r_j(t)$, hence equal hitting times. This comparison is valid up to the first hitting time since r_i is absolutely continuous and positive on $(0, T)$ before crossing. ■

Fig. 2 illustrates sequential convergence under anisotropic weights: larger entries in ε (or E for the single-vehicle case) tighten the corresponding channels so that the induced hitting times satisfy the order constraints. The three colored curves represent different state variables.

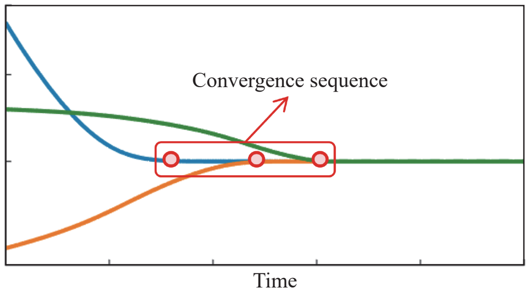

Fig. 2 Illustration of sequential convergence.

Theorem 3 Prescribed sequence synchronization [27] Under Theorem 2, if a sampled version of weighted ratio invariance holds and

$$s_i < s_j \Rightarrow \varepsilon_i > \varepsilon_j \tag{18}$$

with strict initial separation $s_i < s_j \Rightarrow r_i(0) < r_j(0)$, then for any $\delta > 0$, $T_i(\delta) < T_j(\delta)$. The sampled invariance means that at the controller update instants $\{t_k\}$, the one-step map induced by the baseline/residual keeps $r_i(t_{k+1}) \leqslant r_j(t_{k+1})$ whenever $r_i(t_k) \leqslant r_j(t_k)$, ensuring strict order under Formula (18).

B. Residual Actor-Critic and Stability Preservation

We augment the baseline by

$$u_i = u_{i,\text{base}} + \Delta u_i(\theta_i) \tag{19}$$

with critic $V(y) = b_0 + b^{\mathrm{T}} y + \frac{1}{2} y^{\mathrm{T}} \mathrm{diag}(P) y$, actor $\Delta u_i = W_2 \tanh(W_1 y + b_1) + b_2$, cost $c = e^{\mathrm{T}} Q e + u^{\mathrm{T}} R u$, and TD error $\delta = c\Delta t + V(y^+) - V(y)$. The projection constraints are

$$\begin{aligned} &\|\Delta u_i\| \leqslant \gamma \|u_{i,\text{base}}\|, \\ &\langle \Delta u_i, \, \mathrm{sign}_{\varepsilon_i}(\tilde{e}_i) \rangle \geqslant -\rho \|u_{i,\text{base}}\|, \\ &\Delta u_i(0) = 0 \end{aligned} \tag{20}$$

with $\gamma \in (0,1)$ and small $\rho \geqslant 0$. The feasible set in Formula (20) is closed and convex; the Euclidean projection onto it is nonexpansive, and the half-space constraint preserves the baseline descent cone.

Lemma 2 Projection preserves two-exponent decay Assume $\|g_i(x_i)\| \leqslant \bar{g}$ on $\mathcal{X}$. Then there exists $\bar{\gamma} \in (0,1)$ such that for any $\gamma \in (0, \bar{\gamma})$ and sufficiently small $\rho \geqslant 0$

$$\dot{V} \leqslant -\tilde{\alpha} V^{g} - \tilde{\beta} V^{p},$$

with some $\tilde{\alpha}, \tilde{\beta} > 0$, $p = (\ell_2 + 1)/2 \in (0,1)$, and $g = (\ell_1 + 1)/2 > 1$.

Proof 4 From Eqs. (1), (7), and (19), the residual contribution satisfies

$$\begin{aligned} \sum_{i=1}^{N} y_i^{\mathrm{T}} g_i(x_i) \Delta u_i \leqslant \sum_{i=1}^{N} \|y_i\| \|g_i\| \|\Delta u_i\| \leqslant \\ \bar{g} \gamma \sum_{i=1}^{N} \|y_i\| \|u_{i,\text{base}}\| \end{aligned} \tag{21}$$

Using $\|u_{i,\text{base}}\| \leqslant \alpha_1 \|y_i\|^{\ell_1} + \alpha_2 \|y_i\|^{\ell_2}$ and Young's inequality, for any $\eta \in (0,1)$

$$\begin{aligned} \sum_i \|y_i\| \|u_{i,\text{base}}\| \leqslant \eta \sum_i (\alpha_1 \|y_i\|^{\ell_1 + 1} + \alpha_2 \|y_i\|^{\ell_2 + 1}) + \\ C_\eta \sum_i \|y_i\|^2 \end{aligned} \tag{22}$$

where $C_\eta > 0$ depends only on $(\alpha_1, \alpha_2, \eta)$. Pick γ so that $\bar{g}\gamma\eta$ is dominated by the baseline's negative terms; the quadratic remainder $C_\eta(2V)$ is dominated by V^p near the origin and by V^g when V is large. Aggregating constants and reusing the norm inequalities in Theorem 1 yields the claim. The half-space constraint (parameter ρ) only improves the bound. ■

Theorem 4 RL layer preserves FTS and sequence Under Assumptions 1–3 and projection in Formula (20), for sufficiently small γ and ρ, the residual RL preserves the FTS bound of Theorem 1 (up to constants) and the sequence properties in Theorems 2 and 3. Moreover, actor-critic updates monotonically improve the performance objective within the stabilizing policy class.

Proof 5 Lemma 2 gives the same two-exponent decay as Theorem 1; Lemma 1 implies FTS (with possibly modified constants). Because the projection keeps Δu_i within a stabilizing cone aligned with $\mathrm{sign}_{\varepsilon_i}(\tilde{e}_i)$ and $\Delta u_i(0) = 0$, Assumption 2 is not violated, hence the sequence properties persist. Standard continuous-time ADP arguments under Assumption 3 ensure monotone improvement of the value within the projected policy class. ■

Remark 1 Different from event-triggered or observer-based fixed-time consensus designs for MASs [2, 28, 29] and predefined-/fixed-time tracking in aerospace systems [26, 30–32], our method targets single-vehicle affine dynamics with explicit sequential convergence via E and integrates a projected residual actor-critic. Compared with CLF-/CBF-based safe RL [5–9], our design employs a finite-/fixed-time baseline Lyapunov descent together with a constrained residual, thereby retaining contraction while improving performance. Finally, unlike aerospace RL works that primarily emphasize policy training and transfer [12–15], we formalize stability guarantees via a composite Lyapunov argument and encode the convergence order in the controller. To our knowledge, the axis-wise hitting-time evidence for "order preservation under learning" has been seldom quantified in prior RL-based spacecraft control studies [24, 25]. We stress that our guarantees are baseline-certified: learning operates only within a projection that preserves the fixed-time descent cone.

V. SIMULATION RESULTS

This section instantiates the proposed framework on a single affine plant (the translational motion of a spacecraft). We adopt the same weighted variables and controller notation as in Sections II–IV.

A. RL Design

We consider the first-order affine translation model

$$\dot{r}(t) = u(t) + d(t) \tag{23}$$

where $r(t) = [x(t), y(t), z(t)]^{\mathrm{T}}$ with a constant desired position $r_d = [1.0, -0.5, 0.8]^{\mathrm{T}}$. The tracking error is $e(t) = r(t) - r_d$. Following Section II, we use the transformed error

$$y(t) = \varepsilon\tilde{e}(t) = E\,e(t) \tag{24}$$

where $E = \mathrm{diag}(3.0, 1.0, 0.3)$, so that the order $x \succ y \succ z$ is encoded by the anisotropy of E. The fixed-time baseline law adopts the weighted power-sign form in Eq. (7), specialized to the single-agent case

$$u_{\mathrm{base}}(t) = -a_1\,\mathrm{sig}_E^{\ell_1}(e(t)) - a_2\,\mathrm{sig}_E^{\ell_2}(e(t)) \tag{25}$$

with $(a_1, a_2, \ell_1, \ell_2)$ given below. A residual actor-critic augments the baseline according to Eq. (19), and is constrained by the projection in Formula (20) (norm cap and half-space alignment relative to $\mathrm{sign}_E(e)$), so that the two-exponent Lyapunov decay and the order property are preserved (Theorem 4).

B. Details and Parameters

Unless otherwise stated, the initial condition is $r(0) = [4, 3, -2.5]^{\mathrm{T}}$, so that the three channels exhibit distinct transients. The disturbance is

$$d(t) = 0.02\,[\sin(0.7t), \cos(0.5t), \sin(0.9t + 0.5)]^{\mathrm{T}},$$

which is bounded and fits Assumption 1. The remaining settings are as follows:

- Baseline gains and exponents: $a_1 = 1.2$, $a_2 = 0.8$, $\ell_1 = 1.6$, and $\ell_2 = 0.6$.
- Critic: $V(y) = b_0 + b^{\mathrm{T}}y + \frac{1}{2}y^{\mathrm{T}}\mathrm{diag}(P)y$; learning rate 5×10^{-3}.
- Actor: $\Delta u = W_2\tanh(W_1 y + b_1) + b_2$; hidden size $h = 8$; learning rate 10^{-3}.
- Cost: $c(t) = e^{\mathrm{T}}Qe + u^{\mathrm{T}}Ru$ with $Q = I_3$ and $R = 0.05\,I_3$.
- Projection (Formula (20)): $\gamma = 0.25$ and $\rho = 0.20$.
- Discretization: forward Euler with $\Delta t = 0.002$ s, horizon $T = 20$ s.

We report the control energy $J_u = \int_0^T \|u(t)\|^2\mathrm{d}t$ (trapezoidal rule), the peak control $\|u\|_\infty = \max_{t\in[0,T]}\|u(t)\|$, and the axis-wise first-hit times $t_x^\star$, $t_y^\star$, and $t_z^\star$ defined as the first instants when $|e_x|$, $|e_y|$, and $|e_z|$ fall below 10^{-3}. These hitting times are overlaid as vertical dashed lines in the plots for visual certification of the order $x \to y \to z$ (Theorems 2 and 3).

The trajectories (Fig. 3) and error decays (Fig. 4) verify that all states converge to their references and that the axis-wise order $x \to y \to z$ is realized by the constant anisotropic weighting E, exactly as predicted by Theorems 2 and 3. The per-axis controls (Fig. 5) satisfy the projection constraints in Formula (20), which, together with the baseline two-exponent structure, preserves the fixed-time Lyapunov decay (Theorem 1) in the presence of the residual RL in Theorem 4. In Table 1, the RL residual maintains energy and peak control at levels comparable to the baseline while achieving earlier convergence on prioritized axes, illustrating that performance improvements can be attained without compromising the fixed-time and sequence guarantees.

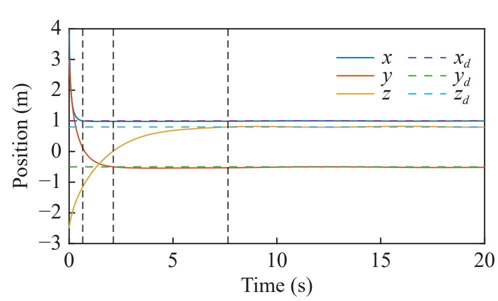

Fig. 3 Trajectory versus desired setpoints.

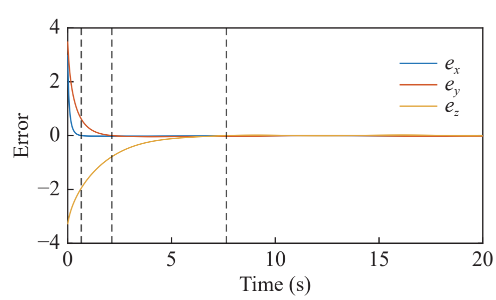

Fig. 4 Tracking errors e_x, e_y, e_z and ordered first-hit times $(t_x^\star, t_y^\star, t_z^\star)$.

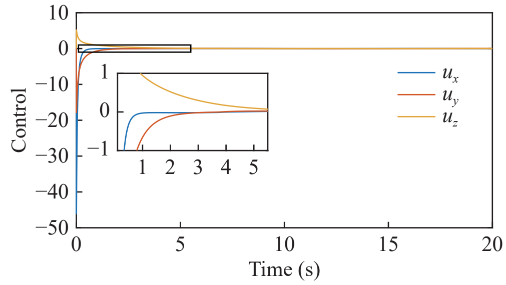

Fig. 5 Per-axis control inputs u_x, u_y, and u_z with the projection envelope implied by Formula (20).

Table 1 reports the energy J_u, peak control $\|u\|_\infty$, and the first-hit times $(t_x^\star, t_y^\star, t_z^\star)$ for the baseline and the baseline+RL cases. In our run (seed fixed for reproducibility), baseline achieves $(J_u, \|u\|_\infty) = (84.20, 49.85)$ with $(t_x^\star, t_y^\star, t_z^\star) = (4.028, 3.622, 7.194)$, while baseline+RL yields $(84.30, 49.69)$ and $(0.654, 2.126, 7.646)$. Two observations follow: (1) both controllers track successfully; (2) the RL residual (with projection) recovers the prescribed order $x \to y \to z$ (whereas baseline shows a mild $x \to y$ swap under disturbance and discretization) and significantly reduces $t_x^\star$ without inflating the control peak. The slight increase in J_u under RL is expected: the residual explicitly spends a small budget to improve transients and order compliance while respecting the stabilizing cone.

Table 1 Quantitative metrics.

Method	J_u	$\|u\|_\infty$	$t_x^\star$	$t_y^\star$	$t_z^\star$
Baseline	84.2000	49.8481	4.028	3.622	7.194
Baseline+RL	84.2952	49.6887	0.654	2.126	7.646

The ordered hitting times observed in Fig. 4 are consistent with Theorems 2 and 3, since $E = \mathrm{diag}(3, 1, 0.3)$ imposes $x \succ y \succ z$. The bounded residual and the stabilization of the critic parameters in Fig. 6 corroborate Theorem 4: the projection keeps u_{res} within a stabilizing cone so that the two-exponent Lyapunov decay is preserved, while the actor-critic improves transient order and speed.

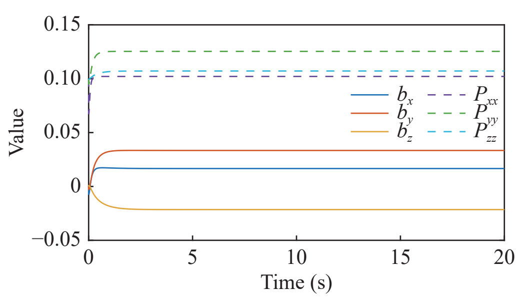

Fig. 6 Critic parameter trajectories (b components and diag(P)).

VI. Conclusion

We presented a residual RL-based controller for sequential synchronization of first-order affine systems under bounded disturbances. An anisotropic weighted power-sign baseline encodes channel priorities and guarantees fixed-time stability via a two-exponent Lyapunov descent, while a lightweight projection keeps the residual actor-critic inside a baseline-certified descent cone so that both fixed-time contraction and the prescribed convergence order are preserved. A spacecraft translation case study confirmed these properties: the desired order was achieved, residual actions remained bounded relative to the baseline, and axis-wise first-hit times improved without inflating control energy or peak inputs. Future work will incorporate explicit constraints and extend to six-DOF dynamics and adaptive weight design.

Acknowledgment

This work was supported by the China Postdoctoral Science Foundation (No. 2024M762602).

References

[1] J. Zhuang, S. Peng, and H. Peng, Finite-time and fixed-time consensus of nonlinear multi-agent systems: A unified two-phase control, *Int. J. Robust Nonlinear Control*, 2024, 34(14), 9438–9455.

[2] C. Ge, L. Ma, and S. Xu, Distributed fixed-time leader-following consensus for multi-agent systems: An event-triggered mechanism, *Actuators*, 2024, 13(1), 40.

[3] L. Hou and X. Lan, Fixed-time consensus multi-agent-systems-based speed cooperative control for multiple permanent magnet synchronous motors with complementary sliding mode control, *Electronics*, 2024, 13(22), 4407.

[4] S. Wang, J. Xiao, X. Hu, Z. Zuo, and T. Chen, Finite-time general function consensus for multi-agent systems over signed digraphs, *J. Franklin Inst.*, 2023, 360(12), 7808–7831.

[5] M. H. Cohen and C. Belta, Safe exploration in model-based reinforcement learning using control barrier functions, *Automatica*, 2023, 147, 110684.

[6] Y. Emam, G. Notomista, P. Glotfelter, Z. Kira, and M. Egerstedt, Safe reinforcement learning using robust control barrier functions, *IEEE Robot. Autom. Lett.*, 2025, 10(3), 2886–2893.

[7] Y. Cheng, P. Zhao, and N. Hovakimyan, Safe and efficient reinforcement learning using disturbance-observer-based control barrier functions, in *Proc. 5th Annual Conference on Learning for Dynamics and Control*, Philadelphia, PA, USA, 2023, 104–115.

[8] K. Garg, J. Usevitch, J. Breeden, M. Black, D. Agrawal, H. Parwana, and D. Panagou, Advances in the theory of control barrier functions: Addressing practical challenges in safe control synthesis for autonomous and robotic systems, *Annu. Rev. Control*, 2024, 57, 100945.

[9] Y. Wang and Z. Wu, Control Lyapunov-barrier function-based safe reinforcement learning for nonlinear optimal control, *AIChE J.*, 2024, 70(3), e18306.

[10] L. Wang, C. Deng, Z. Pan, and Q. Fang, Fixed-time input-to-state stability via generalized homogeneous approximations, *Int. J. Control Autom. Syst.*, 2024, 22(1), 116–126.

[11] H. Wen, W. Zheng, and H. Du, Homogeneous constrained finite-time controller for double integrator systems: Analysis and experiment, *Automatica*, 2021, 134, 109894.

[12] K. Hovell and S. Ulrich, Deep reinforcement learning for spacecraft proximity operations guidance, *J. Spacecr. Rockets*, 2021, 58(2), 254–264.

[13] C. E. Oestreich, R. Linares, and R. Gondhalekar, Autonomous six-degree-of-freedom spacecraft docking with rotating targets via reinforcement learning, *J. Aerosp. Inf. Syst.*, 2021, 18(7), 417–428.

[14] A. Aborizk and N. Fitz-Coy, Multiphase autonomous docking via model-based and hierarchical reinforcement learning, *J. Spacecr. Rockets*, 2024, 61(4), 993–1005.

[15] D. Du, N. Qi, Y. Liu, and W. Pan, Deep Bayesian reinforcement learning for spacecraft proximity maneuvers and docking, arXiv preprint arXiv: 2311.03680, 2023.

[16] Y. Jia and X. Zhou, Policy gradient and actor-critic learning in continuous time and space: Theory and algorithms, *J. Mach. Learn. Res.*, 2022, 23(275), 1–50.

[17] J. Duan, J. Li, Q. Ge, S. Li, M. Bujarbaruah, F. Ma, and D. Zhang, Relaxed actor-critic with convergence guarantees for continuous-time

optimal control of nonlinear systems, *IEEE Trans. Intell. Veh.*, 2023, 8(5), 3299–3311.

[18] M. Han, L. Zhang, J. Wang, and W. Pan, Actor-critic reinforcement learning for control with stability guarantee, *IEEE Robot. Autom. Lett.*, 2020, 5(4), 6217–6224.

[19] M. Han, Y. Tian, L. Zhang, J. Wang, and W. Pan, Reinforcement learning control of constrained dynamic systems with uniformly ultimate boundedness stability guarantee, *Automatica*, 2021, 129, 109689.

[20] T. Johannink, S. Bahl, A. Nair, J. Luo, A. Kumar, M. Loskyll, J. A. Ojea, E. Solowjow, and S. Levine, Residual reinforcement learning for robot control, in *Proc. 2019 International Conference on Robotics and Automation*, Montreal, Canada, 2019.

[21] R. C. B. Rego and F. M. U. de Araújo, Nonlinear control system with reinforcement learning and neural networks based Lyapunov functions, *IEEE Latin Am. Trans.*, 2021, 19(8), 1253–1260.

[22] J. Tan, S. Xue, H. Cao, and H. Li, Nash equilibrium solution based on safety-guarding reinforcement learning in nonzero-sum game, in *Proc. 2023 International Conference on Advanced Robotics and Mechatronics*, Sanya, China, 2023, 630–635.

[23] J. Tan, S. Xue, H. Cao, and H. Li, Safe human-machine cooperative game with level-k rationality modeled human impact, in *Proc. 2023 IEEE International Conference on Development and Learning*, Macao, China, 2023, 188–193.

[24] M. Guerrier, H. Fouad, and G. Beltrame, Learning control barrier functions and their application in reinforcement learning: A survey, arXiv preprint arXiv: 2404.16879, 2024.

[25] A. Wachi, X. Shen, and Y. Sui, A survey of constraint formulations in safe reinforcement learning, in *Proc. 33rd International Joint Conference on Artificial Intelligence*, Jeju, Korea, 2024, 913.

[26] V. K. Saini, D. K. Giri, and R. Varatharajoo, Prescribed-time spacecraft attitude control using time-varying state feedback, *J. Spacecr. Rockets*, 2023, 60(4), 1242–1255.

[27] H. Li, S. Xue, Z. Guo, J. Tan, H. Cao, and D. Li, Sequence-based group consensus for heterogeneous multi-agent systems, *IEEE Trans. Signal Inf. Process. Netw.*, 2025, 11, 1188–1199.

[28] G. Cao, G. Xiong, X. Hong, and X. Tan, Observer-based fixed-time consensus tracking for a class of nonlinear multi-agent systems, *Electronics*, 2024, 13(22), 4473.

[29] L. Lu, B. Xiao, T. Han, X.-F. Zhao, and H. Yan, Observer-based fixed-time consensus tracking for second-order multi-agent systems with disturbances, *Trans. Inst. Meas. Control*, 2024, 46(1), 70–77.

[30] Y. Zhou, J. Liu, G.-P. Jiang, and C. Li, Consensus of multi-agent systems via prescribed-time sliding mode control method, *Trans. Inst. Meas. Control*, 2022, 44(12), 2369–2377.

[31] Q. Huang and Y. Zhang, Fixed-time fault-tolerance attitude tracking control for spacecraft without unwinding, *IET Control Theory Appl.*, 2023, 17(4), 357–367.

[32] M.-L. Zhuang and S.-M. Song, Fixed-time coordinated attitude tracking control for spacecraft formation flying considering input amplitude constraint, *Int. J. Control, Autom. Syst.*, 2022, 20(7), 2129–2147.

Xiaodong Zheng received the BS degree in electrical engineering from Florida Institute of Technology, Melbourne, FL, USA, in 2018, and the MS degree in electrical engineering from University of Southern California, Los Angeles, CA, USA, in 2020. He is currently pursuing the PhD degree in electrical engineering at School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, China. His research interests include multi-agent reinforcement learning and distributed control.

Shuangsi Xue received the BE degree in electrical engineering and automation from Hunan University, Changsha, China, in 2014, and the ME and PhD degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2018 and 2023, respectively. He is currently an assistant professor at School of Electrical Engineering, Xi'an Jiaotong University, China. His research interests include adaptive control and data-driven control of networked systems.

Hui Cao received the BE, ME, and PhD degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2000, 2004, and 2009, respectively. He is currently a professor at School of Electrical Engineering, Xi'an Jiaotong University, China. He was a postdoctoral research fellow at Department of Electrical and Computer Engineering, National University of Singapore, Singapore, from 2014 to 2015. He has authored or coauthored over 30 scientific and technical papers in recent years. His research interests include knowledge representation and discovery. He was a recipient of the Second Prize of the National Technical Invention Award.