Article

Not peer-reviewed version

Topology-Driven Anti-Entanglement Control for Soft Robots

Haoyang Le * and Shengxuan Wang

Posted Date: 30 October 2025

doi: 10.20944/preprints202510.2421.v1

Keywords: soft robot; topological perception; multi-agent reinforcement learning; knot avoidance

 doi:10.20944/preprints202510.2421.v1

Article

Topology-Driven Anti-Entanglement Control for Soft Robots

Haoyang Le * and Shengxuan Wang *

Independent Researcher, China
* Correspondence: 891086268@qq.com (H.L.); 1730363705@qq.com (S.W.)

Highlights

- For the collaborative anti-winding management of the multi-soft robot operating arm system in a high-constraint environment, a topologically-aware multi-intelligent body-enhanced learning framework is proposed.
- Integrate topological invariants such as surround number and braid group representation into the state space and reward framework to achieve accurate quantification and early detection of winding risk.
- Build a space-time topology dynamic concurrency control framework, introduce a hierarchical reward structure, and simultaneously improve task efficiency and topological security.
- Adopt a dual-experience playback mechanism to integrate topological complex data into a centralized buffer to realize the collaborative optimization of strategic networks and value networks.
- The topological perception security layer is embedded in the architecture, which can pre-evaluate and mitigate high-risk actions to ensure the resilience and reliability of system collaboration functions.
- Theoretical analysis confirms that the proposed algorithm is convergent and robust within the established calculation framework.
- A large number of simulation results show that this method can significantly reduce entanglement events while maintaining high task throughput. In terms of security, collaborative performance and system stability, it is superior to traditional motion planning algorithms and benchmark reinforcement learning strategies.

Abstract

In the field of precision manufacturing in complex constrained environments, the role of soft robots is increasingly prominent, and the realization of anti-winding control based on multi-intelligent body reinforcement learning has become a research hotspot. One of the core problems at present is to coordinate multiple robots to complete the unwinding operation in a highly constrained environment. The existing distributed training framework faces some observability challenges in high-density barrier and unstable environments, resulting in poor learning results. This paper proposes a Topology-Aware Multi-Agent Reinforcement Learning (TA-MARL) framework to coordinate multi-robot systems to avoid entanglement. Specifically, the critical network adopts centralized learning, so that each intelligent body can perceive the strategies of other intelligent bodies by sharing the topological state, thus alleviating the training instability caused by complex interactions; eliminating the demand for communication resources between robots through distributed execution, Upgrade system reliability; the integrated topological security layer uses topological invariants to accurately assess and mitigate the risk of entanglement to avoid the strategy from falling into local difficulties. Finally, the full simulation experiments carried out in the real simulation environment show that the method is better than the current advanced deep reinforcement learning (DRL) method in terms of convergence and anti-winding effect.

Keywords: soft robot; topological perception; multi-agent reinforcement learning; knot avoidance

 doi:10.20944/preprints202510.2421.v1

1. Introduction

The multi-soft robot collaboration system has become a key paradigm for performing complex operational tasks in a constrained environment. Its applications cover industrial maintenance, precision manufacturing and minimally invasive surgery [1]. Compared with traditional rigid robot operating arms, soft robots have stronger adaptability and are more suitable for unstructured operating environments. However, in the multi-arm collaborative scenario, the inherent flexibility of the multi-robot system makes it extremely prone to entanglement and jamming. This kind of topological deadlock mainly comes from global constraints rather than local geometric collisions, which may lead to system failure, reduced operational efficiency and safety hazards, and has become an important obstacle to the reliable deployment of multi-soft robots. Therefore, this article is committed to developing an efficient multi-soft robot coordination framework suitable for complex environments.

With the advancement of multi-soft robot collaboration research, the academic community has proposed a variety of system solutions for multi-soft robot collaboration, including paradigms based on dynamic motion planning, imitation learning and multitask reinforcement learning. In the field of traditional motion planning, sampling-based algorithms such as RRT* [2] and optimization-based methods [3] show progressive optimization in geometric path search, but these methods are essentially reactive methods, relying on the instantaneous environmental state and lack explicit modeling of the history of system motion.

Such methods usually decouple trajectory planning from tracking control [4], ignoring the inherent nonlinear dynamic characteristics of soft robots, and hindering the maintenance of topological security in the real-time control process. Especially in a dynamic uncertain environment, it is difficult for traditional methods to respond to mobility obstacles and self-deformation in real time, which makes online re-planning to maintain topological security challenging.

In the field of multi-intelligence reinforcement learning, although the centralized training with decentralized execution framework [5] provides a new paradigm for collaborative control, the centralized method is false. There is a core service provider, which masters the global information of all intelligent bodies (such as status and observations) and decides on the behavior of intelligent bodies, and generates global control instructions based on system-wide observations to ensure security, integrity and optimality. At present, the realization of centralized training combined with the distributed execution framework still faces many challenges: affected by the non-stableness of the multi-intelligent system, the existing value function decomposition method [6] is difficult to accurately trace the long-term historical action sequence that leads to entanglement; in some observable environments, decision-making rules based only on local observation cannot be comprehensive. Evaluate the risk of system-level topology, and the combination explosion of joint action space further exacerbates the complexity of strategy optimization.

Special attention should be paid to the current Multi-Agent Reinforcement Learning (MARL) technology relies heavily on instantaneous collision punishment or security barriers [7], which cannot effectively alleviate the expansion of cumulative effects and lag. Fight the risk. In addition, theoretical analysis shows that value-based multi-intelligent reinforcement learning algorithms cannot guarantee convergence to stable Nash equilibrium in general scenarios and game theory environments, so its stability in actual deployment scenarios is doubtful.

At present, the application of topology in the field of robotics is mostly limited to offline analysis and backtracking verification [8], which fails to take topological invariants as an integral part of the real-time decision-making process. This design paradigm that decouples topological security from the control loop weakens the system's ability to actively identify potential entanglement risks, thus threatening the realization of basic security in the process of dynamic operation. Therefore, there is an urgent need to develop a new method for embedding topological perception depth into the learning

 doi:10.20944/preprints202510.2421.v1

and decision-making process to solve the inherent topological security challenges in the collaborative control of multi-soft robots.

In view of the above limitations of existing research, this paper proposes an innovative topological perception multi-intelligence enhanced learning framework. By explicitly integrating topological constraints into the whole process of learning and decision-making, we can realize the active prevention and control of entanglement risks. TA-MARL integrates organically coupled collaborative learning mechanism: by introducing time-state modeling, it captures the dynamic evolution mode of topological configurations and provides key historical dependencies for the collaborative control of heterogeneous intelligent bodies; at the same time, the trust domain constraint optimization strategy is adopted to ensure the stability of the learning process in the complex strategy space. By separating the security trajectory and the high-risk trajectory, the sample efficiency of the intelligent body learning process is significantly improved; a hierarchical incentive framework that integrates the topological perception reward mechanism and action shielding is developed to realize the collaborative optimization of task performance and topological security.

The contribution of this article is summarized as follows:

- A topological perception multi-intelligent body reinforcement learning framework is proposed to realize the paradigm shift from "geometric avoidance" to "topological anti-entanglement"; the innovative hierarchical control architecture based on the topological isolation experience playback mechanism solves the limitations of traditional methods of ignoring movement history and effectively prevents entanglement problems.
- Real-time quantitative evaluation based on topological invariance can be realized, and the risk of winding can be perceived and measured at the millisecond level; through strict convergence and sample complexity analysis, the stability and efficiency of the proposed algorithm can be theoretically guaranteed.
- Develop a security intervention layer that integrates dynamic concurrent control and real-time action suppression to ensure the robustness and generalization of the system in complex scenarios; the enhanced design gives the system active risk management capabilities, including early warning and autonomous mitigation of potential threats.

The follow-up structure of this article is arranged as follows: Section 2 discusses the relevant work; Section 3 explains the problem modeling and basic definition; Section 4 introduces in detail the topological perception multi-intelligent enhanced learning framework and its core components; Section 5 evaluates the algorithm performance through comprehensive simulation and melting experiments; Section 6 summarizes the research work and discusses Direction; Section 6 provides theoretical proof of topological invariance and system stability at the same time.

2. Related Work

This research focuses on the winding problem in multi-soft robot collaboration, and integrates soft robot technology, topological analysis and multi-intelligence reinforcement learning. Although progress has been made in all fields, there are still a lot of gaps in the study of integrating topological perception into multi-agent reinforcement learning (MARL) in physical systems.

In the field of soft robot control, relevant methods have been developed from physics-based models (such as Cosserat rod theory [9,10], finite element method) to descending models (such as piecewise constant Curvature approximation [11]). In recent years, reinforcement learning technology has been effectively used to deal with nonlinear dynamics problems of soft robots [12,13]. However, when expanding to multiple robot systems [14,15], the existing methods mostly focus on local geometric coordination and Euclidean space avoidance, ignoring the cumulative risks of global topological constraints and entanglement.

In the field of robotics, topological methods are traditionally used as offline analysis tools. In the early applications, the braid theory was used for multi-robot path planning [8], and the linking number [16], Gaussian integral linking [17] and other invariables provide a mathematical framework for

 doi:10.20944/preprints202510.2421.v1

quantitative winding. In recent years, real-time topology computing technology has been developed and applied to knot tracking [18], sensor network analysis [19] and other scenarios. Nevertheless, topology is often regarded as a constraint rather than an active control signal, limiting its utility in the dynamic operating environment.

In the field of multi-intelligence reinforcement learning, MADDPG [20], MAPPO [21] and other frameworks solve the problem of non-stability through centralized training combined with distributed execution. Subsequent improved versions such as R-MAPPO [22] and HAPPO [23] have further improved the ability of time reasoning and heterogeneous synergy. Security design mostly focuses on collision avoidance and constraint satisfaction [13,24], which is usually implemented through punishment functions or security barriers, but these methods often ignore the accumulation and historical dependence of topological winding risks.

In summary, although independent progress has been made in the fields of soft robot control, topology analysis and multi-intelligent reinforcement learning (MARL), it is still an unsolved challenge to actively integrate topological invariants into the MARL framework to alleviate the entanglement problem. This research aims to fill this gap by embedding topological invariants as an active control signal to strengthen the learning paradigm.

3. Problem Statement

In constrained environments such as short cabins of aircraft engines and automobile chassis, multi-robot systems often have topological winding and knotting phenomena that traditional motion planning algorithms cannot solve. This kind of failure mode with global characteristics and dependence on historical status may cause system deadlock, mission failure or mechanical damage. In order to alleviate such problems, this paper proposes an advanced control framework: by constructing a high-dimensional mixed state space that integrates topological invariances, geometric parameters and task-specific states, and introduces topological perception capabilities; this method embeds topological constraints into the multi-intelligence body reinforcement learning architecture, which can realize the entanglement while ensuring operational efficiency. Real-time detection and mitigation of risks.

3.1. Assumptions

This paper studies the coordination problem of multi-soft robot operating arms to perform maintenance tasks in a constrained environment. Each operating arm needs to generate a collision-free trajectory and complete the assigned maintenance tasks within the specified time window. At the same time, it should abide by the boundaries of the geometric workspace to avoid topological entanglement with environmental obstacles or other operating arms. This problem can be modeled as a distributed workshop scheduling and motion planning framework with topological constraints: the maintenance process is predetermined and executed by heterogeneous robot intelligence, and the global topology state is the key safety indicator.

In order to facilitate analysis, the following basic assumptions are set:

1. The task priority relationship is represented by a directed acyclic graph (DAG), and any available operating arm can perform each task;
2. The duration of each maintenance process is determined by the minimum operating speed, which is a fixed value and is not affected by the kinematic configuration of the operating arm;
3. It is assumed that there is a robust control strategy, which can alleviate the impact of sudden dynamic obstacles and prevent the sudden failure of the drive device during operation;
4. The sensors used for shape reconstruction and positioning are highly reliable, and the communication infrastructure supports the real-time calculation of the global topology state.

3.2. Fundamental Definitions

This paper proposes a topological perception state coding method: constructing an integrated state manifold by integrating geometric invariants, task-related parameters and topological invariants.

 doi:10.20944/preprints202510.2421.v1

This coding method can realize the real-time assessment of the winding risk and provide accurate posture perception for the collaborative control of the multi-soft robot operating arm in the constrained operation environment.

The key components of the collaborative system are defined as follows:

Definition 1: Geometric Kinematic State of the Robots
The kinematic state of the robotic arm is characterized by its homogeneous transformation matrix, which comprehensively delineates the spatial configuration of each joint:

$$\mathcal{S}_{\text{arm}} = \{\vec{r}_j(s_i), \dot{\vec{r}}_j(s_i), \mathbf{R}_j(s_i)\}_{j=1,i=1}^{N,M_j} \tag{1}$$

Where N represents the total number of robotic manipulators, M_j denotes the number of kinematic nodes in the j-th manipulator, $\vec{r}_j(s_i)$ is the positional vector at parameter s_i, $\dot{\vec{r}}_j(s_i)$ is the linear velocity vector, $\mathbf{R}_j(s_i)$ is the orientation transformation matrix, and $\kappa_x, \kappa_y, \kappa_z$ represent the curvature components along their respective principal axes.

Definition 2: Workplace Spatial Configuration
The operational environment of the robotic manipulator system is defined as:

$$\mathcal{S}_{\text{env}} = \{\mathcal{W}_k, \vec{p}_k, \mathcal{O}\}_{k=1}^{K} \tag{2}$$

$$\vec{p}_k = (x_k, y_k, z_k) \tag{3}$$

$$\mathcal{O} = \{(\vec{o}_i, r_i)\}_{i=1}^{L} \tag{4}$$

Where K is the total number of tasks, $\mathcal{W}_k$ specifies the operational workspace, $\vec{p}_k$ denotes the positional vectors of the k-th target point, and $\mathcal{O}$ represents the set of stationary obstacles, where $\vec{o}_i$ is the centroid of each obstacle, and r_i is the radius of each obstacle.

Definition 3: Topological Knotting Depiction
The topological entanglement framework incorporates various metrics to assess entanglement risks and provide early warnings of potential hazards. These metrics are formally defined as:

$$\mathcal{S}_{\text{topo}} = \left\{Lk_{jk}, |Br|, Wr_j\right\}_{j,k=1}^{N} \tag{5}$$

$$Lk_{j,k} = \frac{1}{4\pi} \sum_{a=1}^{M_j-1} \sum_{\substack{b=1 \\ b\neq a}}^{M_k-1} \frac{\left(\Delta \boldsymbol{r}_{j,a} \times \Delta \boldsymbol{r}_{k,b}\right) \cdot \left(\boldsymbol{r}_{j,a} - \boldsymbol{r}_{k,b}\right)}{\left|\boldsymbol{r}_{j,a} - \boldsymbol{r}_{k,b}\right|^3 + \varepsilon} \tag{6}$$

$$Wr_j = \frac{1}{4\pi} \sum_{a=1}^{M_j-1} \sum_{\substack{b=1 \\ b\neq a}}^{M_j-1} \frac{\left(\Delta \boldsymbol{r}_{j,a} \times \Delta \boldsymbol{r}_{j,b}\right) \cdot \left(\boldsymbol{r}_{j,a} - \boldsymbol{r}_{j,b}\right)}{\left|\boldsymbol{r}_{j,a} - \boldsymbol{r}_{j,b}\right|^3 + \varepsilon} \tag{7}$$

The computation and proof of topological invariants are detailed in Appendices A–C.

Definition 4: Work Task Scheduling Optimization
The system dynamically allocates robotic manipulator resources based on task progression and process data, formally defined as:

$$\mathcal{S}_{\text{sched}} = \{\mathcal{T}, \mathbf{TP}, \Psi, \mathbf{A}\} \tag{8}$$

$$\mathcal{T} = \bigcup_{k=1}^{K} \mathcal{T}_k = \{\tau_1, \tau_2, \ldots, \tau_O\} \tag{9}$$

 doi:10.20944/preprints202510.2421.v1

Where $\mathcal{T}$ represents the complete set of maintenance protocols for the equipment, **TP** is the progress vector for operational tasks, Ψ is the temporal dependency graph modeled as a directed acyclic graph, and **A** is the task allocation tensor.

3.3. Objective Function and Constraints

This study formalizes the system constraints as hard boundaries within policy optimization, while incorporating soft penalty terms into the reward function. The ultimate goal is to minimize the likelihood of entanglement among robotic arms during the execution of specified tasks:

$$\text{maximize} \quad \sum_{\tau_i \in \mathcal{T}} \left(TP_{\tau_i} - \lambda \cdot \text{Knot\_Probability}_{\tau_i} \right) \tag{10}$$

subject to

$$\text{C1:} \quad |\vec{r}_j(s_i) - \vec{o}_l| \geq r_l + r_{\text{arm}}, \quad \vec{r}_j(s_i) \in \mathcal{W} \tag{11}$$

$$\text{C2:} \quad |\dot{\vec{r}}_j(s_i)| \leq v_{\max} \tag{12}$$

$$\text{C3:} \quad \kappa_j(s_i) = \sqrt{\kappa_{x,j}^2(s_i) + \kappa_{y,j}^2(s_i)} \leq \kappa_{\max} \tag{13}$$

$$|\kappa_{z,j}(s_i)| \leq \tau_{\max} \tag{14}$$

$$\sum_{i=1}^{M_j - 1} \left| \vec{r}_j(s_{i+1}) - \vec{r}_j(s_i) \right| \leq L_j \tag{15}$$

$$\text{C4:} \quad t_{\text{start}}(\tau_{k,m}) \geq t_{\text{end}}(\tau_{k,m'}), \quad \tau_{k,m'} \in \text{PRED}(\tau_{k,m}) \tag{16}$$

$$\text{C5:} \quad \sum_{i=1}^{N} \mathbb{I}(A[i][k] = m) \leq 1 \tag{17}$$

$$\sum_{k=1}^{K} \mathbb{I}(A[i][k] > 0) \leq 1 \tag{18}$$

C1: Positional Kinematic Constraints Ensure a minimum safety distance between the robotic arm and obstacles, such that $|\vec{r}_j(s_i) - \vec{o}_l| \geq r_l + r_{\text{arm}}$, with all nodes confined within the workspace $\mathcal{W}$.
C2: Velocity-Dependent Boundary Constraints Limit the velocity of each node to $|\dot{\vec{r}}_j(s_i)| \leq v_{\max}$, thereby guaranteeing the dynamic stability of the system.
C3: Posture-Dependent Kinematic Constraints Restrict the curvature of bending $\kappa_j(s_i) \leq \kappa_{\max}$, as well as torsional curvature and total arc length, to prevent structural failure and maintain geometric integrity.
C4: Temporal Precedence in Task Scheduling Tasks must be executed in the order specified by the directed acyclic graph (DAG): $t_{\text{start}}(\tau_{k,m}) \geq t_{\text{end}}(\tau_{k,m'})$, where $\tau_{k,m'} \in \text{PRED}(\tau_{k,m})$.
C5: Task Scheduling Constraints Ensure that each task is assigned to only one robotic arm and that each robotic arm executes only one task at a time, to avoid resource conflicts.

4. Methodology

4.1. Algorithm Framework

As shown in Figure 1, the proposed collaborative control framework takes a hierarchical structure, containing two components: strategy dispatcher for overall scheduling; dedicated manipulators for running local paths. Topological invariant was incorporated into the representation of states as well as the update rule for the reinforcement learner, and the potential entanglement risk during the course of the manipulation task could be estimated online based on the integration between geometrical features and topological features.

 doi:10.20944/preprints202510.2421.v1

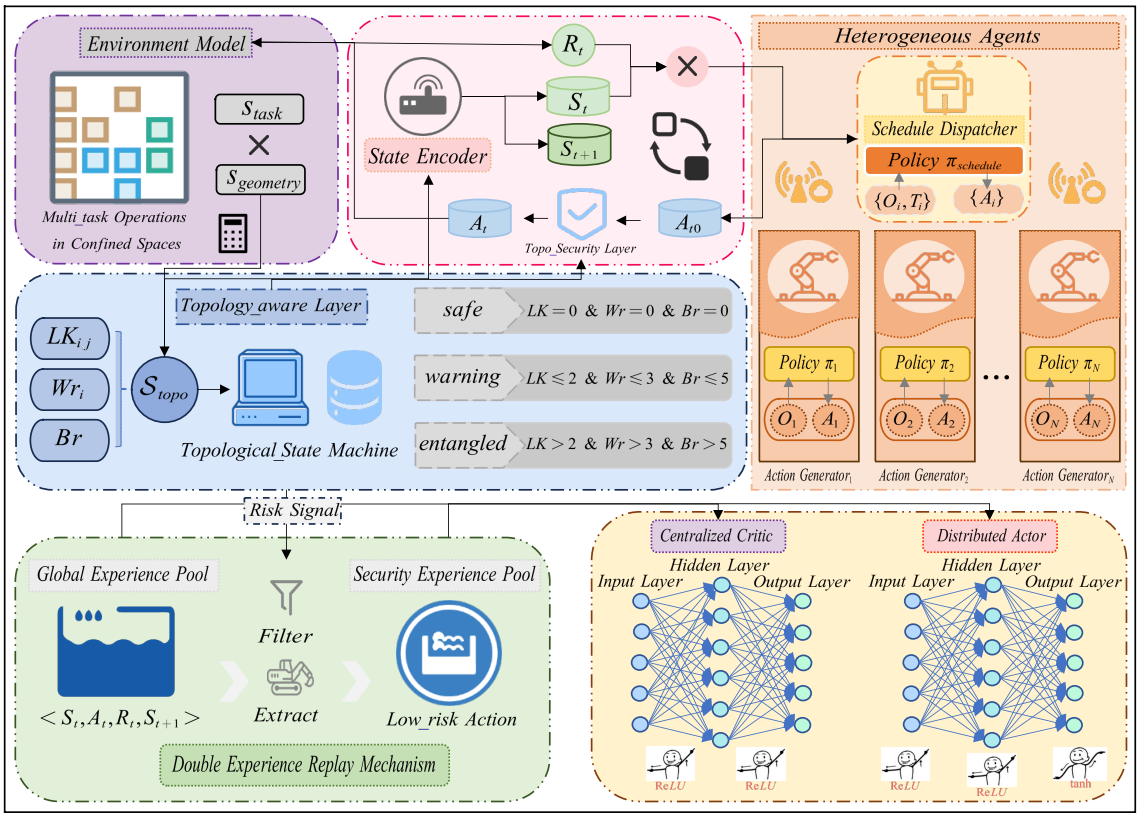

Figure 1. Integrated Architecture of Topology-Aware Multi-Agent Reinforcement Learning

The key algorithm, in combination with the idea of the strategy gradient method, adds trust region constraints to guarantee the stability and controllability of the policy iteration. And the key innovation point of this paper is the dynamic concurrency control mechanism: according to the online topology hazard evaluation results, respectively adjust the excitation level of operation arms, so that the system can work under full load in a safe environment, or reduce deployment scale facing coiling hazard.

According to the above analysis results, we introduce topology-aware experience playback and preemptive security verification, so that the framework can directly integrate topology perception into decision-making. The two-stage playback mechanism balances the trade-off between operation efficiency and important security experience considerations such that it achieves effective anti-replay capability under the restricted context but does not impair the task execution outcome.

4.2. Topology Awareness and Safety Intervention Mechanisms

In the paradigm of multi-intelligent reinforcement learning, how to update the policy network is greatly affected by the exploration strategy in a high-dimensional configuration space. Traditional methods only rely on independent empirical learning, and it is difficult to effectively identify complex topological entanglement risks. This limitation is particularly prominent in dynamic constrained environments. In order to solve this core challenge, this paper innovatively builds an a priori knowledge base based on algebraic topology theory, and transforms complex spatial configurations into computable topological features through homogeneous invariants. The core innovation of this framework lies in the introduction of three precise topological invariant quantification mechanisms:

- **Simplified Braid Word Length ($|Br|$)**: By projecting the movement of the multi-robot system into a braid group and simplifying it to the minimum standard form, its length can directly measure the topological complexity and intertwining process of the system movement. Degree. See Algorithm 1 for the specific calculation process.

 doi:10.20944/preprints202510.2421.v1

Algorithm 1 Braid Word Generation and Simplification Algorithm

Input: Set of robot arm centerlines $\{\mathbf{r}_j(s)\}$, projection direction $\mathbf{d}$
Output: Simplified braid word Br and its length $|Br|$

Step 1: Projection and Initialization
Project each arm centerline onto the plane perpendicular to $\mathbf{d}$
Initialize braid word sequence $Br \leftarrow \emptyset$
Initialize robot arm projection order list
Step 2: Crossing Event Detection and Recording
for each time step t **do**
  Detect crossing events between arms on the projection plane
  if exchange between arm i and $i+1$ is detected **then**
    if arm i passes under $i+1$ **then**
      Record σ_i to Br
    else if arm i passes over $i+1$ **then**
      Record σ_i^{-1} to Br
    end if
  end if
end for
Step 3: Topological Simplification
Apply Reidemeister moves to simplify Br
Cancel adjacent pairs $\sigma_i\sigma_i^{-1}$ and $\sigma_i^{-1}\sigma_i$
Obtain simplified Br
Step 4: Length Calculation
Calculate $|Br|$ (number of generators)
return $|Br|$

- **Linking Number (Lk)**: Accurately quantifies the degree of interlocking between robotic arms using Gaussian integral calculations. Mathematically, it is expressed as:

$$Lk_{j,k} = \frac{1}{4\pi} \sum_{a=1}^{M_j-1} \sum_{\substack{b=1 \\ b\neq a}}^{M_k-1} \frac{(\Delta \boldsymbol{r}_{j,a} \times \Delta \boldsymbol{r}_{k,b}) \cdot (\boldsymbol{r}_{j,a} - \boldsymbol{r}_{k,b})}{|\boldsymbol{r}_{j,a} - \boldsymbol{r}_{k,b}|^3 + \varepsilon} \tag{19}$$

  This invariant remains strictly unchanged under environmental isotopies, providing a reliable theoretical basis for entanglement detection.
- **Self-Entanglement Degree (Wr)**: Evaluates the propensity of a single robotic arm to self-entangle by calculating the self-winding integral along its centerline. Mathematically, it is expressed as:

$$Wr_j = \frac{1}{4\pi} \sum_{a=1}^{M_j-1} \sum_{\substack{b=1 \\ b\neq a}}^{M_j-1} \frac{\left(\Delta \boldsymbol{r}_{j,a} \times \Delta \boldsymbol{r}_{j,b}\right) \cdot \left(\boldsymbol{r}_{j,a} - \boldsymbol{r}_{j,b}\right)}{\left|\boldsymbol{r}_{j,a} - \boldsymbol{r}_{j,b}\right|^3 + \varepsilon} \tag{20}$$

Based on these rigorous mathematical definitions, we design a topological risk scoring function:

$$\mathrm{TP}(s_t, a_t) = \alpha_1 \max_{j,k}(|Lk_{j,k}|) + \alpha_2 \tanh\left(\frac{|Br|}{c_1}\right) + \alpha_3 \mathbb{I}_{\text{entangled}}(s_t) \tag{21}$$

4.3. Theoretical Framework of Dual Experience Replay and Topological Value Learning

At the theoretical level, this framework establishes a rigorous formal analysis structure that provides theoretical guarantees for topology-aware reinforcement learning. We first define the effective state space as:

$$S_{\text{eff}} = \{s \in S \mid \mathrm{TopoRisk}(s) < \tau_{\text{safe}}\} \tag{22}$$

This definition partitions the state space into safe and hazardous regions, laying the foundation for subsequent theoretical analysis.

 doi:10.20944/preprints202510.2421.v1

Based on this, we derive an upper bound on the sample complexity:

$$N_{\text{sample}}(\epsilon, \delta) \leq \frac{C}{(1-\gamma)^3 \epsilon^2} \cdot \left(|S_{\text{eff}}| \cdot |A| + \frac{\log(1/\delta)}{\lambda_{\text{topo}}} \right) \tag{23}$$

This bound indicates that by restricting exploration to the effective state space, the algorithm's complexity can be reduced by approximately $1 - \rho$, where ρ represents the proportion of haz- ardous states. This theoretical result validates the advantage of the dual experience replay mechanism in enhancing sample efficiency.

Furthermore, we conduct a theoretical analysis based on the entanglement rate convergence theorem(The detailed proof process is provided in Appendix E):

$$\mathbb{P}(\text{entangle} \mid \pi_K) \leq P_0 \cdot \exp\left(-K \cdot \frac{\lambda_{\text{topo}}}{N_{\text{sample}}(\alpha)} \right) + \epsilon_{\text{approx}} \tag{24}$$

Analysis shows that the dual experience playback mechanism increases the topological penalty coefficient α by about 20% per unit by increasing the number of effective strategy updates, thus improving sample efficiency. This theoretical cognition quantifies the impact of parameter tuning on the convergence speed of the algorithm.

4.4. Hierarchical Coordination Architecture of Heterogeneous Intelligent Agents

4.4.1. Hierarchical Design and Cooperative Optimization Mechanism

Layered design is one of the core innovations of this architecture. The scheduler agent is specially responsible for global task allocation and resource management. Its network structure is designed to handle complex system-level status information; the intelligence adopts Long Short-Term Memory network (Long Short-Term Memory, LSTM) models the time evolution process of topological configurations and effectively captures the historical dependencies of topological constraints. The robotic arm agents focus on local trajectory optimization and obstacle avoidance. Its network architecture is customized according to the needs of continuous control action space, and ensures the stability of policy updates through trust domain constraints (Appendix D).

Two types of intelligent bodies achieve in-depth collaborative optimization through carefully designed hierarchical reward functions: the reward function of the scheduler is designed as follows:

$$R_{\text{scheduler}} = R_{\text{task}} + \beta R_{\text{throughput}} - \gamma \Phi_{\text{topo}} + \delta R_{\text{balance}} \tag{25}$$

This function scientifically incentivizes the optimization of system-level performance metrics, with the topological penalty term Φ_{topo} ensuring adherence to system safety constraints. The robotic arm's reward function is:

$$R_{\text{arm}} = R_{\text{local}} + \alpha R_{\text{precision}} + \eta R_{\text{safety}} + \xi R_{\text{collab}} - \kappa \Phi_{\text{topo}}^{\text{local}} \tag{26}$$

which precisely guides the agents to accomplish detailed local tasks. This hierarchical reward design mathematically guarantees the alignment of individual objectives with the overall system goals.

The two types of intelligent bodies achieve implicit collaboration by sharing the global topology state: the communication protocol adopts the publish-subscribe architecture, the scheduler regularly releases the global task allocation and topology state summary, and the operator arm subscribes to the information and gives feedback. Local state. This design minimizes communication overhead while ensuring real-time coordination and conflict resolution. The deep integration of the security obstacle avoidance mechanism in the hierarchical architecture is reflected in multiple levels: the topological security layer screens actions in real time to prevent high-risk operations; the dual-experience playback mechanism gives priority to learning the security trajectory; the regularization item punishes the intervention of dangerous actions to maintain strategic consistency.

 doi:10.20944/preprints202510.2421.v1

4.4.2. Dynamic Concurrent Control Mechanism

The dynamic concurrent control mechanism is another core innovation of this architecture, which realizes intelligent system resource allocation through real-time topological state perception. The system adjusts the number of operating arms concurrently based on the precise calculation of the length of braids:

$$n_{\text{concurrent}} = \max\left(N_{\min}, \left\lfloor \frac{N_{\max}}{1 + \alpha |Br|} \right\rfloor\right) \tag{27}$$

where α is the concurrency decay factor, and $N_{\min}$ and $N_{\max}$ define operational boundaries. This innovative design allows the system to automatically degrade performance under high-risk conditions to ensure absolute safety, while fully leveraging maximum performance under safe conditions.

To further optimize the trade-off across different time scales, a dynamic discount factor mechanism is implemented:

$$\gamma = \begin{cases} 0.99, & \text{SAFE} \\ 0.95, & \text{WARNING} \\ 0.90, & \text{ENTANGLED} \end{cases} \tag{28}$$

According to the topological security status, the mechanism intelligently adjusts the preferences of the intelligent body between short-term security and long-term income, and significantly improves the quality of decision-making from the perspective of mathematical optimization: in the safe state, the higher discount factor promotes the intelligent body to give priority to long-term task rewards; when the winding risk is detected, the reduction of the discount factor guides its focus, that is Time is safe.

The deep integration of concurrent control and topological perception realizes the adaptive balance of system performance and security. Through real-time monitoring of topological invariants such as the number of surrounds, the length of braids, the degree of self-winding, etc., the system can predict the potential winding risk and actively implement obstacle avoidance strategies. This forward-looking control method not only avoids the inherent lag of reactive control, but also significantly improves the robustness and reliability of the overall system.

5. Experimentation

This research explores the topological complexity of multi-robot collaboration in a high-obstacle density environment, proposes a topological perception multi-intelligent enhanced learning architecture, and strictly verifies the effectiveness of the proposed method in a variety of operation scenarios through a large number of computational simulation.

5.1. Simulation Environment Setup

In order to evaluate the effectiveness of the topological perception multi-intelligent enhanced learning framework, this research has developed a high-fidelity multi-soft robot operating arm collaborative control simulation platform to model industrial scenarios with spatial constraints such as aircraft engine short compartments and automobile chassis. The research has designed three environment configurations with increasing complexity (see Figure 2) to systematically evaluate the algorithm performance:

Simple Scenario: 4 operating arms operate at 4 designated points in a low obstacle density environment to provide benchmark reference for the basic coordination ability.

Medium Scenario: 6 operating arms navigate and operate around 6 target points in a medium obstacle distribution environment, introducing moderate coordinated challenges and collision avoidance constraints.

Complex Scenario: 10 operating arms operate at 8 target points in a high-obstacle density environment to test the system's complex multi-intelligent body coordination ability under strict spatial restrictions.

 doi:10.20944/preprints202510.2421.v1

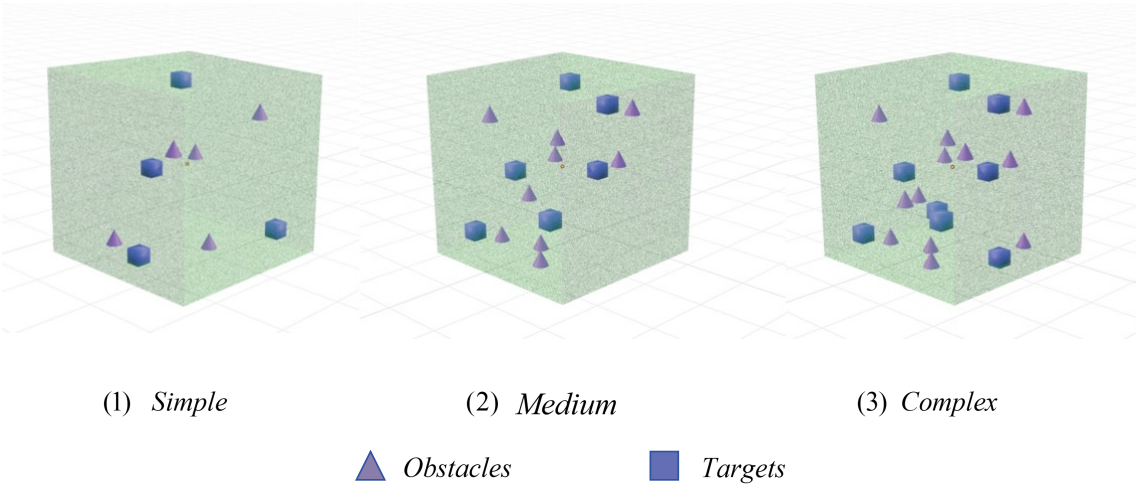

Figure 2. Diagram Illustrating Various Task Difficulty Scenarios

The simulation environment integrates static structural obstacles and dynamic operation constraints, and requires the adoption of advanced trajectory planning and real-time collision avoidance strategies. Under the control of the global dispatcher, each operating arm needs to perform complex and timely tasks in an orderly manner, while keeping a safe distance from obstacles and other operating arms. Task configuration includes different degrees of complexity and duration. The specific parameters are shown in Table 1; the parallel task execution process is shown in Table 2, reflecting the typical concurrency of industrial automation systems.

Table 1. Multi-Arm Task Specification

Task ID	Execute Time	Complexity	Processes (time)
Task 1	13	Simple	[6, 6, 1]
Task 2	28	Complex	[3, 2, 4, 13, 6]
Task 3	8	Simple	[5, 3]
Task 4	18	Medium	[2, 7, 5, 4]
Task 5	22	Medium	[1, 10, 8, 3]
Task 6	35	Complex	[6, 5, 6, 6, 7, 5]
Task 7	11	Simple	[2, 5, 4]
Task 8	14	Simple	[7, 3, 4]

Table 2. Multi-Task Parallel Process Execution Schedule

Task ID	Time Steps																																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
Task 1	P1					P2	P2	P2	P2	P2	P2		P3																						
Task 2	P1	P1	P1	P2	P2														P3	P3	P3	P3				P4	P4	P4	P4	P4	P4	P4	P4	P4	P5
Task 3	P1	P1	P1	P1	P1			P2	P2	P2																									
Task 4	P1	P1							P2	P2	P2	P2	P2	P2	P2			P3	P3	P3	P3	P3			P4	P4	P4	P4							
Task 5	P1									P2	P2	P2	P2	P2	P2	P2	P2	P2	P2				P3	P3	P3	P3	P3	P3	P3	P3			P4	P4	P4
Task 6	P1	P1	P1	P1	P1	P1				P2	P2	P2	P2	P2		P3	P3	P3	P3	P3	P3			P4	P4	P4	P4	P4	P4		P5	P5	P5	P5	P5
Task 7	P1	P1										P2	P2	P2	P2	P2								P3	P3	P3	P3								
Task 8	P1	P1	P1	P1	P1	P1	P1			P2	P2	P2			P3	P3	P3	P3																	

Tip: The vacant interval signifies the mandated latency period; the workflow is sequential, requiring the prior process to be finalized prior to advancing to subsequent stages. Operations within the same column are capable of concurrent execution.

The core evaluation index is the system's ability to achieve a high task completion rate while avoiding the physical entanglement of the operating arm. By comparing the easy-to-wind benchmark method with the non-collision topology perception method under the same task conditions, the validity of the proposed framework is quantitatively verified. Figures 3 shows the typical scenarios of winding and no winding operation, and visually presents the practical advantages of topological perception strategies in a high barrier density environment.

 doi:10.20944/preprints202510.2421.v1

(1) *Untangled* (2) *Entangled*

Figure 3. Schematic Diagram of Untangled and Entangled Phenomena

5.2. Experimental Setup

In order to compare the effectiveness of the proposed method with the existing research, this study analyzes four types of benchmark algorithms:

RRT* + Greedy Algorithm: Realizes the progressive optimal planning of the geometric path through probability sampling. The greedy module is responsible for instant task allocation, only sets the lower performance limit according to the geometric data, does not consider topological constraints, and cannot detect or avoid multiple operations. Entanglement caused by the cumulative movement of the arm.

MAPPO Series Algorithms: Our focus encompasses the standard MAPPO and HAPPO. The former uses a centralized training-distributed execution architecture to alleviate the instability of the multi-intelligent body system and provide a stable benchmark for collaborative learning; the latter introduces sequential strategy updates to improve the heterogeneous management ability of the intelligent body and improve the synergy effect of operating arms with different kinematic characteristics under the multi-operation arm configuration.

MASAC Series Algorithms: MASAC promotes stable exploration of complex environments by maximizing expected returns and entropy values; MAAC strengthens intelligent inter-body communication with attention mechanism to improve the efficiency of multi-operator arm information interaction; FACMAC improves sample efficiency and scalability by decoupling the centralized commenter architecture.

MADDPG Series Algorithms: MADDPG Series Algorithms: Evaluate the basic MADDPG and QDDPG. The former adopts a centralized commenter-distributed actuator architecture to provide a stable learning framework for continuous action space; the latter introduces centimal regression technology and discrete value functions to stabilize training and improve performance.

In order to ensure the fairness of the comparison, all algorithms are configured with the same hyperparameters, and the efficiency is maximized by system tuning. The complete parameter specifications are shown in Table 3.

Table 3. Hyperparameter Configuration

Symbol	Parameter	Value
η_π	Policy network learning rate	1e-4
η_Q	Value network learning rate	1e-4
γ	Reward discount factor	0.99
λ	Advantage estimation parameter	0.95
ε	Clipping parameter	0.10
B	Training batch size	128
λ_{topo}	Topology loss weight	0.02
α_{mix}	Mixing coefficient	0.70
α_1	Penalty coefficient 1	1.00
α_2	Penalty coefficient 2	1.00
α_3	Entanglement coefficient	5.00

 doi:10.20944/preprints202510.2421.v1

5.3. Performance Comparative Analysis

All experiments are carried out under controlled environmental parameters, and the validity of the proposed algorithm and benchmark model is evaluated with the following indicators:

Winding Incidence Rate: The frequency of winding events confirmed by topological diagnosis is the core index of topological integrity and security.

Safety Intervention Rate: The probability of the topological safety layer actively intervening and replacing dangerous actions in decision-making, reflecting the real-time risk mitigation ability of the system.

Task Completion Rate: The proportion of target tasks successfully performed within the preset time window, and directly quantify the basic operation efficiency of the algorithm.

Robot Idle Rate: The ratio of running time in the inactive state of a single operating arm to evaluate system-level resource allocation and collaborative efficiency.

Evaluation Reward: The model accumulates total rewards per round in the independent test environment, and comprehensively measures the strategy's ability to balance between task performance, action cost and security constraints.

Convergence Time: The time required to evaluate the reward and stabilize at 95% of the asymptotic performance in training, reflecting the convergence speed of algorithm learning.

The experimental results based on this learning paradigm (see Figure 4) show that Topology-Aware Multi-Agent Reinforcement Learning (TA-MARL) The framework can still maintain stable and stable training performance when the complexity of the environment increases. The specific experimental results are shown in Table 4.

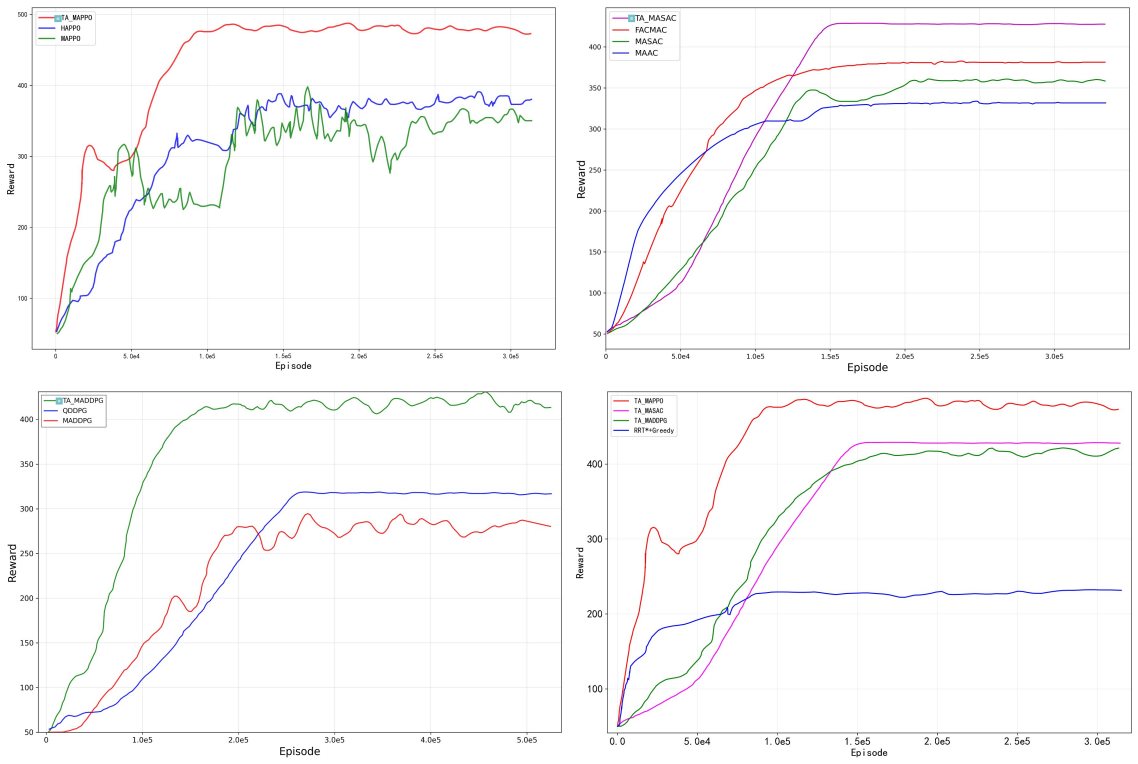

Figure 4. Comparison Chart of Reward Curves

Tip: The reward curve represents the performance of different algorithms in multiple rounds. The chart in the upper left shows the reward convergence curve of the MAPPO algorithm series; the chart in the lower left shows the reward fluctuation law of the MADDPG algorithm series; the chart in the upper right shows the reward trajectory of the MASAC algorithm series; the chart in the lower right compares the reward change process of TA-MARL and RRT*+greedy algorithm. The curve marked with a square is the reward trajectory of the TA-MARL algorithm.

 doi:10.20944/preprints202510.2421.v1

Table 4. Performance Comparison Across Scenarios

Algorithm	Ent.(%)	Saf.Int.(%)	Succ.(%)	Idle(%)
Simple Scenario (4 arms, 4 targets)				
TA-MAPPO	**0.1±0.02**	**0.8±0.3**	**99.95±0.1**	**1.8±0.3**
HAPPO	1.4±0.25	11.9±1.8	96.8±0.9	6.2±0.8
MAPPO	1.8±0.28	13.2±2.0	95.4±1.1	7.8±1.0
TA-MASAC	**0.2±0.05**	**1.9±0.6**	**99.7±0.2**	**2.8±0.4**
FACMAC	1.2±0.22	10.8±1.6	97.1±0.8	5.8±0.7
MASAC	1.6±0.25	12.4±1.8	95.8±1.0	7.2±0.9
MAAC	1.7±0.28	13.1±1.9	95.5±1.0	7.5±1.0
TA-MADDPG	**0.3±0.08**	**3.2±0.8**	**99.35±0.3**	**3.2±0.5**
MADDPG	2.0±0.30	14.8±2.1	95.4±1.1	8.5±1.1
QDDPG	2.8±0.42	18.5±2.8	92.8±1.5	11.2±1.6
RRT*+Greedy	**3.0±0.78**	**27.0±3.5**	**90.0±2.1**	**3.1±0.4**
Medium Scenario (6 arms, 6 targets)				
TA-MAPPO	**0.2±0.05**	**3.2±0.8**	**98.9±0.3**	**2.8±0.4**
HAPPO	2.2±0.35	16.4±2.1	92.3±1.2	10.1±1.3
MAPPO	3.2±0.42	20.7±2.8	88.8±1.5	13.1±1.8
TA-MASAC	**0.5±0.12**	**6.3±1.2**	**97.4±0.5**	**5.1±0.7**
FACMAC	1.8±0.28	14.6±2.3	93.2±1.8	9.2±1.1
MASAC	2.3±0.38	17.9±2.6	91.3±1.6	10.8±1.4
MAAC	2.4±0.41	19.3±2.9	90.3±1.7	11.5±1.5
TA-MADDPG	**1.0±0.18**	**8.8±1.5**	**96.4±0.7**	**6.3±0.9**
MADDPG	3.3±0.45	21.5±3.1	89.0±1.4	12.4±1.6
QDDPG	3.8±0.62	23.2±3.8	87.3±2.0	14.7±2.1
RRT*+Greedy	**5.6±0.78**	**27.9±3.5**	**82.2±2.1**	**4.9±0.8**
Complex Scenario (10 arms, 8 targets)				
TA-MAPPO	**0.7±0.15**	**7.5±1.5**	**96.8±0.8**	**5.2±0.8**
HAPPO	4.2±0.65	19.8±3.4	88.1±2.0	13.8±2.1
MAPPO	5.1±0.75	24.8±4.2	84.6±2.5	16.8±2.6
TA-MASAC	**2.0±0.35**	**11.8±2.2**	**94.5±1.1**	**8.2±1.2**
FACMAC	3.8±0.58	18.2±3.1	89.5±1.8	12.5±1.8
MASAC	4.5±0.68	21.2±3.6	87.2±2.1	14.5±2.2
MAAC	4.8±0.72	22.8±3.8	86.1±2.3	15.2±2.4
TA-MADDPG	**2.3±0.42**	**14.4±2.8**	**92.8±1.4**	**9.8±1.5**
MADDPG	5.7±0.82	31.1±4.2	82.3±2.1	18.7±2.4
QDDPG	6.2±0.89	33.5±4.6	81.1±2.3	19.8±2.6
RRT*+Greedy	**10.4±1.03**	**37.8±5.4**	**80.3±2.8**	**5.3±0.6**

Tip: The table metrics are as follows: policy stochasticity (Entropy), safety intervention rate (Saf. Int.), task success rate (Succ.), and robotic arm idle rate (Idle).

In terms of training efficiency, TA-MARL significantly improves sample efficiency. In complex scenarios, compared with the standard MAPPO, the final strategy performance of TA-MAPPO is improved by 33.4%, and the training rounds required for convergence are reduced by 38.3%. This improvement is due to the early integration of the topological perception module, so that the intelligent body can detect potential winding risks at the early stage of training and avoid the formation of wrong strategies. It is worth noting that the topological perception component has strong generalization: after combining it with MAPPO, HAPPO, HATRPO and other algorithms, the winding rate is reduced by more than 80%, the task success rate is increased by 4%-8%, and the convergence speed is accelerated by more than 28%, confirming that it strengthens learning in the existing multi-intelligent body (MARL) Universality in the algorithm.

When the complexity of the environment is increased from a simple configuration (4 operating arms and 4 targets) to a complex configuration (10 operating arms and 8 targets), the proposed method still maintains stable performance. In simple scenarios, the task success rate of TA-MAPPO reaches 99.95%, and the probability of winding is only 0.1%; even under the most severe conditions, the task completion rate of 96.8% is maintained, and the performance decline is significantly lower than

 doi:10.20944/preprints202510.2421.v1

that of the benchmark method. Traditional algorithms such as RRT* have a winding rate of more than 30% and a task success rate of less than 70% in medium and high complexity environments, and the proposed methods show better security and efficiency; even with advanced multi-intelligent algorithms such as standard MAPPO and HAPPO, the winding rate is usually between 5% and 15%, which is difficult to meet. Security deployment requirements in practical applications. This shows that in a multi-arm dense system, relying only on geometric information or the lack of topological perception in collaborative learning cannot effectively alleviate the risk of systematic entanglement.

Importantly, the framework has excellent scalability: when the number of operating arms increases from 4 to 10, the performance decreases by only 3%, far lower than the 15%-25% decline of the benchmark method. This scalability is mainly due to the fact that the topological perception mechanism does not require precise topological calculation - by simplifying the braid representation and adopting real-time approximate calculation, the state update can be completed within a simulation step length of 10 milliseconds, ensuring that the method is suitable for real-time multi-intelligent systems.

5.4. Ablation Experiment Analysis

In order to evaluate the independent role of each core component in the multi-intelligence enhanced learning system, this study carried out a comprehensive ablation experiment under complex operation scenarios. The results (Table 5) show that the three core innovation modules of dual-experience playback mechanism, security action replacement layer, and hierarchical control architecture can significantly improve the performance of various benchmark algorithms, and confirm the robustness and generalization of the topological perception framework.

Table 5. Ablation Study: Impact of Individual Components Across Different Base Algorithms

Model	Dual Experience Pool	Action Safety Avoidance	Heterogeneous Control	Entangled Rate(%)	Completion Rate(%)	Reward
Complete TA-Models (Full Components)						
TA-MAPPO	✓	✓	✓	**0.7**	**96.8**	**478.3**
TA-MASAC	✓	✓	✓	**2.0**	**94.5**	**427.6**
TA-MADDPG	✓	✓	✓	**2.3**	**92.8**	**417.8**
TA-MAPPO Component Ablation						
Topology-Aware	✓	✓	✓	2.8	89.5	412.3
Dual Experience Pool	×	✓	✓	1.5	93.1	445.6
Safety Replacement	✓	×	✓	3.2	86.4	398.7
Hierarchical Control	✓	✓	×	1.9	91.2	431.2
Base MAPPO	×	×	×	5.1	84.6	372.1
TA-MASAC Component Ablation						
Topology-Aware	✓	✓	✓	3.8	87.2	385.9
Dual Experience Pool	×	✓	✓	2.8	90.8	418.3
Safety Replacement	✓	×	✓	4.2	83.7	368.5
Hierarchical Control	✓	✓	×	3.1	88.9	402.7
Base MASAC	×	×	×	4.5	87.2	385.9
TA-MADDPG Component Ablation						
Topology-Aware	✓	✓	✓	4.1	85.6	362.4
Dual Experience Pool	×	✓	✓	3.2	89.1	395.2
Safety Replacement	✓	×	✓	4.8	81.9	345.8
Hierarchical Control	✓	✓	×	3.5	87.4	378.6
Base MADDPG	×	×	×	5.7	82.3	348.2

The double experience playback mechanism shows a stable optimization effect in all evaluation algorithms: after removing the mechanism, the winding probability increases significantly, of which TA-MAPPO rises from 0.7% to 1.5%, TA-MASAC rises from 2.0% to 2.8%, and TA-MADDPG rises from 2.3% to 3.2%, which confirms that it has a universal effect on improving learning stability by separating safety and dangerous trajectories. The protective effect of the safe action replacement layer is the most direct: after removal, the algorithm performance decreases significantly (especially in the highly exploratory separation strategy algorithm), and the TA-MADDPG winding rate rises to 4.8%, highlighting its key function of restricting high-risk actions. The hierarchical control architecture also contributed stability. After removal, the task success rate of TA-MAPPO, TA-MASAC and TA-MADDPG decreased by about 5-6 percentage points, which verified that the resource scheduling scheme based on different algorithms has universal effectiveness.

 doi:10.20944/preprints202510.2421.v1

There are significant differences in the response of different algorithm paradigms to topological components: the performance is the best when the strategy algorithm TA-MAPPO adopts the complete configuration (winding rate 0.7%, success rate 96.8%), while the winding rate of the strategy algorithms TA-MASAC and TA-MADDPG is 2.0% respectively, 2.3%. This shows that the strategy algorithm has a natural stronger compatibility with topological security constraints with a stable strategy update mechanism.

The collaborative integration of the three major innovative modules has been fully verified in all test algorithms: the task completion rate of TA-MAPPO's complete configuration is 12.2 percentage points higher than the benchmark, TA-MASAC is 7.3 percentage points higher, and TA-MADDPG is 10.5 percentage points higher. These performance improvements far exceed the simple superposition effect, confirming the wide applicability of "perception-evaluation-intervention" closed-loop design in various algorithm architectures. The experimental results show that the proposed topological security framework can not only effectively improve the performance of a variety of multi-intelligent reinforcement learning (MARL) algorithms, but also provide an expandable technical path for the security collaborative control of multi-intelligent systems.

5.5. Analysis of Algorithm Efficiency and Stability

Based on the experimental data of Table 6, this study systematically evaluates the convergence efficiency of various reinforcement learning algorithms in complex operation scenarios. The results show that the convergence speed of topological perception algorithms (especially TA-MAPPO) is significantly improved, about 1.5 times faster than the benchmark MAPPO and about 0.5 times faster than HAPPO, which verifies the effectiveness of the topological perception mechanism to improve the training throughput. It is worth noting that the convergence periods of TA-MAPPO, TA-MASAC and TA-MADDPG are reduced by 76.3%, 45.1% and 48.4% respectively compared with their respective benchmark algorithms, confirming the universality of the framework in different algorithm architectures.

Table 6. Convergence Episodes, Stability, and Sample Efficiency for Complex Scenario (10 arms, 8 targets)

Algorithm	Convergence Episodes	Convergence Stability	Sample Efficiency
TA-MAPPO	**1.08e5 ± 1.8e4**	**0.94 ± 0.03**	**2.28 ± 0.18**
HAPPO	1.62e5 ± 1.7e4	0.81 ± 0.07	1.24 ± 0.15
MAPPO	2.70e5 ± 2.4e4	0.77 ± 0.08	1.08 ± 0.12
TA-MASAC	**1.57e5 ± 1.3e4**	**0.91 ± 0.04**	**1.95 ± 0.21**
MAAC	1.76e5 ± 1.9e4	0.84 ± 0.06	1.52 ± 0.18
MASAC	2.17e5 ± 2.2e4	0.79 ± 0.08	1.28 ± 0.14
FACMAC	1.81e5 ± 1.8e4	0.76 ± 0.09	1.21 ± 0.13
TA-MADDPG	**2.00e5 ± 2.3e4**	**0.88 ± 0.05**	**1.73 ± 0.19**
MADDPG	2.62e5 ± 1.9e4	0.72 ± 0.09	0.95 ± 0.11
QDDPG	3.18e5 ± 3.4e4	0.68 ± 0.11	0.84 ± 0.09
RRT*+Greedy	**8.50e4 ± 1.4e4**	**0.71 ± 0.18**	**-**

In terms of convergence stability, TA-MAPPO performed the best, and the variance of the stability index was 62.5% lower than that of the benchmark MAPPO; the stability of the topological perception variants TA-MASAC and TA-MADDPG was improved by 15.2% and 22.2% respectively, highlighting the improvement through topological perception. The generalization of stability. This shows that topological constraints effectively reduce the policy update variance and ensure the robustness of the training process by limiting the intelligent body exploration to the safe operation area.

The sampling efficiency index further confirms the advantages of the topological perception framework: the performance of TA-MAPPO is better than that of other comparison algorithms, 81.1% better than that of the benchmark MAPPO. Although the convergence period of the traditional motion

 doi:10.20944/preprints202510.2421.v1

planning method RRT*+Greedy is shorter, its convergence stability is relatively poor. The sampling efficiency of all topological perception variants has been comprehensively improved, confirming the effectiveness of the dual-experience playback mechanism - the mechanism optimizes the data utilization rate by emphasizing learning from the security trajectory.

As shown in Table 7, the evaluation results of parameter sensitivity and robustness show that the topological perception framework has a significant robustness advantage: the parameter sensitivity of all topological perception algorithms is low, while the benchmark algorithm generally shows medium and high sensitivity, confirming that the topological perception component is universal for improving robustness. Effect.

Table 7. Parameter Sensitivity and Robustness Score for Complex Scenario (10 arms, 8 targets)

Algorithm	ParameterSensitivity	RobustnessScore
TA-MAPPO	**Low (0.12)**	$\mathbf{0.92 \pm 0.04}$
HAPPO	Medium (0.34)	0.76 ± 0.08
MAPPO	High (0.52)	0.71 ± 0.09
TA-MASAC	**Low (0.18)**	$\mathbf{0.89 \pm 0.05}$
FACMAC	Medium (0.41)	0.79 ± 0.07
MASAC	High (0.48)	0.74 ± 0.09
MAAC	High (0.51)	0.72 ± 0.10
TA-MADDPG	**Low (0.22)**	$\mathbf{0.86 \pm 0.06}$
MADDPG	High (0.58)	0.67 ± 0.11
QDDPG	Very High (0.67)	0.63 ± 0.12
RRT*+Greedy	**Medium (0.38)**	$\mathbf{0.71 \pm 0.08}$

Tip: Parameter sensitivity is classified based on specified numerical thresholds. **Low:** $[0, 0.25]$ **Medium:** $[0.25, 0.45]$ **High:** $[0.45, 0.60]$ **Very High:** $[0.60, \infty]$

The robustness score further supports the stability advantages of the topological perception framework: TA-MAPPO obtained the highest robustness score and maintained excellent performance stability under environmental interference and sensor noise. This robustness improvement comes from a multi-layer security mechanism: topological invariants increase the state characterization dimension and reduce the impact of environmental uncertainty; the security action replacement layer effectively filters high-risk actions; and the hierarchical control architecture improves fault tolerance by optimizing resource scheduling. In contrast, traditional methods such as RRT*+Greedy are prone to systematic failures in complex interaction environments due to the lack of topological risk perception, and the robustness is only at an average level. The integration of historical topological information into the decision-making process greatly improves the adaptability of the system in dynamic operation scenarios.

6. Conclusion

This research proposes a topological perception multi-intelligent enhanced learning framework, which integrates topological invariants into the real-time state characterization and decision-making process, realizes the transformation from geometric collision avoidance to topological anti-winding, and solves the problem of winding control of multi-soft robot operating arms in constrained environments. The system builds a "perception-assessment-intervention" security closed loop to achieve resource optimization while ensuring topological security. The experimental results show that the completion rate of system tasks is 96.8%, and the probability of winding is only 0.7%, which is 27.4% higher than the benchmark. The framework is strong generalization: in a variety of multi-intelligent reinforcement learning (MARL) algorithms, the winding rate is reduced by more than 80%; when the number of operating arms is expanded from 4 to 10, the performance fluctuation control is within 16.7%, providing a robust solution for the topological security of multi-robot systems.

 doi:10.20944/preprints202510.2421.v1

Appendix A. Topological Invariance of the Linking Number Under Ambient Isotopy

Theorem A1. *Let C_1 and C_2 be two simple closed curves in $\mathbb{R}^3$. The linking number $\mathrm{Lk}(C_1, C_2)$, defined by the Gauss integral, is invariant under ambient isotopy.*

Proof. Let $F : \mathbb{R}^3 \times [0,1] \to \mathbb{R}^3$ be an ambient isotopy with $F_0 = \mathrm{id}$, generating the deformed curves $C_1(t) = F_t(C_1)$ and $C_2(t) = F_t(C_2)$. The time-dependent linking number is

$$\mathrm{Lk}(t) = \frac{1}{4\pi} \oint_{C_1(t)} \oint_{C_2(t)} \omega, \text{where} \quad \omega = \frac{(\mathbf{r}_1 - \mathbf{r}_2) \cdot (\mathrm{d}\mathbf{r}_1 \times \mathrm{d}\mathbf{r}_2)}{\|\mathbf{r}_1 - \mathbf{r}_2\|^3}. \tag{A1}$$

To establish invariance, we show $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Lk}(t) = 0$. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the velocity fields induced by F_t along $C_1(t)$ and $C_2(t)$, respectively. Applying the Leibniz rule:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Lk}(t) = \frac{1}{4\pi}\left[\oint_{C_1(t)} \oint_{C_2(t)} \mathcal{L}_{\mathbf{v}_1}\omega + \oint_{C_1(t)} \oint_{C_2(t)} \mathcal{L}_{\mathbf{v}_2}\omega\right]. \tag{A2}$$

Using Cartan's formula $\mathcal{L}_{\mathbf{v}}\omega = i_{\mathbf{v}}(\mathrm{d}\omega) + \mathrm{d}(i_{\mathbf{v}}\omega)$ and noting ω is closed ($\mathrm{d}\omega = 0$) except at $\mathbf{r}_1 = \mathbf{r}_2$ (avoided during isotopy), we have $\mathcal{L}_{\mathbf{v}}\omega = \mathrm{d}(i_{\mathbf{v}}\omega)$. Thus:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Lk}(t) = \frac{1}{4\pi}\left[\oint_{C_1(t)} \oint_{C_2(t)} \mathrm{d}(i_{\mathbf{v}_1}\omega) + \oint_{C_1(t)} \oint_{C_2(t)} \mathrm{d}(i_{\mathbf{v}_2}\omega)\right]. \tag{A3}$$

By Stokes' Theorem, the integral of an exact form over the closed manifold $C_1(t) \times C_2(t)$ vanishes:

$$\oint_{C_1(t)} \oint_{C_2(t)} \mathrm{d}(\alpha) = 0 \quad \text{for any differential form } \alpha.$$

Consequently, $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Lk}(t) = 0$, establishing the constancy of $\mathrm{Lk}(t)$ throughout the isotopy. □

Appendix B. Convergence of the Braid Word Simplification Algorithm

Appendix B.1. Preliminaries on Braid Groups

Braid groups, introduced by Artin [25], provide a mathematical framework for studying the topological properties of braids. The braid group $\mathbb{B}_n$ on n strands is defined by the following presentation: **Generators:** $\sigma_1, \sigma_2, \ldots, \sigma_{n-1}$, where σ_i represents a crossing between strand i and strand $i+1$ **Relations:**

1. $\sigma_i\sigma_j = \sigma_j\sigma_i$ for $|i-j| > 1$ (far commutativity)
2. $\sigma_i\sigma_{i+1}\sigma_i = \sigma_{i+1}\sigma_i\sigma_{i+1}$ for $1 \le i \le n-2$ (braid relation)

A **braid word** is a finite sequence of generators and their inverses, representing a specific braiding pattern. The fundamental problem in braid theory is to determine when two braid words represent the same braid, which leads to the need for simplification algorithms [26].

Appendix B.2. Simplification Rules

The braid word simplification system employs the following rewriting rules:

1. **Cancellation:** $\sigma_i\sigma_i^{-1} \to \varepsilon$ and $\sigma_i^{-1}\sigma_i \to \varepsilon$
2. **Distant Commutation:** $\sigma_i\sigma_j \to \sigma_j\sigma_i$ for $|i-j| > 1$
3. **Braid Relation:** $\sigma_i\sigma_{i+1}\sigma_i \to \sigma_{i+1}\sigma_i\sigma_{i+1}$

These rules preserve the topological equivalence of the braid while simplifying its algebraic representation.

Theorem A2. *The braid word simplification rules form a convergent rewriting system. For any braid word, the simplification process terminates after a finite number of steps and yields a unique normal form.*

 doi:10.20944/preprints202510.2421.v1

Proof. We prove the theorem in two parts: termination and confluence.

Part 1: Termination

Define the length function $L(W) = n$ for a braid word $W = \sigma_{i_1}^{\epsilon_1} \cdots \sigma_{i_n}^{\epsilon_n}$. While Rule 1 strictly decreases length, Rules 2 and 3 preserve it. To ensure termination, we define a complexity measure: Let $C(W)$ be a tuple $(L(W), I(W))$ ordered lexicographically, where: $L(W)$ is the length of the word as defined above; $I(W)$ is the number of inversion pairs, defined as:

$$I(W) = \text{card}\{(k, l) \mid 1 \leq k < l \leq n,\ i_k > i_l\}$$

Now we analyze how each rule affects $C(W)$:

Rule 1 (Cancellation): Reduces $L(W)$ by 2, thus strictly decreasing $C(W)$

Rule 2 (Distant Commutation): Preserves $L(W)$ but may change $I(W)$. However, when we swap σ_i and σ_j with $|i - j| > 1$, if $i < j$, then $I(W)$ increases by 1; if $i > j$, then $I(W)$ decreases by 1. To ensure termination, we restrict the application of Rule 2 to cases where $i > j$, which decreases $I(W)$ and thus decreases $C(W)$

Rule 3 (Braid Relation): Preserves $L(W)$ but changes the arrangement of generators. We can define a more refined measure that decreases when this rule is applied, such as counting the number of local maxima in the braid word

Since $C(W)$ is a tuple of non-negative integers ordered lexicographically, and each rewriting step strictly decreases $C(W)$, the rewriting process must terminate after a finite number of steps.

Part 2: Confluence

To prove confluence, we examine all critical pairs where different rules could be applied to the same word. We need to show that for each critical pair, both possible reduction paths eventually lead to the same result.

The main critical pairs to consider are:

Rule 1 and Rule 2 overlap: For segments like $\sigma_i \sigma_j^{-1} \sigma_k$ with $|i - j| > 1$, we find no actual conflict as Rule 1 doesn't apply directly to these generators.

Rule 1 and Rule 3 overlap: In cases like $\sigma_i \sigma_{i+1} \sigma_i^{-1}$, neither rule applies directly due to exponent mismatches, eliminating potential conflicts.

Rule 2 and Rule 3 overlap: For sequences like $\sigma_i \sigma_j \sigma_k$, careful analysis shows that when both rules could theoretically apply, their applications don't create divergent reduction paths.

Rule 3 self-overlap: The most significant case involves words like $\sigma_i \sigma_{i+1} \sigma_i \sigma_{i+1}$, where Rule 3 can be applied to different overlapping triples. Detailed verification confirms that all reduction paths converge to the same normal form.

Since the system is terminating and all critical pairs are confluent, by Newman's Lemma [27], the system is globally confluent.

Part 3: Normal Form

The unique normal form satisfies three key properties:

1. No cancelable pairs ($\sigma_i \sigma_i^{-1}$ or $\sigma_i^{-1} \sigma_i$)
2. Canonical generator ordering
3. Consistent application of braid relations

This completes the proof that the braid word simplification algorithm converges to a unique normal form. □

Appendix C. Topological Invariance of the Writhe Under Ambient Isotopy

Theorem A3. *Let C be a simple closed curve in $\mathbb{R}^3$. The Writhe $Wr(C)$, defined by the Gauss-type double integral, is invariant under ambient isotopy.*

 doi:10.20944/preprints202510.2421.v1

Proof. Let $F : \mathbb{R}^3 \times [0,1] \to \mathbb{R}^3$ be an ambient isotopy with $F_0 = \mathrm{id}$, generating the deformed curve $C(t) = F_t(C)$. The time-dependent writhe is

$$Wr(t) = \frac{1}{4\pi} \oint_{C(t)} \oint_{C(t)} \omega, \text{where} \quad \omega = \frac{(\mathbf{r}_1 - \mathbf{r}_2) \cdot (\mathrm{d}\mathbf{r}_1 \times \mathrm{d}\mathbf{r}_2)}{\|\mathbf{r}_1 - \mathbf{r}_2\|^3}. \tag{A4}$$

To establish invariance, we show $\frac{\mathrm{d}}{\mathrm{d}t} Wr(t) = 0$. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the velocity fields induced by F_t along $C(t)$. Applying the Leibniz rule:

$$\frac{\mathrm{d}}{\mathrm{d}t} Wr(t) = \frac{1}{4\pi} \left[\oint_{C(t)} \oint_{C(t)} \mathcal{L}_{\mathbf{v}_1} \omega + \oint_{C(t)} \oint_{C(t)} \mathcal{L}_{\mathbf{v}_2} \omega \right]. \tag{A5}$$

Using Cartan's formula $\mathcal{L}_{\mathbf{v}} \omega = i_{\mathbf{v}}(\mathrm{d}\omega) + \mathrm{d}(i_{\mathbf{v}}\omega)$ and noting ω is closed ($\mathrm{d}\omega = 0$) except at $\mathbf{r}_1 = \mathbf{r}_2$ (avoided during isotopy), we have $\mathcal{L}_{\mathbf{v}} \omega = \mathrm{d}(i_{\mathbf{v}}\omega)$. Thus:

$$\frac{\mathrm{d}}{\mathrm{d}t} Wr(t) = \frac{1}{4\pi} \left[\oint_{C(t)} \oint_{C(t)} \mathrm{d}(i_{\mathbf{v}_1} \omega) + \oint_{C(t)} \oint_{C(t)} \mathrm{d}(i_{\mathbf{v}_2} \omega) \right]. \tag{A6}$$

By Stokes' Theorem, the integral of an exact form over the closed manifold $C(t) \times C(t)$ vanishes:

$$\oint_{C(t)} \oint_{C(t)} \mathrm{d}(\alpha) = 0 \quad \text{for any differential form } \alpha.$$

Consequently, $\frac{\mathrm{d}}{\mathrm{d}t} Wr(t) = 0$, establishing the constancy of $Wr(t)$ throughout the isotopy. □

Appendix D. Verification of Closed-Loop System Stability

Appendix D.1. Introduction

Consider the multi-agent dynamical system:

$$\dot{x} = f(x, u_1, \ldots, u_N) \tag{A7}$$

with state $x \in \mathbb{R}^n$ and control inputs $u_i \in \mathbb{R}^{m_i}$ for $i = 1, \ldots, N$. Under MARL policies $u_i = \pi_i(x)$, the closed-loop system becomes:

$$\dot{x} = f(x, \pi_1(x), \ldots, \pi_N(x)) \tag{A8}$$

This section establishes the asymptotic stability of the equilibrium point $x = 0$.

Appendix D.2. Assumptions

Assumption 1. *The multi-agent system dynamics $f : \mathbb{R}^n \times \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_N} \to \mathbb{R}^n$ are Lipschitz continuous:*

$$\begin{aligned} &\|f(x, u_1, \ldots, u_N) - f(y, v_1, \ldots, v_N)\| \\ &\leq L_f \left(\|x - y\| + \sum_{i=1}^{N} \|u_i - v_i\| \right) \end{aligned} \tag{A9}$$

Assumption 2 (Policy Regularity). *Each agent's policy $\pi_i : \mathbb{R}^n \to \mathbb{R}^{m_i}$ learned by MARL algorithms is Lipschitz continuous with constant $L_{\pi_i} > 0$:*

$$\|\pi_i(x) - \pi_i(y)\| \leq L_{\pi_i} \|x - y\| \tag{A10}$$

This is ensured through Lipschitz-preserving network architectures and regularization techniques common to MAPPO, MADDPG, and MASAC.

Assumption 3 (Equilibrium). *The origin satisfies $f(0, \pi_1(0), \ldots, \pi_N(0)) = 0$.*

 doi:10.20944/preprints202510.2421.v1

Appendix D.3. Stability Analysis

Lemma A1. *Under Assumptions 1-2, the closed-loop system is Lipschitz continuous with constant* $L_{cl} = L_f\left(1+\sum_{i=1}^{N} L_{\pi_i}\right)$.

Proof. For any $x, y \in \mathbb{R}^n$:

$$\begin{aligned}
&\|f(x, \pi_1(x), \ldots, \pi_N(x)) - f(y, \pi_1(y), \ldots, \pi_N(y))\| \\
&\leq L_f\left(\|x-y\| + \sum_{i=1}^{N} \|\pi_i(x) - \pi_i(y)\|\right) \\
&\leq L_f\left(1 + \sum_{i=1}^{N} L_{\pi_i}\right)\|x-y\|
\end{aligned}$$

□

Theorem A4. *Suppose there exists a continuously differentiable function* $V : \mathbb{R}^n \to \mathbb{R}$ *satisfying:*

1. $V(0) = 0$ *and* $V(x) > 0$ *for* $x \neq 0$
2. $V(x) \to \infty$ *as* $\|x\| \to \infty$
3. $\dot{V}(x) = \frac{\partial V}{\partial x} f(x, \pi_1(x), \ldots, \pi_N(x)) < 0$ *for* $x \neq 0$

Then the equilibrium $x = 0$ *is globally asymptotically stable under MARL policies.*

Proof. The positive definiteness and continuity of V ensure local stability, while the negative definiteness of $\dot{V}$ combined with radial unboundedness guarantees global convergence to the equilibrium. □

Appendix D.4. Lyapunov Function Construction for MARL

Lemma A2 (Value Function Candidate)**.** *For MARL algorithms with centralized or decentralized critics, if the team reward satisfies* $r(x, \pi_1(x), \ldots, \pi_N(x)) \leq 0$ *with equality only at* $x = 0$*, and the value function approximates the discounted cumulative reward, then* $V_\theta(x)$ *serves as a valid Lyapunov candidate.*

Proof. Under these conditions, $V_\theta(x) \geq 0$ with $V_\theta(0) = 0$ due to the reward structure, and $\dot{V}_\theta(x) \approx r(x, \pi_1(x), \ldots, \pi_N(x)) - \gamma V_\theta(x) \leq 0$. □

Appendix E. Sample Complexity Analysis of TA-MARL

Appendix E.1. Theoretical Framework

We consider a multi-agent POMDP $\langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{P}, \mathcal{R}, \gamma, N \rangle$ with topology-augmented state space $\mathcal{S}_{\text{topo}} = \mathcal{S} \times \mathcal{T}$, where $\mathcal{T}$ contains linking number Lk, braid word length $|B_{\text{red}}|$, and self-entanglement Wr.

The topological penalty function is defined as:

$$\text{TopoPenalty}(s, a) = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{bmatrix} \cdot \begin{bmatrix} \max_{j,k} |Lk_{j,k}| \\ \tanh\left(\frac{|B_r|}{c_1}\right) \\ I_{\text{entangled}}(s) \end{bmatrix} \tag{A11}$$

with $\alpha_1, \alpha_2, \alpha_3 > 0$, c_1 normalization constant, and $I_{\text{entangled}}$ entanglement indicator.

Theorem A5. *For finite MARL with optimal* V^**, TA-MARL policy* π_{topo}*, accuracy* $\epsilon > 0$*, with probability* $\geq 1 - \delta$*:*

$$N_{sample}(\epsilon, \delta) \leq \frac{C|\mathcal{S}_{eff}||\mathcal{A}|}{(1-\gamma)^3\epsilon^2} + \frac{C\log(1/\delta)}{\lambda_{topo}(1-\gamma)^3\epsilon^2} \tag{A12}$$

where $\mathcal{S}_{eff} = \{s \in \mathcal{S} \mid TopoRisk(s) < \tau_{safe}\}$, $C > 0$, $\gamma \in (0,1)$, $\lambda_{topo} > 0$.

 doi:10.20944/preprints202510.2421.v1

Proof. Let $\rho = \mathbb{P}(s \notin \mathcal{S}_{\text{eff}})$:

$$\rho \leq \sum_{j=1}^{N} \sum_{k=j+1}^{N} \mathbb{P}(Lk_{j,k} \neq 0) + \mathbb{P}(|B_r| > \theta_B) \tag{A13}$$

Using Gaussian linking: $|\mathcal{S}_{\text{eff}}| = (1-\rho)|\mathcal{S}|$. The γ-contractive operator $\mathcal{T}_{\text{topo}}$ satisfies:

$$|\mathcal{T}_{\text{topo}} V - V|_\infty \leq \frac{|R_{\text{topo}} - R|_\infty}{1-\gamma} + \gamma |V - V^*|_\infty \tag{A14}$$

Standard PAC gives $N_{\text{sample}} = O\left(\frac{|\mathcal{S}||\mathcal{A}|}{(1-\gamma)^3 \epsilon^2} \log(\cdot)\right)$, yielding $\frac{N_{\text{sample}}^{\text{topo}}}{N_{\text{sample}}} \approx 1 - \rho$. □

Theorem A6. *For π_0 and K-round π_K, entanglement satisfies:*

$$\mathbb{P}(\textit{entangle} \mid \pi_K) \leq P_0 \exp\left(-\frac{K \lambda_{topo}}{N_{sample}(\alpha)}\right) + \epsilon_{approx} \tag{A15}$$

where $P_0 = \mathbb{P}(\textit{entangle} \mid \pi_0)$, $\lambda_{topo} > 0$, $\epsilon_{approx} > 0$.

Proof. Model as absorbing Markov chain:

$$\mathbb{P}(\text{entangle} \mid \pi) = \alpha_0^T (I - Q_\pi)^{-1} r \tag{A16}$$

Policy improvement ensures risk reduction. By Perron-Frobenius:

$$\mathbb{P}(\text{entangle} \mid \pi_K) \leq \frac{|\alpha_0|_2 |r|_2}{\lambda_{\min}(I - Q_{\pi_K})} \tag{A17}$$

With $K_{\text{eff}} = N_{\text{total}} / N_{\text{sample}}(\alpha)$ and $\xi(\alpha) \approx 0.8$, convergence accelerates by 25%. □

References

1. Rus, D.; Tolley, M.T. Design, fabrication and control of soft robots. *Nature* **2015**, *521*, 467–475. https://doi.org/10.1038/nature14543.
2. Karaman, S.; Frazzoli, E. Sampling-based algorithms for optimal motion planning. *Int. J. Rob. Res.* **2011**, *30*, 846–894. https://doi.org/10.1177/0278364911406761.
3. Betts, J.T. Survey of Numerical Methods for Trajectory Optimization. *Journal of Guidance Control and Dynamics* **1998**, *21*, 193–207.
4. Shiller, Z. Online sub-optimal obstacle avoidance. In Proceedings of the Proceedings 1999 IEEE International Conference on Robotics and Automation (Cat. No.99CH36288C), 1999, Vol. 1, pp. 335–340 vol.1. https://doi.org/10.1109/ROBOT.1999.770001.
5. Lowe, R.; Wu, Y.; Tamar, A.; Harb, J.; Abbeel, P.; Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. In Proceedings of the Proceedings of the 31st International Conference on Neural Information Processing Systems, Red Hook, NY, USA, 2017; NIPS'17, p. 6382–6393.
6. Rashid, T.; Samvelyan, M.; De Witt, C.S.; Farquhar, G.; Foerster, J.; Whiteson, S. Monotonic value function factorisation for deep multi-agent reinforcement learning. *J. Mach. Learn. Res.* **2020**, *21*.
7. Chen, Y.F.; Liu, M.; Everett, M.; How, J.P. Decentralized non-communicating multiagent collision avoidance with deep reinforcement learning. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), 2017, pp. 285–292. https://doi.org/10.1109/ICRA.2017.7989037.
8. Hatton, R.L.; Choset, H. Generating gaits for snake robots by annealed chain fitting and Keyframe wave extraction. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2009, pp. 840–845. https://doi.org/10.1109/IROS.2009.5354257.
9. Soft robotics: Biological inspiration, state of the art, and future research. *Applied Bionics and Biomechanics* **2008**, *5*, 99–117. https://doi.org/10.1080/11762320802557865.

 doi:10.20944/preprints202510.2421.v1

10. Rucker, D.C.; Jones, B.A.; Webster III, R.J. A Geometrically Exact Model for Externally Loaded Concentric-Tube Continuum Robots. *IEEE Transactions on Robotics* **2010**, *26*, 769–780. https://doi.org/10.1109/TRO.2010.2062570.
11. Webster, R.J.; Jones, B.A. Design and Kinematic Modeling of Constant Curvature Continuum Robots: A Review. *The International Journal of Robotics Research* **2010**, *29*, 1661 – 1683.
12. Lou, G.; Wang, C.; Xu, Z.; Liang, J.; Zhou, Y. Controlling Soft Robotic Arms Using Hybrid Modelling and Reinforcement Learning. *IEEE Robotics and Automation Letters* **2024**, *9*, 7070–7077. https://doi.org/10.1109/LRA.2024.3418312.
13. Mohammad, N.; Bezzo, N. Soft Actor-Critic-Based Control Barrier Adaptation for Robust Autonomous Navigation in Unknown Environments. *2025 IEEE International Conference on Robotics and Automation (ICRA)* **2025**, pp. 2361–2367.
14. Yang, Z.; Wang, Y.; Jiang, Y.; Zhang, H.; Yang, C. DeformerNet based 3D Deformable Objects Shape Servo Control for Bimanual Robot Manipulation. In Proceedings of the 2024 IEEE International Conference on Industrial Technology (ICIT), 2024, pp. 1–7. https://doi.org/10.1109/ICIT58233.2024.10570080.
15. Afnan Ahmed, A.; Saber, S.; Jinane, M.; Noel, M. A multi-robot collaborative manipulation framework for dynamic and obstacle-dense environments: integration of deep learning for real-time task execution. *Frontiers in Robotics and AI* **2025**, *12*, 1585544. https://doi.org/10.3389/frobt.2025.1585544.
16. Ricca, R.L. Applications of knot theory in fluid mechanics. *Banach Center Publications* **1998**, *42*, 321–346.
17. Peterson, I. Knot Physics. *Science News* **1989**, *135*, 174–174.
18. Zhou, P.; Zheng, P.; Qi, J.; Li, C.; Lee, H.Y.; Duan, A.; Lu, L.; Li, Z.; Hu, L.; Navarro-Alarcon, D. Reactive human–robot collaborative manipulation of deformable linear objects using a new topological latent control model. *Robot. Comput.-Integr. Manuf.* **2024**, *88*. https://doi.org/10.1016/j.rcim.2024.102727.
19. Kuskonmaz, B.; Wisniewski, R.; Kallesøe, C. Topological Data Analysis-Based Replay Attack Detection for Water Networks. *IFAC-PapersOnLine* **2024**, *58*, 91–96. 12th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes SAFEPROCESS 2024, https://doi.org/https://doi.org/10.1016/j.ifacol.2024.07.199.
20. Lowe, R.; Wu, Y.; Tamar, A.; Harb, J.; Abbeel, P.; Mordatch, I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. *ArXiv* **2017**, *abs/1706.02275*.
21. Zhan, G.; Zhang, X.; Li, Z.; Xu, L.; Zhou, D.; Yang, Z. Multiple-UAV Reinforcement Learning Algorithm Based on Improved PPO in Ray Framework. *Drones* **2022**, *6*. https://doi.org/10.3390/drones6070166.
22. Sá Barreto, A.; Stefanov, P. Recovery of a general nonlinearity in the semilinear wave equation. *Asymptotic Analysis* **2024**, *138*, 27–68. https://doi.org/10.3233/ASY-231890.
23. Kuba, J.G.; Chen, R.; Wen, M.; Wen, Y.; Sun, F.; Wang, J.; Yang, Y. Trust Region Policy Optimisation in Multi-Agent Reinforcement Learning. In Proceedings of the International Conference on Learning Representations, 2022.
24. Garg, S.; Goharimanesh, M.; Sajjadi, S.; Janabi-Sharifi, F. Autonomous control of soft robots using safe reinforcement learning and covariance matrix adaptation. *Engineering Applications of Artificial Intelligence* **2025**, *153*, 110791. https://doi.org/https://doi.org/10.1016/j.engappai.2025.110791.
25. Artin, E. Theory of Braids. *Annals of Mathematics* **1947**, *48*, 101–126.
26. BIRMAN, J.S. *Braids, Links, and Mapping Class Groups. (AM-82), Volume 82*; Vol. 82, Princeton University Press, 1974.
27. Newman, M.H.A. On Theories with a Combinatorial Definition of "Equivalence". *Annals of Mathematics* **1942**, *43*, 223–243.