Zhenyi ZHANG, Jie HUANG, Congjie PAN, 2024. Multi-agent reinforcement learning behavioral control for nonlinear second-order systems. *Frontiers of Information Technology & Electronic Engineering,* 25(6):869-886.
https://doi.org/10.1631/FITEE.2300394

Multi-agent reinforcement learning behavioral control for nonlinear second-order systems

Key words: Reinforcement learning; Behavioral control; Second-order systems; Mission supervisor

Corresponding author: Jie HUANG
E-mail: jie.huang@fzu.edu.cn
ORCID: https://orcid.org/0000-0001-7346-5034

Motivation

1. Reinforcement learning mission supervisor can be used for only single-agent applications, since it is modeled through Markov decision processes. As a result, any cooperative behavior cannot be implemented, which greatly limits the swarm intelligence.
2. Although reinforcement learning behavioral control guarantees the convergence of position error signals, it is not enough for second-order systems. Generally, second-order systems require both position and velocity error signals to converge.
3. Reinforcement learning controller (RLC) does not have input saturation constraints, and thus the control input may exceed the physical limit of the actuator. Particularly, if behavior priorities are switched, the problem of excessive control input may be aggravated.

Main idea

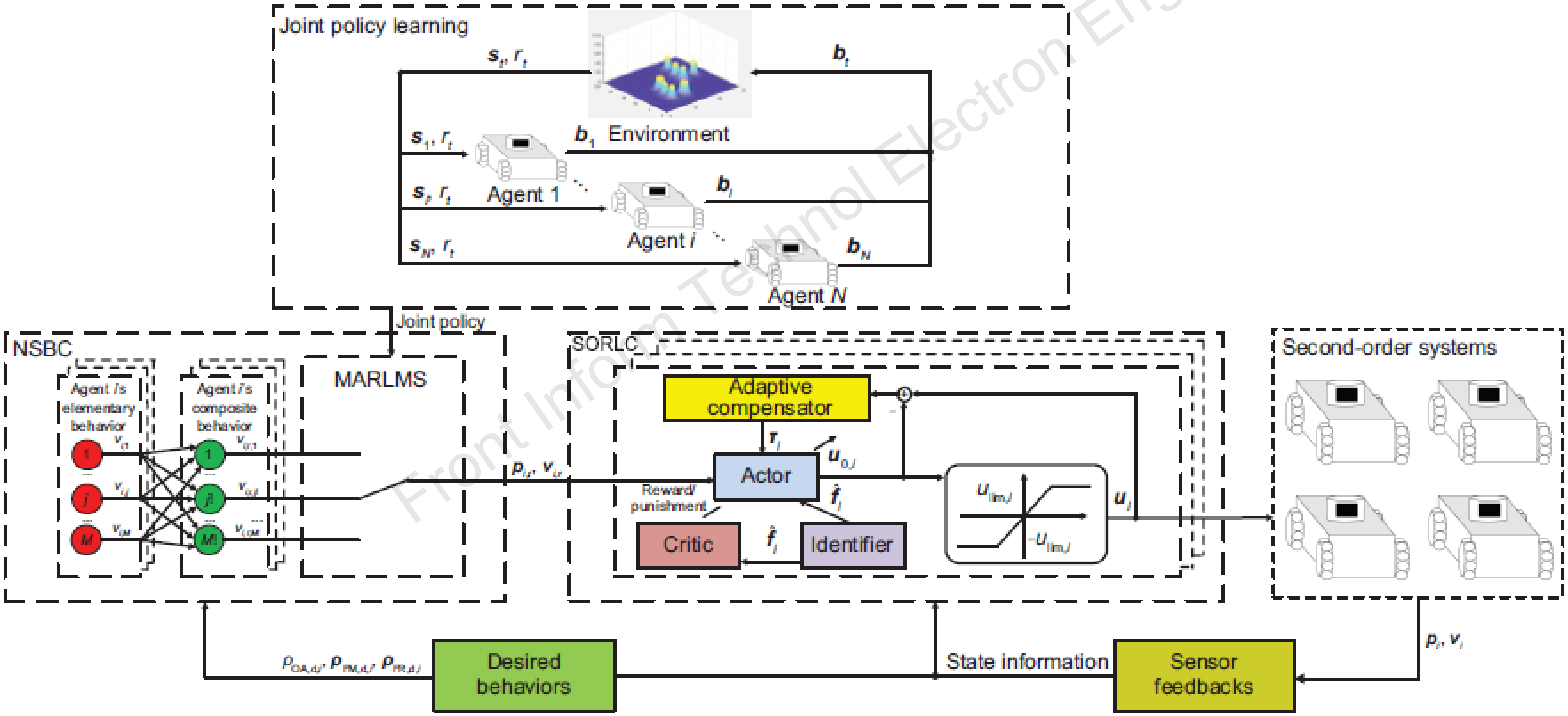
1. A multi-agent reinforcement learning mission supervisor (MARLMS) is proposed to learn the optimal joint behavior priority policy. Through learning a joint behavior priority policy, the switching frequency of behavioral control is reduced significantly.

2. A group of second-order reinforcement learning controllers (SORLCs) with the identifier–actor–critic structure are developed to learn the optimal control policies. As a result, the control cost of behavioral control is reduced significantly.

3. A group of adaptive compensators are designed to maintain the optimal control performance and counteract the saturation effect in real time.

Method

Multi-agent reinforcement learning behavioral control



Major results

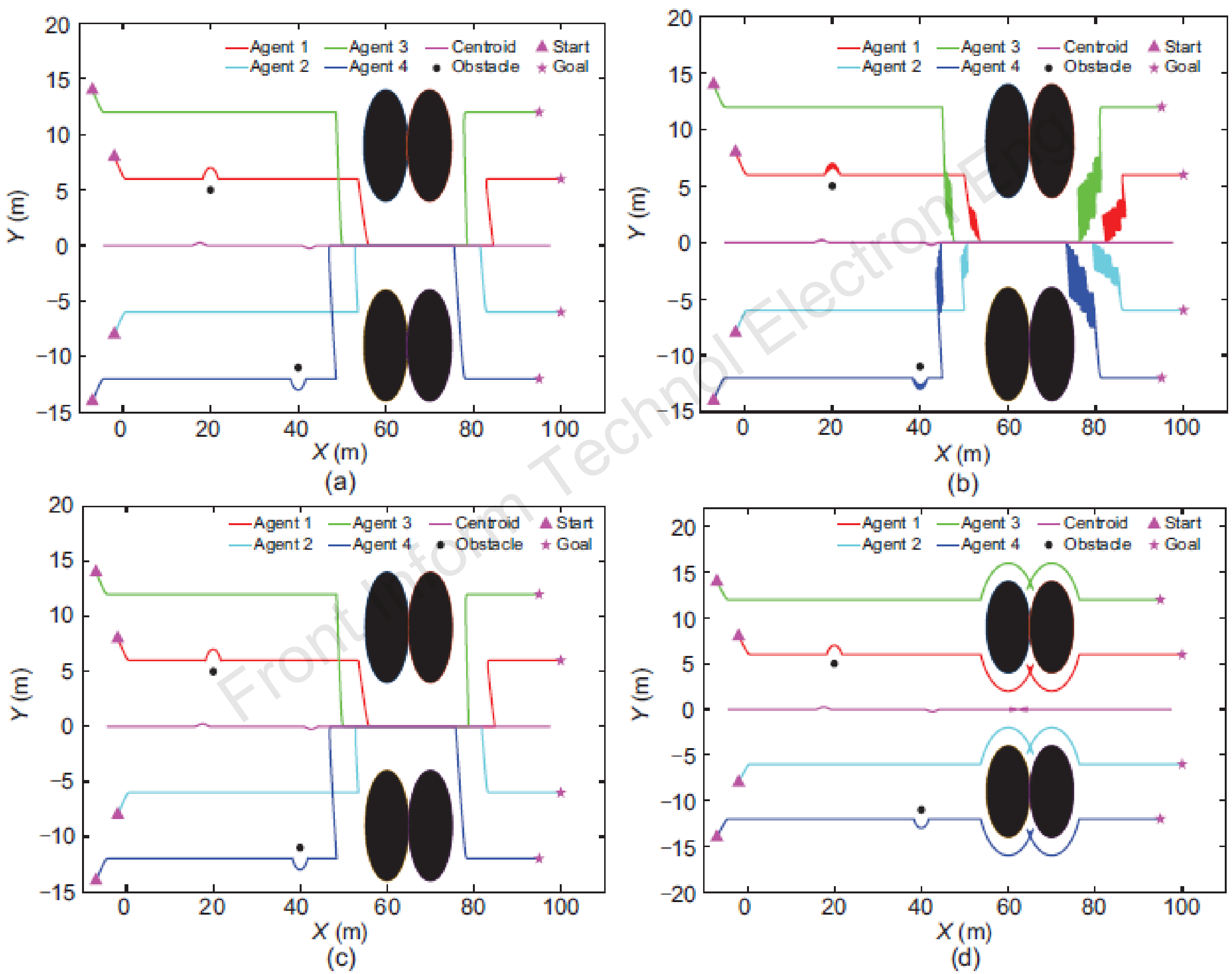


Fig. 2 Trajectories of agents with different mission supervisors: (a) MARLMS; (b) FSAMS; (c) MPCMS; (d) RLMS

Major results (Cont'd)

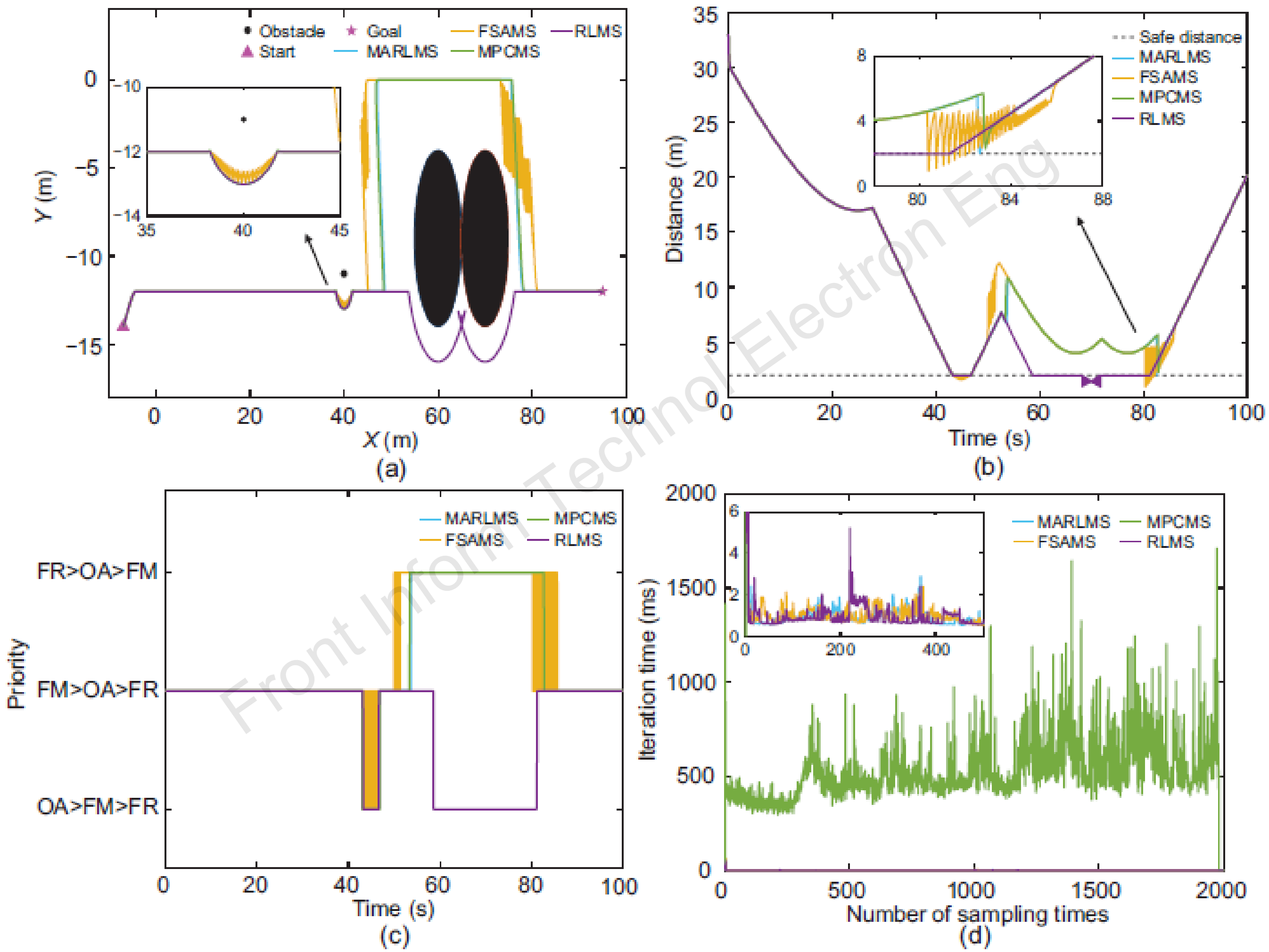


Fig. 3 Results of the 4th agent with different mission supervisors: (a) trajectories; (b) distances between the agent and obstacles; (c) behavior priorities; (d) iteration time

Major results (Cont'd)

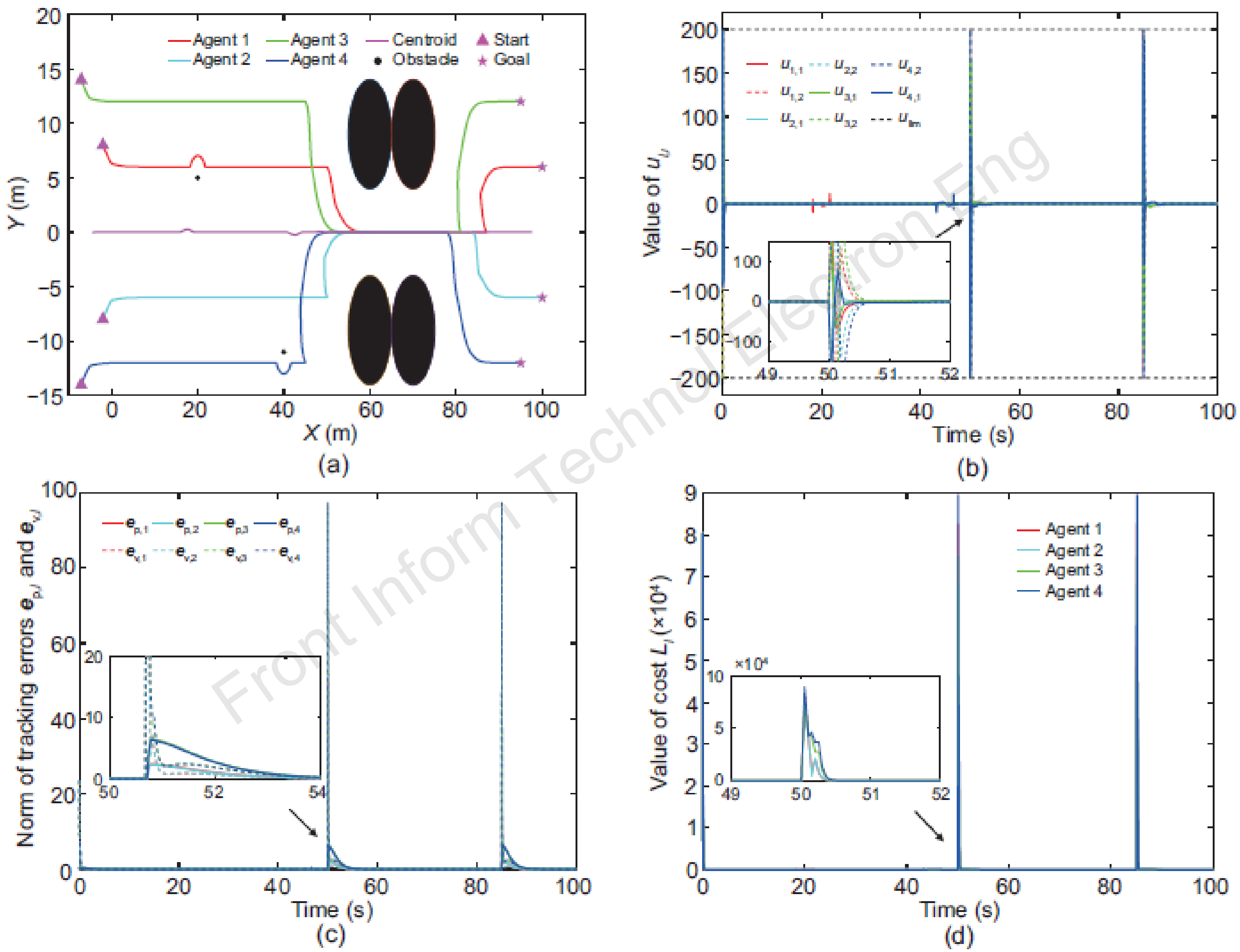


Fig. 6 Control performance of the proposed SORLC: (a) trajectories; (b) control inputs; (c) tracking errors; (d) costs

Major results (Cont'd)

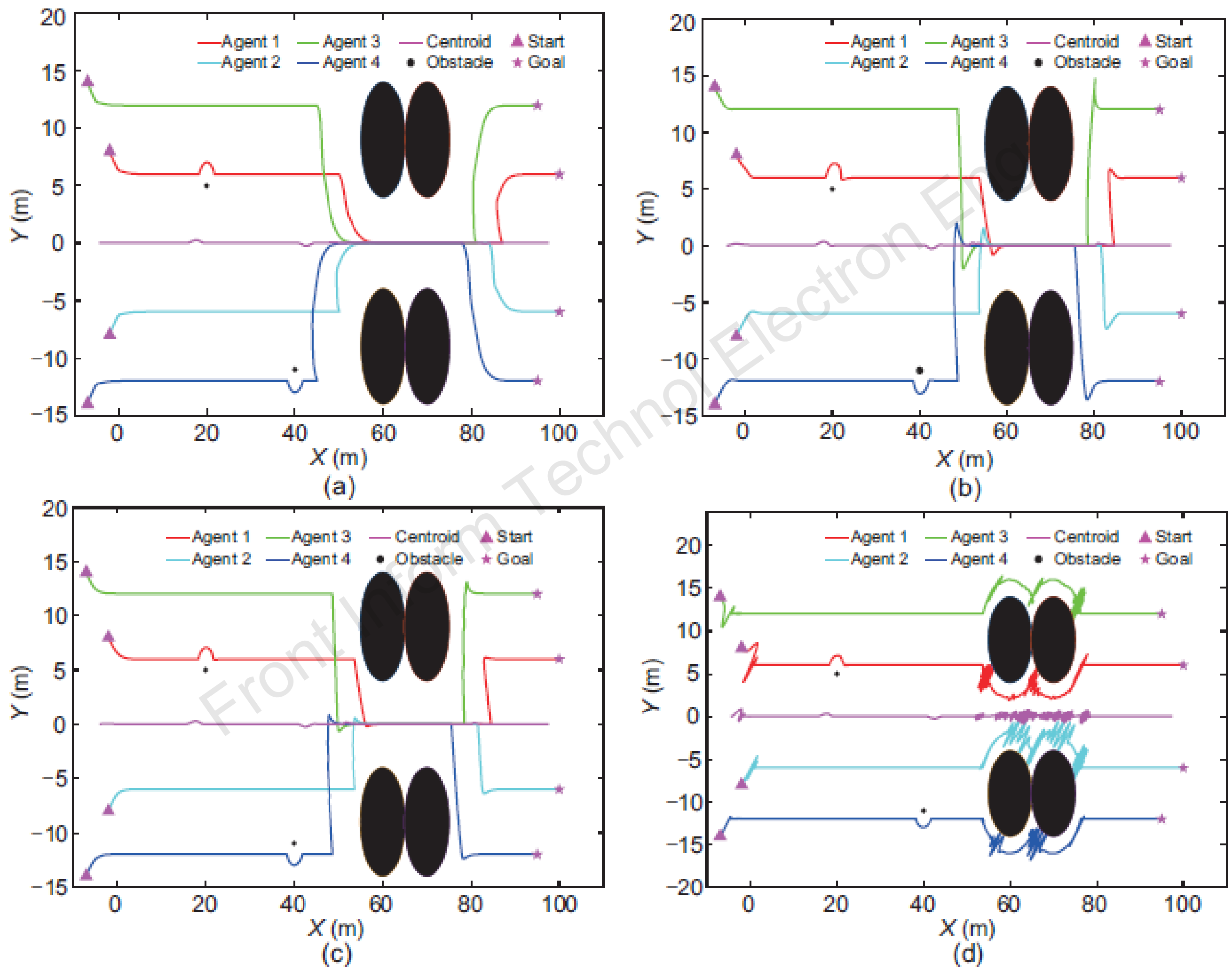


Fig. 8 Trajectories of agents with different NSBC methods: (a) MARLBC; (b) finite-time NSBC; (c) fixed-time NSBC; (d) RLBC

Major results (Cont'd)

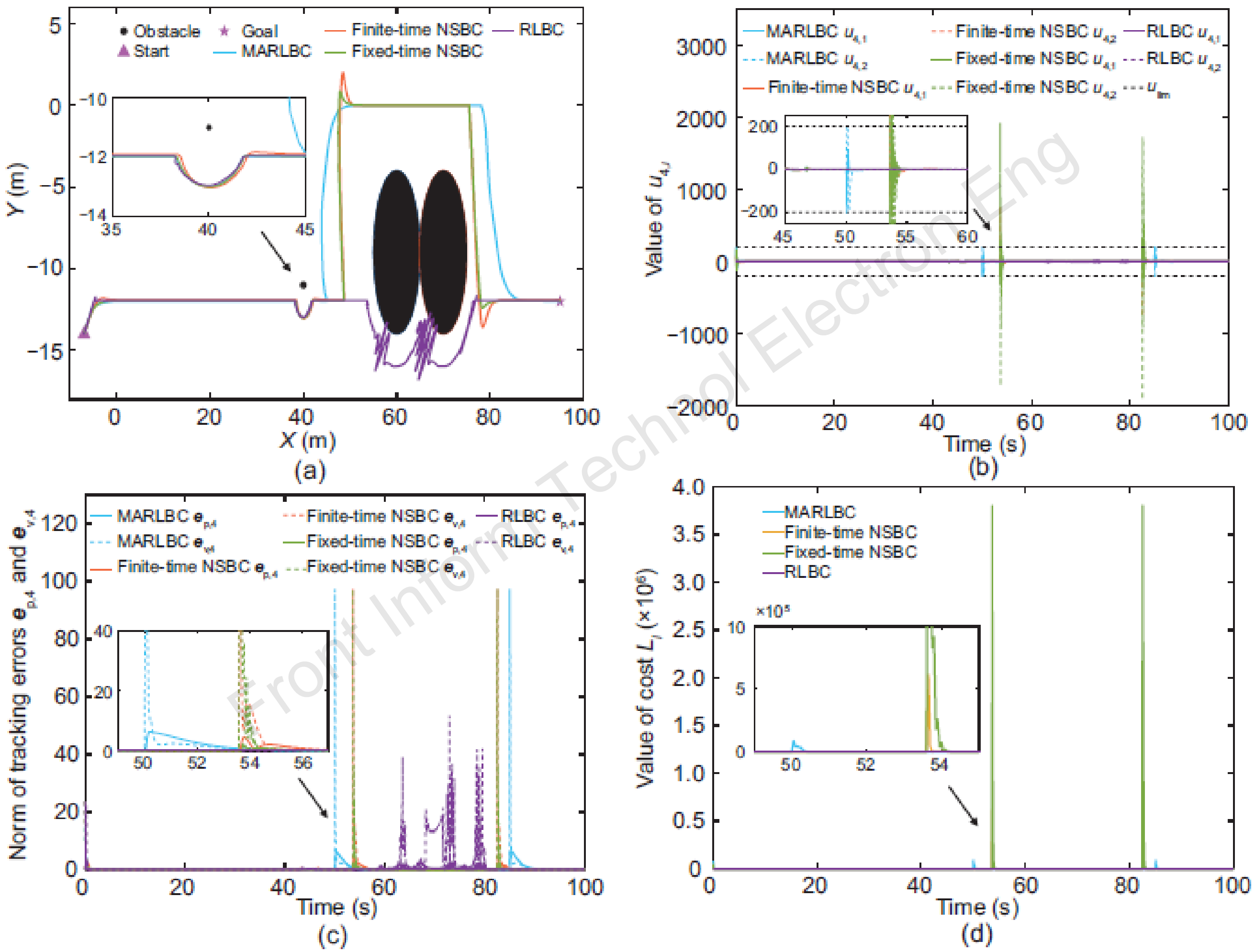


Fig. 9 Control performances of the 4th agent with different NSBC methods: (a) trajectories; (b) control inputs; (c) tracking errors; (d) costs

Major results (Cont'd)

Table 5 Maximum and total control cost values of different NSBC frameworks

Index	Maximum control cost value ($\times 10^4$)			Index	Total control cost value ($\times 10^4$)		
	MARLBC	Finite-time	Fixed-time		MARLBC	Finite-time	Fixed-time
Max-a1	9	17	103	Tot-a1	382	435	518
Max-a2	9	17	115	Tot-a2	364	404	790
Max-a3	9	60	372	Tot-a3	570	1384	1813
Max-a4	9	60	372	Tot-a4	609	1391	1868

Table 6 Maximum and total control input values of different NSBC frameworks

Index	Maximum control input value			Index	Total control input value		
	MARLBC	Finite-time	Fixed-time		MARLBC	Finite-time	Fixed-time
Max-a1	200	384	960	Tot-a1	2210	2756	9127
Max-a2	200	384	1061	Tot-a2	2341	2883	13 370
Max-a3	200	768	1920	Tot-a3	4118	5274	16 112
Max-a4	200	768	1920	Tot-a4	4285	5466	16 646

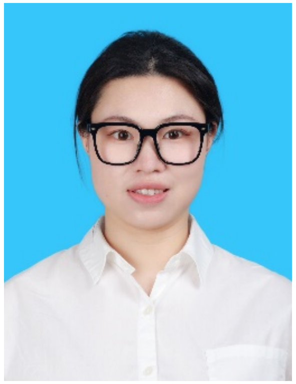
Conclusions

1. Multi-agent reinforcement learning behavioral control has smaller control input and cost values than the existing null-space-based behavioral control methods of second-order systems when the behavior priorities are switched.

2. Compared with the traditional reinforcement learning behavioral control, the proposed multi-agent reinforcement learning behavioral control allows the implementation of cooperative behavior and ensures the convergence of velocity signals.

Zhenyi ZHANG received the B.E. degree in mechanical and electronic engineering, and the M.E. degree in mechanical engineering from Zhejiang Sci-Tech University, Hangzhou, China, in 2016 and 2019, respectively. He is now a Ph.D. student at Fuzhou University, China. His research interests include intelligent robot ethology and multi-agent systems.

Jie HUANG received the B.E. degree in electrical engineering and automation, and the M.E. degree in control engineering from Fuzhou University in 2005 and 2010, respectively, and the Ph.D. degree in control science and engineering from Beijing Institute of Technology in 2015. From 2014 to 2016, he was a postdoctoral researcher with the Faculty of Mathematics and Natural Sciences, University of Groningen, the Netherlands. From 2016 to 2018, he was a docent with the Faculty of Science and Engineering, University of Groningen, the Netherlands. He is currently a Full Professor of robotic and control with the College of Electrical Engineering and Automation, Fuzhou University. He is the Vice-President of the Fujian Automation Association, Fujian Province, China. His research interests include autonomous robots, complex network dynamics, multi-agent systems and industrial Internet.



Congjie PAN received her B.E. degree in electrical engineering and automation engineering from Huzhou Normal University in 2019. She is currently a master's student at Fuzhou University, China. Her research interests include reinforcement learning and multi-agent systems.