Yu SHI, Yongzhao HUA, Jianglong YU, Xiwang DONG, Zhang REN, 2022. Multi-agent differential game based cooperative synchronization control using a data-driven method. *Frontiers of Information Technology & Electronic Engineering*, 23(7):1043-1056. https://doi.org/10.1631/FITEE.2200001

Multi-agent differential game based cooperative synchronization control using a data-driven method

Key words: Multi-agent system; Differential game; Synchronization control; Data-driven; Reinforcement learning.

Corresponding author: Xiwang DONG
E-mail: xwdong@buaa.edu.cn
ORCID: https://orcid.org/0000-0002-4778-248X

Motivation

1. Differential game for multi-agent system (MAS) is an intersection of control and game problems. Typical methods may not necessarily lead to the expected Nash equilibrium due to the coupling of network interactions.

2. It is required to provide a systematized formulation and analysis method for differential game while the corresponding controller can achieve synchronization control.

3. The solution to the differential game is highly coupled and a data-driven solution is a worthy research topic.

Main idea

1. A differential game based frame for MAS synchronization control is proposed, where the optimum or equilibrium is considered along with the overall stability.

2. A systematized scheme consisting of both local and global games is provided with guaranteed Nash equilibrium.

3.The off-policy reinforcement learning (RL) technique is derived for the continuous-time MAS to solve this problem using online data.

Method

1. The distributed and Nash properties of the game solution are discussed in detail. It is proved that they may not hold at the same time under local neighboring interactions.

2. Local zero-sum and global nonzero-sum game based index functions are separately discussed with stability and equilibrium analysis.

3. Inspired by the concept of Q-learning, a data-driven method based on off-policy RL is proposed to solve this coupled differential game problem.

Major results

1. Solvability analysis of coupled differential game

MAS dynamics

$$\xi_i(t) = \sum_{j\in\mathcal{N}_i} w_{ij}\left(x_i(t) - x_j(t)\right) + w_{i0}\left(x_i(t) - x_0(t)\right).$$

$$\dot{\xi}_i(t) = A\xi_i(t) + d_i B u_i - \sum_{j\in\mathcal{N}_i} w_{ij} B u_j.$$

The corresponding Hamilton–Jacobi (HJ) equation derived as follows is hard to solve:

$$\begin{aligned}0 =& (\nabla_{\xi_i} V_i)^{\mathrm{T}}(A\xi_i - d_i^2 B R_{ii}^{-1} B^{\mathrm{T}} \nabla_{\xi_i} V_i) \\ &+ (\nabla_{\xi_i} V_i)^{\mathrm{T}}\left(\sum_{j\in\mathcal{N}_i} w_{ij} d_i B_j R_{jj}^{-1} B_j^{\mathrm{T}} \nabla_{\xi_j} V_j\right) \\ &+ \frac{1}{2}\xi_i^{\mathrm{T}} Q_i \xi_i + \frac{1}{2} d_i^2 (\nabla_{\xi_i} V_i)^{\mathrm{T}} B R_{ii}^{-1} B^{\mathrm{T}} \nabla_{\xi_i} V_i \\ &+ \frac{1}{2}\sum_{j\in\mathcal{N}_i} d_j^2 (\nabla_{\xi_j} V_j)^{\mathrm{T}} B R_{jj}^{-1} R_{ij} R_{jj}^{-1} B^{\mathrm{T}} \nabla_{\xi_j} V_j.\end{aligned}$$

Coupled index functions

$$\begin{aligned}J_i(t)|_{(u_i,u_{-i})} &= V_i\left(\xi_i(t), \xi_{-i}(t)\right) \\ &= \int_t^{\infty} r_i\big(\xi_i(\tau), \xi_{-i}(\tau), u_i(\tau), u_{-i}(\tau)\big)\mathrm{d}\tau,\end{aligned}$$

$$\begin{aligned}&r_i\left(\xi_i, u_i, u_{-i}\right) \\ &= \frac{1}{2}\left(\xi_i^{\mathrm{T}} Q_i \xi_i + u_i^{\mathrm{T}} R_{ii} u_i + \sum_{j\in\mathcal{N}_i} u_j^{\mathrm{T}} R_{ij} u_j\right),\end{aligned}$$

There is a contradiction between distributed control and the global Nash equilibrium.

Major results (Cont'd)

2. Local best response solution using RL

By introducing the local matched controller $u_{ij}^{\#}$, the best response solution is provided by $u_i^{\#} = -R_{ii}^{-1} B^{\mathrm{T}} \nabla_{\xi_i} V_i,\ u_{ij}^{\#} = -w_{ij} R_{ij}^{-1} B^{\mathrm{T}} \nabla_{\xi_i} V_i,$ satisfying

$$J_i^*\left(u_i^*, u_{ij}, \tilde{P}_i^*\right) \le J_i^*\left(u_i^*, u_{ij}^*, \tilde{P}_i^*\right) \le J_i^*\left(u_i, u_{ij}^*, \tilde{P}_i^*\right)$$

$$\tilde{P}_i^* A + A^{\mathrm{T}} \tilde{P}_i^* - d_i \tilde{P}_i^* B R_{ii}^{-1} B^{\mathrm{T}} \tilde{P}_i^* + w_{ij}^2 \tilde{P}_i^* B \sum_{j \in \mathcal{N}_i} R_{ij}^{-1} B^{\mathrm{T}} \tilde{P}_i^* + Q_i = 0$$

Design the off-policy RL for coupled continuous time MAS as

$$\begin{aligned} & V_i^k\left(\xi_i(t+T)\right) - V_i^k\left(\xi_i(t)\right) \\ & + \int_t^{t+T} \sum_{j \in \mathcal{N}_i} \left(u_{ij}^{\#k}\right)^{\mathrm{T}} R_{ij} \left(u_{ij}^0 - u_{ij}^k\right) \mathrm{d}t \\ & - \int_t^{t+T} d_i \left(u_i^{\#k}\right)^{\mathrm{T}} R_{ii} \left(u_i^0 - u_i^k\right) \\ = & - \int_t^{t+T} r_i(\xi_i, u_i^k, u_{ij}^k) \mathrm{d}t. \end{aligned}$$

$$\begin{cases} W_i = \left[W_{i1}^{\mathrm{T}}, \mathrm{vec}(W_{i2}^{\mathrm{T}}), \mathrm{vec}(W_{ij2}^{\mathrm{T}})\right], \\ \varPhi_i = [\varPhi_{i1}, \varPhi_{i2}, \varPhi_{i3}], \\ \varPhi_{i1} = \phi_i(\xi_i(t+T)) - \phi_i(\xi_i(t)), \\ \varPhi_{i2} = -\int_t^{t+T} d_i \left[\left(u_i^0 - u_i^k\right)^{\mathrm{T}} R_{ii} \otimes \varphi_i(\xi_i)\right] \mathrm{d}t, \\ \varPhi_{i3} = \int_t^{t+T} \left[\left(u_{ij}^0 - u_{ij}^k\right)^{\mathrm{T}} R_{ij} \otimes \varphi_{ij}(\xi_i)\right] \mathrm{d}t. \end{cases}$$

$$\begin{cases} \dot{W}_i(t) = -\alpha_i \frac{\varPhi_i(\xi_i) S_i(t)}{\left(1 + \varPhi_i^{\mathrm{T}}(\xi_i) \varPhi_i(\xi_i)\right)^2}, \\ S_i(t) = W_i(\varPhi_i(\xi_i)) + \int_t^{t+T} r_i\left(\xi_i, u_i^k, u_{ij}^k\right) \mathrm{d}t. \end{cases}$$

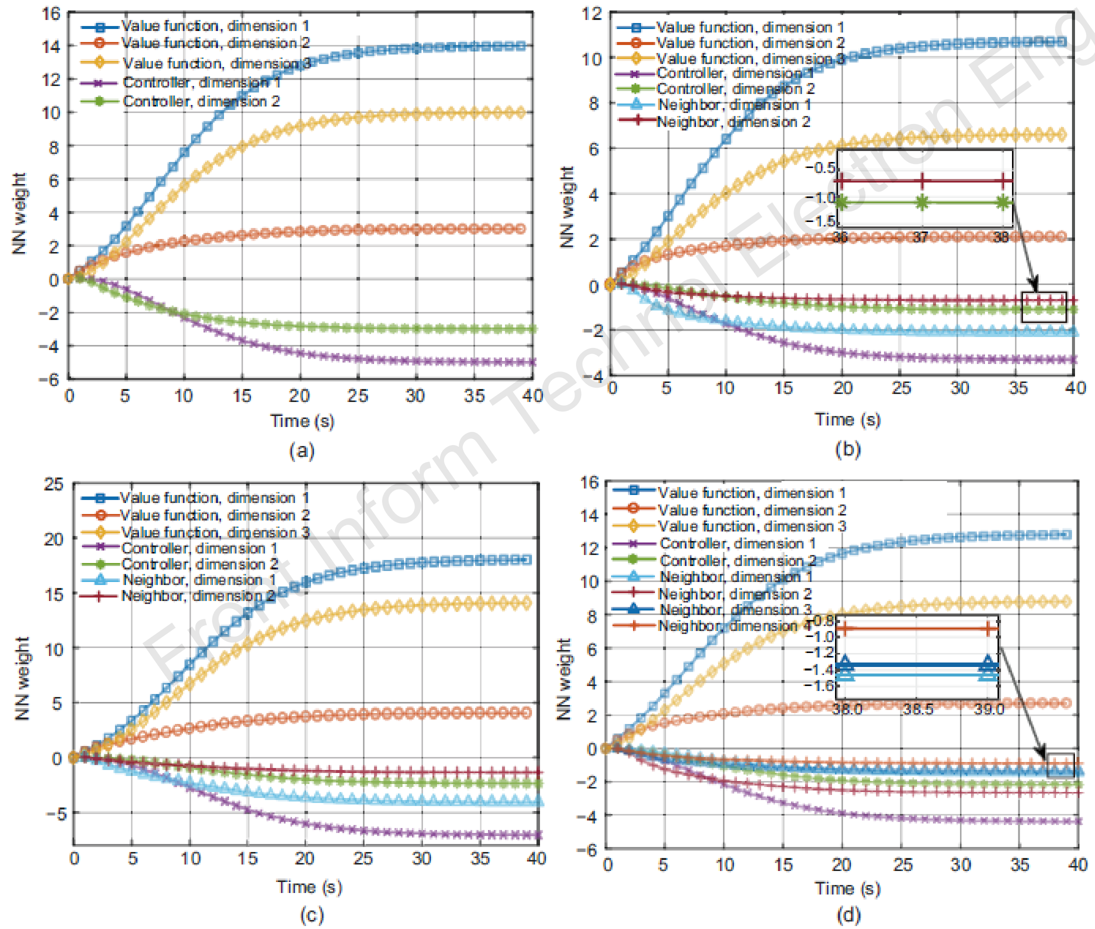
Model-free HJ equation

Neural network approximation

Data-driven update

Major results (Cont’d)

2. Local best response solution using RL



Convergence curves of the neural network (NN) weights in the local best response case: (a) agent 1; (b) agent 2; (c) agent 3; (d) agent 4

Major results (Cont'd)

3. Global Nash solution using RL

Define the following modified index function in a general case:

$$r_i(\xi_i, u_i, u_{-i}) = \frac{1}{2}\left(d_i u_i^{\mathrm{T}} R_{ii} u_i + \sum_{j\in\mathcal{N}_i} u_j^{\mathrm{T}} R_{ij} u_j + \sum_{j\in\mathcal{N}_i} \xi_{ij}^{\mathrm{T}} \tilde{Q}_i \xi_{ij}\right),$$

satisfying

$$\begin{cases} Q_{ij} = -\bar{P}_j^* B R_{jj}^{-1} R_{ij} R_{jj}^{-1} B^{\mathrm{T}} \bar{P}_j^*, \\ S_{ij} = -\bar{P}_i^* B R_{jj}^{-1} B^{\mathrm{T}} \bar{P}_j^*, \end{cases}$$

where $\xi_{ij} = [\xi_i, \xi_j]^{\mathrm{T}}$ and $\tilde{Q}_i = \begin{bmatrix} d_i^{-1} Q_i & S_{ij} \\ S_{ij}^{\mathrm{T}} & Q_{ij} \end{bmatrix}$.

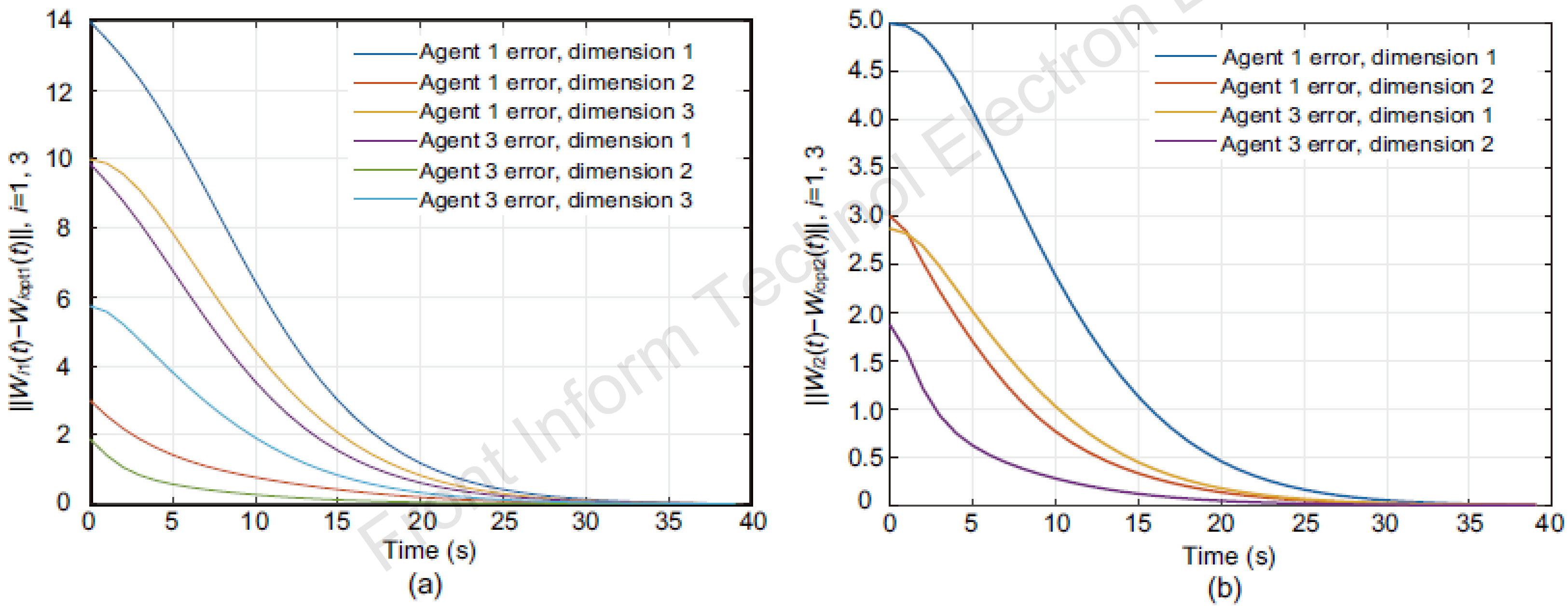
The global Nash equilibrium is guaranteed as

$$\begin{aligned} &\bar{J}_i^*(\xi_i, \xi_{-i}, u_i, u_{-i}^*) = \bar{J}_i^*(\xi_i, \xi_{-i}, u_i^*, u_{-i}^*) \\ &+ \frac{1}{2}\int_0^{\infty} d_i (u_i - u_i^*)^{\mathrm{T}} R_{ii} \left(u_i - u_i^{\#}\right) \mathrm{d}t \\ &\geq \bar{J}_i^*(\xi_i, \xi_{-i}, u_i^*, u_{-i}^*). \end{aligned}$$

The off-policy RL for the coupled Nash equilibrium can be derived in a similar way in the local best response case.

Major results (Cont’d)

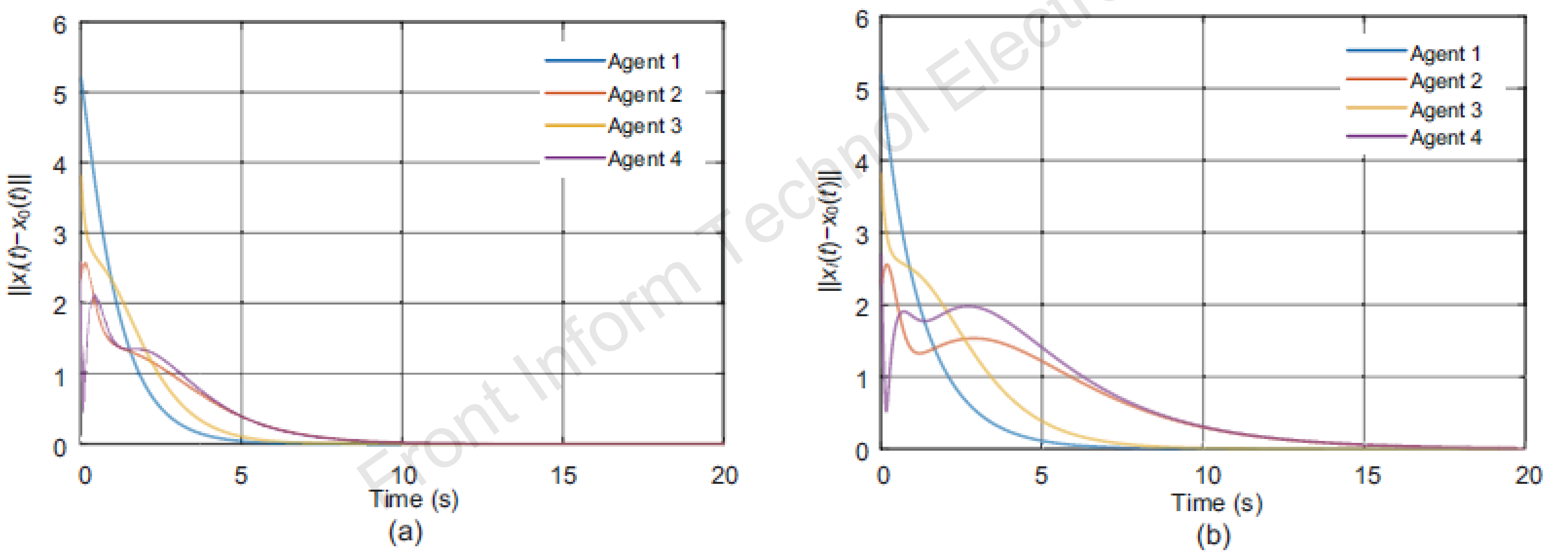
3. Global Nash solution using RL



Convergence error curves of neural network (NN) weights in the global Nash case: (a) value function weight error; (b) controller weight error

Major results (Cont'd)

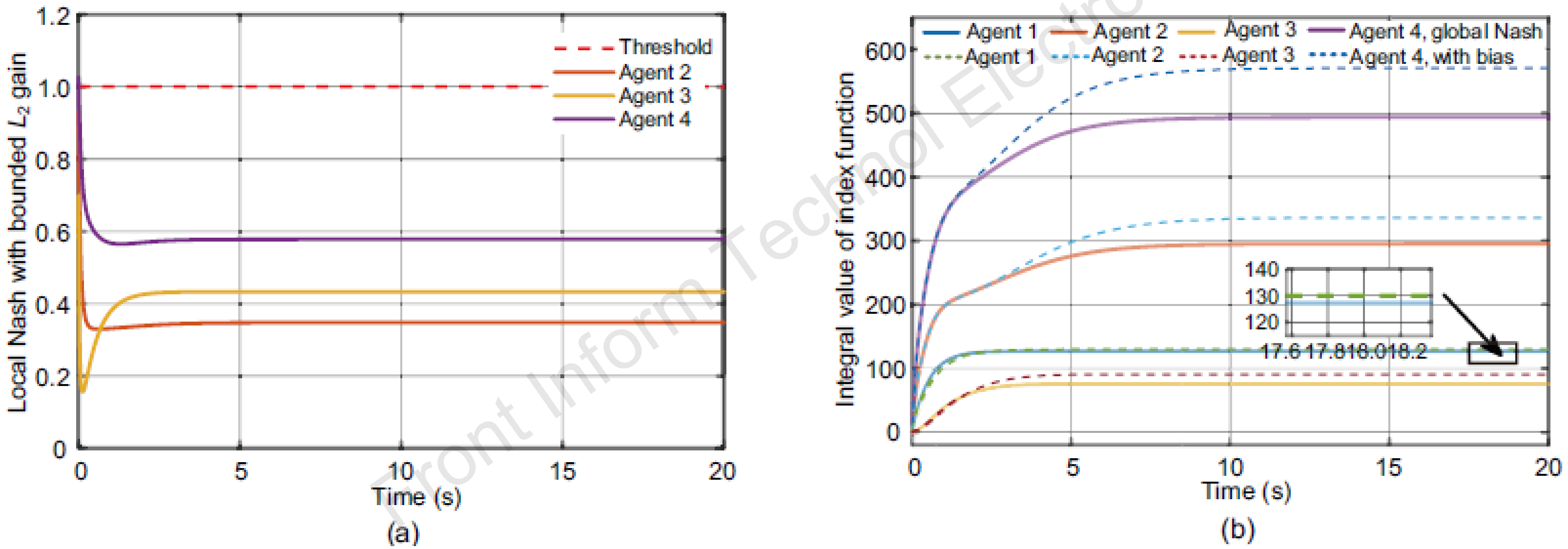
4. Stability verification to the synchronization control



Multi-agent system (MAS) synchronization errors in the local best response case (a) and the global Nash case (b)

Major results (Cont'd)

4. Stability verification to the synchronization control



Evolution of index functions in the local best response case (a) and the global Nash case (b)

Conclusions

1. The cooperative synchronization control problem of MASs was solved from a differential game perspective.

2. The local best response and global Nash controller with modified index functions were investigated to deal with the coupling issues.

3. An off-policy RL method was proposed to solve the problem online in a data-driven manner.

Yu SHI received his BE degree in aircraft design and engineering from Beihang University, Beijing, China, in 2018, where he is currently pursuing his PhD degree with the School of Automation Science and Electrical Engineering. His current research interests include formation control of multi-agent systems and reinforcement learning.

Yongzhao HUA received his BE and PhD degrees in navigation, guidance and control from Beihang University, Beijing, China, in 2014 and 2019, respectively. He is currently an associate professor with the Institute of Artificial Intelligence, Beihang University. His current research interests include distributed control, optimization, and game for multi-agent systems.

Jianglong YU received his BE and PhD degrees in navigation, guidance, and control from Beihang University, Beijing, China, in 2015 and 2020, respectively. He is currently a postdoctoral researcher with the School of Automation Science and Electrical Engineering, Beihang University. His current research interests include nonlinear control, robust control, and formation control of multi-agent systems.

Xiwang DONG received his BE degree in automation from Chongqing University, Chongqing, China, in 2009, and his PhD degree in control science and engineering from Tsinghua University, Beijing, China, in 2014. He is currently a professor and a Deputy Dean in the Institute of Artificial Intelligence, Beihang University. His research interests include consensus control, formation control, and containment control of swarm systems with applications to UAV swarm systems.

Zhang REN received his BE, ME, and PhD degrees in aircraft guidance, navigation, and simulation from Northwestern Polytechnical University, Xi’an, China, in 1982, 1985, and 1994, respectively. He is now a professor with the School of Automation Science and Electrical Engineering, Beihang University, Beijing, China. His research interests include aircraft guidance, navigation and control, fault detection, isolation and recovery, and cooperative control of multi-agent systems.