Note that according to Assumption 1, these residual errors are bounded on the compact set $\Omega$, i.e. $Sup_{x\in\Omega}\|e_{B_i}\| \le \bar{e}_i$ , $i = 1, ..., N$.

The following assumption for the drift and control dynamics is needed for every agent.

**Assumption 3.** *For a given compact set $\Omega \subset \Re^n$ and $i = 1, ..., N$:*

(a) $f_i(x_i) \le b_f\|x_i\|$
(b) *$g_i(x_i)$ is bounded by constant: $\|g_i(x_i)\| \le b_{g_i}$.*
(c) *The critic NNs weights are bounded by known constants $\|W_i\| < W_{i\max}$.*

The ideal weights of the critic NNs, i.e. $W_i, i = 1, ..., N$, which provide the solution to (15) are unknown and must be approximated in real time. Therefore, the output of the critic NNs $\hat{V}_i$ and the approximate Bellman equations (18) can be written as

$$\hat{V}_i = \hat{W}_i^T \sigma_i(\delta_i) \tag{17}$$

$$\begin{aligned} e_{H_i} &= \hat{W}_i^T \nabla\sigma_i[\sum_{j\in N_i} e_{ij}(f_i(x_i) - f_j(x_j)) \\ &+ e_{i0}(f_i(x_i) - f_0(x_0)) + (d_i + e_{i0})g_i(x_i)u_i \\ &- \sum_{j\in N_i} e_{ij}g_j(x_j)u_j] + \frac{1}{2}Q_i(\delta_i) + \frac{1}{2}u_i^T R_{ii}u_i + \frac{1}{2}\sum_{j\in N_i} u_j^T R_{ij}u_j \end{aligned} \tag{18}$$

where $\hat{W}_i \in \Re^{K_i}$ is the current estimated value of the ideal weight $W_i \in \Re^{K_i}$ for every agent.

It is desired to select $\hat{W}_i$ to minimize the following squared residual error:

$$E_i = \frac{1}{2}e_{H_i}^T e_{H_i} \tag{19}$$

Hence, we shall select the tuning law for the critic weights as the following normalized gradient descent algorithm:

$$\begin{aligned} \dot{\hat{W}}_i &= -\alpha_i \frac{\partial E_i}{\partial \hat{W}_i} = -\alpha_i \frac{B_i}{(1 + B_i^T B_i)^2} e_{H_i} = -\alpha_i \frac{\bar{B}_i}{m_{s_i}} e_{H_i} \\ &= -\alpha_i \frac{\bar{B}_i}{m_{s_i}}(B_i^T \hat{\mathrm{W}}_i + \frac{1}{2}Q_i(\delta_i) + \frac{1}{2}u_i^T R_{ii}u_i \\ &+ \frac{1}{2}\sum_{j\in N_i} u_j^T R_{ij}u_j) , \quad i = 1, ..., N \end{aligned} \tag{20}$$

where $B_i = \nabla\sigma_i(\sum_{j\in N_i} e_{ij}(f_i(x_i) - f_j(x_j)) + e_{i0}(f_i(x_i) - f_0(x_0)) + (d_i + e_{i0})g_i(x_i)u_i - \sum_{j\in N_i} e_{ij}g_j(x_j)u_j)$, $m_{s_i} = 1 + B_i^T B_i$, $\bar{B}_i = \frac{B_i}{1 + B_i^T B_i}$. $\alpha_i > 0$ is the learning rate and $(1 + B_i^T B_i)^2$ is used for normalization.

The following definition, is needed before we proceed to Lemma 1.

**Definition 1** (Persistence of Excitation (PE)). The bounded vector signal $\bar{B}_i(t)$, $i = 1, ..., N$ is PE (Zhang et al., 2012a) over the interval $[t, t + T_i]$ if there exists $T_i > 0$, $\gamma_i > 0$ and $\gamma_{i+N} > 0$ such that for all $t$:

$$\gamma_i I \le \int_t^{t+T_i} \bar{B}_i(\tau)\bar{B}_i^T(\tau)d\tau \le \gamma_{i+N} I, \ i = 1, \ldots, N$$

**Lemma 1.** *Let $(u_i, u_{N_i})$, $\forall i$ be a given admissible feedback policy set, let (20) be the tuning of the critic NNs and assume that $\bar{B}_i$ is PE. Then the critic parameter errors converge exponentially to the residual set $\|\tilde{W}_i\| \le \eta_{i_1} e^{-\eta_{i_2} t} + \frac{\alpha_i}{m_{s_i}\eta_{i2}}\bar{e}_i$, $i = 1, ..., N$ for some $\eta_{i_1}, \eta_{i_2} > 0$.*

**Proof.** From the coupled HJ equations we have

$$\begin{aligned} &- W_i^T \nabla\sigma_i[\sum_{j\in N_i} e_{ij}(f_i(x_i) - f_j(x_j)) + e_{i0}(f_i(x_i) - f_0(x_0)) \\ &+ (d_i + e_{i0})g_i(x_i)u_i - \sum_{j\in N_i} e_{ij}g_j(x_j)u_j] \\ &+ e_{B_i} = \frac{1}{2}Q_i(\delta_i) + \frac{1}{2}u_i^T R_{ii}u_i + \frac{1}{2}\sum_{j\in N_i} u_j^T R_{ij}u_j. \end{aligned} \tag{21}$$

Now, substituting (21) into (18) and doing some simple algebra, we can write

$$e_{H_i} = -\tilde{W}_i^T B_i + \mathrm{e}_{B_i} \tag{22}$$

Substituting (22) into (20), then the dynamics of the critic weight error dynamics for every agent becomes

$$\dot{\tilde{W}}_i = -\dot{\hat{W}}_i = -\alpha_i \bar{B}_i \bar{B}_i^T \tilde{W}_i + \alpha_i \frac{\bar{B}_i}{m_{s_i}} e_{B_i} \tag{23}$$

Assuming that (23) is a linear time-varying system with an input given by $e_{B_i}$, $i = 1, ..., N$, then the closed-form solution $\tilde{W}_i$ is given as

$$\tilde{W}_i(t) = \phi_i(t, t_0)\, \tilde{W}_i(0) + \alpha_i \int_{t_0}^{t} \phi_i(\tau, t_0)\, \frac{\bar{B}_i}{m_{s_i}} e_{B_i}\, d\tau \tag{24}$$

where the state transition matrix can be found from

$$\frac{\partial\phi_i(t, t_0)}{\partial t} = -\alpha_i \bar{B}_i \bar{B}_i^T \phi_i(t, t_0), \quad \phi_i(t_0, t_0) = I, \ \ i = 1, ..., N \tag{25}$$

The state transition matrix $\phi_i, i = 1, ..., N$ is exponentially stable provided that $\bar{B}_i$ is PE (Ioannou and Sun, 1996). As $\bar{B}_i$ is PE and the fact that $\|\bar{B}_i\| \le 1$ and $Sup_{x\in\Omega}\|e_{B_i}\| \le \bar{e}_i$ , $i = 1, ..., N$, we finally obtain

$$\|\tilde{W}_i\| \le \eta_{i_1} e^{-\eta_{i_2} t} + \frac{\alpha_i}{m_{s_i}\eta_{i2}}\bar{e}_i, \quad i = 1, ..., N \tag{26}$$

for some $\eta_{i_1}, \eta_{i_2} > 0$.

This completes the proof.

## Control policy approximation by actor neural networks

We shall now define the optimal control policy by using the value function approximation (13)–(14), as follows:

$$u_i = -(d_i + e_{i0})R_{ii}^{-1}g_i^T(x_i)(\nabla\sigma_i^T W_i + \nabla\omega_i), \quad i = 1, ..., N \tag{27}$$

However, as (27) is unknown, we can approximate the control policy with an actor NN as

$$u_i = W_{u_i}^T\sigma_{u_i}(\delta_i) + \omega_{u_i}(\delta_i), \quad i = 1, ..., N \tag{28}$$

and

$$\hat{u}_i = \hat{W}_{u_i}^T\sigma_{u_i}(\delta_i), \quad i = 1, ..., N \tag{29}$$

where $\hat{W}_{u_i} \in \Re^{L\times m}$ is the current estimated value of the ideal NN weight $W_{u_i} \in \Re^{L\times m}$, $\sigma_{u_i}(\delta_i)$ are the activation functions and $L$ is the number of neurons in the hidden layer.

Define the critic and the actor NNs estimation errors respectively as

$$\tilde{W}_i = W_i - \hat{W}_i \tag{30}$$

$$\tilde{W}_{u_i} = W_{u_i} - \hat{W}_{u_i} \tag{31}$$

In order to tune the actor NNs, we shall define the following error signal $e_{u_i} \in \Re^m$:

$$e_{u_i} = \hat{W}_{u_i}^T\sigma_{u_i}(\delta_i) + (d_i + e_{i0})R_{ii}^{-1}g_i^T(x_i)\nabla\sigma_i^T\hat{W}_i \tag{32}$$

Now, the objective function to be minimized is given as

$$E_{e_{u_i}} = \frac{1}{2}e_{u_i}^T e_{u_i} \tag{33}$$

Hence, the weight update law for the actor NN can be found by using a normalized gradient descent algorithm, as follows

$$\dot{\hat{W}}_{u_i} = -\alpha_{u_i}\frac{\bar{\sigma}_{u_i}(\delta_i)}{m_{s_{u_i}}}\left[\sigma_{u_i}(\delta_i)^T\hat{W}_{u_i} + (d_i + e_{i0})R_{ii}^{-1}g_i^T(x_i)\frac{\partial\sigma_i}{\partial\delta_i}^T\hat{W}_i\right] \tag{34}$$

where $m_{s_{u_i}} = 1 + \sigma_{u_i}(\delta_i)^T\sigma_{u_i}(\delta_i), \bar{\sigma}_{u_i} = \dfrac{\sigma_{u_i}(\delta_i)}{1 + \sigma_{u_i}(\delta_i)^T\sigma_{u_i}(\delta_i)}$, $i = 1, ..., N$.

### *Stability and convergence analysis*

The critic and the actor NNs' tuning laws are designed to ensure global synchronization, closed-loop system stability and convergence of the policies to a Nash equilibrium.

**Theorem 2.** *Consider the dynamical system (5) and the multiplayer graphical game formulation. Let the critic NN of each agent be given by (17) and the corresponding control input be given by (29). Consider that the tuning for agent i critic NN is given by (20) and the corresponding actor NN is tuned by (34). Let Assumptions 1–3 hold. Then the closed-loop system states* $\delta_i(t)$, *the critic NN errors* $\tilde{W}_i$, *the actor NN errors* $\tilde{W}_{u_i}$ *are UUB (Uniformly Ultimately Bounded), for a sufficiently large number of NN neurons.*

**Proof.** See Appendix.

## Simulation results

Consider a five-node MAS, as shown in Figure 1 with dynamics given by $\dot{x}_i = f_i(x_i) + g_i(x_i)u_i$, $x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix}$, $i = 1, 2, ..., 5$,

where

$$f_i(x_i) = \begin{pmatrix} x_{i2} \\ -x_{i1} + \varepsilon(1 - x_{i1}^2)x_{i2} \end{pmatrix}, g_1(x_1) = \begin{bmatrix} 0 \\ -0.8x_{11}x_{12} \end{bmatrix},$$

$$g_2(x_2) = \begin{bmatrix} 0 \\ x_{21}x_{22} \end{bmatrix}, g_3(x_3) = \begin{bmatrix} 0 \\ 0.5x_{31}x_{32} \end{bmatrix},$$

$$g_4(x_4) = \begin{bmatrix} 0 \\ -0.2x_{41}x_{42} \end{bmatrix}, g_5(x_5) = \begin{bmatrix} 0 \\ 1.4x_{51}x_{52} \end{bmatrix},$$

with the pinning gains and the edge weights chosen to be 1, $\varepsilon = 0.5$ and the leader node dynamics is $f(x_0) = \begin{pmatrix} x_{02} \\ -x_{01} + \varepsilon(1 - x_{01}^2)x_{02} \end{pmatrix}$.

For $i, j = 1, 2, ..., 5$, we pick $Q_i(\delta_i) = \delta_i^T Q_{ii}\delta_i = \delta_i^T \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \delta_i$, $R_{ii} = 10$, $R_{ij} = 1$, $(i \neq j,\ j \in N_i)$.

and the tuning gains are picked as $\alpha_{u_i} = 1$ and $\alpha_i = 5$, $i = 1, ..., 5$.

The available information vector for each agent, $\delta_i = \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \end{bmatrix}, i = 1, ..., 5$, is restricted by the graph topology. The critic NN activation functions are chosen as follows

$$\begin{aligned}
\sigma_1 &= [\delta_{11}^2, \delta_{11}\delta_{12}, \delta_{12}^2, \delta_{11}^3, \delta_{11}^2\delta_{12}, \delta_{11}\delta_{12}^2, \delta_{12}^3, \delta_{11}^4, \delta_{11}^3\delta_{12}, \delta_{11}^2\delta_{12}^2, \delta_{11}\delta_{12}^3, \delta_{12}^4]\\
\sigma_2 &= [\delta_{21}^2, \delta_{21}\delta_{22}, \delta_{22}^2, \delta_{21}^3, \delta_{21}^2\delta_{22}, \delta_{21}\delta_{22}^2, \delta_{22}^3, \delta_{21}^4, \delta_{21}^3\delta_{22}, \delta_{21}^2\delta_{22}^2, \delta_{21}\delta_{22}^3, \delta_{22}^4],\\
\sigma_3 &= [\delta_{31}^2, \delta_{31}\delta_{32}, \delta_{32}^2, \delta_{31}^3, \delta_{31}^2\delta_{32}, \delta_{31}\delta_{32}^2, \delta_{32}^3, \delta_{31}^4, \delta_{31}^3\delta_{32}, \delta_{31}^2\delta_{32}^2, \delta_{31}\delta_{32}^3, \delta_{32}^4],\\
\sigma_4 &= [\delta_{41}^2, \delta_{41}\delta_{42}, \delta_{42}^2, \delta_{41}^3, \delta_{41}^2\delta_{42}, \delta_{41}\delta_{42}^2, \delta_{42}^3, \delta_{41}^4, \delta_{41}^3\delta_{42}, \delta_{41}^2\delta_{42}^2, \delta_{41}\delta_{42}^3, \delta_{42}^4],\\
\sigma_5 &= [\delta_{51}^2, \delta_{51}\delta_{52}, \delta_{52}^2, \delta_{51}^3, \delta_{51}^2\delta_{52}, \delta_{51}\delta_{52}^2, \delta_{52}^3, \delta_{51}^4, \delta_{51}^3\delta_{52}, \delta_{51}^2\delta_{52}^2, \delta_{51}\delta_{52}^3, \delta_{52}^4],
\end{aligned}$$

and the actor NN activation functions are $\sigma_{u_i} = \nabla\sigma_i$, $i = 1, \ldots, N$.

It should be noted that a small exponential decreasing probing noise is added to the control inputs to ensure excitation. Figures 2 and 3 show the convergence of all the agents' critic weights after using the proposed learning algorithm.

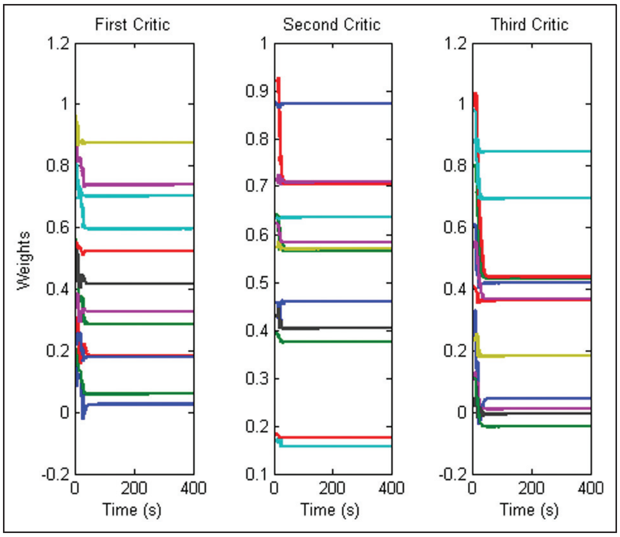


**Figure 2.** Convergence of the critic weights for the first, second and third agents.

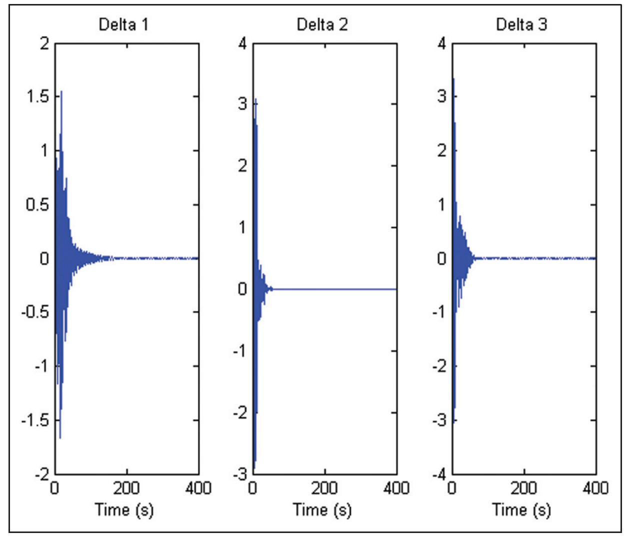


**Figure 4.** Tracking error of first, second and third agents.

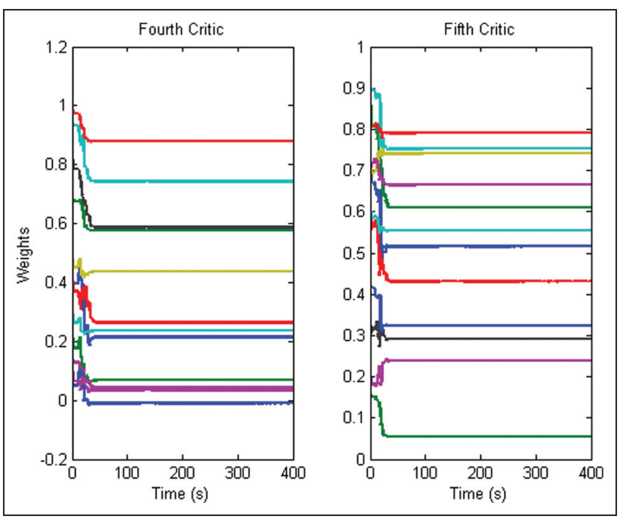


**Figure 3.** Convergence of the critic weights for the fourth and fifth agents.

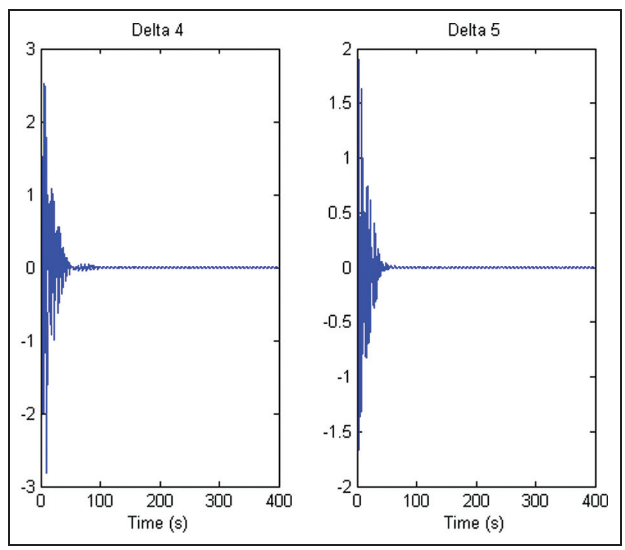


**Figure 5.** Tracking error of fourth and fifth agents.

Figures 4 and 5 show the evolution of the local tracking errors, which approximately converge to zero.

## Conclusion and future work

This paper integrates the abilities of optimal adaptive control, differential games and non-linear MASs to introduce a formulation of leader–follower non-linear differential graphical games. An online distributed optimal adaptive control algorithm based on RL techniques is presented to solve the continuous-time multi-agent non-linear differential graphical games. Each agent uses a critic and an actor NN to learn online respectively the optimal value and optimal control policy. The closed-loop signals are proved to be bounded according to Lyapunov stability and the policies form a Nash equilibrium.

For future work, we intend to extend the approach of this paper to distributed control of non-linear differential graphical games under unknown dynamics.

## Acknowledgement

The authors would like to thank Prof. Frank Lewis for his helpful suggestions and guidance.

## Conflict of interest

The authors declare that there is no conflict of interest.

## Funding

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

## References

Abdessameud A and Tayebi A (2009) attitude synchronization of a group of spacecraft without velocity measurements. *IEEE Transactions on Automatic Control* 54(11): 2642–2648.

Abu-Khalaf M and Lewis FL (2005) Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica* 41: 779–791.

Adams R and Fournier J (2003) *Sobolev Spaces*. New York: Academic Press.

Barto AG, Sutton RS and Anderson CW (1983) Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, and Cybernetics* 13: 834–846.

Cao Y and Ren W (2011) Finite-time consensus for second order multi-agent networks with inherent nonlinear dynamics under an undirected fixed graph. In: *50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, FL.

Chung S-J, Bandyopadhyay S, Chang I, et al. (2013) Phase synchronization control of complex networks of Lagrangian systems on adaptive digraphs. *Automatica* 49(5): 1148–1161.

Cui R and Yan W (2012) Mutual synchronization of multiple robot manipulators with unknown dynamics. *Journal of Intelligent Robot Systems* 68: 105–119.

Defoort M, Floquet T, Kokosy A, et al. (2008) Sliding-mode formation control for cooperative autonomous mobile robots. *IEEE Transactions on Industrial Electronics* 55(11): 3944–3953.

Esparza LG, Torres GM and Saynes Torres LM (2013). A brief introduction to differential games. *International Journal of Physical and Mathematical Sciences* 4(1): 396–411.

Finlayson BA (1990) *The Method of Weighted Residuals and Variational Principles*. New York: Academic Press.

Hong Y, Hu J and Gao L (2006) Tracking control for multi-agent consensus with an active leader and variable topology. *Automatica* 42(7): 1177–1182.

Hornik K, Stinchcombe M and White H (1990) Universal approximation of an unknown mapping and its derivatives using multi layer feedforward networks. *Neural Networks* 3(5): 551–560.

Ioannou P and Fidan B (2006) *Advances in Design and Control*. *Adaptive Control Tutorial*. Philadelphia, PA: SIAM.

Ioannou P and Sun J (1996) *Robust Adaptive Control*. Englewood Cliffs, NJ: Prentice Hall.

Isaacs R (1965) *Differential Games*. New York: John Wiley & Sons.

Kamalapurkar R, Dinh H, Walters P, et al. (2013) Approximate optimal cooperative decentralized control for consensus in a topological network of agents with uncertain nonlinear dynamics. In: *American Control Conference (ACC)*, Washington, DC.

Khalil HK (1996) *Nonlinear Systems*. Englewood Cliffs, NJ: Prentice Hall.

Lewis FL and Vrabie D (2009) Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits and Systems Magazine* 9(3): 32–50.

Li Z, Ren W, Liu X, et al. (2013) Consensus of multi-agent systems with general linear and Lipschitz nonlinear dynamics using distributed adaptive protocols. *IEEE Transactions on Automatic Control* 58(7): 1786–1791.

Li X, Wang X and Chen G (2004) Pinning a complex dynamical network to its equilibrium. *IEEE Transactions on Circuits and Systems* 51(10): 2074–2087.

Lin W (2014) Distributed UAV formation control using differential game approach. *Aerospace Science and Technology* 35: 54–62.

Liu K, Xie G, Ren W, et al. (2013) Consensus for multi-agent systems with inherent nonlinear dynamics under directed topologies. *Systems and Control Letters* 62: 152–162.

Mao D, He Y, Ye X, et al. (2011) Inverse optimal stabilization of cooperative control in networked multi-agent systems. In: *Control and Decision Conference (CCDC)*, pp. 1031–1037.

Mei J, Ren W and Ma G (2012) Cooperative control of nonlinear multi-agent system with only relative position measurements. In: *American Control Conference*, Fairmont Queen Elizabeth, Montreal, Canada.

Meng Z, Lin Z and Ren W (2013) Robust cooperative tracking for multiple non-identical second-order nonlinear systems. *Automatica* 49: 2363–2372.

Murray JJ, Cox CJ, Lendaris GG, et al. (2002) Adaptive dynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics* 32 (2): 140–153.

Powell W (2007) *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. New York: John Wiley & Sons.

Ren W (2007) Distributed attitude alignment in spacecraft formation flying. *International Journals of Adaptive Control and Signal Processing* 21: 95–113.

Ren W, Beard R and Atkins E (2007a) Information consensus in multi vehicle cooperative control. *IEEE Control Systems* 27(2): 71–82.

Ren W, Moore K and Chen Y (2007b) High-order and model reference consensus algorithms in cooperative control of multi vehicle systems. *Journal of Dynamic Systems, Measurement, and Control* 129(5): 678–688.

Rong L, Xu S and Zhang B (2012) On the general second-order consensus protocol in multi-agent systems with input delays. *Transactions of the Institute of Measurement and Control* 34(8): 983–989.

Sastry S and Bodson M (1989) *Adaptive Control: Stability, Convergence, and Robustness*. Englewood Cliffs, NJ: Prentice Hall.

Semsar-Kazerooni E and Khorasani K (2009) Multi-agent team cooperation: a game theory approach. *Automatica* 45: 2205–2213.

Shi G, Johansson KH and Hong Y (2011) Multi-agent systems reaching optimal consensus with directed communication graphs. In: *Proceedings of the American Control Conference*.

Slotine JE and Li W (1991) *Applied Nonlinear Control*. Englewood Cliffs, NJ: Prentice Hall.

Starr A and Ho Y (1969) Nonzero-sum differential games. *Journal of Optimization Theory and Applications* 3(3): 184–206.

Sutton RS and Barto AG (1998) *Reinforcement Learning – an Introduction*. Cambridge, MA: MIT Press.

Tan LN, Thien TN and Minh TH (2014) Reinforcement learning-based intelligent tracking control for wheeled mobile robot. *Transactions of the Institute of Measurement and Control*, first published on March10, 2014.

Tang Z (2014) Leader-following consensus with directed switching topologies. *Transactions of the Institute of Measurement and Control*, first published on July 9, 2014.

Tijs S (2003) *Introduction to Game Theory*. New Delhi: Hindustan Book Agency.

Tolwinski B, Havrie A and Leimann G (1986) Cooperative equilibrium in differential games. *Journal of Mathematical Analysis and Applications* 119: 182–202.

Vamvoudakis KG and Lewis FL (2010) Online actor–critic algorithm to solve the continuous infinite-time horizon optimal control problem. *Automatica* 46: 878–888.

Vamvoudakis KG and Lewis FL (2011) Multi-player non-zero-sum games: online adaptive learning solution of coupled Hamilton–Jacobi equations. *Automatica* 47(8): 1556–1569.

Vamvoudakis KG, Lewis FL and Hudas GR (2012) Multi-agent differential graphical games: online adaptive learning solution for synchronization with optimality. *Automatica* 48: 1598–1611.

Vamvoudakis KG and Lewis FL (2012) An online integral reinforcement learning algorithm to solve N-player Nash games. In: *IEEE International Symposium on Intelligent Control*, Dubrovnik, Croatia.

Vrabie D and Lewis FL (2010) Integral reinforcement learning for online computation of feedback Nash strategies of nonzero-sum differential games. In: *49th IEEE Conference on Decision and Control*, Atlanta, GA.

Vrabie D, Lewis F and Abu-Khalaf M (2008) Biologically inspired scheme for continuous-time approximate dynamic programming. *Transactions of the Institute of Measurement and Control* 30(3–4): 207–223.

Wang X and Chen G (2002) Pinning control of scale-free dynamical net-works. *Physica A* 310(3–4): 521–531.

Wang J and Xin M (2012) Distributed optimal cooperative tracking control of multiple autonomous robots. *Robotics and Autonomous Systems* 60: 572–583.

Wen G, Rahmani A and Yu Y (2011) Consensus tracking for multi-agent systems with nonlinear dynamics under fixed communication topologies. In: *Proceedings of the World Congress on Engineering and Computer Science*, 19–21 October, San Francisco, CA.

Werbos PJ (1992) Approximate dynamic programming for real time control and neural modeling. In White DA and Sofge DA (eds) *Handbook of Intelligent Control*. New York: Multiscience Press.

Xie D, Yuan D, Lu J, et al. (2013) Consensus control of second-order leader–follower multi-agent systems with event-triggered strategy. *Transactions of the Institute of Measurement and Control* 35(4): 426–436.

Xie D and Chen J (2013) Consensus problem of data-sampled networked multi-agent systems with time-varying communication delays. *Transactions of the Institute of Measurement and Control* 35(6): 753–763.

Zhang H, Cui L and Luo Y (2012a) Near-optimal control for nonzero-sum differential games of continuous-time nonlinear systems using single-network ADP. *IEEE Transactions on Systems, Man, and Cybernetics* 43(1): 206–216.

Zhang H, Lewis F and Qu Z (2012b) Lyapunov, adaptive, and optimal design techniques for cooperative systems on directed communication graphs. *IEEE Transactions on Industrial Electronics* 59: 3026–3041.

Zhuand W and Cheng D (2010) Leader-following consensus of second-order agents with multiple time-varying delays. *Automatica* 46(12): 1994–1999.


## Appendix

### *Proof of Theorem 1*

Consider the following Lyapunov function

$$L(t) = \sum_{i=1}^{N}\{\,V_i(t) + \gamma_i(\frac{1}{2}\tilde{W}_i^T\alpha_i^{-1}\tilde{W}_i) + \lambda_i(\frac{1}{2}\tilde{W}_{u_i}^T\alpha_{u_i}^{-1}\tilde{W}_{u_i})\,\} \tag{A.1}$$

where $V_i(\mathrm{t}) = W_i^T\sigma_i + \omega_i, i = 1, \ldots, N$ are the approximate solutions to (12).

The time derivative of the Lyapunov function (A.1) is given by

$$\dot{L}(t) = \sum_{i=1}^{N}\{\,\dot{V}_i(t) - \gamma_i(\tilde{W}_i^T\alpha_i^{-1}\dot{\hat{W}}_i) - \lambda_i(\tilde{W}_{u_i}^T\alpha_{u_i}^{-1}\dot{\hat{W}}_{u_i})\,\} \tag{A.2}$$

Using (7), the first term of (A.2) is written as

$$\sum_{i=1}^{N}\dot{V}_i(t) = \sum_{i=1}^{N} -\frac{1}{2}\{Q_i(\delta_i) + \hat{u}_i^T R_{ii}\hat{u}_i + \sum_{j\in N_i}\hat{u}_j^T R_{ij}\hat{u}_j\} \tag{A.3}$$

We shall note that after, using (29)–(31), $\dot{V}_i$ can be written as

$$\begin{aligned}\sum_{i=1}^{N}\dot{V}_i(t) &= \sum_{i=1}^{N} -\frac{1}{2}\{Q_i(\delta_i) + \sigma_{u_i}^T\hat{W}_{u_i}R_{ii}\hat{W}_{u_i}^T\sigma_{u_i}\\
&\quad + \sum_{j\in N_i}\sigma_{u_j}^T\hat{W}_{u_j}R_{ij}\hat{W}_{u_j}^T\sigma_{u_j}\} = \sum_{i=1}^{N}\{-\frac{1}{2}Q_i(\delta_i)\\
&\quad -\frac{1}{2}\sigma_{u_i}^T(W_{u_i}-\tilde{W}_{u_i})R_{ii}(W_{u_i}-\tilde{W}_{u_i})^T\sigma_{u_i}\\
&\quad -\frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T(W_{u_j}-\tilde{W}_{u_j})R_{ij}(W_{u_j}-\tilde{W}_{u_j})^T\sigma_{u_j}\}\\
&= \sum_{i=1}^{N}\{\dot{\mathrm{L}}_{V_i} - \frac{1}{2}Q_i(\delta_i) - \frac{1}{2}\sigma_{u_i}^T\tilde{W}_{u_i}R_{ii}\tilde{W}_{u_i}^T\sigma_{u_i}\\
&\quad + \frac{1}{2}\sigma_{u_i}^T W_{u_i}R_{ii}\tilde{W}_{u_i}^T\sigma_{u_i} + \frac{1}{2}\sigma_{u_i}^T\tilde{W}_{u_i}R_{ii}W_{u_i}^T\sigma_{u_i}\\
&\quad -\frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T\tilde{W}_{u_j}R_{ij}\tilde{W}_{u_j}^T\sigma_{u_j} + \frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T W_{u_j}R_{ij}\tilde{W}_{u_j}^T\sigma_{u_j}\\
&\quad + \frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T\tilde{W}_{u_j}R_{ij}W_{u_j}^T\sigma_{u_j}\}\end{aligned} \tag{A.4}$$

where

$$\dot{\mathrm{L}}_{V_i} = -\frac{1}{2}\sigma_{u_i}^T W_{u_i}R_{ii}W_{u_i}^T\sigma_{u_i} - \frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T W_{u_j}R_{ij}W_{u_j}^T\sigma_{u_j}.$$

In order to simplify the second term of (A.2), we know that

$$\dot{\tilde{W}}_i = -\dot{\hat{W}}_i = -\alpha_i\bar{B}_i\bar{B}_i^T\tilde{W}_i + \alpha_i\frac{\bar{B}_i}{m_{s_i}}e_{B_i} \tag{A.5}$$

Then, accordingly for the summation of $\sum_{i=1}^{N}\dot{L}_i = -\sum_{i=1}^{N}\gamma_i(\tilde{W}_i^T\alpha_i^{-1}\dot{\hat{W}}_i)$ we can write:

$$\begin{aligned}\sum_{i=1}^{N}\dot{L}_i &= -\sum_{i=1}^{N}\gamma_i(\tilde{W}_i^T\alpha_i^{-1}\dot{\hat{W}}_i)\\
&= \sum_{i=1}^{N}\{-\gamma_i\tilde{W}_i^T\bar{B}_i\bar{B}_i^T\tilde{W}_i + \gamma_i\tilde{W}_i^T\frac{\bar{B}_i}{m_{s_i}}e_{B_i}\}\end{aligned} \tag{A.6}$$

Before we proceed to the third term of (A.2), we shall use (27) and (28) to write

$$W_{u_i}^T\sigma_{u_i}(\delta_i) + \omega_{u_i}(\delta_i) + (d_i + e_{i0})R_{ii}^{-1}g_i^T(x_i)(\nabla\sigma_i^T W_i + \nabla\omega_i) = 0 \tag{A.7}$$

Thus, we can write the actor error dynamics using (34) and (A.7) as

$$\dot{\tilde{W}}_{u_i} = -\dot{\hat{W}}_{u_i} = \alpha_{u_i}\frac{\bar{\sigma}_{u_i}(\delta_i)}{m_{s_{u_i}}}$$
$$[-\sigma_{u_i}(\delta_i)^T\tilde{W}_{u_i} - (d_i + e_{i0})R_{ii}^{-1}g_i^T(x_i)\frac{\partial\sigma_i}{\partial\delta_i}^T \tilde{W}_i + \omega_{i,u_i}]$$
$$= \alpha_{u_i}[-\bar{\sigma}_{u_i}\bar{\sigma}_{u_i}^T\tilde{W}_{u_i} - (d_i + e_{i0})\frac{\bar{\sigma}_{u_i}}{m_{s_{u_i}}}R_{ii}^{-1}g_i^T(x_i)\frac{\partial\sigma_i}{\partial\delta_i}^T \tilde{W}_i + \frac{\bar{\sigma}_{u_i}}{m_{s_{u_i}}}\omega_{i,u_i}] \tag{A.8}$$

where $\omega_{i,u_i} = -\omega_{u_i}(\delta_i) - (d_i + e_{i0})R_{ii}^{-1}g_i^T(x_i)\nabla\omega_i$.

The summation of the third term of (A.2) $\sum_{i=1}^{N}\dot{L}_{u_i} = -\sum_{i=1}^{N}\lambda_i(\tilde{W}_{u_i}^T\alpha_{u_i}^{-1}\dot{\hat{W}}_{u_i})$ can be written as

$$\sum_{i=1}^{N}\dot{L}_{u_i} = -\sum_{i=1}^{N}\lambda_i(\tilde{W}_{u_i}^T\alpha_{u_i}^{-1}\dot{\hat{W}}_{u_i}) = \sum_{i=1}^{N}$$
$$\left\{-\lambda_i\tilde{W}_{u_i}^T\bar{\sigma}_{u_i}\bar{\sigma}_{u_i}^T\tilde{W}_{u_i} - (d_i + e_{i0})\lambda_i\tilde{W}_{u_i}^T\frac{\bar{\sigma}_{u_i}}{m_{s_{u_i}}}R_{ii}^{-1}g_i^T(x_i)\frac{\partial\sigma_i}{\partial\delta_i}^T \tilde{W}_i + \lambda_i\tilde{W}_{u_i}^T\frac{\bar{\sigma}_{u_i}}{m_{s_{u_i}}}\omega_{i,u_i}\right\} \tag{A.9}$$

Now for the total Lyapunov function, we shall use (A.4), (A.6) and (A.9) to write (A.2) as

$$\begin{aligned}\dot{L}(t) = \sum_{i=1}^{N}\{&\dot{\mathrm{L}}_{V_i} - \frac{1}{2}Q_i(\delta_i) - \frac{1}{2}\sigma_{u_i}^T\tilde{W}_{u_i}R_{ii}\tilde{W}_{u_i}^T\sigma_{u_i}\\
&+ \frac{1}{2}\sigma_{u_i}^T W_{u_i}R_{ii}\tilde{W}_{u_i}^T\sigma_{u_i}\\
&+ \frac{1}{2}\sigma_{u_i}^T\tilde{W}_{u_i}R_{ii}W_{u_i}^T\sigma_{u_i} - \frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T\tilde{W}_{u_j}R_{ij}\tilde{W}_{u_j}^T\sigma_{u_j}\\
&+ \frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T W_{u_j}R_{ij}\tilde{W}_{u_j}^T\sigma_{u_j}\\
&+ \frac{1}{2}\sum_{j\in N_i}\sigma_{u_j}^T\tilde{W}_{u_j}R_{ij}W_{u_j}^T\sigma_{u_j} - \gamma_i\tilde{W}_i^T\bar{B}_i\bar{B}_i^T\tilde{W}_i\\
&+ \gamma_i\tilde{W}_i^T\frac{\bar{B}_i}{m_{s_i}}e_{B_i} - \lambda_i\tilde{W}_{u_i}^T\bar{\sigma}_{u_i}\bar{\sigma}_{u_i}^T\tilde{W}_{u_i}\\
&- (d_i + e_{i0})\lambda_i\tilde{W}_{u_i}^T\frac{\bar{\sigma}_{u_i}}{m_{s_{u_i}}}R_{ii}^{-1}g_i^T(x_i)\frac{\partial\sigma_i}{\partial\delta_i}^T \tilde{W}_i\\
&+ \lambda_i\tilde{W}_{u_i}^T\frac{\bar{\sigma}_{u_i}}{m_{s_{u_i}}}\omega_{i,u_i}\}.\end{aligned} \tag{A.10}$$

As $Q_i(\delta_i) > 0,\ i = 1, ..., N$, there exists $q_i > 0,\ \forall i$ such that $\delta_i^T q_i\delta_i < Q_i(\delta_i)$, therefore $-\delta_i^T q_i\delta_i > -Q_i(\delta_i)$.

In order to write in a compact form, we shall define $\tilde{Z}_i = \left[\delta_i^T,\ \tilde{W}_i{}^T,\ (\sigma_{u_i}^T\tilde{W}_{u_i})^T,\ (\sigma_{u_j}^T\tilde{W}_{u_j})_{j\in N_i}^T\right]^T$, and hence (A.2) can be written as

$$\dot{L}(t) = \sum_{i=1}^{N}\{-\tilde{Z}_i^T M_i\tilde{Z}_i + D_i\tilde{Z}_i\} \tag{A.11}$$

where $M_i = \begin{bmatrix} m_{11}^i & m_{12}^i & m_{13}^i & m_{14}^i\\ m_{21}^i & m_{22}^i & m_{23}^i & m_{24}^i\\ m_{31}^i & m_{32}^i & m_{33}^i & m_{34}^i\\ m_{41}^i & m_{42}^i & m_{43}^i & m_{44}^i\end{bmatrix}, D_i = \left[d_1^i, d_2^i, d_3^i, d_4^i\right]$,
$m_{11}^i = \frac{1}{2}q_i$ , $m_{22}^i = +\gamma_i\bar{B}_i\bar{B}_i^T$ , $m_{13}^i = m_{14}^i = m_{41}^i{}^T = m_{31}^i{}^T = 0$, $m_{24}^i = m_{42}^i{}^T = 0$, $m_{34}^i = m_{43}^i{}^T = 0, m_{12}^i = m_{21}^i{}^T = 0, m_{23}^i = \frac{\lambda_i}{2m_{s_{u_i}}}$ $(d_i + e_{i0})\frac{\partial\sigma_i}{\partial\delta_i}g_i(\mathrm{x}_i)\mathrm{R}_{ii}^{-T} = m_{32}^T, m_{33}^i = \frac{1}{2}R_{ii} + \frac{\lambda_i}{m_{s_{u_i}}^2}$ , $m_{44}^i = \frac{1}{2}\sum_{j\in N_i}R_{ij}$, $d_1^i = 0$ , $d_2^i = \gamma_i\frac{\bar{B}_i^T}{m_{s_i}}e_{B_i}$ , $d_3^i = R_{ii}W_{u_i}^T\sigma_{u_i} + \frac{\lambda_i}{m_{s_{u_i}}^2}\omega_{i,u_i}, d_4^i = \sum_{j\in N_i}R_{ij}W_{u_j}^T\sigma_{u_j}, D_i \le D_{i\max}$

Let the parameters be chosen such that $M_i > 0$ with

$$M_i = \begin{bmatrix} M_{11}^i & 0 & 0\\ 0 & M_{22}^i & 0\\ 0 & 0 & M_{33}^i\end{bmatrix}, M_{11}^i = \left[m_{11}^i\right],$$
$$M_{22}^i = \begin{bmatrix} m_{22}^i & m_{23}^i\\ m_{32}^i & m_{33}^i\end{bmatrix}, M_{33}^i = \left[m_{44}^i\right] \tag{A.12}$$

where the following properties must hold:

(a) $M_{11}^i = \frac{1}{2}q_i > 0$.
(b) $M_{22}^i > 0$, which requires that $m_{33}^i > 0$ and $m_{22}^i - m_{23}^i m_{33}^i{}^{-1}m_{32}^i > 0$(Schur complement for $m_{33}^i$), which hold after proper selection of $\gamma_i$ and $\lambda_i$.
(c) $M_{33}^i = \frac{1}{2}\sum_{j\in N_i}R_{ij} > 0$.

Finally, (A.11) becomes

$$\dot{L} < \sum_{i=1}^{N}\{-\|\tilde{Z}_i\|^2\sigma_{\min}(M_i) + D_{i\max}\|\tilde{Z}_i\|\} \tag{A.13}$$

and it is negative as long as

$$\|\tilde{Z}_i\| > \frac{D_{i\max}}{\sigma_{\min}(M_i)} \equiv B_{\tilde{Z}_i} \tag{A.14}$$

It is shown that if (A.14) exceeds a certain bound, then $\dot{L}$ is negative. Therefore, according to the standard Lyapunov extension theorem the analysis above demonstrates that the states and the weights are UUB (Khalil, 1996). Condition (A.14) holds if the norm of any component of $\tilde{Z}_i$ exceeds the bound, i.e. specifically $\delta_i > B_{\tilde{Z}_i}$, $\tilde{W}_i > B_{\tilde{Z}_i}$, $\sigma_{u_i}^T\tilde{W}_{u_i} > B_{\tilde{Z}_i}$,$\sigma_{u_j}^T\tilde{W}_{u_j} > B_{\tilde{Z}_i}$. Note that according to Assumption 1, the actor NN activation functions $\sigma_{u_i}$ and $\sigma_{u_j}$ are also bounded.