Advances in Applied Mathematics 应用数学进展, 2022, 11(1), 158-179
Published Online January 2022 in Hans. http://www.hanspub.org/journal/aam
https://doi.org/10.12677/aam.2022.111022

Hans汉斯

基于零和博弈的部分未知线性离散系统多智能体分布式最优跟踪控制

熊天娇[1]，王朝立[2*]

[1]上海理工大学理学院，上海
[2]上海理工大学光电信息与计算机工程学院，上海

收稿日期：2021年12月13日；录用日期：2022年1月3日；发布日期：2022年1月19日

摘　要

本文考虑了具有外部扰动的不确定线性离散系统分布式最优跟踪控制问题。现有的研究要求系统动力学已知且未证明最优解就是纳什均衡解。由于控制策略和干扰之间的竞争关系，该问题首先转变为多智能体零和博弈。本文根据所提出的新性能指标，采用内外循环算法对哈密顿雅可比艾萨克斯(HJI)方程进行迭代求解，并验证了收敛性。此外，它表明该算法得到的最优解是零和博弈的纳什均衡解。本文进一步表明，每当系统不完全已知时，单层神经网络可用于近似实值函数，与现有的三层网络相比，这可以降低计算复杂性。最后，通过仿真验证了该方法的有效性。

关键词

零和博弈，L_2-增益，纳什均衡，线性离散系统

The Multiagent Distributed Optimal Tracking Control of Partially Unknown Linear Discrete Systems Based on Zero-Sum Games

Tianjiao Xiong[1], Chaoli Wang[2*]

[1]School of Science, University of Shanghai for Science and Technology, Shanghai
[2]School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai

Received: Dec. 13th, 2021; accepted: Jan. 3rd, 2022; published: Jan. 19th, 2022

*通讯作者。

文章引用: 熊天娇, 王朝立. 基于零和博弈的部分未知线性离散系统多智能体分布式最优跟踪控制[J]. 应用数学进展, 2022, 11(1): 158-179. DOI: 10.12677/aam.2022.111022

Abstract

The paper studies the distributed optimal tracking control problem by considering linear discrete systems with unknown disturbances. The existing research requires that the system dynamics are known and have not proved that the optimal solution is the Nash equilibrium. Such a problem is first transformed into a multiagent zero-sum game due to the competitive situation among inputs and disturbances. According to the proposed new performance index, the internal and external loop algorithm is adopted to solve the Hamilton Jacobi Isaacs (HJI) equations iteratively, and the convergence is also proven. In addition, it shows that the optimal solution obtained by the algorithm is the Nash equilibrium of the zero-sum game. This paper further shows that, whenever the system is not fully known, the single-layer neural network could be used to approximate the real value function, which can reduce the computational complexity compared with the prevalent three-layer networks. Finally, simulations are provided to show the effectiveness of the method.

Keywords

Zero-Sum Game, L_2-Gain, Nash Equilibrium, Linear Discrete Systems

1. 引言

近年来，多智能体系统控制因其在各种工程系统中的潜在应用而受到广泛关注，且分布式控制在无人机、移动机器人等领域也有很多应用[1]。目前，控制系统的结构越来越复杂，实际的工业系统大多数由多个控制器控制，每个控制器都是一个玩家，在这些玩家的合作和竞争之间作出权衡，可以看作是一场博弈。博弈论为多人博弈决策提供了一种有效的方法来获得最优策略[2] [3] [4] [5]。多玩家博弈可以分为零和博弈和非零和博弈两种，这两类研究结果目前在工程动力学、经济学和社会学上都有了广泛的研究[6] [7]。

分布式控制的目标是为每个智能体设计一个控制协议，仅依赖于本地邻居信息。在无领导者的协作调节器共识问题中，所有智能体根据其初始条件收敛到一个不可控的共同值。而在合作跟踪共识问题中，所有智能体同步到控制智能体状态或者是一个领导者[8] [9]。但是在上述论文所提到的算法中不存在最优性保证，则使用博弈论框架来克服这个问题。每个智能体为了找到最优策略，它们通过独立优化其性能指标，而最终得到结果就是纳什均衡解。为了找到纳什均衡解，必须求解耦合哈密顿雅可比(HJ)方程，该方程在线性二次型(LQR)下简化为耦合黎卡提方程[10]，这些耦合的 HJ 方程求解很困难且依赖于全局信息。况且智能体动力学系统还可能受到外界干扰，当考虑系统中存在干扰的情况，就是二人零和博弈问题，即控制策略和干扰策略是博弈双方，这需要求解 HJI 方程，但是 HJI 方程是非常难或不可能求解的。自适应动态规划(ADP)方法[11]-[21]被提出，它是处理各种最优控制问题的最有效方法之一，而后随着强化学习(RL) [22]和神经网络学科的发展，与执行－评价神经网络结合的在线迭代算法[23] [24]和值迭代[25]相继被提出来，在 RL 框架下[26]，为一类电压源逆变器设计了一种事件触发输出反馈控制器，解决了线性事件驱动 H_∞控制问题。随后，利用 RL 求解 H_∞控制问题[27]，获得连续时间非线性输入约束系统的鲁棒事件驱动控制。而这些是针对单个智能体，对本文中考虑的多智能体难以适用。在这些学习控

制方案中成功应用的两种技术是值函数逼近和 Q 函数[28]逼近。在多智能体框架下，[29] [30]研究了非零和博弈问题，并使用了与执行－评价神经网络结合的策略迭代算法进行求解，但是并没考虑到干扰的存在。[31] [32]给出了在系统动力学部分未知时，求解最优控制的有效方法，但是并没有给出最优解和纳什均衡关系的证明。

本文考虑的分布式控制跟踪问题，目标是为每个智能体设计一个仅依赖于邻居信息的控制协议。通过对一定数量的利益达成一致来保证所有智能体的同步行为，使得性能指标最优，跟踪误差 L_2 有界。本文的贡献主要有以下几个方面：第一，相比[29] [30]，都是线性系统，但是本文考虑了系统存在干扰的情况下，如何求得一组纳什均衡解使得性能指标最优，文献没有考虑干扰。第二，相比[30]，我们考虑干扰的同时使用单层网络进行逼近，而文献采用三层神经网络进行逼近，所以减小了计算的复杂度。第三，相比[31] [32]，没有给出纳什均衡的证明，本文给出了求得的最优解满足纳什均衡定义的不等式的证明。

2. 通信图与同步

在本节中，介绍了通信图，并给出了多智能体系统受外部干扰时的同步问题。

2.1. 图论

有向图 Gr 定义为 $Gr=(V,\zeta)$ 与 N 个非空有限集 $V=(v_1,\cdots,v_N)$ 和一组边 $\varsigma\subseteq V\times V$，定义连通矩阵 E，使得 $E=[a_{ij}]$，如果 $(v_j,v_i)\in\zeta$，则 $e_{ij}>0$，否则，$e_{ij}=0$。每个节点 v_i 的领域集合为 $N_i=\{v_j:(v_j,v_i)\in\zeta\}$。定义入度矩阵 D 为对角矩阵 $D=diag\{d_i\}$，$d_i=\sum_{j\in N_i}a_{ij}$ 为节点的加权入度，图拉普拉斯矩阵 L 定义为 $L=D-E$。

从节点 v_0 到节点 v_r 的有向路径定义为边 $v_0,v_1,\cdots,v_r$ 的序列 $(v_i,v_{i+1})\in\varsigma,i\in 0,1,\cdots,r-1$。对图中所有 $v_i,v_j\in V$，如果从 v_i 到 v_j 存在一个有向路径，则这个有向图强连接的，反之亦然。

2.2. 同步和跟踪误差动力学

考虑通信图 $Gr=(V,\zeta)$ 有 N 个智能体，每个智能体的局部动态由

$$x_i(k+1)=Ax_i(k)+B_iu_i(k)+D_iw_i(k) \tag{1}$$

其中 $x_i(k)\in R^n,u_i(k)\in R^{m_i},w_i\in R^{q_i}$ 分别是节点 i 的状态、控制输入和外部干扰。A,B_i,D_i 是具有适当维数的系统矩阵。为了便于表示，我们将 $x_i(k)$ 写成 x_{ik}，$u_i(k)$ 写成 u_{ik}，$w_i(k)$ 写成 w_{ik}。领导者节点的状态为 $x_{0k}\in R^n$，假设它满足动力学

$$x_{0(k+1)}=Ax_{0k}. \tag{2}$$

假设 1：领导者连接图中的一小部分节点。

目的是设计控制输入 u_{ik}，只使用来自邻居节点的信息，使所有智能体的状态同步到领导者的状态，也就是 $\lim_{k\to\infty}\|x_{ik}-x_{0k}\|=0,\forall i\in N$。

对于每个节点 i，定义局部邻域跟踪误差 $\varepsilon_{ik}\in R^n$ 为

$$\varepsilon_{ik}=\sum_{j\in N_i}a_{ij}(x_{ik}-x_{jk})+g_i(x_{ik}-x_{0k}) \tag{3}$$

其中 $g_i\geq 0$ 为节点 i 的固定增益。如果节点 i 耦合到控制节点 x_{0k}，则非零。

所有节点的总体跟踪误差向量为

$$\varepsilon_k=((L+G)\otimes I_n)x_k-((L+G)\otimes I_n)\underline{x_{0k}}=((L+G)\otimes I_n)(x_k-\underline{x_{0k}}) \tag{4}$$

其中 $x_k=\left[x_{1k}^{\mathrm{T}}x_{2k}^{\mathrm{T}}\cdots x_{Nk}^{\mathrm{T}}\right]$ 和 $\varepsilon_k=\left[\varepsilon_{1k}^{\mathrm{T}}\varepsilon_{2k}^{\mathrm{T}}\cdots\varepsilon_{Nk}^{\mathrm{T}}\right]$ 分别为全局节点状态向量和全局跟踪误差向量。

可以将式(4)重新表示成

$$\varepsilon_k=\left((L+G)\otimes I_n\right)\eta_k \tag{5}$$

全局同步误差向量

$$\eta_k=\left(x_k-\underline{x}_{0k}\right)\in R^{nN}. \tag{6}$$

其中 $\underline{x}_0=\underline{I}x_0$，$I=\underline{1}\otimes I_n$，$I_n$ 是 $n\times n$ 单位矩阵，$\underline{1}$ 是 n 个 1 组成的向量，$G=diag\{g_i\}\in R^{N\times N}$ 是固定增益的对角矩阵。

矩阵的最大奇异值和最小奇异值分别表示为 $\bar{\delta}(.)$，$\underline{\delta}(.)$。

下面的引理表明，小的局部邻域同步误差等价于小的全局同步误差。

引理 1 [23]：设 $(L+G)$ 为非奇异，那么同步误差的界限是

$$\|\eta_k\|\le\|\varepsilon_k\|/\underline{\delta}(L+G). \tag{7}$$

备注 1：如果对于根节点 $g_i\neq 0$，并且图中包含一棵生成树，则 $(L+G)$ 是非奇异的。

由(1)和(3)，可以得到节点 i 的局部邻域跟踪误差动态为

$$\begin{aligned}\varepsilon_{i(k+1)}&\equiv f_i\left(\varepsilon_{ik},u_{ik},u_{jk},w_{ik},w_{jk}\right)\\&=A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\end{aligned} \tag{8}$$

其中 u_{jk} 是智能体 i 的邻居的控制策略，w_{jk} 是智能体 i 的邻域的外部干扰。这是一个具有多个控制输入和外界干扰的动态系统。

因此，本文的目标是找到控制策略 u_{ik}，当干扰输入不存在，即 $w_{ik}=0$ 时，保证局部邻域同步误差(8)渐近收敛。此外，当干扰输入存在，即 $w_{ik}\neq 0$ 时，控制策略 u_{ik} 保证性能输出满足下列有界 L_2-增益，并预先确定一个常数 $\gamma>0$

$$\begin{aligned}\sum_{k=0}^{\infty}\|z_{ik}\|^2&=\sum_{k=0}^{\infty}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+u_{ik}^{\mathrm{T}}R_{ii}u_{ik}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}\right)\\&\le\gamma^2\sum_{k=0}^{\infty}\left(w_{ik}^{\mathrm{T}}T_{ii}w_{ik}+\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)+\beta\left(\varepsilon_i(0)\right)\end{aligned} \tag{9}$$

对于某些有界函数，使 $\beta(0)=0$，其中 $Q_{ii}>0,R_{ii}>0,R_{ij}>0,T_{ii}>0,T_{ij}>0$。$\gamma^*$ 为满足干扰衰减条件(9)的 γ 的最小值，其中 z_{ik} 是性能输出。

3. 多智能体零和博弈

在本节中，我们将定义每个智能体的局部性能指标函数和值函数，并给出满足纳什均衡的条件。利用最优控制原理建立了零和博弈的哈密顿函数和贝尔曼方程，并利用离散哈密顿雅可比理论给出了二者之间的关系。

3.1. 零和博弈

零和博弈是基于每个智能体 i 对图中其他智能体的响应。

定义智能体 i 的邻居的控制策略为

$$u_{-i}=\left\{u_j\mid j\in N_i\right\}. \tag{10}$$

定义智能体 i 的邻居的外部扰动为

$$w_{-i}=\left\{w_j \mid j\in N_i\right\}. \tag{11}$$

为每个代理定义以下本地性能指标

$$\begin{aligned}&J_i\left(\left\{\varepsilon_{ik},u_{ik},u_{-ik},w_{ik},w_{-ik}\right\}_{k\geq 0}\right)\\&=\sum_{k=0}^{\infty}U_i\left(\varepsilon_{ik},u_{ik},u_{-ik},w_{ik},w_{-ik}\right)\\&=\frac{1}{2}\sum_{k=0}^{\infty}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+u_{ik}^{\mathrm{T}}R_{ii}u_{ik}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\sum_{k=0}^{\infty}w_{ik}^{\mathrm{T}}T_{ii}w_{ik}+\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\end{aligned} \tag{12}$$

其中 $Q_{ii}>0\in R^{n_i\times n_i}$，$R_{ii}>0\in R^{m_i\times m_i}$，$R_{ij}>0\in R^{m_j\times m_j}$，$T_{ii}>0\in R^{q_i\times q_i}$，$T_{ij}>0\in R^{q_j\times q_j}$ 是对称时不变加权矩阵，$\gamma>0$ 满足有界 L_2 -增益。

动力学(8)和性能指标(12)依赖于图拓扑 $Gr=(V,\varsigma)$。

将智能体 i 及其邻居的固定策略给出，则每个智能体 i 的值函数为

$$V_i\left(\varepsilon_{ik}\right)=\sum_{l=k}^{\infty}U_i\left(\varepsilon_{il},u_{il},u_{-il},w_{il},w_{-il}\right). \tag{13}$$

备注 2：给出了控制策略和外部干扰，其值函数只与局部邻域同步跟踪误差 ε_{ik} 相关。

3.2. 零和博弈的贝尔曼方程

取(13)关于 k 的一阶差分得到贝尔曼方程

$$V_i\left(\varepsilon_{ik}\right)=U_i\left(\varepsilon_{il},u_{il},u_{-il},w_{il},w_{-il}\right)+V_i\left(\varepsilon_{i(k+1)}\right) \tag{14}$$

$V_i(0)=0$ 是初始条件。

因此，有界 L_2 -增益同步问题可以转化为以下的多智能体零和博弈问题.多智能体零和博弈问题的目标是找到 $\bar{u}_i=\left\{u_{ik}\right\}_{k=0}^{\infty}$，$\bar{w}_i=\left\{w_{ik}\right\}_{k=0}^{\infty}$，$\forall i\in N$，

$$V_i^*\left(\varepsilon_{ik}\right)=\min_{\bar{u}_i}\max_{\bar{w}_i}\left(V_i\left(\varepsilon_{ik}\right)\right)=\max_{\bar{w}_i}\min_{\bar{u}_i}\left(V_i\left(\varepsilon_{ik}\right)\right) \tag{15}$$

其中，控制输入 u_{ik} 试图最小化 $V_i\left(\varepsilon_{ik}\right)$，而干扰输入 w_{ik} 试图最大化 $V_i\left(\varepsilon_{ik}\right)$。

根据式(14)和(15)可以得到

$$V_i^*\left(\varepsilon_{ik}\right)=\min_{u_{ik}}\max_{w_{ik}}U_i\left(\varepsilon_{il},u_{il},u_{-il},w_{il},w_{-il}\right)+V_i^*\left(\varepsilon_{i(k+1)}\right). \tag{16}$$

定义 1：多智能体零和博弈最优控制策略和外部扰动策略 $\left(u_{ik}^*,w_{ik}^*\right)$ 对于 $\forall i\in N$ 如果满足以下不等式，那么具有全局纳什均衡解。

$$J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik},w_{-ik}^*\right)\leq J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right)\leq J_i\left(\varepsilon_{ik},u_{ik},u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right) \tag{17}$$

因此，每个智能体 i 的最优控制策略表示为

$$u_{ik}^*=-\left(d_i+g_i\right)R_{ii}^{-1}B_i^{\mathrm{T}}\nabla V_i^*\left(\varepsilon_{i(k+1)}\right) \tag{18}$$

每个智能体 i 的最优外部扰动为

$$w_{ik}^*=\frac{1}{\gamma^2}\left(d_i+g_i\right)T_{ii}^{-1}D_i^{\mathrm{T}}\nabla V_i^*\left(\varepsilon_{i(k+1)}\right). \tag{19}$$

将(18)和式(19)代入式(16)得到零和博弈贝尔曼最优方程

$$\begin{aligned}V_i^*\left(\varepsilon_{ik}\right)=&V_i^*\left(\varepsilon_{i(k+1)}\right)+\frac{1}{2}\Bigg(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(d_i+g_i\right)^2\nabla V_i^*\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}}B_iR_{ii}^{-1}B_i^{\mathrm{T}}\nabla V_i^*\left(\varepsilon_{i(k+1)}\right)\\&+\sum_{j\in N_i}\left(d_i+g_i\right)^2\nabla V_j^*\left(\varepsilon_{j(k+1)}\right)^{\mathrm{T}}B_jR_{jj}^{-1}R_{ij}R_{jj}^{-1}B_j^{\mathrm{T}}\nabla V_j^*\left(\varepsilon_{j(k+1)}\right)\\&-\frac{1}{\gamma^2}\left(d_i+g_i\right)^2\nabla V_i^*\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}}D_iT_{ii}^{-1}D_i^{\mathrm{T}}\nabla V_i^*\left(\varepsilon_{i(k+1)}\right)\\&-\frac{1}{\gamma^2}\sum_{j\in N_i}\left(d_i+g_i\right)^2\nabla V_j^*\left(\varepsilon_{j(k+1)}\right)^{\mathrm{T}}D_jT_{jj}^{-1}T_{ij}T_{jj}^{-1}D_j^{\mathrm{T}}\nabla V_j^*\left(\varepsilon_{j(k+1)}\right)\Bigg)\end{aligned}\tag{20}$$

3.3. 零和博弈的哈密顿函数

考虑节点误差动力学(8)和性能指标(12)，我们可以将每个智能体 i 的哈密顿方程定义为

$$\begin{aligned}&H_i\left(\varepsilon_{ik},\lambda_{i(k+1)},u_{ik},u_{-ik},w_{ik},w_{-ik}\right)\\&=\lambda_{i(k+1)}^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\right)\\&+\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+u_{ik}^{\mathrm{T}}R_{ii}u_{ik}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2w_{ik}^{\mathrm{T}}T_{ii}w_{ik}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\end{aligned}\tag{21}$$

其中 $\lambda_{ik}=\lambda_i\left(k\right)$ 是每个智能体 i 的伴随变量。由最优性的必要条件，我们得到

$$\partial H_i/\partial\varepsilon_{ik}=\lambda_{ik}\Rightarrow\lambda_{ik}=A^{\mathrm{T}}\lambda_{i(k+1)}+Q_{ii}\varepsilon_{ik}.\tag{22}$$

最优控制策略由平稳条件 $\partial H_i/\partial u_{ik}=0,\partial H_i/\partial w_{ik}=0$ 给出：

$$u_{ik}^*=-\left(d_i+g_i\right)R_{ii}^{-1}B_i^{\mathrm{T}}\lambda_{i(k+1)}^*,\tag{23}$$

$$w_{ik}^*=\frac{1}{\gamma^2}\left(d_i+g_i\right)T_{ii}^{-1}D_i^{\mathrm{T}}\lambda_{i(k+1)}^*.\tag{24}$$

3.4. 离散哈密顿雅可比理论：哈密顿方程与贝尔曼最优方程的等价性

定义值函数 $V_i\left(\varepsilon_{ik}\right)$ 的第一差值为

$$\Delta V_i\left(\varepsilon_{ik}\right)=V_i\left(\varepsilon_{i(k+1)}\right)-V_i\left(\varepsilon_{ik}\right).\tag{25}$$

定义值函数 $V_i\left(\varepsilon_{ik}\right)$ 的梯度为

$$\nabla V_i\left(\varepsilon_{i(k+1)}\right)=\partial V_i\left(\varepsilon_{i(k+1)}\right)\Big/\partial\varepsilon_{i(k+1)}.\tag{26}$$

然后我们用值函数定义伴随的变量为

$$\lambda_{i(k+1)}=\nabla V_i\left(\varepsilon_{i(k+1)}\right).\tag{27}$$

备注 3：有关协态变量和值函数梯度的等价关系，可以参考[29] [33] [34]。

3.5. 耦合 HJI 方程

将控制策略(18)和干扰策略(19)代入(21)得到耦合的 HJI 方程

$$\begin{aligned}&\nabla V_i\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}} A_i^c+\frac{1}{2}\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\frac{1}{2}\left(d_i+g_i\right)^2\nabla V_i\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}} B_iR_{ii}B_i^{\mathrm{T}}\nabla V_i\left(\varepsilon_{i(k+1)}\right)\\&+\frac{1}{2}\sum_{j\in N_i}\left(d_j+g_j\right)^2\nabla V_j\left(\varepsilon_{j(k+1)}\right)^{\mathrm{T}} B_jR_{jj}^{-1}R_{ij}R_{jj}^{-1}\nabla V_j\left(\varepsilon_{j(k+1)}\right)\\&-\frac{1}{2\gamma^2}\left(d_i+g_i\right)^2\nabla V_i\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}} D_iT_{ii}^{-1}D_i^{\mathrm{T}}\nabla V_i\left(\varepsilon_{i(k+1)}\right)\\&-\frac{1}{2\gamma^2}\sum_{j\in N_i}\left(d_j+g_j\right)^2\nabla V_j\left(\varepsilon_{j(k+1)}\right)^{\mathrm{T}} D_jT_{jj}^{-1}T_{ij}T_{jj}^{-1}\nabla V_j\left(\varepsilon_{j(k+1)}\right)=0\end{aligned}\tag{28}$$

边界条件$V_i(0)=0$。

$$\begin{aligned}A_i^c=&A\varepsilon_{ik}-\left(d_i+g_i\right)^2B_iR_{ii}^{-1}B_i^{\mathrm{T}}\nabla V_i\left(\varepsilon_{i(k+1)}\right)+\sum_{j\in N_i}a_{ij}\left(d_j+g_j\right)B_jR_{jj}^{-1}B_j^{\mathrm{T}}\nabla V_j\left(\varepsilon_{j(k+1)}\right)\\&+\frac{1}{\gamma^2}\left(d_i+g_i\right)^2D_iT_{ii}^{-1}D_i^{\mathrm{T}}\nabla V_i\left(\varepsilon_{i(k+1)}\right)-\frac{1}{\gamma^2}\sum_{j\in N_i}a_{ij}\left(d_j+g_j\right)D_jT_{jj}^{-1}D_j^{\mathrm{T}}\nabla V_j\left(\varepsilon_{j(k+1)}\right).\end{aligned}$$

为智能体 i 对应的闭环系统。

因此(28)可以改写为

$$\begin{aligned}&H_i\left(\varepsilon_{ik},\nabla V_i^*\left(\varepsilon_{i(k+1)}\right),u_{ik}^*,u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right)=0,\\&V_i(0)=0.\end{aligned}\tag{29}$$

引理 2 [35]：选择$\gamma>\gamma^*$。假设$V_i^*\left(\varepsilon_{ik}\right)>0$，$i\in N$是耦合 HJI 方程(25)的正定解。让智能体 i 的邻域控制策略u_{jk}已经最优，然后闭环系统在平衡点

$$\varepsilon_{i(k+1)}=A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^*-\sum_{j\in N_i}a_{ij}B_ju_{jk}^*\tag{30}$$

渐近稳定，在控制输入$u_{ik}=u_{ik}\left(V_i^*\left(\varepsilon_{ik}\right)\right)$的条件下，其中$V_i^*\left(\varepsilon_{ik}\right)$由式(28)给出。此外，在所有干扰$w_{ik}$存在的情况下控，制输$u_{ik}$入将满足有界$L_2$-增益条件(9)。

证明：引理的证明参考[36]中定理 1 的证明。

引理 3 [36]：假设引理 2 中的假设满足。如果图具有生成树，则全局同步误差η_k为L_2-有界。

由于上述方程的解析解很难得到，我们将寻求其迭代解。

4. 求解 HJI 方程的多智能体策略迭代算法

4.1. 最佳回应

定义 2：假设存在稳定控制策略u_{ik}和外界扰动w_{ik}，给定固定邻居控制策略u_{-ik}和扰动策略w_{-ik}。然后，满足以下不等式的一组策略$\left(u_{ik}^*,w_{ik}^*\right)$是最佳响应

$$J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik},w_{ik},w_{-ik}\right)\le J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik},w_{ik}^*,w_{-ik}\right)\le J_i\left(\varepsilon_{ik},u_{ik},u_{-ik},w_{ik}^*,w_{-ik}\right).\tag{31}$$

控制$u_{ik}=u_{ik}^*$由(23)给出，扰动$w_{ik}=w_{ik}^*$由(24)给出，给定任意策略u_{-ik}，w_{-ik}，定义最佳响应 HJI 方程

$$\begin{aligned}&\nabla V_i\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}} A_i^c+\frac{1}{2}\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\frac{1}{2}\left(d_i+g_i\right)^2\nabla V_i\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}} B_iR_{ii}B_i^{\mathrm{T}}\nabla V_i\left(\varepsilon_{i(k+1)}\right)\\&+\frac{1}{2}\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\frac{1}{2\gamma^2}\left(d_i+g_i\right)^2\nabla V_i\left(\varepsilon_{i(k+1)}\right)^{\mathrm{T}} D_iT_{ii}^{-1}D_i^{\mathrm{T}}\nabla V_i\left(\varepsilon_{i(k+1)}\right)\\&-\frac{1}{2\gamma^2}\sum_{j\in N_i}w_{jk}^{\mathrm{T}}R_{ij}w_{jk}=0\end{aligned}\tag{32}$$

边界条件$V_i(0)=0$。

4.2. 策略迭代求解 HJI 方程

本节提出一种策略迭代(PI)算法求解最优响应 HJI 方程(32)和 HJI 方程(28)，PI 算法有内部循环和外部循环。在内环中，控制策略是固定不变的。通过求解哈密顿方程(21)，用(19)更新扰动。在外部循环中，使用(18)进行更新迭代控制策略。最后，两个定理证明了算法的收敛性。同时，我们证明出，耦合 HJI 方程(28)的解就是零和博弈的纳什均衡解。

耦合 HJI 方程(28)和最优响应 HJI 方程(32)的策略迭代算法 1 的收敛性在下面的两个定理中给出。在定理 1 中，只有智能体 i 更新自己的策略，而其他邻居的策略是固定的。在定理 2 中，所有的智能体同时更新它们的策略。

本文所设计的策略迭代(PI)算法如表 1 所示：

Table 1. Multiagent learning policy iteration to solve the HJI equations
表 1. 求解 HJI 方程的多智能体策略迭代

算法 1：求解 HJI 方程的多智能体策略迭代

要求： 设$u_{ik}^0,\forall i=1,\cdots,N$为任何稳定的初始控制策略。

$p=0,1,\cdots$，给定$u_{ik}^p,\forall i=1,\cdots,N$

重复 p

$l=0,1,\cdots$，给定$w_{ik}=0$

重复 l

求解$V_i\left(\varepsilon_{i(k+1)}^{(p,l)}\right)$使用零和贝尔曼方程

$$\begin{aligned}&\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^p-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}^l-\sum_{j\in N_i}a_{ij}D_jw_{jk}\right)\\&+\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^p\right)^{\mathrm{T}}R_{ii}u_{ik}^p+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)=0\end{aligned}\tag{33}$$

更新w_{ik}^{l+1}

$$w_{ik}^{l+1}=\frac{1}{\gamma^2}\left(d_i+g_i\right)T_{ii}^{-1}D_i^{\mathrm{T}}\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\tag{34}$$

直到$\left\|V_i\left(\varepsilon_{i(k+1)}^{(p,l+1)}\right)-V_i\left(\varepsilon_{i(k+1)}^{(p,l)}\right)\right\|\le\varepsilon$收敛

结束

更新u_{ik}^{p+1}

$$u_{ik}^{p+1}=-\left(d_i+g_i\right)R_{ii}^{-1}B_i^{\mathrm{T}}\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\tag{35}$$

直到$\left\|V_i\left(\varepsilon_{i(k+1)}^{(p+1,l)}\right)-V_i\left(\varepsilon_{i(k+1)}^{(p,l)}\right)\right\|\le\varepsilon$收敛

结束

定理 1：假设邻居的策略是固定的，并且存在一个稳定控制策略和扰动策略，其中$\gamma\ge\gamma^*$，假设智能体 i 执行算法 1。那么值函数$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$收敛到最优值函数$V_i^*\left(\varepsilon_{ik}\right)$，其中$V_i^*\left(\varepsilon_{ik}\right)$是最佳回应 HJI 方程(32)

的解。

证明：首先证明了内环的收敛性。通过$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$对 l 沿着误差系统的解求差分

$$\varepsilon_{i(k+1)}=A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^p-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}^{l+1}-\sum_{j\in N_i}a_{ij}D_jw_{jk} \tag{36}$$

因此

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[U_i\left(\varepsilon_{ik},u_{ik}^p,u_{-ik},w_{ik}^l,w_{-ik}\right)-U_i\left(\varepsilon_{ik},u_{ik}^p,u_{-ik},w_{ik}^{l+1},w_{-ik}\right)\right]\\&=\sum_k^{\infty}\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^p\right)^{\mathrm{T}}R_{ii}u_{ik}^p+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\right.\\&\left.-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^p\right)^{\mathrm{T}}R_{ii}u_{ik}^p+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l+1}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\right]\end{aligned} \tag{37}$$

由于只有智能体 i 更新它的策略，因此

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^p-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}^{l+1}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\right)\right.\\&\left.+\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^p-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}^{l+1}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\right)\right]\end{aligned} \tag{38}$$

将式(33)代入式(38)可得

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^p\right)^{\mathrm{T}}R_{ii}u_{ik}^p+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\right.\\&-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^p\right)^{\mathrm{T}}R_{ii}u_{ik}^p+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2w_{ik}^{(l+1)\mathrm{T}}T_{ii}w_{ik}^{l+1}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\\&\left.-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{l+1}+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^l\right]\end{aligned} \tag{39}$$

从(39)我们可以得到

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[\frac{1}{2}\gamma^2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l+1}-\frac{1}{2}\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{l+1}\right.\\&\left.+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^l\right]\end{aligned} \tag{40}$$

然后将(34)代入(40)，并做一些运算

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[\frac{1}{2}\gamma^2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l+1}+\frac{1}{2}\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\gamma^2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}\right]\\&=-\frac{1}{2}\gamma^2\sum_k^{\infty}\left[\left(w_{ik}^{l+1}-w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}\left(w_{ik}^{l+1}-w_{ik}^{l}\right)\right]\le 0.\end{aligned}\tag{41}$$

因此，有$V_i^{(p,l)}\left(\varepsilon_{ik}\right)\le V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)$。因此值函数$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$对于 l 是单调递增的，根据公式(9)得到$V_i\left(\varepsilon_{ik}^{(p,l)}\right)$是有上界的，因此它收敛到一个极限，因此$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$对于 l 收敛。这就完成了内环收敛性的证明。

为了证明外环的收敛性，通过$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$对 p 沿着误差系统的解求差分

$$\varepsilon_{i(k+1)}=A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p+1}-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}^{l}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\tag{42}$$

因此

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[U_i\left(\varepsilon_{ik},u_{ik}^{p},u_{-ik},w_{ik}^{l},w_{-ik}\right)-U_i\left(\varepsilon_{ik},u_{ik}^{p+1},u_{-ik},w_{ik}^{l},w_{-ik}\right)\right]\\&=\sum_k^{\infty}\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\right.\\&\left.-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p+1}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p+1}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\right]\end{aligned}\tag{43}$$

由于只有智能体 i 更新它的策略，因此

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p+1}-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}^{l}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\right)\right.\\&\left.+\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p+1}-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}^{l}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\right)\right]\end{aligned}\tag{44}$$

将式(33)代入式(44)可得

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\&=\sum_k^{\infty}\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\right.\\&-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p+1}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p+1}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\\&\left.-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{p+1}+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{p}\right]\end{aligned}\tag{45}$$

从(45)我们可以得到

$$\begin{aligned} & V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right) \\ & =\sum_k^{\infty}\left[-\frac{1}{2}\left(\left(u_{ik}^{p+1}\right)^{\mathrm{T}} R_{ii}u_{ik}^{p+1}-\left(u_{ik}^{p}\right)^{\mathrm{T}} R_{ii}u_{ik}^{p}\right)-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial \varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)B_iu_{ik}^{p+1}\right. \\ & \left.+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial \varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)B_iu_{ik}^{p}\right] \end{aligned} \tag{46}$$

然后将(35)代入(46)，并做一些运算

$$V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)=\frac{1}{2}\sum_k^{\infty}\left[\left(u_{ik}^{p+1}-u_{ik}^{p}\right)^{\mathrm{T}} R_{ii}\left(u_{ik}^{p+1}-u_{ik}^{p}\right)\right]\geq 0. \tag{47}$$

因此，有$V_i^{(p,l)}\left(\varepsilon_{ik}\right)\geq V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)$。因此值函数$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$对于 p 是单调递减的，根据$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$的定义，我们知道$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$是一个下界，因此它收敛到一个极限，因此$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$对于 p 收敛。所以对外环收敛性的证明就完成了。

因此，根据上述结果和最优性原理，执行算法 1 得到的$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$收敛到最佳响应 HJI 方程(32)的解$V_i^{*}\left(\varepsilon_{ik}\right)$。

备注 3：在接下来的证明中，定义$\rho_{ij}=\bar{\sigma}\left(R_{jj}^{-1}R_{ij}\right),\beta_{ij}=\bar{\sigma}\left(T_{jj}^{-1}T_{ij}\right)$的相对权值。

定理 2：假设智能体 i 执行算法 1，且所有邻居的策略正在更新，并且存在稳定控制策略和干扰策略。对于小权值a_{ij}，相对权值β_{ij},ρ_{ij}，值函数$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$收敛到最优值函数$V_i^{*}\left(\varepsilon_{ik}\right)$，其中$V_i^{*}\left(\varepsilon_{ik}\right)$是 HJI 方程(28)的解。

证明：首先证明了内环的收敛性。通过$V_i^{(p,l)}\left(\varepsilon_{ik}\right)$对 l 沿着误差系统的解求差分

$$\varepsilon_{i(k+1)}=A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p}-\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p}+\left(d_i+g_i\right)D_iw_{ik}^{l+1}-\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l} \tag{48}$$

考虑到所有智能体都会更新他们的策略

$$\begin{aligned} & V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right) \\ & =\sum_k^{\infty}\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p}\right)^{\mathrm{T}} R_{ii}u_{ik}^{p}+\sum_{j\in N_i}\left(u_{jk}^{p}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p}-\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l}\right)\right. \\ & \left.-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p}\right)^{\mathrm{T}} R_{ii}u_{ik}^{p}+\sum_{j\in N_i}\left(u_{jk}^{p}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p}-\gamma^2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l+1}-\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l}\right)\right] \end{aligned} \tag{49}$$

因此，我们可以有

$$\begin{aligned} & V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right) \\ & =\sum_k^{\infty}\left[-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial \varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p}-\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p}+\left(d_i+g_i\right)D_iw_{ik}^{l+1}-\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l}\right)\right. \\ & \left.+\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial \varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p}-\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p}+\left(d_i+g_i\right)D_iw_{ik}^{l+1}-\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l}\right)\right] \end{aligned} \tag{50}$$

将式(33)代入式(50)可得

$$\begin{aligned}
&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\
&=\sum_k^{\infty}\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p}+\sum_{j\in N_i}\left(u_{jk}^{p}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p}-\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l}\right)\right.\\
&\quad-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p}+\sum_{j\in N_i}\left(u_{jk}^{p}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p}-\gamma^2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l+1}-\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l+1}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l+1}\right)\\
&\quad-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{l+1}+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{l}\\
&\quad\left.-\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l}+\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l+1}\right]
\end{aligned}\tag{51}$$

从(51)我们可以得到

$$\begin{aligned}
&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\
&=\sum_k^{\infty}\left[\frac{1}{2}\gamma^2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l+1}+\frac{1}{2}\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l+1}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l+1}-\frac{1}{2}\gamma^2\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}-\frac{1}{2}\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l}\right.\\
&\quad-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{l+1}+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)D_iw_{ik}^{l}\\
&\quad\left.-\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l}+\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l+1}\right]
\end{aligned}\tag{52}$$

然后将(34)代入(52)，并做一些运算

$$\begin{aligned}
&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)\\
&=\sum_k^{\infty}\left[-\frac{1}{2}\gamma^2\left(\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l+1}-2\left(w_{ik}^{l+1}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}+\left(w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}w_{ik}^{l}\right)+\frac{1}{2}\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l+1}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l+1}\right.\\
&\quad\left.-\frac{1}{2}\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l}\right)^{\mathrm{T}}T_{ij}w_{jk}^{l}-\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l}+\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}D_jw_{jk}^{l+1}\right]\\
&=\sum_k^{\infty}\left[-\frac{1}{2}\gamma^2\left(w_{ik}^{l+1}-w_{ik}^{l}\right)^{\mathrm{T}}T_{ii}\left(w_{ik}^{l+1}-w_{ik}^{l}\right)-\frac{1}{2}\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l+1}-w_{jk}^{l}\right)^{\mathrm{T}}T_{ij}\left(w_{jk}^{l+1}-w_{jk}^{l}\right)\right.\\
&\quad\left.+\gamma^2\sum_{j\in N_i}\left(w_{jk}^{l+1}\right)^{\mathrm{T}}T_{ij}\left(w_{jk}^{l+1}-w_{jk}^{l}\right)+\sum_{j\in N_i}\left(\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}a_{ij}D_j\left(w_{jk}^{l+1}-w_{jk}^{l}\right)\right]
\end{aligned}\tag{53}$$

得到了 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)\le V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)$ 的充分条件

$$\frac{1}{2}\gamma^2\underline{\sigma}\left(T_{ij}\right)\left\|\Delta w_j\right\|>\left(d_j+g_j\right)\beta_{ij}\left\|\frac{\partial V_j^{(p,l)}\left(\varepsilon_{j(k+1)}\right)}{\partial\varepsilon_{j(k+1)}}\right\|\cdot\left\|D_j\right\|+a_{ij}\left\|\frac{\partial V_i^{(p,l+1)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right\|\cdot\left\|D_j\right\| \tag{54}$$

其中 $\Delta w_j=w_j^{l+1}-w_j^l$，$\underline{\sigma}\left(T_{ij}\right)$ 是 T_{ij} 的最小奇异值。当 a_{ij} 和 β_{ij} 足够小时，$V_i^{(p,l)}\left(\varepsilon_{ik}\right)\le V_i^{(p,l+1)}\left(\varepsilon_{ik}\right)$ 成立。因此值函数 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 对于 l 是单调递增的，根据 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 的定义，我们知道 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 是有上界的，因此它收敛于一个极限，因此 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 对于 l 收敛。这就完成了内环收敛性的证明。(参考[37]中定理 5 的证明)

为了证明外环的收敛性，通过 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 对 p 沿着误差系统的解求差分

$$\varepsilon_{i(k+1)}=A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p+1}-\sum_{j\in N_i}a_{ij}B_ju_{jk}^p+\left(d_i+g_i\right)D_iw_{ik}^l-\sum_{j\in N_i}a_{ij}D_jw_{jk}^l \tag{55}$$

考虑到所有智能体都会更新他们的策略

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\&=\sum_k^\infty\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^p\right)^{\mathrm{T}}R_{ii}u_{ik}^p+\sum_{j\in N_i}\left(u_{jk}^p\right)^{\mathrm{T}}R_{ij}u_{jk}^p-\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\gamma^2\sum_{j\in N_i}\left(w_{jk}^l\right)^{\mathrm{T}}T_{ij}w_{jk}^l\right)\right.\\&\left.-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p+1}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p+1}+\sum_{j\in N_i}\left(u_{jk}^p\right)^{\mathrm{T}}R_{ij}u_{jk}^p-\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\gamma^2\sum_{j\in N_i}\left(w_{jk}^l\right)^{\mathrm{T}}T_{ij}w_{jk}^l\right)\right]\end{aligned} \tag{56}$$

因此，我们可以有

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\&=\sum_k^\infty\left[-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p+1}-\sum_{j\in N_i}a_{ij}B_ju_{jk}^p+\left(d_i+g_i\right)D_iw_{ik}^l-\sum_{j\in N_i}a_{ij}D_jw_{jk}^l\right)\right.\\&\left.+\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(A\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}^{p+1}-\sum_{j\in N_i}a_{ij}B_ju_{jk}^p+\left(d_i+g_i\right)D_iw_{ik}^l-\sum_{j\in N_i}a_{ij}D_jw_{jk}^l\right)\right]\end{aligned} \tag{57}$$

将式(33)代入式(57)可得

$$\begin{aligned}&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\&=\sum_k^\infty\left[\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^p\right)^{\mathrm{T}}R_{ii}u_{ik}^p+\sum_{j\in N_i}\left(u_{jk}^p\right)^{\mathrm{T}}R_{ij}u_{jk}^p-\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\gamma^2\sum_{j\in N_i}\left(w_{jk}^l\right)^{\mathrm{T}}T_{ij}w_{jk}^l\right)\right.\\&-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^{p+1}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p+1}+\sum_{j\in N_i}\left(u_{jk}^{p+1}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p+1}-\gamma^2\left(w_{ik}^l\right)^{\mathrm{T}}T_{ii}w_{ik}^l-\gamma^2\sum_{j\in N_i}\left(w_{jk}^l\right)^{\mathrm{T}}T_{ij}w_{jk}^l\right)\\&-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)B_iu_{ik}^{p+1}+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)B_iu_{ik}^p\\&\left.-\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}B_ju_{jk}^p+\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p+1}\right]\end{aligned} \tag{58}$$

从(58)我们可以得到

$$\begin{aligned}
&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\
&=\sum_k^{\infty}\left[-\frac{1}{2}\left(\left(u_{ik}^{p+1}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p+1}+\sum_{j\in N_i}\left(u_{jk}^{p+1}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p+1}-\sum_{j\in N_i}\left(u_{jk}^{p}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p}-\left(u_{ik}^{p}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p}\right)\right.\\
&-\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)B_iu_{ik}^{p+1}+\left(\frac{\partial V_i^{(p,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\left(d_i+g_i\right)B_iu_{ik}^{p}\\
&\left.-\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p}+\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p+1}\right]
\end{aligned}\tag{59}$$

然后将(35)代入(59)，并做一些运算

$$\begin{aligned}
&V_i^{(p,l)}\left(\varepsilon_{ik}\right)-V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)\\
&=\sum_k^{\infty}\left[\frac{1}{2}\left(\left(u_{ik}^{p+1}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p+1}-2\left(u_{ik}^{p+1}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p}+\left(u_{ik}^{p}\right)^{\mathrm{T}}R_{ii}u_{ik}^{p}\right)\right.\\
&+\frac{1}{2}\left(\sum_{j\in N_i}\left(u_{jk}^{p+1}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p+1}-\sum_{j\in N_i}\left(u_{jk}^{p}\right)^{\mathrm{T}}R_{ij}u_{jk}^{p}\right)\\
&\left.-\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p}+\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}\sum_{j\in N_i}a_{ij}B_ju_{jk}^{p+1}\right]\\
&=\sum_k^{\infty}\left[\frac{1}{2}\left(u_{ik}^{p+1}-u_{ik}^{p}\right)^{\mathrm{T}}R_{ii}\left(u_{ik}^{p+1}-u_{ik}^{p}\right)+\frac{1}{2}\sum_{j\in N_i}\left(u_{jk}^{p+1}-u_{jk}^{p}\right)^{\mathrm{T}}R_{ij}\left(u_{jk}^{p+1}-u_{jk}^{p}\right)\right.\\
&\left.-\sum_{j\in N_i}\left(u_{jk}^{p+1}\right)^{\mathrm{T}}R_{ij}\left(u_{jk}^{p+1}-u_{jk}^{p}\right)+\sum_{j\in N_i}\left(\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right)^{\mathrm{T}}a_{ij}B_j\left(u_{jk}^{p+1}-u_{jk}^{p}\right)\right]
\end{aligned}\tag{60}$$

得到了 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)\geq V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)$ 的充分条件

$$\frac{1}{2}\gamma^2\underline{\sigma}\left(R_{ij}\right)\left\|\Delta u_j\right\|>\left(d_j+g_j\right)\rho_{ij}\left\|\frac{\partial V_j^{(p,l)}\left(\varepsilon_{j(k+1)}\right)}{\partial\varepsilon_{j(k+1)}}\right\|\cdot\left\|B_j\right\|+a_{ij}\left\|\frac{\partial V_i^{(p+1,l)}\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\right\|\cdot\left\|B_j\right\|\tag{61}$$

其中 $\Delta u_j=u_j^{p+1}-u_j^{p}$，$\underline{\sigma}\left(R_{ij}\right)$ 是 R_{ij} 的最小奇异值。当 a_{ij} 和 ρ_{ij} 足够小时，$V_i^{(p,l)}\left(\varepsilon_{ik}\right)\geq V_i^{(p+1,l)}\left(\varepsilon_{ik}\right)$ 成立。因此值函数 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 对于 p 是单调递减的，根据 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 的定义，我们知道 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 是有下界的，因此它收敛于一个极限，因此 $V_i^{(p,l)}\left(\varepsilon_{ik}\right)$ 对于 p 收敛。这就完成了外环收敛性的证明。

因此，根据上述结果和最优性原理，我们可以说该算法收敛于耦合 HJI 方程(28)的解 $V_i\left(\varepsilon_{ik}^{*}\right)$。

在接下来的定理中，我们将证明耦合 HJI 方程(28)的解就是零和博弈的纳什均衡解。

定理 3: (稳定性和纳什均衡解)让 $\gamma\geq\gamma^{*}$。所有的智能体的控制策略由(23)给出，干扰策略由(24)给出。各智能体 i 的最优性能指标为 $J_i^{*}\left(\varepsilon_{il},u_{il}^{*},u_{-il}^{*},w_{il}^{*},w_{-il}^{*}\right)=V_i^{*}\left(\varepsilon_{il}\right)$，所有智能体均处于纳什均衡状态。

证明：我们有

$$J_i\left(\varepsilon_{il},u_{il},u_{-il},w_{il},w_{-il}\right)=V_i^*\left(\varepsilon_{il}\right)+\sum_{k=l}^{\infty}\left[U_i\left(\varepsilon_{il},u_{il},u_{-il},w_{il},w_{-il}\right)-U_i^*\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right)\right] \tag{62}$$

由值函数定义可得

$$\begin{aligned}
&U_i\left(\varepsilon_{ik},u_{ik},u_{-ik},w_{ik},w_{-ik}\right)-U_i^*\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right)\\
&=\frac{1}{2}\Bigg(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+u_{ik}^{\mathrm{T}}R_{ii}u_{ik}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2w_{ik}^{\mathrm{T}}T_{ii}w_{ik}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\\
&\quad-\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+\left(u_{ik}^*\right)^{\mathrm{T}}R_{ii}u_{ik}^*-\sum_{j\in N_i}\left(u_{jk}^*\right)^{\mathrm{T}}R_{ij}u_{jk}^*+\gamma^2\left(w_{ik}^*\right)^{\mathrm{T}}T_{ii}w_{ik}^*+\gamma^2\sum_{j\in N_i}\left(w_{jk}^*\right)^{\mathrm{T}}T_{ij}w_{jk}^*\Bigg)\\
&=\frac{1}{2}\left(u_{ik}-u_{ik}^*\right)^{\mathrm{T}}R_{ii}\left(u_{ik}-u_{ik}^*\right)+\left(u_{ik}^*\right)^{\mathrm{T}}R_{ii}\left(u_{ik}-u_{ik}^*\right)+\frac{1}{2}\sum_{j\in N_i}\left(u_{jk}-u_{jk}^*\right)^{\mathrm{T}}R_{ij}\left(u_{jk}-u_{jk}^*\right)\\
&\quad+\sum_{j\in N_i}\left(u_{jk}^*\right)^{\mathrm{T}}R_{ij}\left(u_{jk}-u_{jk}^*\right)-\frac{1}{2}\gamma^2\left(w_{ik}-w_{ik}^*\right)^{\mathrm{T}}T_{ii}\left(w_{ik}-w_{ik}^*\right)-\gamma^2\left(w_{ik}^*\right)^{\mathrm{T}}T_{ii}\left(w_{ik}-w_{ik}^*\right)\\
&\quad-\frac{1}{2}\gamma^2\sum_{j\in N_i}\left(w_{jk}-w_{jk}^*\right)^{\mathrm{T}}T_{ij}\left(w_{jk}-w_{jk}^*\right)-\gamma^2\sum_{j\in N_i}\left(w_{jk}^*\right)^{\mathrm{T}}T_{ij}\left(w_{jk}-w_{jk}^*\right)
\end{aligned} \tag{63}$$

使用(63)到(62)得到

$$\begin{aligned}
&J_i\left(\varepsilon_{il},u_{il},u_{-il},w_{il},w_{-il}\right)\\
&=V_i^*\left(\varepsilon_{il}\right)+\sum_{k=l}^{\infty}\Bigg[\frac{1}{2}\left(u_{ik}-u_{ik}^*\right)^{\mathrm{T}}R_{ii}\left(u_{ik}-u_{ik}^*\right)+\left(u_{ik}^*\right)^{\mathrm{T}}R_{ii}\left(u_{ik}-u_{ik}^*\right)\\
&\quad+\frac{1}{2}\sum_{j\in N_i}\left(u_{jk}-u_{jk}^*\right)^{\mathrm{T}}R_{ij}\left(u_{jk}-u_{jk}^*\right)+\sum_{j\in N_i}\left(u_{jk}^*\right)^{\mathrm{T}}R_{ij}\left(u_{jk}-u_{jk}^*\right)\\
&\quad-\frac{1}{2}\gamma^2\left(w_{ik}-w_{ik}^*\right)^{\mathrm{T}}T_{ii}\left(w_{ik}-w_{ik}^*\right)-\gamma^2\left(w_{ik}^*\right)^{\mathrm{T}}T_{ii}\left(w_{ik}-w_{ik}^*\right)\\
&\quad-\frac{1}{2}\gamma^2\sum_{j\in N_i}\left(w_{jk}-w_{jk}^*\right)^{\mathrm{T}}T_{ij}\left(w_{jk}-w_{jk}^*\right)-\gamma^2\sum_{j\in N_i}\left(w_{jk}^*\right)^{\mathrm{T}}T_{ij}\left(w_{jk}-w_{jk}^*\right)\Bigg]
\end{aligned} \tag{64}$$

利用(64)得到最优性能指标 J_i^*，使

$$J_i^*\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right)=V_i^*\left(\varepsilon_{il}\right). \tag{65}$$

从(62)我们可以得到

$$J_i\left(\varepsilon_{il},u_{il},u_{-il}^*,w_{il}^*,w_{-il}^*\right)=V_i^*\left(\varepsilon_{il}\right)+\sum_{k=l}^{\infty}\left[U_i\left(\varepsilon_{il},u_{il},u_{-il}^*,w_{il}^*,w_{-il}^*\right)-U_i^*\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right)\right] \tag{66}$$

由值函数定义可知

$$\sum_{k=l}^{\infty}\left[U_i\left(\varepsilon_{il},u_{il},u_{-il}^*,w_{il}^*,w_{-il}^*\right)-U_i^*\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right)\right]\geq 0 \tag{67}$$

因此

$$J_i\left(\varepsilon_{il},u_{il},u_{-il}^*,w_{il}^*,w_{-il}^*\right)\geq J_i\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right) \tag{68}$$

从(68)可以得到

$$J_i\left(\varepsilon_{ik},u_{ik},u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right)\geq J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right) \tag{69}$$

从(62)我们可以得到

$$J_i\left(\varepsilon_{il},u_{il},u_{-il}^*,w_{il},w_{-il}^*\right)=V_i^*\left(\varepsilon_{il}\right)+\sum_{k=l}^{\infty}\left[U_i\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il},w_{-il}^*\right)-U_i^*\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right)\right] \tag{70}$$

由值函数定义可知

$$\sum_{k=l}^{\infty}\left[U_i\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il},w_{-il}^*\right)-U_i^*\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right)\right]\le 0 \tag{71}$$

因此

$$J_i\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il},w_{-il}^*\right)\le J_i\left(\varepsilon_{il},u_{il}^*,u_{-il}^*,w_{il}^*,w_{-il}^*\right) \tag{72}$$

从(72)可以得到

$$J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik},w_{-ik}^*\right)\le J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right) \tag{73}$$

从(69)和(73)，我们可以得到

$$J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik},w_{-ik}^*\right)\le J_i\left(\varepsilon_{ik},u_{ik}^*,u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right)\le J_i\left(\varepsilon_{ik},u_{ik},u_{-ik}^*,w_{ik}^*,w_{-ik}^*\right). \tag{74}$$

这就完成了证明。

在算法 1 中，我们需要完整的动力学知识。在下一节中，我们给出了一种解决系统动力学部分未知时的最优控制和最优干扰的方法。

5. 神经网络实现多智能体零和博弈在线解

在本节中，由于系统动力学部分未知，HJI 方程不能求解，因此接下来采用单层神经网络来逼近值函数，和三层神经网络相比，可以降低计算复杂度。

根据维尔斯特拉斯高阶近似定理，值函数$V_i\left(\varepsilon_{i(k+1)}\right)$可以近似为

$$V_i\left(\varepsilon_{i(k+1)}\right)=W_i^{\mathrm{T}}\varphi\left(\varepsilon_{i(k+1)}\right)+\delta_{i(k+1)} \tag{75}$$

其中W_i是权值，其中$\varphi\left(\varepsilon_{i(k+1)}\right)\in R^h$为评价神经网络的激活函数向量，神经网络隐藏层中的神经元数为h。根据维尔斯特拉斯高阶近似定理，神经网络近似误差$\delta_{i(k+1)}$当$h\to\infty$一致收敛到零。假设当前权重估计$\hat{W}_i$，则神经网络的输出为

$$\hat{V}_i\left(\varepsilon_{i(k+1)}\right)=\hat{W}_i^{\mathrm{T}}\varphi\left(\varepsilon_{i(k+1)}\right). \tag{76}$$

控制策略和外部干扰近似为

$$\hat{u}_{ik}=-\left(d_i+g_i\right)R_{ii}^{-1}B_i^{\mathrm{T}}\nabla\hat{V}_i\left(\varepsilon_{i(k+1)}\right), \tag{77}$$

$$\hat{w}_{ik}=\frac{1}{\gamma^2}\left(d_i+g_i\right)T_{ii}^{-1}D_i^{\mathrm{T}}\nabla\hat{V}_i\left(\varepsilon_{i(k+1)}\right). \tag{78}$$

利用神经网络将产生以下残差：

$$\begin{aligned}e_{Hi}=&\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+u_{ik}^{\mathrm{T}}R_{ii}u_{ik}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2w_{ik}^{\mathrm{T}}T_{ii}w_{ik}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right)\\&+\hat{W}_i\frac{\partial\varphi\left(\varepsilon_{i(k+1)}\right)}{\partial\varepsilon_{i(k+1)}}\left(\varepsilon_{ik}+\left(d_i+g_i\right)B_iu_{ik}-\sum_{j\in N_i}a_{ij}B_ju_{jk}+\left(d_i+g_i\right)D_iw_{ik}-\sum_{j\in N_i}a_{ij}D_jw_{jk}\right)\\=&\left(\hat{W}_i\right)^{\mathrm{T}}\left(\phi\left(\varepsilon_{ik}\right)-\phi\left(\varepsilon_{i(k+1)}\right)\right)-\frac{1}{2}\left(\varepsilon_{ik}^{\mathrm{T}}Q_{ii}\varepsilon_{ik}+u_{ik}^{\mathrm{T}}R_{ii}u_{ik}+\sum_{j\in N_i}u_{jk}^{\mathrm{T}}R_{ij}u_{jk}-\gamma^2w_{ik}^{\mathrm{T}}T_{ii}w_{ik}-\gamma^2\sum_{j\in N_i}w_{jk}^{\mathrm{T}}T_{ij}w_{jk}\right).\end{aligned}$$

为了使残差平方最小，需要找到$\hat{W}_i$

$$E_1 = \frac{1}{2}e_{Hi}^2. \tag{80}$$

采用基于梯度的自适应方法更新权值

$$\hat{W}_i^{l+1} = \hat{W}_i^l - \alpha_m \frac{\partial E_1}{\partial \hat{W}_i^l}. \tag{81}$$

算法 2 是神经网络的实现过程，本文所设计的算法 2 如表 2 所示：

Table 2. Neural network-based online tuning
表 2. 评价网络调优

算法 2：评价网络调优

第一步：选择一个初始权向量$\hat{W}_{ik}^0$，它可以产生允许的策略

第二步：当$l = 0,1,\cdots$时，计算e_{Hi}^l使用

$$\begin{aligned} e_{Hi}^l = &\left(\hat{W}_i^l\right)^T \left(\varphi\left(\varepsilon_{ik}\right) - \varphi\left(\varepsilon_{i(k+1)}\right)\right) - \frac{1}{2}\Bigg(\varepsilon_{ik}^{\mathrm{T}} Q_{ii} \varepsilon_{ik} + \left(\hat{u}_{ik}^l\right)^{\mathrm{T}} R_{ii} \hat{u}_{ik}^l + \sum_{j\in N_i} \left(\hat{u}_{jk}^l\right)^{\mathrm{T}} R_{ij} \hat{u}_{jk}^l \\ &- \gamma^2 \left(\hat{w}_{ik}^l\right)^{\mathrm{T}} T_{ii} \hat{w}_{ik}^l - \gamma^2 \sum_{j\in N_i} \left(\hat{w}_{jk}^l\right)^{\mathrm{T}} T_{ij} \hat{w}_{jk}^l \Bigg) \end{aligned} \tag{82}$$

第三步：我们使用(81)更新权重$\hat{W}_i^{l+1}$。

第四步：如果$\left\|\hat{W}_i^{l+1} - \hat{W}_i^l\right\| \le \xi$，$\xi$是一个小的实数，停止并输出$\hat{W}_i^l$；否则，设置$l = l+1$，通过下列两个式子更新控制策略$\hat{u}_{ik}^l$和干扰策略$\hat{w}_{ik}^l$，返回步骤 2。

$$\hat{u}_{ik}^l = -\left(d_i + g_i\right) R_{ii}^{-1} B_i^{\mathrm{T}} \nabla \hat{V}_i^l \left(\varepsilon_{i(k+1)}\right)$$

$$\hat{w}_{ik}^l = \frac{1}{\gamma^2}\left(d_i + g_i\right) T_{ii}^{-1} D_i^{\mathrm{T}} \nabla \hat{V}_{ik}^l \left(\varepsilon_{i(k+1)}\right)$$

备注 5：在算法 2 中，我们更新神经网络权值的方法是梯度下降法。假设梯度下降法在每次迭代中都是完全收敛的，则算法 2 的权值是收敛的，仿真结果表明了该算法的有效性。

备注 6：权值得到一致最终有界。证明过程与[5] [36] [37]相似。

6. 仿真结果

在本节中，我们利用仿真来验证理论的正确性。考虑有向图中有 5 个智能体，如图 1 所示。

我们考虑以下节点动态

$$x_i\left(k+1\right) = Ax_i\left(k\right) + B_i u_i\left(k\right) + D_i w_i\left(k\right)$$

其中智能体动态为：$A = \begin{bmatrix} 0.995 & 0.09983 \\ -0.09983 & 0.995 \end{bmatrix}$，$B_1 = \begin{bmatrix} 0.47 \\ 0.3 \end{bmatrix}$，$B_2 = \begin{bmatrix} 0.07 \\ 0.4 \end{bmatrix}$，$B_3 = \begin{bmatrix} 0.047 \\ 0.4 \end{bmatrix}$，$B_4 = \begin{bmatrix} 0.07 \\ 0.2 \end{bmatrix}$，$B_5 = \begin{bmatrix} 0.047 \\ 0.4 \end{bmatrix}$，$D_1 = \begin{bmatrix} 0.21 \\ 0.0984 \end{bmatrix}$，$D_2 = \begin{bmatrix} 0.21 \\ 0.084 \end{bmatrix}$，$D_3 = \begin{bmatrix} 0.1 \\ 0.0984 \end{bmatrix}$，$D_4 = \begin{bmatrix} 0.21 \\ 0.0984 \end{bmatrix}$，$D_5 = \begin{bmatrix} 0.21 \\ 0.0984 \end{bmatrix}$。

性能指标加权矩阵：$Q_{11} = Q_{22} = Q_{33} = Q_{44} = Q_{55} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$。

加权入度：$d_2=1$，$d_1=d_3=d_4=d_5=2$。

固定增益：$g_1=g_2=g_4=g_5=0$，$g_3=1$。

图 1 显示了一个考虑 5 个智能体和一个领导者的有向图，其中领导者连接到一个智能体。图 2 和图 3 为系统动力学未知情况下，权值和残差随迭代次数增加的变化情况。图 2 显示了五个智能体的权重迭代。从图中可以看出，5 个权重最终收敛。图 3 为神经网络估计的值函数与真实值函数之间的误差。可以看出，最终的误差收敛到零。当系统动力学信息已知时，我们也可以得到所有代理的状态变化。图 4、图 5、图 6 分别是一个领导者和五个智能体的状态图。从图中还可以看出，跟踪误差消失，所有智能体同步到领导者。图 6 是状态的三维图。

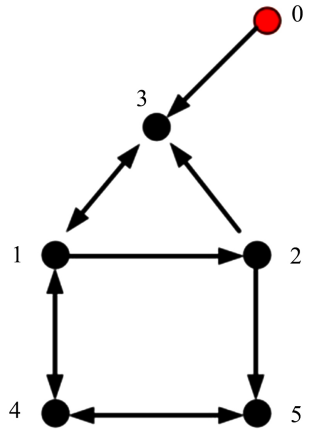

Figure 1. Communication graph
图 1. 通信图

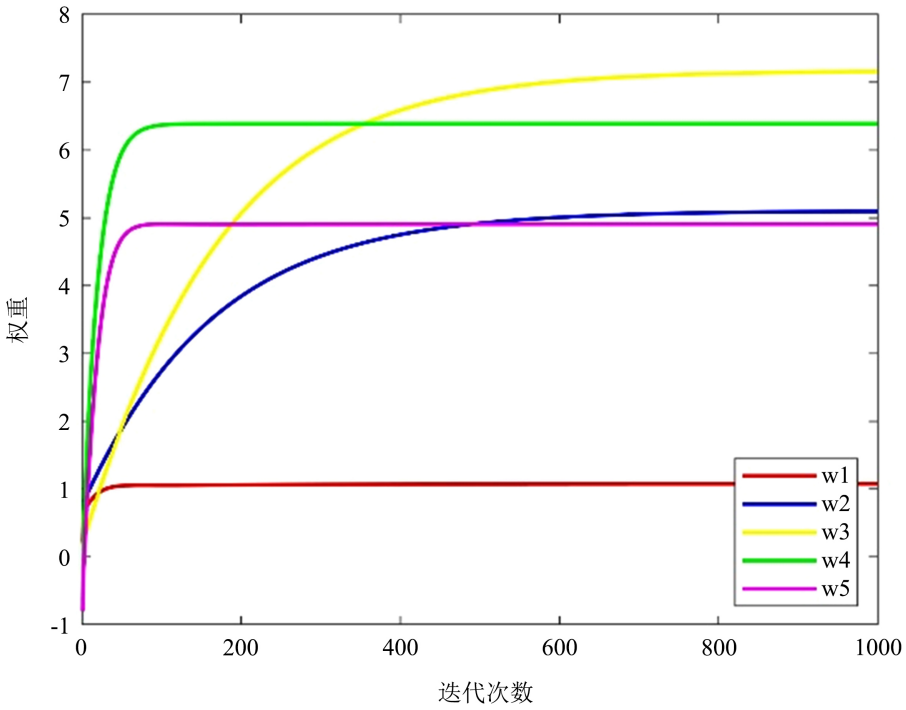

Figure 2. Weights update of agents
图 2. 权重随着迭代次数变化图

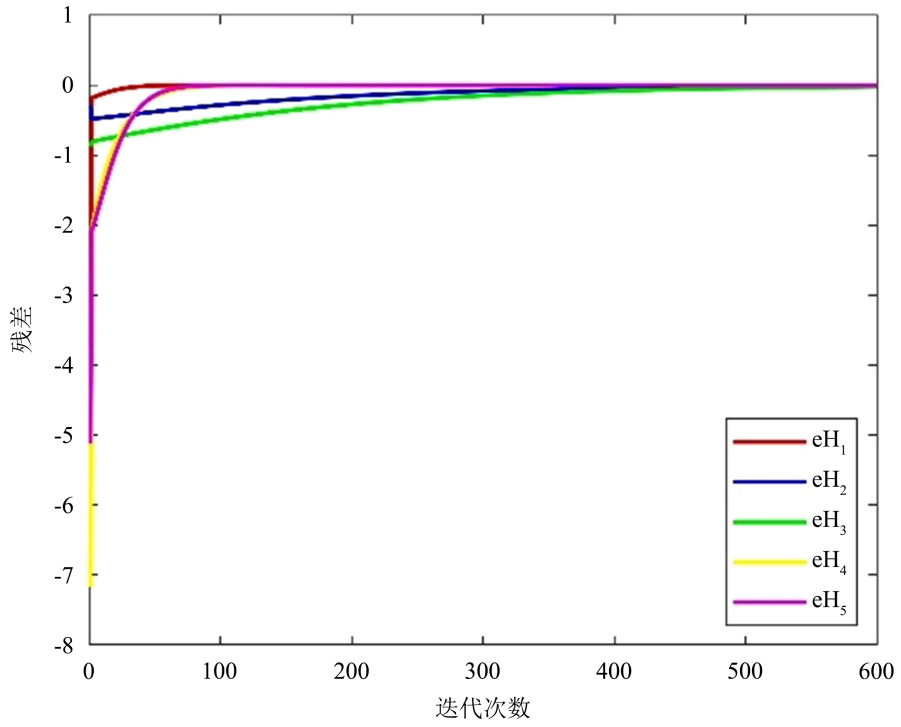

Figure 3. Residual errors
图 3. 残差随着迭代次数变化图

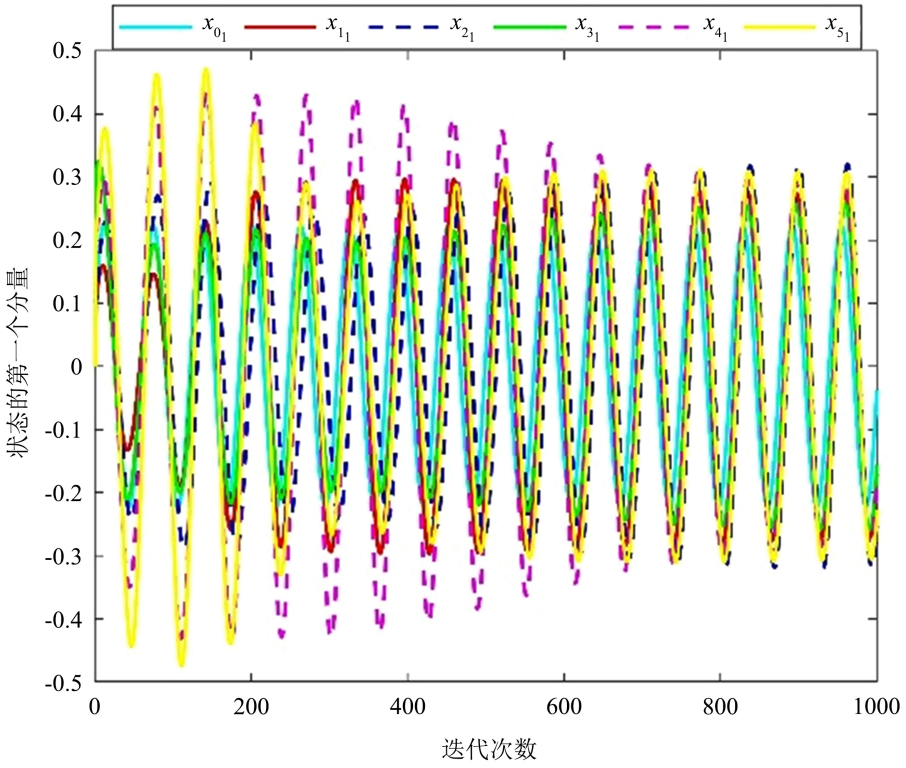

Figure 4. The states of $x_{0_1}, x_{1_1}, x_{2_1}, x_{3_1}, x_{4_1}, x_{5_1}$
图 4. 状态的第一个分量随着迭代次数变化图

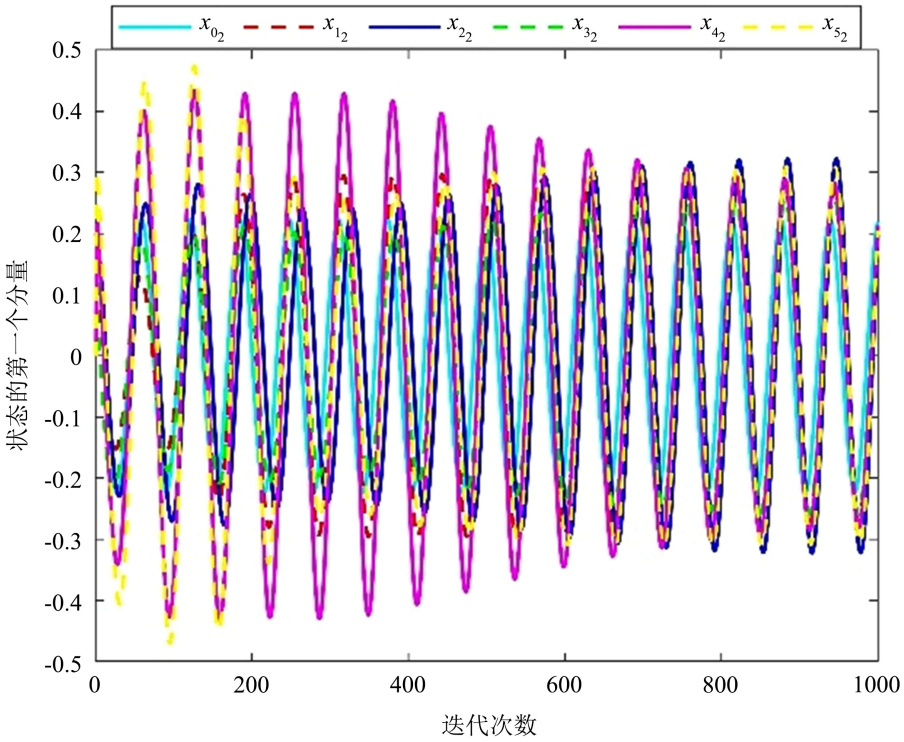

Figure 5. The states of $x_{0_2}, x_{1_2}, x_{2_2}, x_{3_2}, x_{4_2}, x_{5_2}$

图 5. 状态的第二个分量随着迭代次数变化图

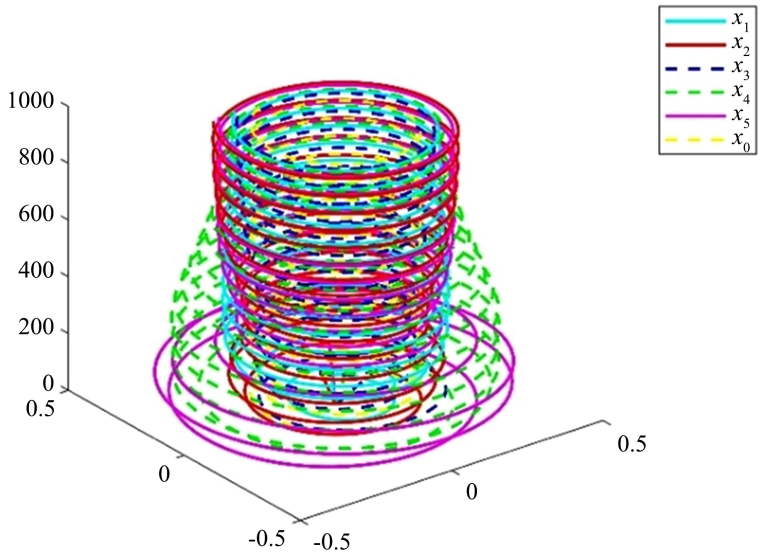

Figure 6. The states of $x_0, x_1, x_2, x_3, x_4, x_5$ with respect to iteration steps

图 6. 状态随着迭代次数变化的三维立体图

7. 结论

本文研究了受外界干扰的多智能体分布式最优跟踪控制问题，即控制策略与干扰策略的零和博弈问题。目标是求得一组最优控制和最佳干扰解，从而使得跟随者可以跟踪上领导者。而求最优解的本质问

题在于求解带有扰动的耦合 HJI 方程，由于 HJI 方程是偏微分方程，很难求解，所以本文设计了迭代算法求解 HJI 方程。

在算法 1 中，我们设计了内循环和外循环来求得最优解，在内循环中，控制策略是固定不变的，更新迭代干扰策略，在外部循环中，更新迭代控制策略，最后可以求得一组最优控制策略和干扰策略。定理 1 和定理 2 证明了算法 1 的收敛性，从而定理 3 也证明了最优解就是零和博弈的纳什均衡解，而由于算法 1 需要知道系统动力学知识，本文是部分未知线性离散系统，所以在 A 未知时，算法 1 就不能适用。针对这一问题，提出算法 2，本文利用单层神经网络逼近值函数，与传统的三层神经网络分别逼近值函数、控制策略和干扰策略相比，降低了计算的复杂度，算法 2 在梯度下降法是收敛的前提下则收敛。最后给出的仿真结果进一步验证了理论的正确性。图 1 显示了一个考虑 5 个智能体和一个领导者的有向图，从图 2 和图 3 可以看出权值最终收敛以及残差收敛到 0。从图 4、图 5 和图 6 可以看出，5 个智能体跟随者跟踪上领导者。本文考虑的是线性离散系统，在后续的研究中，我们将会研究更加复杂的动力学系统，以及进一步减少迭代次数。

基金项目

国防基金(JCKY2019413D001)；国家自然科学基金(6217023627)；上海市自然基金(19ZR1436000)。

参考文献

[1] Mu, S.M., Chu, T.G. and Wang, L. (2005) Coordinated Collective Motion in a Motile Particle Group with a Leader. *Physica A*: *Statistical Mechanics & Its Applications*, **351**, 211-226. https://doi.org/10.1016/j.physa.2004.12.054

[2] Nash, J.F. (1950) Two-Person Cooperative Games. *Econometrica*, **21**, 128-140. https://doi.org/10.2307/1906951

[3] Nash, J.F. (1951) Non-Cooperative Games. *Annals of Mathematics*, **54**, 286-295. https://doi.org/10.2307/1969529

[4] Starr, A.W. and Ho. Y.C. (1969) Nonzero-Sum Differential Games. *Journal of Optimization Theory and Applications*, **3**, 184-206. https://doi.org/10.1007/BF00929443

[5] Vamvoudakis, K.G. and Lewis, F.L. (2011) Multi-Player Non-Zero-Sum Games: Online Adaptive Learning Solution of Coupled Hamilton-Jacobi Equations. *Automatica*, **47**, 1556-1569. https://doi.org/10.1016/j.automatica.2011.03.005

[6] Yang, D.S., Pang, Y.H. and Zhou, B.W. (2019) Fault Diagnosis for Energy Internet Using Correlation Processing-Based Convolutional Neural Networks. *IEEE Transactions on Systems, Man, and Cybernetics*: *Systems*, **49**, 1739-1748. https://doi.org/10.1109/TSMC.2019.2919940

[7] Yang, X.F. and Gao, J.W. (2016) Linear-Quadratic Uncertain Differential Game with Application to Resource Extraction Problem. *IEEE Transactions on Fuzzy Systems*: *A Publication of the IEEE Neural Networks Council*, **24**, 819-826. https://doi.org/10.1109/TFUZZ.2015.2486809

[8] Hong, Y.G., Hu, J.P. and Gao, L.X. (2008) Tracking Control for Multi-Agent Consensus with an Active Leader and Variable Topology. *Automatica*, **42**, 1177-1182. https://doi.org/10.1016/j.automatica.2006.02.013

[9] Ren, W., Moore, K.L. and Chen, Y.Q. (2006) High-Order and Model Reference Consensus Algorithms in Cooperative Control of Multivehicle Systems. *Journal of Dynamic Systems Measurement and Control*, **129**, 678-688. https://doi.org/10.1115/1.2764508

[10] Freiling, G., Jank, G. and Abou-Kandil, H. (2002) On Global Existence of Solutions to Coupled Matrix Riccati Equations in Closed-Loop Nash Games. *IEEE Transactions on Automatic Control*, **41**, 264-269. https://doi.org/10.1109/9.481532

[11] Abu-Khalaf, M., Lewis, F.L. and Huang, J. (2007) Policy Iterations on the Hamilton-Jacobi-Isaacs Equation for H_∞ State Feedback Control with Input Saturation. *IEEE Transactions on Automatic Control*, **51**, 1989-1995. https://doi.org/10.1109/TAC.2006.884959

[12] Lewis, F.L. and Vrabie, D. (2009) Reinforcement Learning and Adaptive Dynamic Programming for Feedback Control. *IEEE Circuits & Systems Magazine*, **9**, 32-50. https://doi.org/10.1109/MCAS.2009.933854

[13] He, H.B., Ni, Z. and Fu. J. (2012) A Three-Network Architecture for On-Line Learning and Optimization Based on Adaptive Dynamic Programming. *Neurocomputing*, **78**, 3-13. https://doi.org/10.1016/j.neucom.2011.05.031

[14] Dierks, T. and Jagnnathan, S. (2012) Online Optimal Control of Affine Nonlinear Discrete-Time Systems with Unknown Internal Dynamics by Using Timebased Policy Update. *IEEE Transactions on Neural Networks & Learning*

Systems, **23**, 1118-1129. https://doi.org/10.1109/TNNLS.2012.2196708

[15] Wei, L.Q., Wang, F.Y. and Liu, D.R. (2014) Finite-Approximation-Error-Based Discrete-Time Iterative Adaptive Dynamic Programming. *IEEE Transactions on Cybernetics*, **44**, 2820-2833. https://doi.org/10.1109/TCYB.2014.2354377

[16] Ni, Z., He, H.B. and Zhao, D.B. (2015) GrDHP: A General Utility Function Representation for Dual Heuristic Dynamic Programming. *IEEE Transactions on Neural Networks & Learning Systems*, **26**, 614-627. https://doi.org/10.1109/TNNLS.2014.2329942

[17] Wei, Q.L., Liu, D.R. and Lin, H.Q. (2016) Value Iteration Adaptive Dynamic Programming for Optimal Control of Discrete-Time Nonlinear Systems. *IEEE Transactions on Cybernetics*, **46**, 840-853. https://doi.org/10.1109/TCYB.2015.2492242

[18] Gao, W.N. and Jiang, Z.P. (2016) Adaptive Dynamic Programming and Adaptive Optimal Output Regulation of Linear Systems. *IEEE Transactions on Automatic Control*, **61**, 4164-4169. https://doi.org/10.1109/TAC.2016.2548662

[19] Zhang, H.G., Liang, H.J. and Wang, Z.S. (2017) Optimal Output Regulation for Heterogeneous Multiagent Systems via Adaptive Dynamic Programming. *IEEE Transactions on Neural Networks & Learning Systems*, **28**, 18-29. https://doi.org/10.1109/TNNLS.2015.2499757

[20] Yang, Y.L., Wunsch, D. and Yin, Y.X. (2017) Hamiltonian-Driven Adaptive Dynamic Programming for Continuous Nonlinear Dynamical Systems. *IEEE Transactions on Neural Networks & Learning Systems*, **28**, 1929-1940. https://doi.org/10.1109/TNNLS.2017.2654324

[21] Sun, J.L. and Long, T. (2020) Event-Triggered Distributed Zero-Sum Differential Game for Nonlinear Multi-Agent Systems Using Adaptive Dynamic Programming. *ISA Transactions*, **110**, 39-52.

[22] 罗傲, 肖文彬, 周琪, 等. 基于强化学习的一类具有输入约束非线性系统最优控制[J/OL]. 控制理论与应用, 2021.

[23] Zhu, Y.H., Zhao, D.B. and Li, X.J. (2017) Iterative Adaptive Dynamic Programming for Solving Unknown Nonlinear Zero-Sum Game Based on Online Data. *IEEE Transactions on Neural Networks & Learning Systems*, **28**, 714-725. https://doi.org/10.1109/TNNLS.2016.2561300

[24] Yasini, S., Sistani, M.B. and Karimpour, A. (2014) Approximate Dynamic Programming for Two-Player Zero-Sum Game Related to H_∞ Control of Unknown Nonlinear Continuous-Time Systems. *International Journal of Control, Automation and Systems*, **13**, 99-109. https://doi.org/10.1007/s12555-014-0085-5

[25] Song, R. and Zhu, L. (2019) Stable Value Iteration for Two-Player Zero-Sum Game of Discrete-Time Nonlinear Systems Based on Adaptive Dynamic Programming. *Neurocomputing*, **340**, 180-195.

[26] Vamvoudakis, K.G., Safaei, F.R.P. and Hespanha, J.P. (2019) Robust Event-Triggered Output Feedback Learning Algorithm for Voltage Source Inverters with Unknown Load and Parameter Variations. *International Journal of Robust and Nonlinear Control*, **29**, 3502-3517. https://doi.org/10.1002/rnc.4565

[27] Yang, D.S., Li, T. and Zhang, H.G. (2019) Event-Trigger-Based Robust Control for Nonlinear Constrained-Input Systems Using Reinforcement Learning Method. *Neurocomputing*, **340**, 158-170.

[28] 张正义, 赵学艳. 基于 Q 学习算法的随机离散时间系统的随机线性二次最优追踪控制[J]. 南京信息工程大学学报, 2020, 13(5): 548-555.

[29] Abouheaf, M.L., Lewis, F.L. and Vamvoudakis, K.G. (2014) Multi-Agent Discrete-Time Graphical Games and Reinforcement Learning Solutions. *Automatica*, **50**, 3038-3053.

[30] Yang, N., Xiao, J.W. and Wang, Y.W. (2018) Non-Zero Sum Differential Graphical Game: Cluster Synchronisation for Multi-Agents with Partially Unknown Dynamics. *International Journal of Control*, **92**, 2408-2419. https://doi.org/10.1080/00207179.2018.1441550

[31] Jiang, H., Zhang, H.G. and Han, J. (2018) Iterative Adaptive Dynamic Programming Methods with Neural Network Implementation for Multiplayer Zero-Sum Games. *Neurocomputing*, **307**, 54-60.

[32] Liu, D.R., Li, H.L. and Wang, D. (2013) Neural-Network-Based Zero-Sum Game for Discrete-Time Nonlinear Systems via Iterative Adaptive Dynamic Programming Algorithm. *Neurocomputing*, **110**, 92-100.

[33] 李传江, 马广富. 最优控制[M]. 北京: 科学出版社, 2011: 216-218.

[34] 吴受章. 最优控制理论与应用[M]. 北京: 机械工业出版社, 2007: 193-194.

[35] Luy, N.T. (2017) Distributed Cooperative H_∞ Optimal Tracking Control of Mimo Nonlinear Multi-Agent Systems in Strict-Feedback Form via Adaptive Dynamic Programming. *International Journal of Control*, **91**, 952-968. https://doi.org/10.1080/00207179.2017.1300685

[36] Jiao, Q., Modares, H. and Xu, S.Y. (2016) Multi-Agent Zero-Sum Differential Graphical Games for Disturbance Rejection in Distributed Control. *Automatica*, **69**, 24-34.

[37] Vamvoudakis, K.G., Lewis, F.L. and Hudas, G.R. (2012) Multi-Agent Differential Graphical Games: Online Adaptive Learning Solution for Synchronization with Optimality. *Automatica*, **48**, 1598-1611.