****

本科生毕业设计(论文)附件

外文文献原文和译文

无人矿卡编队算法

题 目： \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

研究与实现

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

熊一帆

作 者： \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

42004178

学 号： \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

机械工程学院

学 院： \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

车辆工程

专 业： \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

2024年 06 月

目 录

[附件A：外文文献的中文译文 1](#_Toc4764743)

[附件B：外文文献的外文原文 3](#_Toc4764744)

附件A：外文文献的中文译文

|  |
| --- |
| **外文文献原文的文献著录信息：**  Zhang Y, Mu C, Zhang Y, et al. Heuristic dynamic programming-based learning control for discrete-time disturbed multi-agent systems[J]. Control Theory and Technology, 2021, 19: 339-353. |
| **外文文献著录信息的中文翻译：**  张瑶，穆朝旭，张勇，杨鹤峰，等. 基于启发式动态规划的离散时间扰动多智能体系统学习控制[J]. 控制理论与技术，2021，19：339-353. |

摘 要

由于在许多领域的广泛应用，同步问题在多智能体系统中得到了广泛研究。多智能体系统的同步是一个关键问题，它意味着在设计的控制策略下，系统的输出或每个智能体的状态都能与领航者保持一致。本文旨在研究一种基于启发式动态编程（HDP）的离散时间多智能体系统学习跟踪控制方法，同时考虑系统干扰以实现同步。此外，由于耦合汉密尔顿-雅各比-贝尔曼方程难以解析求解，本文提出了一种改进的 HDP 学习控制算法，通过动作批判神经网络实现领航者与所有跟随者之间的同步。通过引入辅助动作网络，利用动作和批判神经网络分别学习最优控制策略和成本函数。最后，介绍了两个数值示例和移动机器人的实际应用，以展示基于 HDP 的学习控制算法的控制性能。

关键词： 多智能体系统，启发式动态编程，学习控制，神经网络，同步性

1. 引 言

近年来，随着人工智能技术的飞速发展，多智能体系统逐渐成为一个热门话题，引起了研究人员的广泛讨论[1-3]。多智能体系统的控制在理论和实践研究中都得到了广泛关注，如编队控制[4]、共识控制[5]和群集行为控制[6]等。在这些研究课题中，由于共识控制在工程中的广泛应用，许多学者特别关注共识控制，例如无人机的编队控制、海底机器人的协同控制和卫星的姿态控制等。共识问题是指在智能体之间局部耦合的情况下，领航者的状态能够被所有跟随者同步。

对于多智能体系统，由于每个智能体的行为都是由其邻居和自身共同决定的，因此建立了一个耦合哈米隆-贾可比-贝尔曼（HJB）方程。因此，解决共识控制的关键在于找到耦合 HJB 方程的解。然而，由于存在偏微分项，直接求取 HJB 方程的解十分困难。因此，人们扩展了许多有效的算法来解决这一问题。

最近，强化学习（RL）在人工智能领域取得了显著进展[7-9]。学习过程大致分为两步。首先，通过与环境的交互构建系统奖励。其次，利用反馈机制获得最佳控制策略[10,11]。自适应动态编程（ADP）是 RL 的一个重要分支，其突出作用是有效逼近 HJB 方程的最优解[12-14]。神经网络的理论研究进一步推动了 ADP 方法的发展[15-17]。ADP 方法通常包括两个过程：离线迭代[18]和在线实现[19,20]。ADP 方法主要包括三种基本类型：启发式动态程序设计（HDP）、双重启发式程序设计（DHP）和全局化双重启发式程序设计（GDHP）。最近，ADP 方法被广泛应用于多智能体系统的共识控制中。在文献[21]中，ADP 技术被用于寻找单智能体而非多智能体连续时间线性系统的最优控制器。文献[22]结合合作控制、RL 和博弈论，提出了团队博弈的多智能体在线算法来解决同步控制问题。文献[23]提出了一种基于模糊 ADP 算法的多智能体共识问题最优协调控制方案。该方案结合了博弈论、广义模糊双曲模型和 ADP 方法。然而，上述研究都没有考虑系统中存在的干扰。具体来说，如果考虑到系统中的干扰，基于这些方法的控制性能可能会下降。

在实际应用中，由于环境的复杂性和多变性，多智能体系统的控制问题经常会受到各种干扰的影响，如具体系统模型无法确定所导致的建模不确定性干扰、模型参数干扰，以及风、噪声、温度等多种因素导致的外部干扰。这些干扰的存在不利于系统的稳定性，最终导致难以实现控制目标。因此，在现代控制理论中，如何处理系统中的干扰成为一个重要问题[24-26]。在[27]中，通过循环小增益定理，在扰动系统中设计分散最优控制策略可实现渐近稳定性。Lin[28]采用投影法来处理干扰。[28]中的思想在[29]中被进一步用于考虑了非匹配干扰的非线性系统，使用了 ADP 方法。此外，在多智能体共识控制的研究中，对有扰动系统的研究更为重要。Cao 等人[30]提出了一种分布式扩展状态观测器。其最终目的是实现具有相同线性动力学和未知外部扰动的多智能体系统的共识。文献[31]设计了一种扰动观测器，用于研究不匹配不确定性下二阶多智能体系统的滑模控制。在[32]中，通过估计未知频率和拒绝有界干扰，获得了具有不确定外部参数的多智能体系统的精确优化解。在[33]中，提出了一种分布式优化控制器来消除由一组已知频率正弦信号组成的有界干扰。

所提出的方法在应用范围上有一定的局限性，主要应用于连续时间多智能体系统。然而，对于应用范围更广的离散时间多智能体系统，很少有研究关注用 HDP 方法求解的带扰动离散时间多智能体系统的共识问题；同时，扰动多智能体系统作为控制对象具有实际应用意义。此外，HDP 算法还有许多独特的优点。因此，本文提出了一种新型的带干扰离散时间多智能体系统学习控制方案。其最终目标是使所有跟随智能体与通信图下的领航者同步。本文的贡献如下：

(1) 针对离散时扰动多智能体系统提出了一种学习控制方法，该方法本质上是一种近似最优控制方案，并从部分邻域通信中学习最优控制策略。

(2) 发展了改进的 HDP 算法，以估计迭代控制策略和神经网络实现的代价函数的方式获得最优控制。

(3) 提出了离散时扰动多智能体系统学习控制的理论保证。

(4) 将所提出的 HDP 算法与 LQR 方法在快速性和准确性方面进行了比较，并通过仿真结果证明了 HDP 算法的优越性。

本文的组织结构如下。第2节中，建立了带扰动的离散-时间多智能体系统的初步模型。第3节，HDP 算法的设计方法、动作批判神经网络的实现。第3节提出了 HDP 算法的设计方法、动作批判神经网络的实现以及稳定性分析。第4节通过两个数值示例和一个实际应用证实了上述方法的有效性。第5节对本文进行了总结。

1. 问题的提出
   1. 代数图论和同步问题

有向图的定义是 .它由以下三个元素组成： 表示顶点、 表示边， 是加权邻接矩阵，其中 。当且仅当 则 则表示节点 则表示节点 否则 。节点 的邻居定义为 。内度矩阵 用对角元素表示，即 表示节点的加权内度。表示节点的加权内度 .拉普拉斯矩阵定义为 .有向路径描述为一连串的边 到节点 到节点 ，使得 。

对于一个通信图 ，所研究的包含 的多智能体系统一般描述如下：

leader的定义如下：

其中 是领航者的目标状态。

备注 1 引导者应产生发散信号或正弦参考轨迹，从而使所有的 的所有特征值都应位于单位圆盘的外侧或边缘。原因是，如果指令轨迹稳定，则指令轨迹最终会收敛。 是稳定的。因此，针对不稳定的指令轨迹设计控制策略更有意义。

由于在探索最优控制策略的过程中，只涉及到智能体本身和相应的相邻智能体。因此，同步问题可以描述为 对于任意智能体 .则每个智能体的部分邻域误差 定义为

其中 是传感器的引脚增益 和 。注意 即 当且仅当信息由领航者直接传递给智能体时 否则 。全局跟踪误差 的 个智能体的全局跟踪误差定义为

其中 表示拉普拉斯矩阵； 是由对角线元素 是 Kronecker 乘积算子。我们定义 和 即 ，其中 是一个 -维向量，且 是一个 为同一矩阵。等式 (4) 可改写为

其中 是全局同步误差向量，且 为非奇异值，条件是图包含一棵生成树，且智能体 与领航者直接相连。

定理 1 如果 为非正弦，则全局同步误差 为

其中 是 的最小有符号数。

根据定理 1，如果全局跟踪误差 收敛为零，则全局同步误差 也收敛为零。因此，如果全局跟踪误差趋于足够小，系统就会实现同步。

显然，扰动包含在部分邻域误差中，这意味着智能体之间在信息交流过程中可能存在偏差。下文将研究受干扰多智能体系统的共识控制。

* 1. 受干扰多智能体系统的共识控制公式

对于受干扰的多智能体系统 (1)，我们将干扰 分解为匹配分量和非匹配分量之和，将其投影到 投影到矩阵 。因此可得出

其中 是 的伪逆矩阵，而 是标识矩阵。那么，具有相同通信网络的辅助系统 设计如下：

为引入的辅助控制策略。那么辅助系统 (11) 的部分邻域误差可进一步表示为

设计每个智能体的控制输入时，需要邻近智能体的信息。 的控制策略描述为 的控制策略描述为

因此，考虑受干扰的多智能体系统共识问题、最优控制策略 和最优辅助控制策略 通过最小化局部性能指标

其中 是干扰的相关指数。 是上界，满足

是一个正常数。效用函数满足 , .效用函数通常用二次方形式表示如下：

与 , 和 为对称正定矩阵。具有成本函数 (15) 的 (11) 名义系统的最优控制可用于受干扰的多智能体系统 (1) [28, 34]。因此，哈密顿函数表述为

其中 .最优成本函数 根据贝尔曼最优性原理可得

其中 .将（12）代入（17），最优控制策略 求解为 而辅助控制策略 求得 ，因此 和 分别表述如下

1. 基于 HDP 的受干扰多智能体系统学习控制
   1. 迭代算法的收敛性

定理 1 假设图中包含一棵生成树，让 满足 (19)，且最优控制策略满足 (20)。则部分邻域误差 是渐近稳定的，可以实现同步目标。

证明 首先，定义 及其梯度定义如下

定理 2 根据哈密顿方程 (18)，局部性能指标满足以下离散时间哈密顿-雅可比方程：

如果 满足（18）且 表示为 (20)、 .根据定理 2，我们可以得到

根据 (18)、(22) 和 (23)，可以进一步推导出 (26)。

因此 作为系统（9）的李亚普诺夫函数，可以证明邻域误差系统是渐近稳定的，即 收敛为零，当 .这意味着，当多智能体系统考虑邻域误差中的扰动时，系统是渐近稳定的。根据定理 1，全局同步误差向量 。因此，所有智能体都会与leader同步。定理 1 由此得到证明。

* 1. 利用启发式动态编程实现多智能体系统学习控制

开发了一种行动批判神经网络，用于在受干扰的多机器人系统中使用 HDP 算法实施学习控制。

批判网络旨在近似估计每个智能体的成本函数 而行动网络则是为了近似估计控制策略。众所周知 并不是存在干扰的系统的真正控制策略，而是有助于近似最优控制策略的辅助控制策略 。输出 和 分别表示为

其中 和 分别是批判、行动和辅助行动网络权重。 是智能体 及其邻居的总数。 表示智能体 及其邻居在 可表示为 。目标成本函数的计算公式为

接下来，批评者网络的近似误差表示为，近似的目标函数表示为

根据梯度下降法，批判者网络的权重更新如下：

其中 是批评者网络的学习率。

目标控制策略 和 分别为

同样，行动网络和辅助行动网络的近似误差分别表示为 和 ，反向传播的误差函数表示如下：

然后，动作网络和辅助动作网络的权值更新过程如下：

其中 和 分别是行动网络和辅助行动网络的学习率。

由神经网络实现的 HDP 结构下的算法程序基于以下步骤进行：

步骤 1 设置参数 和 .步骤 2 初始化批评网络权重 和动作网络权重 .

步骤 3 初始化系统状态 .

步骤 4 由 (28) 计算。

第 5 步 按 (29) 计算。

步骤 6 和 将 和 引入（12），得到下一个系统动态 .

步骤 7 使用 (27)。

步骤 8 利用 (40) 训练动作网络。

步骤 9 利用 (41) 训练辅助行动网络。

步骤 10 计算 和 分别由 (27) 和 (30) 计算得出。

步骤 11 利用 (33) 训练最大步长的批判网络 使用 (33)。

采用 HDP 结构的近似最优跟踪控制框架如图 1 所示。从本质上讲，受扰多智能体系统的学习控制是一个最优控制问题，其最终目标是最小化成本函数。通过对不确定性的变形，构建与受扰多智能体系统相关的辅助系统来解决控制问题。通过引入辅助动作网络，利用批判神经网络和动作神经网络分别学习代价函数和控制策略。值得注意的是，与一般的神经网络不同，辅助控制策略 有助于获得实际的最优控制策略 ，因此 并不是真正的控制策略。

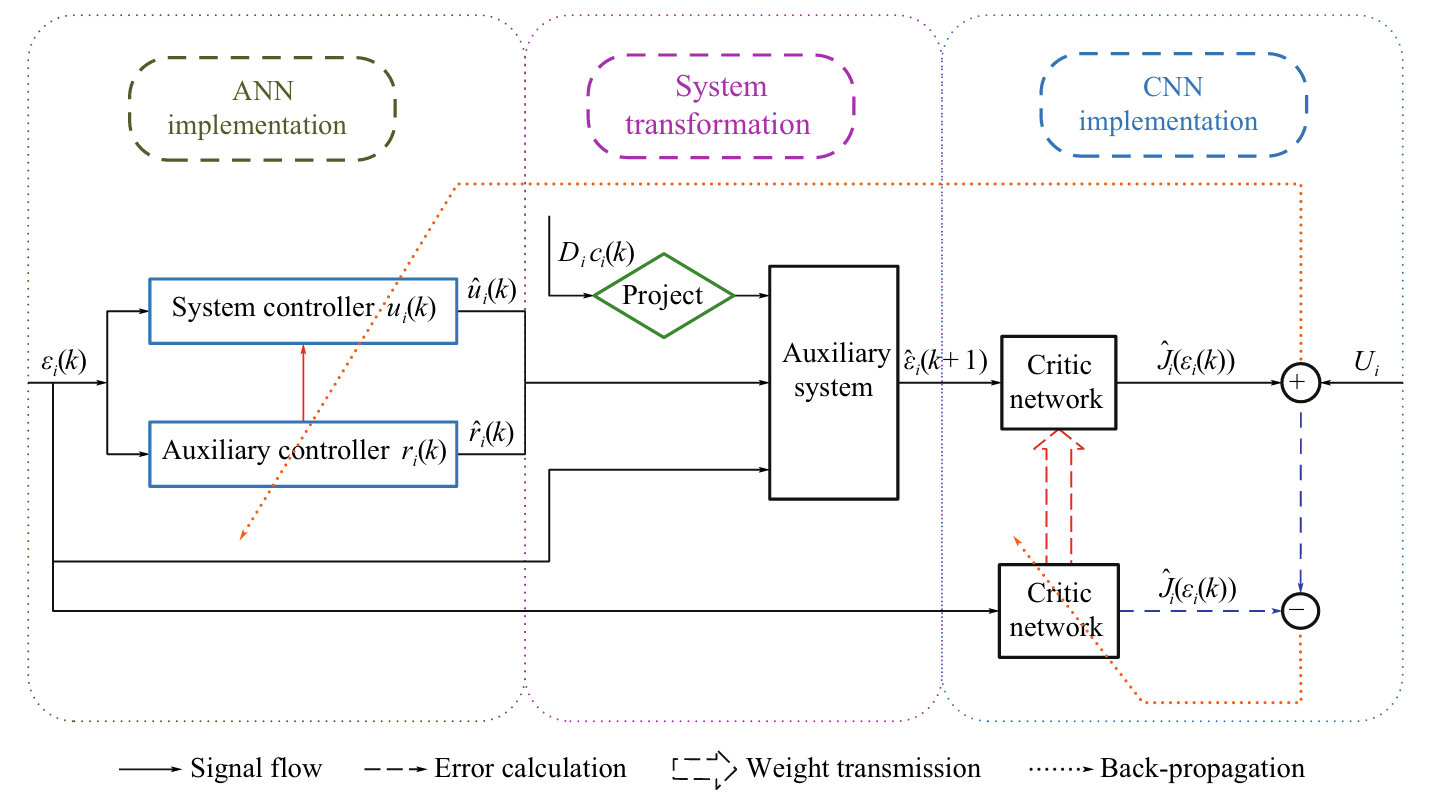


图 1 对受干扰的多智能体系统采用基于学习的 HDP 控制结构

* 1. 稳定性分析

估计权重 和 分别用于批判网络、行动网络和辅助行动网络。最优权重向量表示为 和 ，因此权重误差表述为 , 和 。目标成本函数 可定义为

其中 表示任意小的重建误差。

假设 1 权向量 和 分别有界，即 和 .同时，重建误差 , 和 是有界的，即 和 .

定理 2 如果所有系统状态 都是可测的，则最优控制策略可近似为 (28)。权重估计误差 和 由假设 1 一致终界。

证明 Lyapunov 函数选为

然后，我们定义

首先，我们将 为

根据 (27)、(32) 和 (42)，我们得到

定义 .应用考希-施瓦茨不等式，(44) 可进一步表述为

之差被视为

然后将 (40) 代入 (47)，可得

同样，我们定义

利用考希-施瓦茨不等式、可表示为

最后 得到

然后，将（41）代入（52），可以得到

定义

根据考希-施瓦茨不等式、 可进一步表述为

根据 (43)，我们可以得出结论

定义

其中 和 是 , 和 的上界。随着

可以获得。因此 和 是均匀最终有界的。定理 2 得到证明。

1. 模拟研究

本节将研究三个典型的模拟实例。

* 1. 四智能体系统

图 2 给出了一个四智能体系统的有向图通信结构。带干扰的四智能体系统模型选为

leader的模型是

其中系统矩阵的选择如下

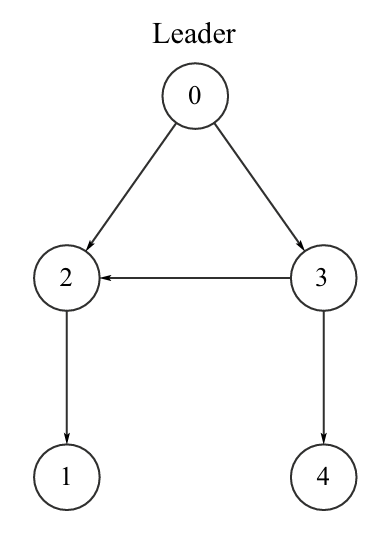


图 2 四个智能体的网络结构

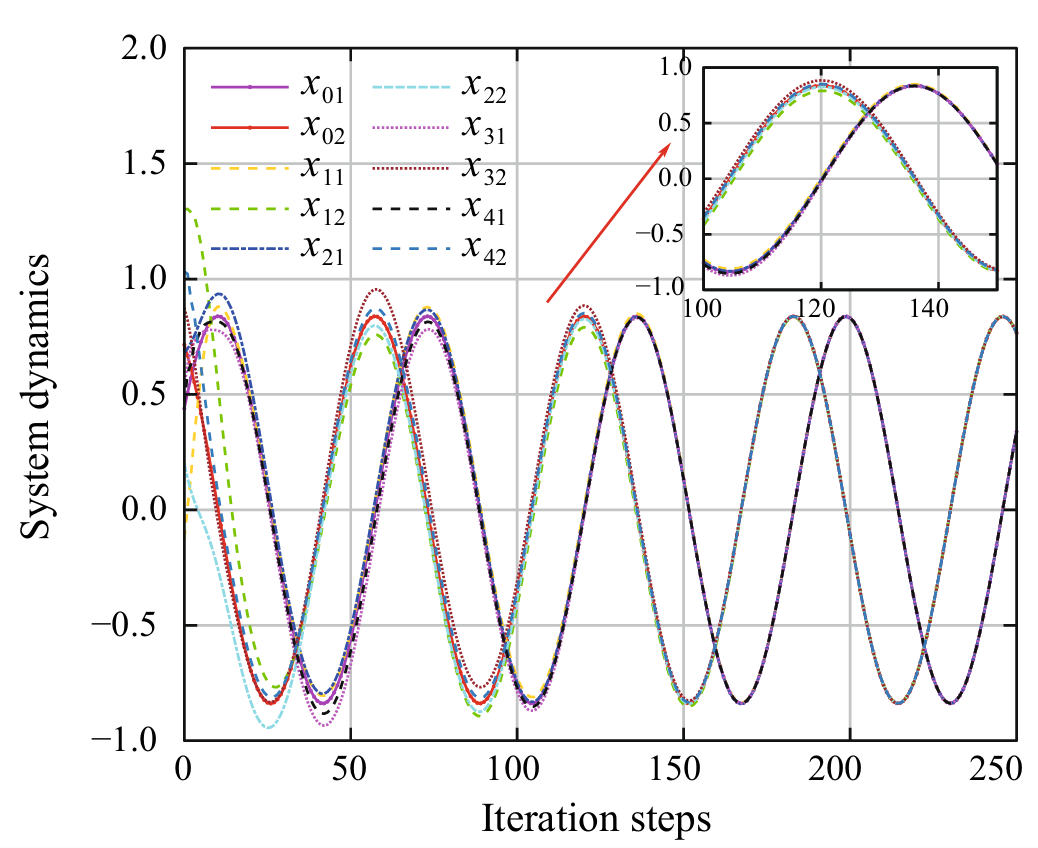


图 3 智能体状态与迭代步骤的关系

系统干扰是

销钉增益为 .边缘权重选取为 。选择性能指标权重矩阵 , .leader和每个智能体的初始状态都是随机设定的。选择学习率为 。最大步长 选为 1500，这个步长足以让所有智能体与leader保持同步。

图 3 显示了智能体和leader的动态。图 4 显示了系统的相平面图。从图中可以看出，所有智能体都能准确地跟踪leader。图 5 反映了共识控制策略。

* 1. 与线性二次调节器（LQR）的算法比较

接下来，我们将进一步研究三智能体系统，有向图通信结构如图 6 所示。

系统矩阵的选择如下：

选择干扰如下：

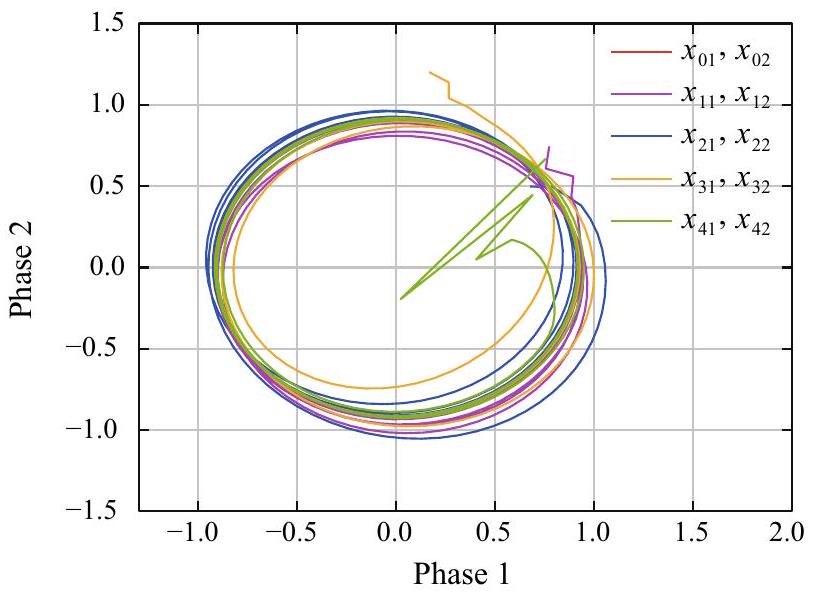


图 4 相平面图

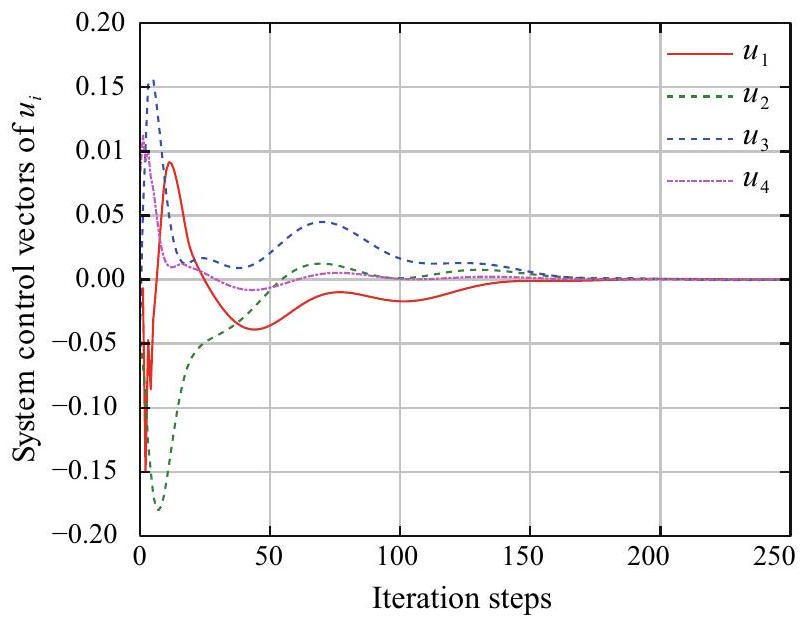


图 5 四智能体系统的共识控制策略

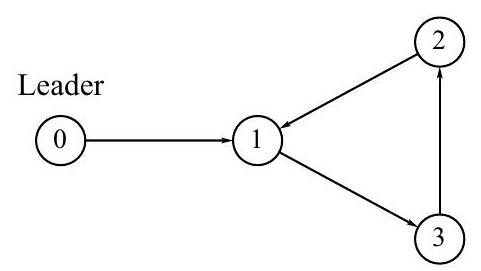


图 6 有三个智能体的网络结构

其中 是系统的未知参数。在训练过程中，未知参数 被选中，而 。

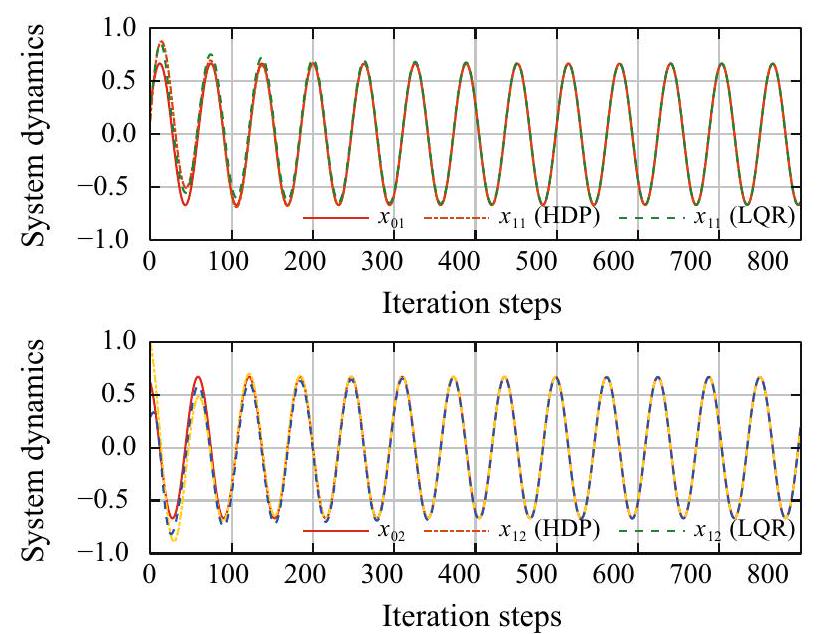


图 7 智能体 1 的跟踪性能

我们选择销钉增益 .边缘权重选为 , .性能指标权重矩阵选为 , , .leader和每个智能体的初始状态都是从 .设置学习率为 分别为

为了比较性能，分别采用 HDP 算法和 LQR 方法获得控制策略，仿真结果如下：各智能体跟踪leader的性能和跟踪误差动态分别如图 7、图 8、图 9、图 10、图 11 和图 12 所示。控制策略 如图 13 所示。

为了进一步说明算法的控制性能，表 1 列出了 LQR 方法和 HDP 算法的均方根误差、绝对平均误差和迭代步数，以便更清晰、更严谨地从快速性和准确性两个方面比较两种算法的性能指标。

比较结果表明，LQR 方法和 HDP 算法都能实现所有智能体与leader之间的同步。但一方面，基于 HDP 控制方法的均方根误差和绝对平均误差均小于 LQR 方法，这表明 HDP 算法在同步精度上优于 LQR 方法。另一方面，收敛精度定义为 当误差收敛到一定范围时，就表示步数。在 LQR 方法下，所有智能体都能在大约 700 个迭代步后跟踪leader，而在 HDP 算法下，大约 300 个迭代步后就能实现同步。与 LQR 方法相比，它表明

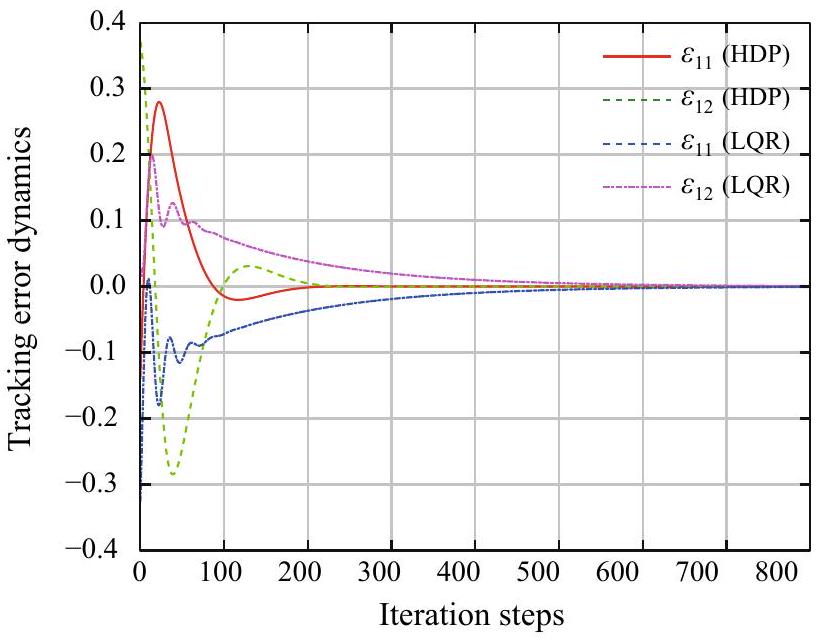


图 8 智能体 1 的跟踪误差

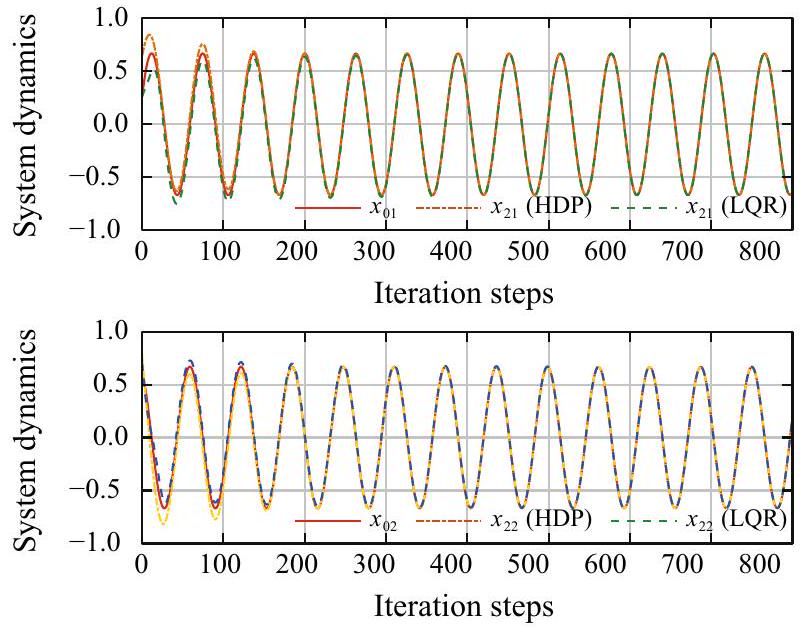


图 9 智能体 2 的跟踪性能

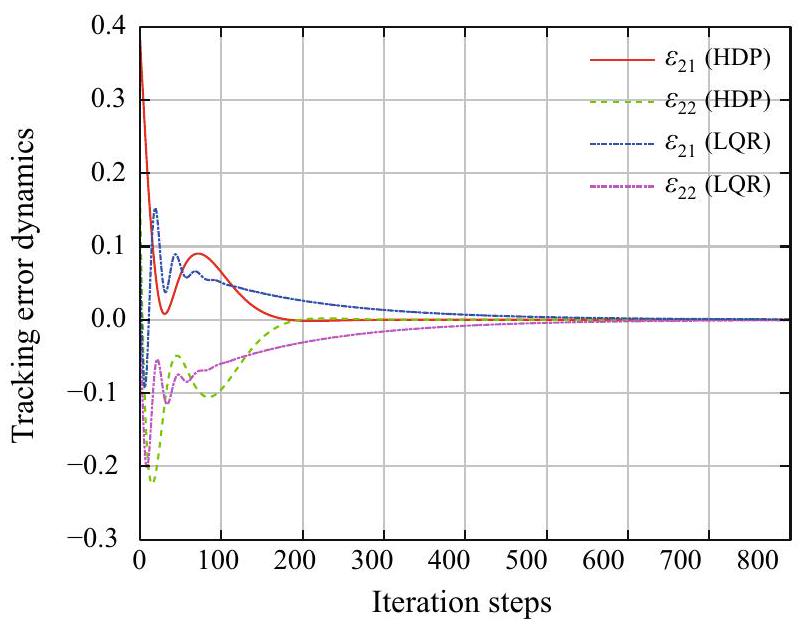


图 10 智能体 2 的跟踪误差

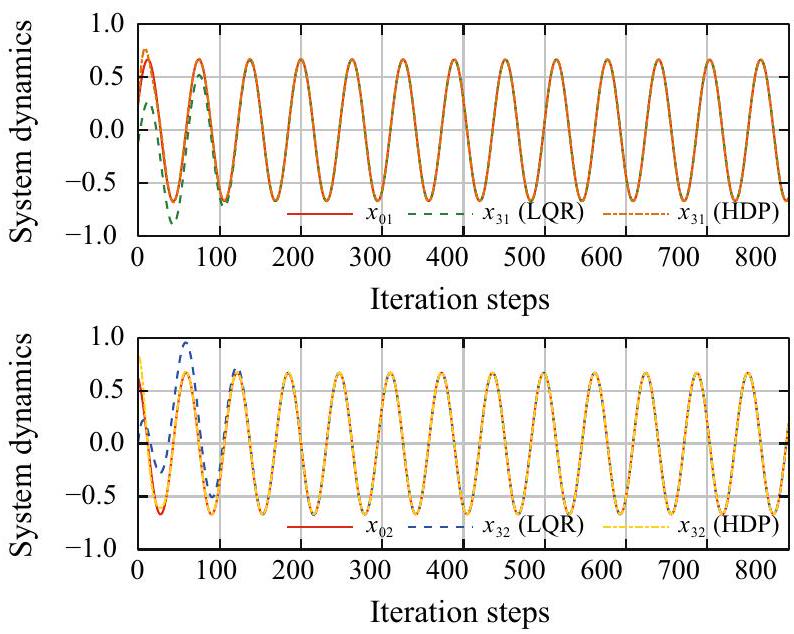


图 11 智能体 3 的跟踪性能

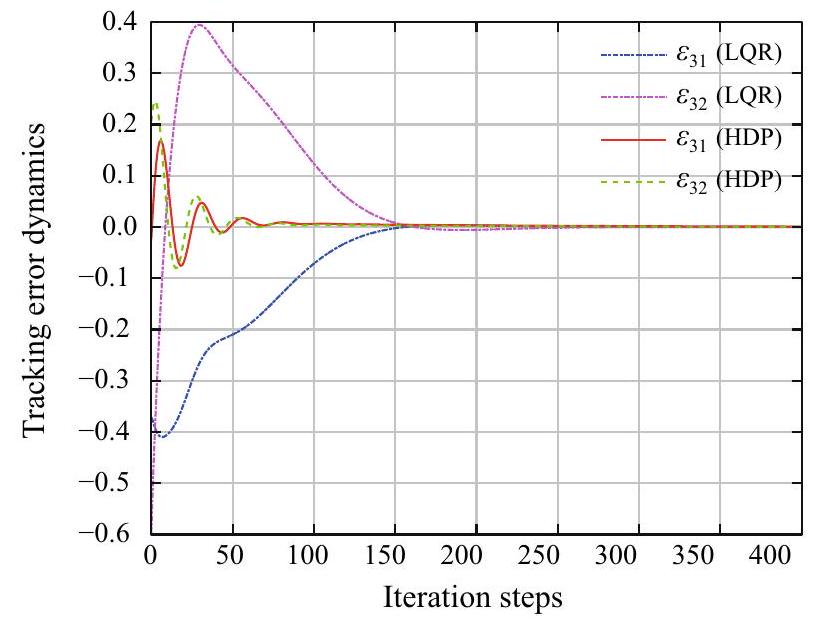


图 12 智能体 3 的跟踪误差

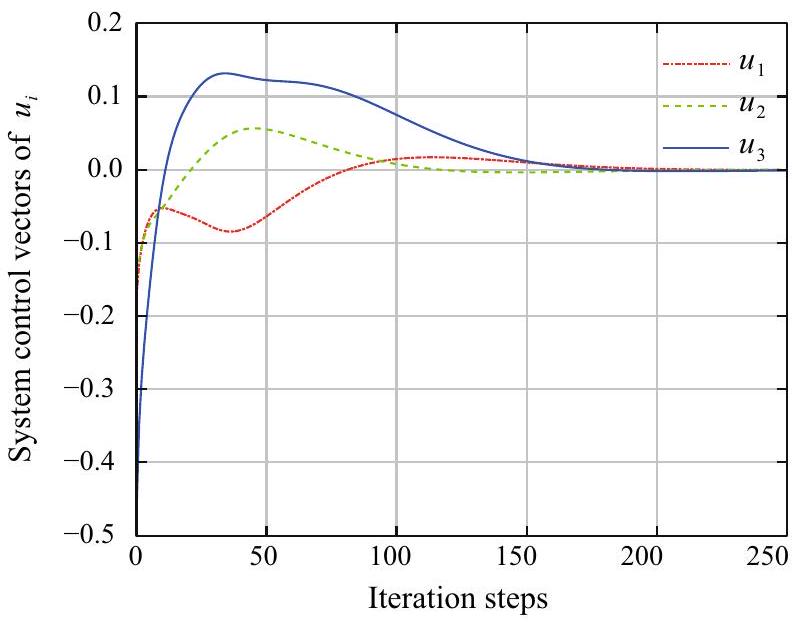


图 13 三智能体系统的控制策略

表 1 HDP 算法与 LQR 方法的性能比较

|  |  |  |
| --- | --- | --- |
| 评估参数 | LQR | HDP |
| 智能体 1 的迭代步骤 | 796 | 312 |
| 智能体 2 的迭代步骤 | 765 | 298 |
| 3 号智能体的迭代步骤 | 730 | 281 |
| 绝对平均误差 |  |  |
| 均方根误差 |  |  |

在考虑干扰因素时，HDP 方法在收敛速度方面具有优势。

总的来说，在对三项性能指标进行比较后，我们可以得出结论：在准确性和快速性方面，建议的 HDP 算法得出的控制策略比 LQR 方法具有更好的跟踪性能。

此外，还考虑了多智能体系统中的参数不确定性，以显示拟议控制方案的有效性。

考虑一个线性多智能体系统，该系统由 个智能体，其中 个智能体的动力学可以表示为

其中 是实矩阵函数，代表多智能体系统中的时变参数不确定性。不确定性是模型线性化的结果，通常假定其形式为

其中 是实数已知常数矩阵，代表了 进入名义矩阵 是未知的真实时变矩阵，其 Lebesgue 可测元素满足

相关的系统矩阵如下：

其他参数设置与 4.2 相同。

图 14 是给定参数下系统的相应跟踪曲线。不难看出，所有智能体的状态都与leader完全同步。此外，图 15 还给出了相应的跟踪误差动态。此外，我们还研究了该方法和具有不确定参数的 LQR 方法在相同条件下的跟踪效果。跟踪误差如图 16 所示。可以明显看出，在

在参数不确定的前提下，采用 HDP 算法的多机器人系统的跟踪效果在快速性方面仍然优于 LQR 方法。

* 1. 多智能体系统的实际应用

为了验证上述理论结果在实际应用场景中的有效性，我们提出了一个应用多智能体系统，该系统由三个移动机器人和一个领导机器人组成。

机器人在一维欧几里得空间中运动，目的是最终实现状态和速度的同步。

三个从动机器人分为两个子系统。第一个子系统如下所述：

第二个子系统的说明如下：

其中 和 是机器人在时间瞬时的状态矢量、速度矢量和控制策略 的控制策略 和 分别是状态系数和采样间隔系数。 和 是设计参数。 表示扰动，具体描述如下

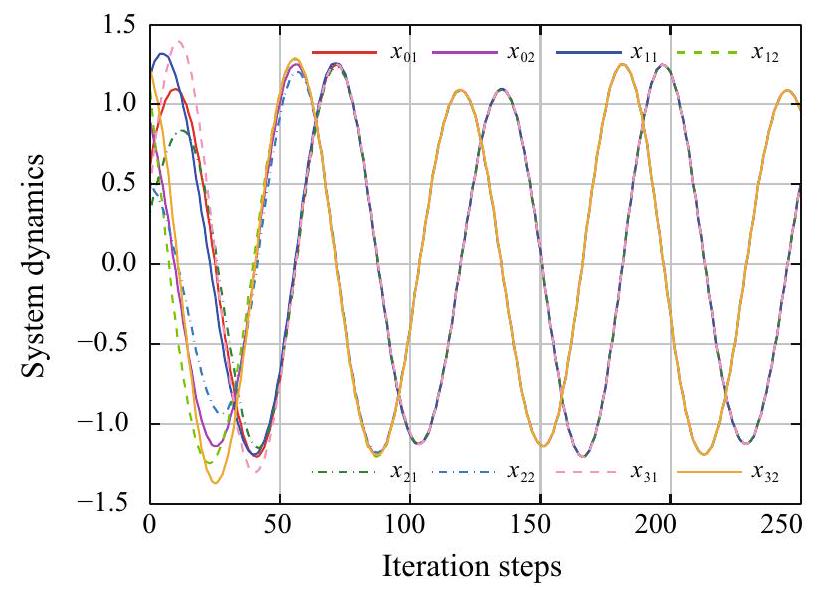


图 14 智能体状态与迭代步骤的关系

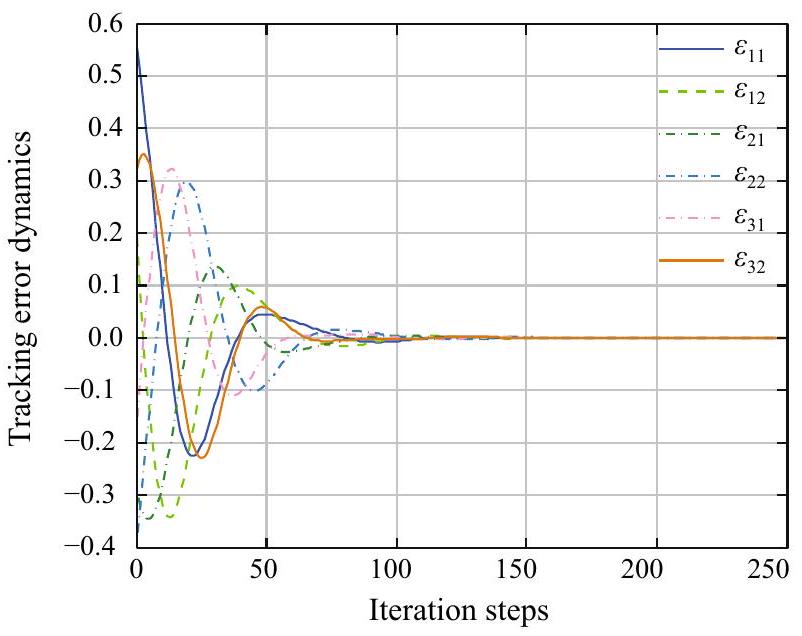


图 15 HDP 算法下的跟踪误差

其中未知参数 选择为 .其他参数和初始化器与例 2 相同。领头机器人的动力学特性如下：

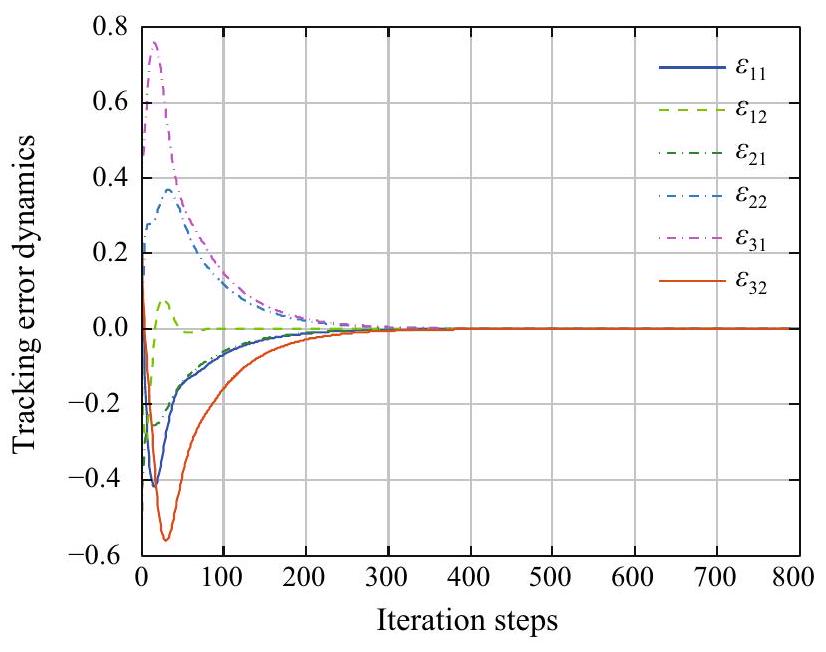


图 16 LQR 方法下的跟踪误差

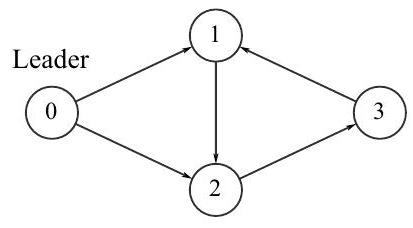


图 17 三个跟随型移动机器人和一个领导型机器人的通信结构

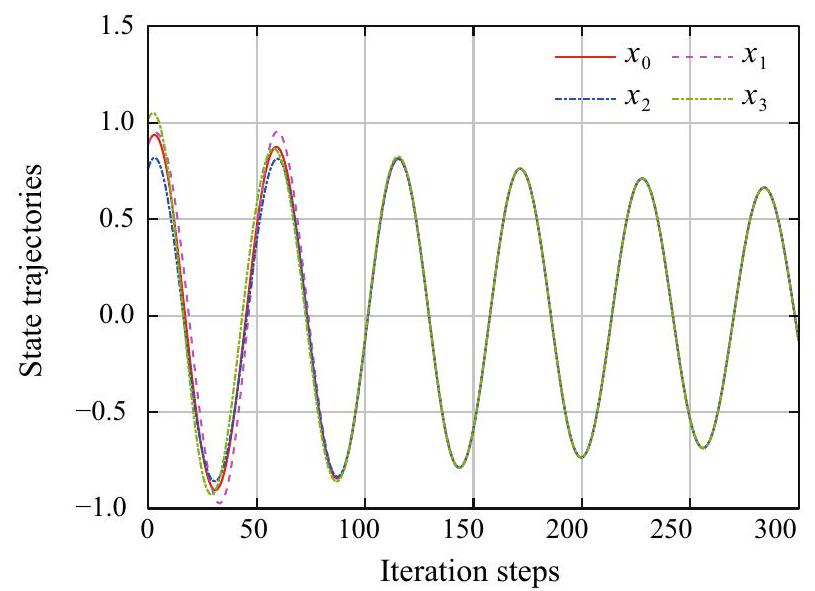


图 18 三个机器人和一个leader的状态响应

三个移动机器人和一个领导机器人的通信结构如图 17 所示。三个跟随移动机器人和一个领导机器人的状态和速度响应分别如图 18 和图 19 所示。三个跟随移动机器人和一个领导机器人的状态和速度误差分别如图 20 和图 21 所示。

1. 结论

在本文中，学习跟踪控制的目标是实现leader和所有跟随者之间的同步。带干扰的多智能体系统衍生出其标称系统的跟踪控制。应用 HDP 算法并通过神经网络实现。同时，介绍了动作批判神经网络的稳定性。最后，通过三个具有代表性的仿真研究，证明了基于 HDP 的学习跟踪控制策略的正确性和优越性。很明显，本文提到的改进型 HDP 算法在跟踪速度和考虑干扰的效果方面都有很好的表现。同时，在实际应用中，该算法也能很好地实现同步。并将一些有意义的工作扩展到了非线性多智能体系统。

参考文献

1. Abouheaf M I, Lewis F L, Vamvoudakis K G, et al. Multi-agent discrete-time graphical games and reinforcement learning solutions[J]. Automatica, 2014, 50(12): 3038-3053.Song, R., Lewis, F. L., & Wei, Q. (2017).
2. Song R, Lewis F L, Wei Q. Off-policy integral reinforcement learning method to solve nonlinear continuous-time multiplayer nonzero-sum games[J]. IEEE transactions on neural networks and learning systems, 2016, 28(3): 704-713.
3. Mu C, Zhao Q, Sun C. Optimal model-free output synchronization of heterogeneous multiagent systems under switching topologies[J]. IEEE Transactions on Industrial Electronics, 2019, 67(12): 10951-10964.
4. Yan J, Guan X, Luo X, et al. Formation and obstacle avoidance control for multiagent systems[J]. Journal of Control Theory and Applications, 2011, 9(2): 141-147.
5. Wu Y, Hu B, Guan Z H. Exponential consensus analysis for multiagent networks based on time-delay impulsive systems[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 49(6): 1073-1080.
6. Lu Y, Guo Y, Dong Z. Multiagent flocking with formation in a constrained environment[J]. Journal of Control Theory and Applications, 2010, 8: 151-159.
7. Mu C, Wang K. Single‐network ADP for near optimal control of continuous‐time zero‐sum games without using initial stabilising control laws[J]. IET Control Theory & Applications, 2018, 12(18): 2449-2458.
8. Liu L, Wang Z, Zhang H. Adaptive fault-tolerant tracking control for MIMO discrete-time systems via reinforcement learning algorithm with less learning parameters[J]. IEEE Transactions on Automation Science and Engineering, 2016, 14(1): 299-313.
9. Chen X, Chen G, Cao W, et al. Cooperative learning with joint state value approximation for multi-agent systems[J]. Journal of Control Theory and Applications, 2013, 11(2): 149-155.
10. Sokolov Y, Kozma R, Werbos L D, et al. Complete stability analysis of a heuristic approximate dynamic programming control design[J]. Automatica, 2015, 59: 9-18.
11. Sun X, Mao T, Ray L, et al. Hierarchical state-abstracted and socially augmented Q-Learning for reducing complexity in agent-based learning[J]. Journal of Control Theory and Applications, 2011, 9: 440-450.
12. Mu C, Zhao Q, Sun C, et al. An ADDHP-based Q-learning algorithm for optimal tracking control of linear discrete-time systems with unknown dynamics[J]. Applied Soft Computing, 2019, 82: 105593.
13. Wang F Y, Zhang H, Liu D. Adaptive dynamic programming: An introduction[J]. IEEE computational intelligence magazine, 2009, 4(2): 39-47.
14. Mu C, Ni Z, Sun C, et al. Air-breathing hypersonic vehicle tracking control based on adaptive dynamic programming[J]. IEEE transactions on neural networks and learning systems, 2016, 28(3): 584-598.
15. Wei Q, Liu D. Finite horizon optimal control of discrete-time nonlinear systems with unfixed initial state using adaptive dynamic programming[J]. Journal of Control Theory and Applications, 2011, 9(3): 381-390.
16. Wen S, Zeng Z, Chen M Z Q, et al. Synchronization of switched neural networks with communication delays via the event-triggered control[J]. IEEE transactions on neural networks and learning systems, 2016, 28(10): 2334-2343.
17. Ding J, Balakrishnan S N. Approximate dynamic programming solutions with a single network adaptive critic for a class of nonlinear systems[J]. Journal of Control Theory and Applications, 2011, 9(3): 370-380.
18. Zhang H, Luo Y, Liu D. Neural-network-based near-optimal control for a class of discrete-time affine nonlinear systems with control constraints[J]. IEEE Transactions on Neural Networks, 2009, 20(9): 1490-1503.
19. Vrabie D, Pastravanu O, Abu-Khalaf M, et al. Adaptive optimal control for continuous-time linear systems based on policy iteration[J]. Automatica, 2009, 45(2): 477-484.
20. Vamvoudakis K G, Lewis F L. Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem[J]. Automatica, 2010, 46(5): 878-888.
21. Jiang Y, Jiang Z P. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics[J]. Automatica, 2012, 48(10): 2699-2704.
22. Vamvoudakis K G, Lewis F L, Hudas G R. Multi-agent differential graphical games: Online adaptive learning solution for synchronization with optimality[J]. Automatica, 2012, 48(8): 1598-1611.
23. Zhang H, Zhang J, Yang G H, et al. Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming[J]. IEEE Transactions on Fuzzy Systems, 2014, 23(1): 152-163.
24. Yang H, Wang F, Zhang Z, et al. Consensus of multiagent systems based on disturbance observer[J]. Journal of Control Theory and Applications, 2010, 8(2): 145-150.
25. Li Y, Liu L, Feng G. Robust adaptive output feedback control to a class of non-triangular stochastic nonlinear systems[J]. Automatica, 2018, 89: 325-332.
26. Xiao B, Yang X, Karimi H R, et al. Asymptotic tracking control for a more representative class of uncertain nonlinear systems with mismatched uncertainties[J]. IEEE Transactions on Industrial Electronics, 2019, 66(12): 9417-9427.
27. Jiang Y, Jiang Z P. Robust adaptive dynamic programming for large-scale systems with an application to multimachine power systems[J]. IEEE Transactions on Circuits and Systems II: Express Briefs, 2012, 59(10): 693-697.
28. Lin F. An optimal control approach to robust control design[J]. International Journal of control, 2000, 73(3): 177-186.
29. Mu C, Zhang Y, Gao Z, et al. ADP-based robust tracking control for a class of nonlinear systems with unmatched uncertainties[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2019, 50(11): 4056-4067.
30. Cao W, Zhang J, Ren W. Leader–follower consensus of linear multi-agent systems with unknown external disturbances[J]. Systems & Control Letters, 2015, 82: 64-70.
31. Wang Q, Sun C, Chai X, et al. Disturbance observer-based sliding mode control for multi-agent systems with mismatched uncertainties[J]. Assembly Automation, 2018, 38(5): 606-614.
32. Wang X, Hong Y, Yi P, et al. Distributed optimization design of continuous-time multiagent systems with unknown-frequency disturbances[J]. IEEE transactions on cybernetics, 2017, 47(8): 2058-2066.
33. Wang X, Hong Y, Ji H. Distributed optimization for a class of nonlinear multiagent systems with disturbance rejection[J]. IEEE transactions on Cybernetics, 2015, 46(7): 1655-1666.
34. Lin F, Brandt R D, Sun J. Robust control of nonlinear systems: Compensating for uncertainty[J]. International Journal of Control, 1992, 56(6): 1453-1459.

附件B：外文文献的外文原文

|  |
| --- |
| **外文文献原文的文献著录信息：**  Zhang Y, Mu C, Zhang Y, et al. Heuristic dynamic programming-based learning control for discrete-time disturbed multi-agent systems[J]. Control Theory and Technology, 2021, 19: 339-353. |
| **外文文献著录信息的中文翻译：**  张瑶，穆朝旭，张勇，杨鹤峰，等. 基于启发式动态规划的离散时间扰动多智能体系统学习控制[J]. 控制理论与技术，2021，19：339-353. |