国家自然科学基金资助项目批准通知

（包干制项目）

张勇威　　先生/女士：

根据《国家自然科学基金条例》、相关项目管理办法规定和专家评审意见，国家自然科学基金委员会（以下简称自然科学基金委）决定资助您申请的项目。项目批准号：62303122，项目名称：通信约束下多智能体系统自学习云优化控制，资助经费：30.00万元，项目起止年月：2024年01月至 2026年 12月，有关项目的评审意见及修改意见附后。

请您尽快登录科学基金网络信息系统（https://grants.nsfc.gov.cn），**认真阅读《国家自然科学基金资助项目计划书填报说明》并按要求填写《国家自然科学基金资助项目计划书》（以下简称计划书）**。对于有修改意见的项目，请您按修改意见及时调整计划书相关内容；如您对修改意见有异议，须在电子版计划书报送截止日期前向相关科学处提出。

请您将电子版计划书通过科学基金网络信息系统（https://grants.nsfc.gov.cn）提交，由依托单位审核后提交至自然科学基金委。自然科学基金委审核未通过者，将退回的电子版计划书修改后再行提交；审核通过者，打印纸质版计划书（一式两份，双面打印）并在项目负责人承诺栏签字，由依托单位在承诺栏加盖依托单位公章，且将申请书纸质签字盖章页订在其中一份计划书之后，一并报送至自然科学基金委项目材料接收工作组。纸质版计划书应当保证与审核通过的电子版计划书内容一致。**自然科学基金委将对申请书纸质签字盖章页进行审核，对存在问题的，允许依托单位进行一次修改或补齐。**

向自然科学基金委提交电子版计划书、报送纸质版计划书并补交申请书纸质签字盖章页截止时间节点如下：

1. **2023年9月7日16点：**提交电子版计划书的截止时间；

2. **2023年9月14日16点：**提交修改后电子版计划书的截止时间；

3. **2023年9月21日：**报送纸质版计划书（一式两份，其中一份包含申请书纸质签字盖章页）的截止时间。

4. **2023年10月7日：**报送修改后的申请书纸质签字盖章页的截止时间。

请按照以上规定及时提交电子版计划书，并报送纸质版计划书和申请书纸质签字盖章页，逾期不报计划书或申请书纸质签字盖章页且未说明理由的，视为自动放弃接受资助；未按要求修改或逾期提交申请书纸质签字盖章页者，将视情况给予暂缓拨付经费等处理。

　　附件：项目评审意见及修改意见表

<div align="right">

国家自然科学基金委员会

2023年8月24日

</div>

受理编号：c2120191024000000603

项目编号：2021A1515110022

文件编号：粤基金字〔2021〕24号

广东省基础与应用基础研究基金项目

合同书

项目名称： 面向多智能体系统的动态事件触发分布式优化容错控制方法

项目类别： 区域联合基金-青年基金项目

项目起止时间： 2021-10-01　至　2024-09-30

管理单位（甲方）： 广东省基础与应用基础研究基金委员会

依托单位（乙方）： 广东工业大学

通讯地址： 广东省广州市越秀区东风东路729号大院

邮政编码： 510006　　　　　　　　　**单位电话：** 020-39322711-000

项目负责人： 张勇威　　　　　　　　　**联系电话：** 13247669804

（广东科技微信公众号）　　　　（查看合同书信息）　　　　（受理纸质材料二维码）

广东省基础与应用基础研究
基金委员会
二〇二〇年制

填写说明

一、项目合同书/任务书内容原则上要求与申报书相关内容保持一致，不得无故修改。

二、项目承担单位通过广东省科技业务管理阳光政务平台下载项目合同书/任务书，按要求完成签名盖章后提交至省科技厅受理窗口。

三、签名盖章说明。请分别在单位工作分工及经费分配情况页、人员信息页、签约各方页等地方按要求签字或盖章，签章不合规或错漏将不予受理。其中，人员信息页要求所有参与人员本人亲笔签名，代签或印章无效，漏签将不予受理。

四、本合同书/任务书自签字并加盖公章之日起生效，各方均应负本任务书的法律责任，不应受机构、人事变动影响。

一、主要研究内容和要达到的目标

本项目主要研究带执行器故障和传感器故障的多智能体系统一致性问题以及编队控制问题。首先，本项目拟采用Schur分解以及线性矩阵不等式技术为每个智能体设计分布式故障观测器，精确的获得每个智能体执行器和传感器的故障信息，并从理论上分析观测误差动态的稳定性。其次，针对多智能体系统的优化容错一致性问题和编队控制问题，根据图论、参考的编队动态以及系统转换得出局部领域协同误差方程。为了处理系统故障，根据分布式故障观测器观测的执行器故障和传感器故障、局部领域协同误差以及相邻智能体的协同控制律，为每个智能体设计能表征故障模态的局部性能指标函数，从而将容错控制问题转换为优化控制问题，并从理论上证明问题转换的等价性。为了节省计算和通信资源，设计动态事件触发条件，并从理论上分析该条件能避免Zeno现象以及能保证闭环系统的稳定性。为了获得优化容错控制器，本项目将采用单评判网络近似性能指标函数，并设计分布式策略迭代算法求解耦合的HJB方程，进而得到每个智能体的动态事件触发分布式优化容错控制器。在稳定性分析方面，本项目将从理论上分析分布式策略迭代算法的收敛性和最优性、神经网络权重误差动态和局部领域协同误差动态的稳定性。最后，本项目将利用Python、Matlab、C++等软件平台，结合NAO机器人、无人机和智能小车等实物平台。设计不同的协同任务，如一致性跟踪、编队飞行以及编队驾驶等，对所提出的动态事件触发分布式优化容错一致性控制方法以及编队控制方法进行实物验证。

一、主要研究内容和要达到的目标

二、项目预期获得的研究成果及形式

<table>
<tr><td rowspan="3">论文及专著情况</td><td>国家统计源刊物以上刊物发表论文（篇）</td><td>5</td><td>科技报告（篇）</td><td colspan="3">2</td></tr>
<tr><td>其中被SCI/EI/ISTP收录论文数（篇）</td><td>5</td><td>培养人才（人）</td><td colspan="3"></td></tr>
<tr><td>专著（册）</td><td></td><td>引进人才（人）</td><td colspan="3"></td></tr>
<tr><td rowspan="3">专利情况(项)</td><td colspan="2">发明专利</td><td colspan="2">实用新型专利</td><td>外观设计专利</td><td>国外专利</td></tr>
<tr><td>申请</td><td>授权</td><td>申请</td><td>授权</td><td>申请　授权</td><td>申请　授权</td></tr>
<tr><td></td><td>2</td><td></td><td></td><td></td><td></td></tr>
<tr><td>其他</td><td colspan="6"></td></tr>
</table>

6

三、项目进度和阶段目标

（一）项目起止时间： 2021-10-01 至 2024-09-30

（二）项目实施进度及阶段主要目标：

开始日期	结束日期	主要工作内容
2021-10-01	2022-12-01	利用Schur分解技术和线性矩阵不等式技术设计分布式故障观测器，得到每个智能体精确的故障信息。给出定理详细分析观测误差的稳定性。依据观测器得出的故障信息，设计能表征故障的性能指标函数，设计分布式策略迭代算法，得出最优容错控制器实现MASs的一致性控制和编队飞行控制。
2023-01-01	2023-12-01	选取合适的内部变量，设计动态事件触发条件，给出定理详细分析该触发条件下闭环系统的稳定性以及Zeno行为，进而得出动态事件触发最优容错控制器。
2024-01-01	2024-09-30	对提出的控制方法，利用Python、Matlab语言编写算法程序进行数值仿真，并利用C++语言编写NAO机器人、无人机和智能小车的接口程序，在实物上验证算法的有效性。最后整理科研成果，完成总结报告。

四、项目总经费及省基金委经费预算

（一）省基金委经费下达总额： （大写）壹拾万圆整；（小写 ）10万元；

（二）省基金委经费年度下达计划：

年度	2021 年	年	年	年	年
经费(万元)	10.00				

（三）总经费及省基金委经费开支预算计划：

经费筹集情况： （单位：万元）

省基金委经费	自筹资金				合计
	自有资金	贷款	地方政府投入	其它	
10.00	0.00	0.00	0.00	0.00	10.00

政府部门、境外资金及其他资金投入情况说明：	无

与本项目相关的其他经费来源 （单位：万元）

其他计划资助经费：	0.00
单位配套经费：	0.00
其他经费资助：	0.00
其他经费来源合计：	

注：青年基金项目试点实施"包干制"，经费支出不设科目比例限制，由项目研究团队自主调剂使用，按照省科研项目经费"包干制"管理有关规定执行，同时应符合以下要求：

（1）经费支出应实际用于项目研究支出，使用范围限于设备费、材料费、测试化验加工费、燃料动力费、差旅/会议/国际合作与交流费、出版/文献/信息传播/知识产权事务费、劳务费、专家咨询费、依托单位管理费用、绩效支出以及其他合理支出。

（2）经费支出应按照省级财政科研项目资金开支范围和标准使用；

（3）间接经费支出比例按照省级财政科研项目资金管理有关规定执行；

（4）不得列支基建费；

（5）项目验收时应提交经费决算表。

五、人员信息

项目负责人								
姓名	证件号码	年龄	性别	职称	学历	在项目中承担的任务	所在单位	签名
张勇威	441481199401215891	28	男	未取得	博士研究生	项目负责人	广东工业大学	

10

六、工作分工及经费分配

承担/参与单位名称 （盖章）	工作分工	总经费分摊 （万元）	省基金委经费分配 （万元）
广东工业大学	项目主持人	10.00	10
	合计	10.00	10

11

七、合同条款

第一条 甲方与乙方根据《中华人民共和国民法典》及国家有关法规和规定，按照《广东省科学技术厅关于广东省基础与应用基础研究基金（省自然科学基金、联合基金等）项目管理的实施细则（试行）》《广东省省级科技计划项目验收结题工作规程（试行）》等规定，为顺利完成（2021）年面向多智能体系统的动态事件触发分布式优化容错控制方法专项项目（文件编号：粤基金字〔2021〕24号 ）经协商一致，特订立本任务书，作为甲乙双方在项目实施管理过程中共同遵守的依据。

第二条 甲方的权利义务：

1.按任务书规定进行经费核拨的有关工作协调。

2.根据甲方需要，在不影响乙方工作的前提下，定期或不定期对乙方项目的实施情况和经费使用情况进行检查或抽查。

3.根据《广东省科研诚信管理办法(试行)》等规定对乙方进行科技计划信用管理。

第三条 乙方的权利义务：

1. 确保落实自筹经费及有关保障条件。

2. 按任务书规定，对甲方核拨的经费实行专款专用，单独列账，并随时配合甲方进行监督检查。

3. 应按照国家和省有关规定，制定经费使用"包干制"内部管理规定；项目经费支出应实际用于研发活动相关支出，使用范围限于设备费、材料费、测试化验加工费、燃料动力费、差旅/会议/国际合作与交流费、出版/文献/信息传播/知识产权事务费、劳务费、专家咨询费、依托单位管理费用、绩效支出以及其他合理支出；管理费用根据实际管理支出情况与项目负责人协商确定；绩效支出由项目负责人根据实际科研需要和相关薪酬标准自主确定，单位按照现行工资制度进行管理；其余用途经费无额度限制，由项目负责人根据实际需要自主决定使用；项目验收时应提交经费决算表。

4. 对项目负责人按计划开展项目研究和规范使用资金进行监督管理，经费使用按照《广东省财政厅 广东省审计厅关于省级财政科研项目资金的管理监督办法》等规定进行管理。

5. 使用财政资金采购设备、原材料等，按照《广东省实施〈中华人民共和国招标投标法〉办法》有关规定，符合招标条件的须进行招标。

6. 项目任务书任务完成后，或任务书规定的任务、指标及经费投入等提前完成的，乙方可提出验收结题申请，并按甲方要求做好项目验收结题工作。

7. 若项目发生需要终止结题的情况，乙方须提出终止结题申请，并按甲方要求做好项目终止结题工作。

8. 在每年规定时间内向甲方如实提交上年度工作情况报告，报告内容包含上年度项目进展情况、经费决算和取得的成果等。

9. 按照国家和省有关规定，提交科技报告及其他材料。

10. 利用甲方的经费获得的研究成果，项目负责人和参与者应当注明获得"广东省基础与应用基础研究基金（英文：Guangdong Basic and Applied Basic Research Foundation）（项目编号）"资助或作有关说明。

11. 乙方要恪守科学道德准则，遵守科研活动规范，践行科研诚信要求，不得抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论；不得购买、代写、代投论文，虚构同行评议专家及评议意见；不得违反论文署名规范，擅自标注或虚假标注获得科技计划（专项、基金等）等资助；不得弄虚作假，骗取科技计划（专项、基金等）项目、科研经费以及奖励、荣誉等；不得有其他违背科研诚信要求的行为。

12. 确保本项目开展的研究工作符合我国科研伦理管理相关规定。

第四条 在履行本任务书的过程中，如出现广东省相关政策法规重大改变等不可抗力情况，甲方有权对所核拨经费的数量和时间进行相应调整。

第五条 在履行本任务书的过程中，当事人一方发现可能导致项目整体或部分失败的情形时，应及时通知另一方，并采取适当措施减少损失，没有及时通知并采取适当措施，致使损失扩大的，应当就扩大的损失承担责任。

第六条 本项目技术成果的归属、转让和实施技术成果所产生的经济利益的分享，除双方另有约定外，按国家和广东省有关法规执行。

第七条 根据项目具体情况，经双方另行协商订立的附加条款，作为本任务书正式内容的一部分，与本任务书具有同等效力。

第八条 本任务书一式三份，各份具有同等效力。甲、乙方及项负责人各执一份，三方签字、盖章后即生效，有效期至项目结题后一年内。各方均应负任务书的法律责任，不应受机构、人事变动的影响。

第九条 乙方必须接受甲方聘请的本项目任务书监理单位的监督和管理。监理单位按照甲方赋予的权利对本项目任务书的履行进行审核、进度调查，对项目任务书变更、经费使用情况进行监督管理及组织项目验收。
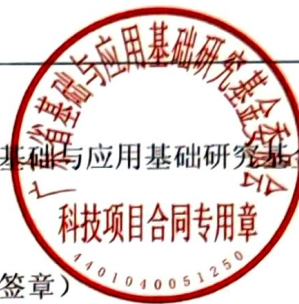
　　说明：1. 本合同书中，凡是当事人约定无需填写的内容，应在空白处划（／）。
　　　　　2. 委托代理人签订本合同书的，应出具合法、有效的委托书。

八、本合同签约各方

管理单位（甲方）：	广东省基础与应用基础研究基金委员会 （盖章）
法定代表人（或法人代理）：	（签章）
	2022 年 01 月 20 日
依托单位（乙方）： 广东工业大学	（盖章）
法定代表人（或法人代理）： 邱学青	（签章）
联系人（项目主管）姓名： 穆淼	（签章）
Email： 438184494@qq.com	
电话： 020-39322711 ／ 15902060363	
开户单位名称： 广东工业大学	
开户银行名称： 工商银行广州三支行	
开户银行帐号： 3602028909000196985	
	2022 年 1 月 25 日
联系人（项目负责人）姓名： 张勇威	（签名）
Email： 925628515@qq.com	
电话： 13247669804	
	2022 年 1 月 25 日

广州市科技计划项目
申 报 书

项目名称：	无人农机集群预设时间协同优化控制方法研究
申报单位：	华南农业大学
项目负责人：	张勇威
计划类别：	基础研究计划
专题名称：	2025年度基础与应用基础研究专题
支持方向：	青年博士"启航"项目
组织单位：	华南农业大学
起止时间：	2025-01-01 至 2026-12-31
主管处室：	引进智力管理处（科技人才处）

广州市科学技术局制

二〇二四年

填写说明

一、请申报单位认真阅读指南，所申报的项目研究内容须对应指南、符合指南的要求。

二、项目名称应清晰、准确反映研究内容，项目名称不宜宽泛，只能由中文、英文字符组成，不超过50中文字。

三、本申报书通过"广州科技大脑"在线填写、报送，不需要线下提交纸质材料。

四、申报书中的单位名称，请按规范全称填写，并与单位公章一致。

五、涉密项目请在"广州科技大脑"下载申报书的电子版模板，按保密要求离线填写、报送。

六、本申报书中凡是无需填写的内容，应在空白处划"/"，或用"无"表示。

七、申报书内容须按照项目申报书据实填写，要遵循实事求是原则，无需凑够字数。

一、基本信息

项目基本信息	项目名称	无人农机集群预设时间协同优化控制方法研究	
	学科领域1	信息综合处-自动化-控制理论与方法- 自适应与学习控制	
	学科领域2	信息综合处-自动化-人工智能与知识工程- 智能系统及应用	
	指南发布日	2024年4月15日	
	申请市级财政 金额	5万元	研究期限	2025年1月1日- 2026年12月31日
项目摘要	本项目针对复杂约束下无人农机集群编队控制问题，提出基于 ADP 的预设时间协同优化控制方法，以高效、节能的方式完成无人农场中的实际作业任务。研究内容如下：1）无人农机集群预设时间协同优化编队控制。2）状态和输入约束下无人农机集群预设时间优化安全编队控制。3）无人农机集群事件触发预设时间协同优化编队控制。本项目为无人农机协同控制提供了新思路，有助于实现无人农机绿色化、高效化和智能化。		

二、申报单位情况

<table>
<tr><td rowspan="10">项目承担单位</td><td>单位名称</td><td>华南农业大学</td><td>统一社会信用代码</td><td>124400004554165634</td></tr>
<tr><td>注册时间</td><td>1952-01-01</td><td>单位类型</td><td>高等院校</td></tr>
<tr><td>注册地址</td><td colspan="3">广东省广州市天河区五山路483号</td></tr>
<tr><td>办公地址</td><td colspan="3">广东省广州市天河区五山路483号</td></tr>
<tr><td rowspan="3">联系人</td><td>姓名</td><td colspan="2">倪慧群</td></tr>
<tr><td>手机号码</td><td colspan="2">13711345768</td></tr>
<tr><td>电子邮箱</td><td colspan="2">kjcgxk@scau.edu.cn</td></tr>
<tr><td>开户银行</td><td colspan="3">广东广州工行五山支行</td></tr>
<tr><td>开户户名</td><td colspan="3">华南农业大学</td></tr>
<tr><td>银行账号</td><td colspan="3">3602002609000310520</td></tr>
</table>

三、项目负责人信息

姓名	张勇威	证件类型	身份证
证件号码	441481199401215891	性别	男
出生年月	1994-01-21	民族	汉族
国籍	中国	学历	博士研究生
学位	博士	学位授予国家（或地区）	中国
职务	无	职称	中级
所学专业	控制科学与工程	手机号码	13247669804
办公电话	020-85285393	电子邮箱	925628515@qq.com

四、项目经费信息

本项目总投入：¥（5）万元，其中，市财政科技经费：¥（5）万元，自筹经费：¥（0）万元。

1. 经费下达计划			
资金来源	小计	市财政科技经费	自筹经费
2025	5	5	0
总计	5	5	0

（单位：万元）

注：本专题纳入"包干制"，市财政科技经费按市科技计划项目经费"包干制"相关规定执行。

五、预期代表性成果

项目负责人在项目实施期内，以该项目作为资助项目获得以下5种情形之一且经费使用符合规定的，由组织单位审核后通过验收。

（一）项目实施期内，以第一作者/通讯作者发表论文1篇或以上（须标注资助项目编号）；

（二）项目实施期内，以第一完成人申请或授权专利、软件著作1项或以上；

（三）项目实施期内，获省级以上科技计划项目或人才项目支持1项或以上；

（四）项目实施期内，获省级以上科技奖励（含列入获奖团队成员名单）1项或以上；

（五）项目实施期内，获得职称晋升。

六、承诺函

申 请 人：	张勇威
承担申报单位：	华南农业大学
项目名称：	无人农机集群预设时间协同优化控制方法研究
专题方向：	2025年度基础与应用基础研究专题-青年博士"启航"项目

申请人承诺：

　　　　本人根据项目申报指南的要求自愿提交项目（课题）申报书，**在此郑重承诺**：严格遵守《关于进一步加强科研诚信建设的若干意见》《关于进一步弘扬科学家精神　加强作风和学风建设的意见》等有关规定，杜绝《科学技术活动违规行为处理暂行规定》（科学技术部令第19号）所列违规行为，所申报材料和相关内容真实有效，不存在违背科研诚信要求的行为；已按要求落实了科研作风学风和科研诚信主体责任；不以任何形式实施请托行为，申报材料符合《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规，符合指南各项申报要求；在参与广州市科技计划项目申报、评审和实施全过程中，恪守职业规范和科学道德，遵守评审规则和工作纪律，杜绝以下行为：

（一）抄袭、剽窃他人科研成果或者伪造、篡改研究数据、研究结论或实施其他侵犯他人知识产权的行为；

（二）购买、代写、代投论文，虚构同行评议专家及评议意见；

（三）违反论文署名规范，擅自标注或虚假标注获得科技计划等资助；

（四）违反科研伦理规范；

（五）弄虚作假，骗取科技计划项目、科研经费以及奖励、荣誉等；

（六）在申报书中以高指标通过评审，在任务书签订时故意篡改降低任务书中相应指标；

（七）以任何形式探听尚未公布的评审专家名单及其他评审过程中的保密信息；

（八）本人或委托他人通过各种方式及各种途径联系有关专家进行请托、游说，违规到评审会议驻地游说评审专家和工作人员、询问评审或尚未正

式向社会公布的信息等干扰评审或可能影响评审公正性的活动；

（九）向评审工作人员、评审专家等提供任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包，或提供宴请、旅游、娱乐健身等任何可能影响评审公正性的活动；

（十）其它违反财经纪律和相关管理规定的行为。

如有违反，本人愿接受项目管理机构和相关部门做出的各项处理决定，包括但不限于取消项目（课题）承担资格，追回项目（课题）经费，向社会通报违规情况，取消一定期限广州市科技计划项目申报资格，记入科研诚信严重失信行为数据库以及接受相应的党纪政纪处理等。

签字：张勇威

日期：2024年07月05日

承担单位承诺：

本单位根据项目申报指南的任务需求，严格履行承担单位职责，自愿审核提交申报书，**在此郑重承诺**：

严格遵守《关于进一步加强科研诚信建设的若干意见》《关于进一步弘扬科学家精神　加强作风和学风建设的意见》等有关规定和其它科研诚信要求的行为，已按要求落实了科研作风学风和科研诚信主体责任；不以任何形式实施请托行为，申报材料符合《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规，符合指南各项申报要求；在参与项目申报和评审活动全过程中，遵守有关评审规则和工作纪律，杜绝以下行为：

（一）采取贿赂或变相贿赂、造假、剽窃、故意重复申报等不正当手段获取科技计划项目承担资格；

（二）以任何形式探听未公开的评审专家名单及其他评审过程中的保密信息；

（三）组织或协助项目团队向评审工作人员、评审专家等提供任何形式的礼品、礼金、有价证券、支付凭证、商业预付卡、电子红包等；宴请评审组织者、评审专家，或向评审组织者、评审专家提供旅游、娱乐健身等可能影响评审公正性的活动；

（四）包庇、纵容项目团队虚假申报项目，甚至骗取国家科技计划项目；

（五）包庇、纵容项目团队，甚至帮助项目团队采取"打招呼"等方式，影响评审公正；

（六）在正式申报书中以高指标通过评审，在任务书签订时故意篡改降低任务书中相应指标；

（七）其它违反财经纪律和相关管理规定的行为。

如有违反，本单位愿接受项目管理机构和相关部门做出的各项处理决定，包括但不限于停拨或核减经费，追回项目（课题）经费，取消一定期限广州市科技计划项目申报资格，记入科研诚信严重失信行为数据库等。

承担单位：*华南农业大学*

日期：*2024年07月05日*

七、单位审核

承担单位意见：
通过

日 期：2024年07月08日

组织单位意见：
通过

日 期：2024年07月11日

项目编号： <u>2024KTSCX100</u>

广东省普通高校省级重大科研项目（青年创新人才、特色创新）合同书（自然科学类）

项目类别： <u>特色创新项目(自然科学)</u>

项目名称： <u>复杂约束下植保无人机集群自学习协同优化控制方法研究</u>

项目负责人： <u>张勇威</u>

承担单位（学院）： <u>数学与信息学院、软件学院</u>

项目起止时间： <u>**2024** 年 **9** 月至 **2026** 年 **9** 月</u>

填表日期： <u>**2024** 年 **9** 月 **04** 日</u>

华南农业大学科研院

二〇二四年制

填 写 说 明

一、管理单位（甲方）为华南农业大学；承担单位（乙方）是指该项目具体实施的依托学院（单位）。

二、填写内容涉及到外文名称，要写清全称和缩写字母。

三、合同书正文用宋体小四号字，用 A4 纸双面打印，必须每份签章。

四、凡不填写内容的栏目，均用"/"表示。

一、项目基本信息

<table>
<tr><td rowspan="4">项目负责人</td><td>姓名</td><td>张勇威</td><td>性别</td><td colspan="2">男</td><td>出生年月</td><td colspan="2">1994.01</td></tr>
<tr><td>文化程度</td><td colspan="2">博士</td><td colspan="2">职称</td><td colspan="3">助理研究员</td></tr>
<tr><td>E-mail</td><td colspan="3">YongweiZhang@scau.edu.cn</td><td>移动电话</td><td colspan="3">13247669804</td></tr>
</table>

项目联系人	姓名	张勇威		职称	助理研究员		
	E-mail	YongweiZhang@scau.edu.cn		移动电话	13247669804		

其他合作单位	单位名称	参加形式 合作/协作	单位性质 科研院所/高校/企业

	姓名	性别	出生年月	学历	职称	任务分工	所在单位	签名
项目组成员	邓金	男	1994.09	博士	中级	自学习优化算法设计	华南农业大学	
	骆威	男	1986.01	博士	副教授	编队模型建立、自学习优化算法设计	华南农业大学	
	林壮壁	男	1996.02	博士	中级	控制器设计、稳定性分析	华南农业大学	
	彭超达	男	1990.09	博士	中级	植保无人机模型建立、控制器设计	华南农业大学	
	黄华琛	男	1999.10	学士	无	仿真平台搭建	华南农业大学	

二、主要研究内容

（一）主要研究内容：

本项目针对复杂约束下植保无人机集群的编队控制问题，提出基于 ADP 的协同优化编队控制方法，以高效、节能的方式完成实际的植保作业任务。研究内容如下：1）考虑信息年龄的植保无人机集群协同优化编队控制。2）状态和输入约束下植保无人机集群协同优化安全编队控制。首先，对于不同的约束，设计能切实反映通信成本、控制成本以及不同约束影响的代价函数。接着，提出不同约束下的自学习优化算法，给出收敛性和最优性分析方案，获得最优通信和编队控制策略。最后，搭建植保无人机集群协同控制平台，结合仿真和实物验证各方法的有效性。本项目的研究丰富了自适应动态规划的理论成果，为植保无人机协同控制提供新的研究思路，有助于降低植保无人机的控制成本，提高植保无人机的作业效率，实现植保无人机绿色化、高效化和智能化，具有重要的理论意义和实际应用价值。

（二）拟解决的关键问题：

1）代价函数设计问题

本项目旨在研究不同约束下植保无人机集群的协同优化编队控制问题。因此，针对不同约束，如何设计合适的代价函数，是提出基于 ADP 的协同优化编队控制方法的前提，也是本项目需要解决的技术难点之一。具体如下：对于信息年龄，需要给出能表征通信成本和控制成本的代价函数，体现通信策略对植保无人机集群控制性能的影响。对于状态和输入约束，需要设计能反映状态和输入约束影响以及控制成本的代价函数。

2）新型控制障碍函数设计问题

本项目考虑了植保无人机集群在实际作业中存在状态和输入约束问题。因此，如何设计能同时体现状态和输入约束的控制障碍函数，建立控制输入与安全性之间的联系，是获得能保证植保无人机集群安全性和最优性的编队控制器的关键，也是本项目亟需解决技术难点之一。

3）不同约束下新型自学习优化算法设计及其理论分析方案制定问题

考虑到信息年龄下需要求解时变的 HJB 方程，状态和输入约束下需要求解带约束的 HJB 方程，故传统的 ADP 算法将不再适用。因此，针对不同的约束，如何设

4

计相应的自学习优化算法，并在理论上给出算法的收敛性和最优性分析方案，确保迭代策略的可容许性、代价函数的收敛性和最优性以及通信和控制策略的最优性，是采用 ADP 理论解决植保无人机集群协同优化编队控制问题的核心，也是整个研究过程中需要解决的技术难点之一。

（三）创新点：

本项目基于最优控制理论、ADP 理论和博弈理论，主要研究了植保无人机集群的协同优化编队控制问题，其主要创新点总结如下：

1）本项目解决了能完成实际植保作业的无人机集群的编队控制问题。从目前掌握的资料来看，无人机编队控制方法的研究取得了初步的结果，但面临着规模小、尚未考虑实际植保作业场景等问题。本项目的研究能扩展无人机协同控制方法的应用范围，符合实际植保作业的需求，是对其在广度上的扩展。

2）本项目提出了基于 ADP 的协同优化编队控制方法，丰富了无人机协同控制的理论成果，提高了控制方法的实用性。传统的无人机控制方法仅考虑了系统的稳定性。而本项目所提出的方法能在保证植保无人机集群稳定性的同时进一步降低控制成本，以高效、节能的方式完成实际植保作业任务。此外，本项目所提出的方法能有效应对信息年龄、状态以及输入等实际约束，并维持植保无人机集群的期望队形。因此，本项目的研究丰富和完善了植保无人机协同控制方法的理论成果，提高了控制方法的实用性，是对其在深度上的扩展。

3）本项目提出了新型的自学习优化算法，建立了完善的理论分析方案，进一步丰富了 ADP 的理论成果，提高其实用性。本项目提出了适用于随机时变非线性系统以及能求解带约束 HJB 方程的自学习优化算法，极大地提高了 ADP 算法的实用性，并且适用于植保无人机集群，是对 ADP 理论在深度和广度上的扩展。

三、经费预算表

预算科目	自筹经费	备注（计算依据与说明）
一、科研业务费		
1、专利申请及代理费	0.7万元	1项*7000元/项 = 0.7万元；
2、会议费、差旅费	0.7万元	参加学术会议 1 人次*7000 元/人次 = 0.7 万元(包含会议注册费)
二、试验材料费	0万元	
三、仪器设备费		
1、计算机配件若干	0.1万元	鼠标、键盘、移动硬盘等计算机配件
四、劳务费	2万元	用于参与项目研究生的劳务费用，500 元/人月*2 人月/年*1 年 = 1.2 万元；专家咨询费 4000 元/人 * 2 次 = 0.8 万元
五、其他费用		
1、间接经费	1.5万元	
合计	5万元	
与本项目有关的其他经费来源	其他计划资助经费	0
	其他经费资助	0
	其他经费合计	0

项目负责人签名：

6

四、研究成果及形式

论文（篇）	总数	2	
	其中：CSCD 核心期刊	0	
	三大索引收录	2	
专著（部）		0	
研究报告（篇）		0	
专利（件）	数量（件）	申请	2
		授权	
	其中发明专利	申请	2
		授权	
鉴定成果（项）		0	
软件登记（项）		0	
新产品（种）（或新装备、新药等）		0	
新技术（项）（或新工艺等）		0	
其他		0	

五、进度和阶段目标

序号	起止时间	阶段性研究工作进展	阶段性目标
1	2024.10-2025.10	1. 建立植保无人机集群的系统模型、通信模型以及编队模型，并定义编队误差。2. 设计能表征通信成本与控制成本的代价函数。3. 建立信息年龄下的合作博弈模型，并给出帕累托最优解存在的充要条件。4. 提出随机时变自学习优化算法并建立收敛性和最优性分析方案，得出最优通信策略和最优编队控制。	发表 SCI 论文 1 篇,申请发明专利 1 项
2	2025.10-2026.10	1.构造能体现状态和输入约束的新型控制障碍函数。2.设计能表征状态和输入约束影响以及控制成本的代价函数。3.提出新型自学习优化算法，求解带约束的 HJB 方程，从而得到最优安全编队控制策略。4.搭建植保无人机集群协同控制平台，验证所提出的基于 ADP 的协同优化编队控制方法的有效性。	发表 SCI 论文 1 篇,申请发明专利 1 项
3			

六、本合同签约各方

项目负责人（签名）：欧阳武

2024 年 9 月 5 日

乙方：学院（公章）

负责人（签名）：黄心

2024 年 9 月 5 日

甲方：华南农业大学（公章）

代表人（签名）：薛红艳

年　月　日

SCAULIB202521998

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威**(学科类型:自然科学) 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Model-Free Game-Based Dynamic Event-Driven Safety-Critical Control of Unknown Nonaffine Systems	IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I-REGULAR PAPERS 出版年：2025 出版日期：SEP 卷期：72 9 页码：5019-5032 文献类型：Article	1	A 类	华南农业大学	SCI	IF2-year=5.2 IF5-year=5.3 （2024）	工程技术 2 区 Top 期刊：是 OA 期刊：否 （2025）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：刘汉忠

华南农业大学图书馆

2025-09-18

SCAULIB202522000

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威**(学科类型:自然科学) 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Event-triggered cooperative robust formation control of multi-agent systems via reinforcement learning	APPLIED INTELLIGENCE 出版年：2024 出版日期：SEP 卷期：54 17-18 页码：8367-8383 文献类型：Article	1	B 类	华南农业大学	SCI	IF2-year=3.5 IF5-year=3.8 （2024）	计算机科学　3区 Top 期刊：否 OA 期刊：否 （2025）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员： 刘汉忠

华南农业大学图书馆

2025-09-18

SCAULIB202519194

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Reinforcement Learning-Based Distributed Robust Bipartite Consensus Control for Multispacecraft Systems With Dynamic Uncertainties	IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS 出版年：2024 出版日期：NOV 卷期：20 11 页码：13341-13351 文献类型：Article	1	T2类	华南农业大学	SCI	IF2-year=9.9 IF5-year=10.7 （2024）	计算机科学 1区 Top 期刊：是 （2023）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：欧群

华南农业大学图书馆

2025-07-15

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Integral sliding mode-based event-triggered optimal fault tolerant tracking control of continuous-time nonlinear systems ☆	EUROPEAN JOURNAL OF CONTROL 出版年：2024 出版日期：SEP 卷期：79 页码：- 文献号：101021 文献类型：Article	1	B类	华南农业大学	SCI	IF2-year=2.6 IF5-year=2.5 （2024）	计算机科学 4区 Top 期刊：否 （2025）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明：如未盖章，报告无效

检索员：欧群

华南农业大学图书馆

2025-07-15

SCAULIB202519193

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Distributed Fault Tolerant Consensus Control of Nonlinear Multiagent Systems via Adaptive Dynamic Programming	IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS 出版年：2024 出版日期：JUL 卷期：35 7 页码：9041-9053 文献类型：Article	1	T2类	广东工业大学	SCI	IF2-year=8.9 IF5-year=11.1 （2024）	计算机科学 1 区 Top 期刊：是 （2025）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：欧群

华南农业大学图书馆

2025-07-15

SCAULIB202519196

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Adaptive Dynamic Programming-Based Event-Triggered Robust Control for Multiplayer Nonzero-Sum Games With Unknown Dynamics	IEEE TRANSACTIONS ON CYBERNETICS 出版年：2023 出版日期：AUG 卷期：53 8 页码：5151-5164 文献类型：Article	第一作者	T2类	广东工业大学	SCI	IF2-year=9.4 IF5-year=10.3 （2023）	计算机科学 1区 Top 期刊：是 （2023）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：邓智心

华南农业大学图书馆

2025-07-15

SCAULIB202519197

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Event-Triggered Control of Discrete-Time Zero-Sum Games via Deterministic Policy Gradient Adaptive Dynamic Programming	IEEE TRANSACTIONS ON SYSTEMS MAN CYBERNETICS-SYSTEMS 出版年：2022 出版日期：AUG 卷期：52 8 页码：4823-4835 文献类型：Article	第一作者	T2类	广东工业大学	SCI	IF2-year=8.7 IF5-year=9.5 （2022）	计算机科学 1区 Top 期刊：是 （2022）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：邓智心

华南农业大学图书馆

2025-07-15

SCAULIB202519195

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Riemannian Mean Shift-Based Data Fusion Scheme for Multi-Antenna Cooperative Spectrum Sensing	IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING 出版年：2022 出版日期：MAR 卷期：8 1 页码：47-56 文献类型：Article	第一作者	A类	广东工业大学	SCI	IF2-year=8.6 IF5-year=7.7 （2022）	计算机科学 2 区 Top 期刊：否 （2022）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明：如未盖章，报告无效

检索员：邓智心

华南农业大学图书馆

2025-07-15

SCAULIB202519199

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Dynamic event-triggered neuro-optimal control for uncertain nonlinear systems with unknown dead-zone constraint	COMMUNICATIONS IN NONLINEAR SCIENCE AND NUMERICAL SIMULATION 出版年：2024 出版日期：DEC 卷期：139 页码：- 文献号：108308 文献类型：Article	通讯作者	A 类	华南农业大学	SCI	IF2-year=3.8 IF5-year=3.4 （2024）	数学 2区 Top 期刊：否 （2025）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：邓智心
华南农业大学图书馆
2025-07-15

SCAULIB202519198

检索证明

根据委托人提供的论文材料，委托人**华南农业大学数学与信息学院 张勇威** 1 篇论文收录情况如下表。

序号	论文名称	发表刊物及发表的年月卷期/页码等	作者排名	论文等级	作者文中单位	收录情况	影响因子	中科院大类分区
1	Integral sliding mode-based event-triggered nearly optimal tracking control for uncertain nonlinear systems	INTERNATIONAL JOURNAL OF ROBUST AND NONLINEAR CONTROL 出版年：2024 出版日期：MAR 10 卷期：34 4页码：2639-2658 文献类型：Article	通讯作者	B 类	华南农业大学	SCI	IF2-year=3.2 IF5-year=3.4 （2024）	计算机科学 4 区 Top 期刊：否 （2025）

说明：论文等级和中科院大类分区按《华南农业大学学术论文评价方案（试行）》划分。

报告免责声明:如未盖章,报告无效

检索员：邓智心

华南农业大学图书馆

2025-07-15

Model-Free Game-Based Dynamic Event-Driven Safety-Critical Control of Unknown Nonaffine Systems

Yongwei Zhang, Weifeng Zhong, Guoxu Zhou, Lihua Xie, *Fellow, IEEE*, and Shengli Xie, *Fellow, IEEE*

Abstract— In this paper, the model-free dynamic event-driven safe (MFDEDS) control of unknown nonaffine systems with state and input constraints is investigated via adaptive dynamic programming. To begin with, by introducing a dynamic compensator and performing system transformation, the safe control problem with state and input constraints is transformed into an optimal regulation problem of an unconstrained system. Afterwards, an integral reinforcement learning algorithm is applied to the unconstrained system to derive an optimal safe control policy independent of the original system model, which achieves model-free approximate optimal control for the original system. To conserve computing and communication resources, a novel game-based dynamic event-driven mechanism is established, which models the control policy and the event-driven error as players in a zero-sum game, with the aim of obtaining the worst event-driven error to maximize the triggering interval. Furthermore, an approximate solution to the Hamilton-Jacobi-Bellman equation is derived by constructing a single-critic learning structure, which results in an approximate optimal safe control policy. Theoretical analysis demonstrates that the proposed MFDEDS control scheme ensures the closed-loop system is asymptotically stable. Ultimately, the efficacy of the developed approach is corroborated through two simulation examples.

Index Terms— Adaptive dynamic programming, safe-critical control, dynamic event-driven control, neural networks.

I. INTRODUCTION

IN MODERN engineering applications, safety-critical systems (SCSs) are prevalent across various fields such as aerospace, automotive, healthcare, and industrial automation,

Yongwei Zhang is with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China (e-mail: YongweiZhang@scau.edu.cn).

Weifeng Zhong, Guoxu Zhou, and Shengli Xie are with the School of Automation, the Key Laboratory of Intelligent Detection and Manufacturing Internet of Things, Ministry of Education, and Guangdong-Hong Kong-Macao Joint Laboratory for Smart Discrete Manufacturing, Guangdong University of Technology, Guangzhou 510006, China (e-mail: wfzhongs@gdut.edu.cn; gx.zhou@gdut.edu.cn; shlxie@gdut.edu.cn).

Lihua Xie is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: elhxie@ntu.edu.sg).

where the stability and the safety of these systems are directly linked to the security of human life and property. For instance, in the aerospace sector, the flight trajectory of a spacecraft must consistently remain within the designated safe set, and any deviation from this safety range could potentially result in a collision with other objects. In the realm of industrial robotics, robots must constantly avoid entering hazardous zones while collaborating with human workers to prevent accidental collisions and personal injuries. However, numerous SCSs exhibit complex characteristics, such as strong nonlinearity, unknown dynamics, and saturation constraints, which pose significant challenges for traditional control methodologies to guarantee the stability and safety of these systems throughout their operational processes.

In recent years, an increasing number of researchers have been dedicated to developing effective safety control methods. Sun et al. [1] addressed the safety-critical control problem of both continuous and sampled-data systems affected by time-varying disturbances by developing a composite controller that includes disturbance compensation and state feedback components. Lu et al. [2] introduced a universal barrier function to transform the state-constrained system into an equivalent unconstrained form and developed a switched-type auxiliary controller to guarantee the tracking performance. Wang et al. [3] tackled the finite-time tracking control problem for switched systems with full state constraints by adopting backstepping technique and barrier Lyapunov functions. Based on the aforementioned research findings, it is evident that the introduction of barrier functions can effectively ensure that the states of SCSs remain within designated safe sets. However, a growing number of practical systems are now equipped with microprocessors that possess limited communication and computational capabilities. Consequently, there is an urgent need to present efficient and energy-saving control methods that can guarantee the safety and stability of SCSs while minimizing computational, communication, and control costs to the greatest extent possible.

Over the past two decades, adaptive dynamic programming (ADP) has attracted considerable attention from researchers as an effective methodology to tackle optimal control problems for complex nonlinear systems [4], [5], [6], [7]. It has been successfully applied to a range of challenges, including optimal regulation [8], [9], trajectory tracking [10], [11], fault-tolerant control [12], [13], and differential game [14], [15]. By integrating iterative algorithms with neural networks, ADP

facilitates the online derivation of optimal control policies for nonlinear system, thereby ensuring system stability while simultaneously minimizing control costs. For SCSs, early research predominantly tackled the challenges associated with state and control input constraints through the utilization of barrier functions and the design of nonquadratic performance index functions. For instance, Zhao et al. [16] designed a feedforward neural network compensator to address optimal regulation problem for an unknown nonlinear system with uncertain input constraints. Xue et al. [18] developed a discounted nonquadratic performance index function to handle the nonzero-sum game problem with asymmetry input constraints. Qin et al. [17] introduced a novel barrier function to address the multiplayer Stackelberg—Nash games problem subject to time-varying state constraints. The aforementioned methods primarily focus on a single type of constraint.

It is noteworthy that real-world systems are typically subjected to both state and control input constraints simultaneously. For instance, in autonomous driving systems, the speed and acceleration of the vehicle, regarded as state variables, are constrained by road conditions and traffic regulations. Additionally, control inputs, such as throttle and brake signals, also encounter limitations due to mechanical characteristics and safety standards. To date, only a limited number of researchers have employed the ADP technique to address safe control issues that encompass both state and control input constraints. For example, Yang et al. [19] solved the H_∞ control problem of SCSs with state and input constraints by employing the barrier function-based system transformation method and developing a nonquadratic performance index function. It is noteworthy that the majority of existing optimal safe control (OSC) methods rely on system models. In practice, obtaining an accurate system model is often challenging. Furthermore, even if a model is available, the prolonged operation of the system in complex environments inevitably leads to model uncertainties. Therefore, it is essential to conduct further research on model-free OSC approaches. Currently, in situations when the system model is unknown, a neural network-based identifier is constructed to approximate the system model. Nonetheless, the incorporation of these identifiers introduces additional complexity into the control methodologies. In addition, researchers have proposed model-free iterative algorithms, such as Q-learning [20] and policy gradient [21], which can realize model-free control since the designed control policy does not contain the control input function. However, existing methods have not yet taken safety into account.

It is widely recognized that event-driven control effectively reduces the update frequency of controllers, which mitigates the computational and communication burdens [22], [23], [24]. Currently, researchers have developed ADP-based control policies that operate under static event-driven, dynamic event-driven, and self-driven mechanisms to address various types of control issues. Peng et al. [25] developed a reinforcement learning-based event-driven control scheme to investigate the distributed tracking control of multiagent systems. Xia et al. [26] explored the input-constrained synchronization problem of heterogeneous multiagent systems

under the dynamic event-driven mechanism. Zhao et al. [27] developed a self-driven optimal neuro-control approach for nonlinear systems to avoid continuous monitoring of the system state. The existing methods typically formulate the event-driven condition based on Lyapunov stability principle to determine the instants at which events occur. In practical applications, to significantly reduce computational and communication burdens, researchers aspire to maximize the event-driven error while ensuring system stability and control performance, thereby achieving a larger triggering time interval. In order to achieve this requirement, a modest number of scholars have developed optimal triggering thresholds within the game framework to obtain the worst-case triggering interval [28], [29]. However, the exploration of this method is still in its nascent stages. From the above discussion, it can be seen that the existing OSC approaches are limited by considering only a single type of constraint and relying on precise system models. Moreover, there is a lack of research on event-triggered safe control methods. In practical applications, the precise model of the SCS is difficult to obtain and is prone to being affected by compound constraints. Therefore, we aim to propose an OSC method that can handle complex constraints under the event-triggered mechanism, which will minimize the consumption of computational and communication resources while ensuring the stability and safety of the SCS. This also motivates our research.

This article presents an ADP-based model-free dynamic event-driven safe (MFDEDS) control scheme to address the OSC problem of unknown SCSs with state and control input constraints. The innovations and contributions of this paper are outlined as follows.

1) Compared with existing model-based OSC methods [16], [17], [18], this paper proposes a model-free OSC approach by combining the dynamic compensator and the integral reinforcement learning algorithm. Through appropriate system transformations, the OSC of unknown SCSs with both state and input constraints is converted into an optimal regulation of an unconstrained system.

2) Unlike existing methods [18], [19] that are only applicable to affine SCSs, the proposed ADP-based MFDEDS approach can be applied to nonaffine forms, which effectively expands the applicability of the control approach and enhances its practicality.

3) This paper introduces a novel game-based dynamic event-driven mechanism, wherein the control policy and the event-driven error are modeled as players in a zero-sum game, aiming to obtain the worst event-driven error. In contrast to conventional event-driven control methods [22], [23], [24], the developed dynamic event-driven approach endeavors to maximize the triggering time interval while maintaining the stability of the SCSs, thereby effectively conserving computational and communication resources.

The structure of the subsequent sections of this paper is organized as follows. Section II presents the problem statement. Section III introduces the game-based dynamic

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG et al.: MODEL-FREE GAME-BASED DYNAMIC EVENT-DRIVEN SAFETY-CRITICAL CONTROL
3

event-driven mechanism and outlines the OSC policy derived from this mechanism. Furthermore, the neural network implementation process is described in detail, along with a theoretical analysis of the stability of the SCS and a discussion of Zeno behavior. Section IV showcases simulation experiments conducted to validate the properties of the ADP-based MFDEDS control approach. Finally, Section V provides concluding remarks.

II. PROBLEM STATEMENT

Consider the unknown nonaffine SCS as

$$\dot{\mathcal{X}}(t) = \mathcal{F}(\mathcal{X}(t), \mu_o(t)), \tag{1}$$

where $\mathcal{X}(t) \in \mathcal{X}_s \subset \mathbb{R}^n$ is the constrained system state, $\mu_o(t) \in \mathcal{U}_s \subset \mathbb{R}^m$ is the constrained control policy, $\mathcal{F}(\cdot) \in \mathbb{R}^n$ is an unknown nonlinear system function, $\mathcal{X}_s$ and $\mathcal{U}_s$ denote the safe sets of the system state and the control policy, which are defined as

$$\mathcal{X}_s \triangleq \left\{ [\mathcal{X}_1, \ldots, \mathcal{X}_n]^\mathsf{T} \in \mathbb{R}^n \mid \right.$$
$$\left. \underline{\alpha}_{x,i} \le \mathcal{X}_i \le \overline{\alpha}_{x,i}, i = 1, 2, \ldots, n \right\},$$
$$\mathcal{U}_s \triangleq \left\{ [\mu_{o,1}, \ldots, \mu_{o,m}]^\mathsf{T} \in \mathbb{R}^m \mid \right.$$
$$\left. \underline{\alpha}_{\mu,j} \le \mu_{o,j} \le \overline{\alpha}_{\mu,j}, j = 1, 2, \ldots, m \right\}, \tag{2}$$

where $\underline{\alpha}_{x,i}$ and $\overline{\alpha}_{x,i}$ represent the lower and upper bounds of the system state components, while $\underline{\alpha}_{\mu,i}$ and $\overline{\alpha}_{\mu,i}$ denote the lower and upper bounds of the control policy components.

Objective 1: This paper aims to propose an ADP-based MFDEDS control approach to guarantee the stability of the unknown nonaffine SCS in an efficient and resource-saving manner, while ensuring that the system state and the control policy remain within the safe sets.

To relax the requirement for precise system information, we introduce a dynamic compensator as

$$\dot{\mu}_o(t) = \mathcal{G}(\mathcal{X}(t), \mu_o(t)), \tag{3}$$

where $\mathcal{G}(\cdot, \cdot) \in \mathbb{R}^m$ is a Lipschitz continuous function and satisfies $\mathcal{G}(0, 0) = 0$. Let $\mathcal{Z} = [\mathcal{X}, \mu_o]^\mathsf{T} \in \mathbb{R}^{n+m}$ be the new system state, an augmented SCS is constructed as

$$\dot{\mathcal{Z}}(t) = \mathcal{F}_z(\mathcal{Z}) + \mathcal{G}_z(\mathcal{Z})\mu_a, \tag{4}$$

where $\mu_a = \dot{\mu}_o$,

$$\mathcal{F}_z(\mathcal{Z}) = \begin{bmatrix} \mathcal{F}(\mathcal{X}, \mu_o) \\ \mathbf{0}_m \end{bmatrix}, \mathcal{G}_z(\mathcal{Z}) = \begin{bmatrix} \mathbf{0}_{n \times m} \\ \mathbf{I}_{m \times m} \end{bmatrix}. \tag{5}$$

Based on the findings presented in [30] and [31], we can infer that if the augmented SCS (4) is stable, then the stability of nonaffine SCS (1) can be ensured. Therefore, Objective 1 is transformed into the following Objective 2.

Objective 2: Devising a control policy μ_a that not only preserves the stability of (4) but also ensures that the new augmented state remains within the designated safe set. Consequently, the stability of the nonaffine SCS (1) is guaranteed, and both the state and the control policy μ_o can be maintained within the safe sets.

To achieve Objective 2, we will introduce the barrier function to further transform the augmented SCS (4). To begin with, the definition and properties of the barrier function are given as follows.

Definition 1: We introduce the barrier function $\mathcal{B}(z; w, \mathcal{W})$, where $z \in (w, \mathcal{W})$, w and $\mathcal{W}$ are two constants satisfying $w < \mathcal{W}$. Note that the barrier function has the following properties

$$\mathcal{B}(z; w, \mathcal{W}) = \ln\left(\frac{\mathcal{W}}{w} \frac{w - z}{\mathcal{W} - z}\right), \mathcal{B}(0; w, \mathcal{W}) = 0,$$
$$\lim_{z \to w^+} \mathcal{B}(z; w, \mathcal{W}) = -\infty, \lim_{z \to \mathcal{W}^-} \mathcal{B}(z; w, \mathcal{W}) = +\infty. \tag{6}$$

The inverse function and the derivative of the barrier function are formulated as

$$\mathcal{B}^{-1}(z; w, \mathcal{W}) = w\mathcal{W} \frac{e^{\frac{z}{2}} - e^{-\frac{z}{2}}}{we^{\frac{z}{2}} - \mathcal{W}e^{-\frac{z}{2}}}, \tag{7}$$
$$\frac{d\mathcal{B}^{-1}(z; w, \mathcal{W})}{dz} = \frac{\mathcal{W}w^2 - w\mathcal{W}^2}{w^2 e^z - 2w\mathcal{W} + \mathcal{W}^2 e^{-z}}. \tag{8}$$

Based on the barrier function, we perform the following transformation on the augmented SCS state as

$$\mathcal{S}_i = \mathcal{B}(\mathcal{Z}_i; w_i, \mathcal{W}_i), \tag{9}$$
$$\mathcal{Z}_i = \mathcal{B}^{-1}(\mathcal{S}_i; w_i, \mathcal{W}_i), \tag{10}$$

where $\mathcal{Z}_i$ denotes the ith component of the augmented SCS state, $\mathcal{S}_i$ refers to the corresponding unconstrained SCS state, w_i and $\mathcal{W}_i$ indicate the lower and upper bounds of the safe set, which satisfy

$$\mathcal{W}_i = \overline{\alpha}_{x,i}, 1 \le i \le n, \quad \mathcal{W}_i = \overline{\alpha}_{\mu,i}, n + 1 \le i \le m + n,$$
$$w_i = \underline{\alpha}_{x,i}, 1 \le i \le n, \quad w_i = \underline{\alpha}_{\mu,i}, n + 1 \le i \le m + n.$$

Subsequently, the dynamics of the unconstrained SCS state is expressed as

$$\dot{\mathcal{S}}_i = \frac{d\mathcal{Z}_i}{dt} \bigg/ \frac{d\mathcal{Z}_i}{d\mathcal{S}_i}$$
$$= \frac{\mathcal{F}_{z,i} + \mathcal{G}_{z,i}\mu_a}{\frac{d\mathcal{B}^{-1}(z; w_i, \mathcal{W}_i)}{dz}\big|_{z=\mathcal{S}_i}}$$
$$= (\mathcal{F}_{z,i} + \mathcal{G}_{z,i}\mu_a)\mathcal{E}_i(\mathcal{S}_i, w_i, \mathcal{W}_i), \tag{11}$$

where $\mathcal{F}_{z,i}$ and $\mathcal{G}_{z,i}$ represent the ith row of $\mathcal{F}_z(\mathcal{Z})$ and $\mathcal{G}_z(\mathcal{Z})$, and

$$\mathcal{E}_i(\mathcal{S}_i, w_i, \mathcal{W}_i) = \frac{\mathcal{W}_i^2 e^{-\mathcal{S}_i} - 2w_i\mathcal{W}_i + w_i^2 e^{\mathcal{S}_i}}{\mathcal{W}_i w_i^2 - w_i \mathcal{W}_i^2}. \tag{12}$$

Let $\mathcal{S} = [\mathcal{S}_1, \ldots, \mathcal{S}_{n+m}]^\mathsf{T}$, the unconstrained SCS is established as

$$\dot{\mathcal{S}} = \mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\mu_a, \tag{13}$$

where

$$\mathcal{F}_s(\mathcal{S}) = [\mathcal{E}_1(\mathcal{S}_1, w_1, \mathcal{W}_1)\mathcal{F}_{z,1}, \ldots,$$
$$\mathcal{E}_n(\mathcal{S}_{n+m}, w_{n+m}, \mathcal{W}_{n+m})\mathcal{F}_{z,n+m}]^\mathsf{T},$$
$$\mathcal{G}_s(\mathcal{S}) = [\mathcal{E}_1(\mathcal{S}_1, w_1, \mathcal{W}_1)\mathcal{G}_{z,1}, \ldots,$$
$$\mathcal{E}_n(\mathcal{S}_{n+m}, w_{n+m}, \mathcal{W}_{n+m})\mathcal{G}_{z,n+m}]^\mathsf{T}.$$

By employing the barrier function-based system transformation approach, the state-constrained augmented system (4) is

reformulated as the unconstrained SCS system (13). Therefore, Objective 2 is further refined into Objective 3 as follows.

Objective 3: Developing a control policy μ_a that ensures the unconstrained SCS (13) is stable, thereby guaranteeing the stability of the augmented SCS system (4), while maintaining the system state within the designated safe set.

Remark 1: The first system transformation involves incorporating a dynamic compensator (3), where the control input and the original system state are treated as a new state of the transformed system. This transformation aims to convert a nonaffine system into an affine form. Furthermore, by applying an integral reinforcement learning algorithm to the transformed system, an optimal control policy can be obtained even when the original system model is unknown. The second system transformation introduces a barrier function, with the aim of converting the constrained nonlinear system into an unconstrained form, thereby addressing the system state and control input constraints simultaneously. Through these two system transformations, the original optimal safe control problem of an unknown nonaffine system with state and control input constraints is transformed into an optimal regulation problem of an unconstrained affine system.

Remark 2: It is worth mentioning that barrier functions are typically designed based on the natural logarithm function $\ln(\cdot)$, that is, $\ln(\cdot) = \log_e(\cdot)$. For instance, $b_1(x, k) = \frac{1}{2}\ln(\frac{k^2}{k^2-x^2})$ and $b_2(x, l, h) = \ln[(e^x - e^l)/(e^h - e^x)]$ in [19], where x is the system state or control input, k, l, and h are restricted boundaries. The purpose of barrier functions is to ensure that the system state or control input remains within specified bounds. However, traditional methods introduce the barrier function into the Lyapunov function, which can only handle a single type of constraint. To address both the constrained system state and control input simultaneously, this paper adopts a barrier function-based system transformation approach, which transforms the constrained system into an unconstrained one. Therefore, in order to ensure the transformed system conforms to the properties of traditional nonlinear systems, the designed barrier function needs to satisfy the following properties. 1) $\mathcal{B}(0; w, \mathcal{W}) = 0$, which ensures that the origin of the transformed system remains a zero equilibrium point. 2) The inverse function of $\mathcal{B}(\cdot)$ exists and is unique, ensuring that the original system state and the transformed system state are in a one-to-one correspondence. 3) When the system state or control input approaches the boundary, the value of the barrier function tends to infinity. Moreover, when the system state or control input is within the boundary, the value of the barrier function remains finite. Only when these two properties are satisfied, the stability of the closed-loop system can be guaranteed while the system state or control input is ensured within the specified boundary. It is worth noting that traditional barrier functions, such as $b_1(\cdot)$ and $b_2(\cdot)$, cannot simultaneously satisfy all three properties. For instance, $b_1(\cdot)$ only satisfies Properties 1 and 3, while $b_2(\cdot)$ only satisfies Properties 2 and 3. Consequently, these traditional barrier functions can only address either a single system state or control input constraint problem, but not both at the same time. In contrast, the barrier function designed in this paper satisfies all three properties, enabling the problem

of system state and control input constraint can be solved simultaneously by using the barrier function-based system transformation method.

III. ADP-BASED MODEL-FREE DYNAMIC EVENT-DRIVEN SAFE CONTROL DESIGN

A. Model-Free Dynamic Event-Driven Safe (MFDEDS) Control Policy Design

In this section, we aim to develop an OSC policy for the unconstrained SCS (13) under a novel dynamic event-driven mechanism. Denote $\{\mathcal{T}_k\}_{k=0}^{\infty}$ as a sequence of triggering moments. The sampled state of the unconstrained SCS is presented as

$$\bar{\mathcal{S}}_k(t) = \mathcal{S}(\mathcal{T}_k), \mathcal{T}_k \le t < \mathcal{T}_{k+1}. \quad (14)$$

Therefore, the sampled control policy is symbolized as $\bar{\mu}_{a,k} = \mu_a(\mathcal{T}_k)$. The event-driven error of the system state and the control policy are defined as

$$e_{s,k} = \bar{\mathcal{S}}_k(t) - \mathcal{S}(t), e_{\mu,k} = \bar{\mu}_{a,k}(t) - \mu_a(t). \quad (15)$$

Afterwards, the event-based unconstrained SCS is built as

$$\dot{\mathcal{S}} = \mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\mu_a + \mathcal{G}_s(\mathcal{S})e_{\mu,k}. \quad (16)$$

In practical applications, a larger event-driven error implies fewer control policy updates, which can save more computational and communication resources. Hence, we aim to maximize the event-driven error as much as possible. However, a larger event-driven error may affect the stability of the unconstrained SCS. In this context, for the unconstrained SCS (13), μ_a and $e_{\mu,k}$ can be viewed as two players in a zero-sum game, that is, μ_a aims to ensure the system stability while minimizing the performance index function, and $e_{\mu,k}$ attempts to maximize the performance index function.

The performance index function of (16) is formulated as

$$
\begin{aligned}
\mathcal{P}(\mathcal{S}) &= \int_t^{\infty} \mathcal{C}\big(\mathcal{S}(\varkappa), \mu_a(\varkappa), e_{\mu,k}(\varkappa)\big) d\varkappa \\
&= \int_t^{\infty} \Big(\mathcal{S}^{\mathsf{T}}(\varkappa)\mathcal{M}_1\mathcal{S}(\varkappa) + \mu_a^{\mathsf{T}}(\varkappa)\mathcal{M}_2\mu_a(\varkappa) \\
&\quad - \gamma^2 e_{\mu,k}^{\mathsf{T}}e_{\mu,k}\Big) d\varkappa,
\end{aligned}
\quad (17)
$$

where γ is a positive constant, $\mathcal{M}_1 \in \mathbb{R}^{n \times n}$ and $\mathcal{M}_2 \in \mathbb{R}^{m \times m}$ are positive definite matrices.

The Hamiltonian of the system (16) is defined as

$$
\begin{aligned}
\mathcal{H}\big(\mathcal{S}, \nabla\mathcal{P}(\mathcal{S}), \mu_a, e_{\mu,k}\big) &= \mathcal{C}(\mathcal{S}, \mu_a) \\
&+ \nabla\mathcal{P}^{\mathsf{T}}(\mathcal{S})\big(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\mu_a + \mathcal{G}_s(\mathcal{S})e_{\mu,k}\big),
\end{aligned}
\quad (18)
$$

Therefore, the optimal performance index function

$$
\begin{aligned}
\mathcal{P}^*(\mathcal{S}) = \min_{\mu_a} \max_{e_{\mu,k}} \int_t^{\infty} \Big(&\mathcal{S}^{\mathsf{T}}(\kappa)\mathcal{M}_1\mathcal{S}(\kappa) + \mu_a^{*\mathsf{T}}(\kappa)\mathcal{M}_2\mu_a^*(\kappa) \\
&- \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*\Big) d\kappa
\end{aligned}
$$

satisfies

$$\min_{\mu_a} \max_{e_{\mu,k}} \mathcal{H}\big(\mathcal{S}, \nabla\mathcal{P}(\mathcal{S}), \mu_a, e_{\mu,k}\big) = 0. \quad (19)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG et al.: MODEL-FREE GAME-BASED DYNAMIC EVENT-DRIVEN SAFETY-CRITICAL CONTROL 5

Thus, the OSC policy and the worst event-driven error are derived as

$$\mu_a^* = -\frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\mathcal{P}^*(\mathcal{S}), \tag{20}$$

$$e_{\mu,k}^* = \frac{1}{2\gamma^2}\mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\mathcal{P}^*(\mathcal{S}). \tag{21}$$

Based on (20) and (21), the Hamilton-Jacobi-Bellman equation is given by

$$0 = \mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* - \gamma e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*$$
$$+ \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\Big(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})(\mu_a^* + e_{\mu,k}^*)\Big). \tag{22}$$

It is noteworthy that (22) is a complex partial differential equation, which poses challenges in obtaining a direct solution. Furthermore, the primary dynamics $\mathcal{F}_s(\mathcal{S})$ of the unconstrained SCS (13) is necessary to solve the (22), which means that all information in the nonaffine SCS (1) is required. To facilitate model-free control, we will employ integral reinforcement learning technique on the unconstrained SCS (13), thereby deriving an OSC policy that relies exclusively on the constant matrix $\mathcal{G}_s$, thus eliminating the need of system information of the nonaffine SCS (1).

For a positive interval $\mathfrak{T}$, the performance index function (17) is represented as

$$\mathcal{P}(\mathcal{S}(t)) = \int_t^{t+\mathfrak{T}} \mathcal{C}(\mathcal{S}(\varkappa), \mu_a(\varkappa), e_{\mu,k}(\varkappa))d\varkappa$$
$$+ \mathcal{P}(\mathcal{S}(t+\mathfrak{T})). \tag{23}$$

Afterwards, the optimal performance index function satisfies

$$\mathcal{P}^*(\mathcal{S}(t+\mathfrak{T})) - \mathcal{P}^*(\mathcal{S}(t))$$
$$+ \int_t^{t+\mathfrak{T}} \mathcal{C}(\mathcal{S}(\varkappa), \mu_a^*(\varkappa), e_{\mu,k}^*(\varkappa))d\varkappa = 0. \tag{24}$$

Guided by [18], we can infer that (22) and (24) are equivalent, which implies that the OSC policy can be derived by addressing (24), thus the nonaffine SCS information is not required.

According to the worst event-driven error (21), a novel dynamic event-driven condition is formulated as

$$\mathcal{T}_{k+1} = \inf\{t > \mathcal{T}_k : \beta\mathcal{D}_p(t) + \mathcal{D}_c(\mathcal{S}, e_{\mu,k}) \le 0\}, \tag{25}$$

where $\beta > 0$, $\mathcal{D}_p(t)$ is a dynamic variable and

$$\mathcal{D}_c(\mathcal{S}, e_{\mu,k}) = \frac{1}{4\gamma^4}\nabla\mathcal{P}^{*\mathsf{T}}\mathcal{G}_s\mathcal{G}_s^\mathsf{T}\nabla\mathcal{P}^* - \gamma^2 e_{\mu,k}^\mathsf{T}e_{\mu,k}.$$

Note that the dynamic variable is generated by

$$\dot{\mathcal{D}}_p(t) = -\alpha\mathcal{D}_p(t) + \mathcal{D}_c(\mathcal{S}, e_{\mu,k}), \tag{26}$$

where α is a positive constant and $\mathcal{D}_p(0) > 0$. Inspired by [33] and [34], if the dynamic variable is generated in accordance with (26), it will remain strictly positive.

Remark 3: Note that the zero-sum game and the H_∞ problem are equivalent, that is, finding a control policy that not only

ensures the stability of (13) but also guarantees the existence of an L_2 gain no larger than γ, that is

$$\int_0^\infty \Big(\mathcal{S}^\mathsf{T}(\kappa)\mathcal{M}_1\mathcal{S}(\kappa) + \mu_a^\mathsf{T}(\kappa)\mathcal{M}_2\mu_a(\kappa)\Big)d\kappa$$
$$\le \gamma^2 \int_0^\infty \|e_{\mu,k}\|^2 d\kappa. \tag{27}$$

In other words, if condition (27) holds, we say that the system (13) has an L_2 gain not exceeding γ. Furthermore, based on the Theorem 16 and the Remark 19 in [32], we can conclude that $\mathcal{P}^*(\mathcal{S})$ is positive definite.

Theorem 1: Consider the unconstrained SCS (13), the OSC control policy (20) and the worst event-driven error (21), if the dynamic event-driven condition (25) is valid and the following inequation holds

$$\lambda_{\min}(\mathcal{M}_1) > \frac{1}{4\gamma^4}c_1^2\bar{\mathcal{G}}_s^2, \tag{28}$$

where c_1 and $\bar{\mathcal{G}}_s$ are positive constants, then the unconstrained SCS is guaranteed to be asymptotically stable.

Proof. The candidate function of the Lyapunov's method is chosen as

$$\mathcal{L}_1 = \mathcal{P}^*(\mathcal{S}) + \mathcal{D}_p(t). \tag{29}$$

By computing the derivative of (29) and combining it with (13), we can derive

$$\dot{\mathcal{L}}_1 = \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\mu_a(\mathcal{T}_k)) + \dot{\mathcal{D}}_p(t)$$
$$= \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\mu_a^* + \mathcal{G}_s(\mathcal{S})e_{\mu,k}) + \dot{\mathcal{D}}_p(t)$$
$$= \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\Big(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\mu_a^* + \mathcal{G}_s(\mathcal{S})e_{\mu,k}^*$$
$$- \mathcal{G}_s(\mathcal{S})e_{\mu,k}^* + \mathcal{G}_s(\mathcal{S})e_{\mu,k}\Big) + \dot{\mathcal{D}}_p(t). \tag{30}$$

In accordance with (21) and (22), the subsequent results can be derived

$$\nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\Big(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\mu_a^* + \mathcal{G}_s(\mathcal{S})e_{\mu,k}^*\Big)$$
$$= -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*, \tag{31}$$
$$2\gamma^2 e_{\mu,k}^* = \mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\mathcal{P}^*(\mathcal{S}). \tag{32}$$

By incorporating (31) and (32) into (30), we additionally obtain

$$\dot{\mathcal{L}}_1 = -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*$$
$$+ \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})(e_{\mu,k} - e_{\mu,k}^*) + \dot{\mathcal{D}}_p(t)$$
$$= -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*$$
$$+ 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}(e_{\mu,k} - e_{\mu,k}^*) + \dot{\mathcal{D}}_p(t)$$
$$= -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*$$
$$+ 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k} - 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* + \dot{\mathcal{D}}_p(t)$$
$$= -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* - \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*$$
$$+ 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k} + \dot{\mathcal{D}}_p(t)$$
$$\le -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* - \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*$$
$$+ \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* + \gamma^2 e_{\mu,k}^\mathsf{T}e_{\mu,k} + \dot{\mathcal{D}}_p(t)$$
$$\le -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^\mathsf{T}e_{\mu,k} + \dot{\mathcal{D}}_p(t)$$

$$\leq -\mathcal{S}^{\mathsf{T}}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{\mathsf{T}}e_{\mu,k} - \alpha\mathcal{D}_p(t)$$
$$+ \frac{1}{4\gamma^4}\nabla\mathcal{P}^{*\mathsf{T}}\mathcal{G}_s\mathcal{G}_s^{\mathsf{T}}\nabla\mathcal{P}^* - \gamma^2 e_{\mu,k}^{\mathsf{T}}e_{\mu,k}$$
$$\leq -\mathcal{S}^{\mathsf{T}}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* - \alpha\mathcal{D}_p(t)$$
$$+ \frac{1}{4\gamma^4}\nabla\mathcal{P}^{*\mathsf{T}}\mathcal{G}_s\mathcal{G}_s^{\mathsf{T}}\nabla\mathcal{P}^*$$
$$\leq -\mathcal{S}^{\mathsf{T}}\mathcal{M}_1\mathcal{S} + \frac{1}{4\gamma^4}\nabla\mathcal{P}^{*\mathsf{T}}\mathcal{G}_s\mathcal{G}_s^{\mathsf{T}}\nabla\mathcal{P}^*$$
$$\leq -\lambda_{\min}(\mathcal{M}_1)\|\mathcal{S}\|^2 + \frac{1}{4\gamma^4}\bar{\mathcal{G}}_s^2 c_1^2\|\mathcal{S}\|^2, \tag{33}$$

where $\bar{\mathcal{G}}_s$ is the norm bound of $\mathcal{G}_s(\mathcal{S})$ and $\|\nabla\mathcal{P}^*\| \leq c_1\|\mathcal{S}\|$ with a positive constant c_1. Thus, drawing upon the finding of (33), one can ascertain that if condition (28) is fulfilled, the unconstrained SCS (13) is assured to exhibit asymptotic stability. The proof is concluded.

B. Neural Network Implementation

Taking into account that the optimal performance index function remains unknown, we shall subsequently develop a single-critic learning framework to derive its approximate counterpart, concurrently obtaining the approximate OSC policy and the approximate worst event-driven error. Utilizing the critic neural network, the optimal performance index function and its approximate version are established as

$$\mathcal{P}^*(\mathcal{S}) = \zeta_c^{*\mathsf{T}}\psi_c(\mathcal{S}) + \varsigma_c(\mathcal{S}), \tag{34}$$
$$\hat{\mathcal{P}}(\mathcal{S}) = \hat{\zeta}_c^{\mathsf{T}}\psi_c(\mathcal{S}), \tag{35}$$

where $\zeta_c^* \in \mathbb{R}^{n_c}$ signifies the target weight, $\hat{\zeta}_c$ represents an approximate value, $\psi_c(\mathcal{S}) \in \mathbb{R}^{n_c}$ characterizes the activation function, $\varsigma_c(\mathcal{S}) \in \mathbb{R}$ designates the approximation error, and n_c indicates the quantity of hidden layer neurons.

Synthesizing (20), (21), (34) and (35), the event-driven OSC policy and the worst event-driven error are reformulated as

$$\mu_a^*(\mathcal{T}_k) = -\frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^{\mathsf{T}}(\bar{\mathcal{S}}_k)(\nabla\psi_c^{\mathsf{T}}(\bar{\mathcal{S}}_k)\zeta_c^* + \nabla\varsigma_c), \tag{36}$$
$$e_{\mu,k}^*(\mathcal{T}_k) = \frac{1}{2\gamma^2}\mathcal{G}_s^{\mathsf{T}}(\bar{\mathcal{S}}_k)(\nabla\psi_c^{\mathsf{T}}(\bar{\mathcal{S}}_k)\zeta_c^* + \nabla\varsigma_c). \tag{37}$$

The approximate versions of (36) and (37) are presented as

$$\hat{\mu}_a(\mathcal{T}_k) = -\frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^{\mathsf{T}}(\bar{\mathcal{S}}_k)\nabla\psi_c^{\mathsf{T}}(\bar{\mathcal{S}}_k)\hat{\zeta}_c, \tag{38}$$
$$\hat{e}_{\mu,k}(\mathcal{T}_k) = \frac{1}{2\gamma^2}\mathcal{G}_s^{\mathsf{T}}(\bar{\mathcal{S}}_k)\nabla\psi_c^{\mathsf{T}}(\bar{\mathcal{S}}_k)\hat{\zeta}_c. \tag{39}$$

The temporal difference error is described as

$$e_c = \hat{\zeta}_c^{\mathsf{T}}\Big(\psi_c\big(\mathcal{S}(t + \mathfrak{T})\big) - \psi_c\big(\mathcal{S}(t)\big)\Big)$$
$$+ \int_t^{t+\mathfrak{T}} \mathcal{C}\big(\mathcal{S}(\tau), \hat{\mu}_a(\tau), \hat{e}_{\mu,k}(\tau)\big)d\tau. \tag{40}$$

By employing gradient descent approach to minimize the temporal difference error, the weight adjustment rule is derived as

$$\dot{\hat{\zeta}}_c = -\frac{\alpha_c\Theta}{(1 + \Theta^{\mathsf{T}}\Theta)^2}\Big(\hat{\zeta}_c^{\mathsf{T}}\Theta + \mathcal{B}\big(\mathcal{S}, \hat{\mu}_a(\mathcal{T}_k)\big)\Big), \tag{41}$$

where $\alpha_c > 0$ is the learning rate, and $\Theta = \psi_c\big(\mathcal{S}(t + \mathfrak{T})\big) - \psi_c\big(\mathcal{S}(t)\big)$ and $\mathcal{B}\big(\mathcal{S}, \hat{\mu}_a(\mathcal{T}_k)\big) = \int_t^{t+\mathfrak{T}} \mathcal{C}\big(\mathcal{S}(\tau), \hat{\mu}_a(\tau), \hat{e}_{\mu,k}(\tau)\big)d\tau$.

Building upon the existing findings [35], [38], the weight adjustment rule (41) guarantees the weight estimation error $\tilde{\zeta}_c = \zeta_c^* - \hat{\zeta}_c$ for the critic neural network remains uniformly ultimately bounded (UUB), which indicates that the approximate weight is capable of converging to the target weight.

Within the neural network framework, the dynamic event-driven condition is turned as

$$\mathcal{T}_{k+1} = \inf\{t > \mathcal{T}_k : \beta\mathcal{D}_p(t) + \hat{\mathcal{D}}_c(\mathcal{S}, e_{\mu,k}) \leq 0\}, \tag{42}$$

where

$$\hat{\mathcal{D}}_c(\mathcal{S}, e_{\mu,k}) = \frac{1}{4\gamma^4}\hat{\zeta}_c^{\mathsf{T}}\nabla\psi_c(\mathcal{S})\mathcal{G}_s\mathcal{G}_s^{\mathsf{T}}\nabla\psi_c^{\mathsf{T}}(\mathcal{S})\hat{\zeta}_c - \gamma^2 e_{\mu,k}^{\mathsf{T}}e_{\mu,k}.$$

C. Stability Analysis

Next, we establish a theoretical paradigm to rigorously examine the stability of the unconstrained SCS (13) in the context of the dynamic event-driven mechanism (42).

Assumption 1: $\tilde{\zeta}_c$, ζ_c^*, $\nabla\psi_c(\mathcal{S})$, and $\nabla\varsigma_c(\mathcal{S})$ satisfy

$$\|\tilde{\zeta}_c\| \leq \bar{\zeta}_c, \ \|\zeta_c^*\| \leq \bar{\zeta}_{cm}, \ \|\nabla\psi_c(\mathcal{S})\| \leq \bar{\psi}_c, \ \|\nabla\varsigma_c(\mathcal{S})\| \leq \bar{\varsigma}_c,$$

where $\bar{\zeta}_c$, $\bar{\zeta}_{cm}$, $\bar{\psi}_c$, and $\bar{\varsigma}_c$ are positive constants [36], [37], [38].

Remark 4: In Assumption 1, $\tilde{\zeta}_c$ is the critic neural network estimation error. According to the existing findings [35], [38], the weight adjustment rule (41) guarantees the weight estimation error is UUB. Therefore, it is reasonable to assume that $\tilde{\zeta}_c$ is norm-bounded. Furthermore, ζ_c^* is the optimal weight, $\psi_c(\mathcal{S})$ represents the activation function and $\varsigma_c(\mathcal{S})$ denotes the neural network reconstruction error. Since they cannot be infinite in practice, the assumption of norm-boundedness is reasonable.

Theorem 2: For the unconstrained SCS (13), the event-driven OSC policy (36) and the worst event-driven error (37), and Assumption 1, if the dynamic event-driven condition (42) is met, then the unconstrained SCS is guaranteed to be stable.

Proof. The Lyapunov function candidate is chosen as

$$\mathcal{L}_{2T} = \mathcal{L}_{2T1} + \mathcal{L}_{2T2} = \mathcal{P}^*(\mathcal{S}) + \mathcal{P}^*(\bar{\mathcal{S}}_k) + \mathcal{D}_p(t). \tag{43}$$

Part 1: The event is not taken, specifically, $t \in [\mathcal{T}_k, \mathcal{T}_{k+1})$. By performing the time differentiation of equation (43) and adopting (13), one obtains

$$\dot{\mathcal{L}}_{2T} = \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\hat{\mu}_a(\mathcal{T}_k)) + \dot{\mathcal{D}}_p(t)$$
$$= \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})(\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\hat{\mu}_a + \mathcal{G}_s(\mathcal{S})e_{\mu,k}) + \dot{\mathcal{D}}_p(t)$$
$$= \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{F}_s(\mathcal{S}) + \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})\hat{\mu}_a$$
$$+ \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})e_{\mu,k} + \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})\mu_a^*$$
$$+ \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})e_{\mu,k}^* - \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})\mu_a^*$$
$$- \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})e_{\mu,k}^* + \dot{\mathcal{D}}_p(t)$$
$$= -\mathcal{S}^{\mathsf{T}}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^*$$
$$+ \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})\hat{\mu}_a + \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})e_{\mu,k}$$
$$- \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})\mu_a^* - \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})e_{\mu,k}^* + \dot{\mathcal{D}}_p(t)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG et al.: MODEL-FREE GAME-BASED DYNAMIC EVENT-DRIVEN SAFETY-CRITICAL CONTROL 7

$$
\begin{aligned}
= & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* \\
& + \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})(\hat{\mu}_a - \mu_a^*) \\
& + \nabla\mathcal{P}^{*\mathsf{T}}(\mathcal{S})\mathcal{G}_s(\mathcal{S})(e_{\mu,k} - e_{\mu,k}^*) + \dot{\mathcal{D}}_p(t) \\
= & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* \\
& - 2\mu_a^{*\mathsf{T}}\mathcal{M}_2(\hat{\mu}_a - \mu_a^*) + 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}(e_{\mu,k} - e_{\mu,k}^*) + \dot{\mathcal{D}}_p(t) \\
= & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* - 2\mu_a^{*\mathsf{T}}\mathcal{M}_2\hat{\mu}_a \\
& + 2\mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* + 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k} - 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* + \dot{\mathcal{D}}_p(t) \\
= & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} - 2\mu_a^{*\mathsf{T}}\mathcal{M}_2\hat{\mu}_a + \mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* \\
& + 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k} - \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* + \dot{\mathcal{D}}_p(t).
\end{aligned} \tag{44}
$$

By applying a straightforward transformation to the second and third terms in (44), we can get

$$
\begin{aligned}
\mu_a^{*\mathsf{T}}\mathcal{M}_2\mu_a^* - 2\mu_a^{*\mathsf{T}}\mathcal{M}_2\hat{\mu}_a = & (\mu_a^* - \hat{\mu}_a)^\mathsf{T}\mathcal{M}_2(\mu_a^* - \hat{\mu}_a) \\
& - \hat{\mu}_a^\mathsf{T}\mathcal{M}_2\hat{\mu}_a.
\end{aligned} \tag{45}
$$

By synthesizing (44) and (45), we can further deduce

$$
\begin{aligned}
\dot{\mathcal{L}}_{2T} = & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + (\mu_a^* - \hat{\mu}_a)^\mathsf{T}\mathcal{M}_2(\mu_a^* - \hat{\mu}_a) \\
& - \hat{\mu}_a^\mathsf{T}\mathcal{M}_2\hat{\mu}_a - \gamma e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* \\
& + 2\gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k} + \dot{\mathcal{D}}_p(t) \\
\leq & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + (\mu_a^* - \hat{\mu}_a)^\mathsf{T}\mathcal{M}_2(\mu_a^* - \hat{\mu}_a) \\
& + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* + \gamma^2 e_{\mu,k}^\mathsf{T}e_{\mu,k} + \dot{\mathcal{D}}_p(t) \\
\leq & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + (\mu_a^* - \hat{\mu}_a)^\mathsf{T}\mathcal{M}_2(\mu_a^* - \hat{\mu}_a) \\
& + \gamma^2 e_{\mu,p}^{*\mathsf{T}}e_{\mu,k}^* + \gamma^2 e_{\mu,p}^\mathsf{T}e_{\mu,k} - \alpha\mathcal{D}_p(t) \\
& + \frac{1}{4\gamma^4}\nabla\hat{\mathcal{P}}^\mathsf{T}\mathcal{G}_s\mathcal{G}_s^\mathsf{T}\nabla\hat{\mathcal{P}} - \gamma^2 e_{\mu,k}^\mathsf{T}e_{\mu,k} \\
\leq & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + (\mu_a^* - \hat{\mu}_a)^\mathsf{T}\mathcal{M}_2(\mu_a^* - \hat{\mu}_a) \\
& + \gamma^2 e_{\mu,k}^{*\mathsf{T}}e_{\mu,k}^* + \frac{1}{4\gamma^4}\nabla\hat{\mathcal{P}}^\mathsf{T}\mathcal{G}_s\mathcal{G}_s^\mathsf{T}\nabla\hat{\mathcal{P}} \\
\leq & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + \|\mathcal{M}_2\|\|\mu_a^* - \hat{\mu}_a\|^2 \\
& + \gamma^2 \frac{1}{4\gamma^4}\nabla\mathcal{P}^{*\mathsf{T}}\mathcal{G}_s\mathcal{G}_s^\mathsf{T}\nabla\mathcal{P}^* \\
& + \frac{1}{4\gamma^4}\nabla\hat{\mathcal{P}}^\mathsf{T}\mathcal{G}_s\mathcal{G}_s^\mathsf{T}\nabla\hat{\mathcal{P}}.
\end{aligned} \tag{46}
$$

Utilizing (36) and (38), the second term in (46) is extended as

$$
\begin{aligned}
\|\mu_a^* - \hat{\mu}_a\|^2 = & \left\| -\frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\mathcal{P}^*(\mathcal{S}) \right. \\
& \left. + \frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\hat{\mathcal{P}}(\mathcal{S}) \right\|^2 \\
= & \left\| -\frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^\mathsf{T}(\mathcal{S})(\nabla\psi_c^\mathsf{T}(\mathcal{S})\zeta_c^* + \nabla\varsigma_c) \right. \\
& \left. + \frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\psi_c^\mathsf{T}(\mathcal{S})\hat{\zeta}_c \right\|^2 \\
= & \left\| -\frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\varsigma_c \right. \\
& \left. - \frac{1}{2}\mathcal{M}_2^{-1}\mathcal{G}_s^\mathsf{T}(\mathcal{S})\nabla\psi_c^\mathsf{T}(\mathcal{S})\tilde{\zeta}_c \right\|^2 \\
= & \lambda_1 + \lambda_2.
\end{aligned} \tag{47}
$$

where $\lambda_1 = \frac{1}{2}\|\mathcal{M}_2^{-1}\|^2\bar{\mathcal{G}}_s^2\bar{\varsigma}_c^2$ and $\lambda_2 = \frac{1}{2}\|\mathcal{M}_2^{-1}\|^2\bar{\mathcal{G}}_s^2\bar{\psi}_c^2\bar{\zeta}_c$. By incorporating (47) into (46), $\dot{\mathcal{L}}_{2T}$ is transformed into

$$
\begin{aligned}
\dot{\mathcal{L}}_{2T} \leq & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + \|\mathcal{M}_2\|(\lambda_2 + \lambda_1) \\
& + \frac{1}{4\gamma^2}\bar{\mathcal{G}}_s^2\|\nabla\mathcal{P}^*(\mathcal{S})\|^2 + \frac{1}{4\gamma^4}\bar{\mathcal{G}}_s^2\|\nabla\hat{\mathcal{P}}(\mathcal{S})\|^2 \\
\leq & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + \|\mathcal{M}_2\|(\lambda_2 + \lambda_1) \\
& + \frac{1}{4\gamma^2}\bar{\mathcal{G}}_s^2\|\nabla\psi_c^\mathsf{T}(\mathcal{S})\zeta_c^* + \nabla\varsigma_c\|^2 \\
& + \frac{1}{4\gamma^4}\bar{\mathcal{G}}_s^2\|\nabla\psi_c^\mathsf{T}(\mathcal{S})\hat{\zeta}_c\|^2 \\
\leq & -\mathcal{S}^\mathsf{T}\mathcal{M}_1\mathcal{S} + \|\mathcal{M}_2\|(\lambda_2 + \lambda_1) + \frac{1}{2\gamma^2}\bar{\mathcal{G}}_s^2\bar{\psi}_c^2\bar{\zeta}_{cm}^2 \\
& + \frac{1}{2\gamma^2}\bar{\mathcal{G}}_s^2\bar{\varsigma}_c^2 + \frac{1}{4\gamma^4}\bar{\mathcal{G}}_s^2\bar{\psi}_c^2\bar{\zeta}_c^2 \\
\leq & -\eta_1^2\lambda_{\min}(\mathcal{M}_1)\|\mathcal{S}\|^2 - (1 - \eta_1^2)\lambda_{\min}(\mathcal{M}_1)\|\mathcal{S}\|^2 \\
& + \Psi_1,
\end{aligned} \tag{48}
$$

where $\eta_1 > 0$ and $\Psi_1 = \|\mathcal{M}_2\|(\lambda_2 + \lambda_1) + \frac{1}{2\gamma^2}\bar{\mathcal{G}}_s^2\bar{\psi}_c^2\bar{\zeta}_{cm}^2 + \frac{1}{2\gamma^2}\bar{\mathcal{G}}_s^2\bar{\varsigma}_c^2 + \frac{1}{4\gamma^4}\bar{\mathcal{G}}_s^2\bar{\psi}_c^2\bar{\zeta}_c^2$. Hence, $\dot{\mathcal{L}}_{2T} < 0$ if the system state $\mathcal{S}$ lies outside the compact set

$$
\Omega_\mathcal{S} = \left\{ \mathcal{S} : \|\mathcal{S}\| \leq \sqrt{\frac{\Psi_1}{(1 - \eta_1^2)\lambda_{\min}(\mathcal{M}_1)}} \right\}. \tag{49}
$$

Part 2: The moment of event occurrence, that is, $t = \mathcal{T}_k$. Building upon the findings presented in Part 1 and Theorem 2 in [33], we can readily deduce that when the event transpires, $\dot{\mathcal{L}}_{2T}$ persists in being less than 0. The proof is concluded.

In the following, we will demonstrate that under the novel dynamic event-driven mechanism, the Zeno behavior will be precluded.

Theorem 3: For the unconstrained SCS (13), the event-driven OSC policy given by (36), and the dynamic event-driven condition specified in (42), then the minimum interval between adjacent triggering moments is greater than zero.

Proof. According to (15), we have

$$
\dot{e}_{s,k} = \dot{\bar{\mathcal{S}}}_k - \dot{\mathcal{S}} = -\dot{\mathcal{S}}. \tag{50}
$$

By taking the norm on both sides of the equation, we get

$$
\begin{aligned}
\|\dot{e}_{s,k}\| = & \|\dot{\mathcal{S}}\| \\
= & \|\mathcal{F}_s(\mathcal{S}) + \mathcal{G}_s(\mathcal{S})\hat{\mu}_a\| \\
\leq & \|\mathcal{F}_s(\mathcal{S})\| + \|\mathcal{G}_s(\mathcal{S})\hat{\mu}_a\| \\
\leq & \mathcal{K}_f\|\mathcal{S}\| + \mathcal{K}_g \\
\leq & \mathcal{K}_f\|\bar{\mathcal{S}}_k - e_{s,k}\| + \mathcal{K}_g \\
\leq & \mathcal{K}_f\|\bar{\mathcal{S}}_k\| + \mathcal{K}_f\|e_{s,k}\| + \mathcal{K}_g,
\end{aligned} \tag{51}
$$

where $\mathcal{F}_s(\mathcal{S})$ is bounded by the norm as $\mathcal{F}_s(\mathcal{S}) \leq \mathcal{K}_f\|\mathcal{S}\|$ with $\mathcal{K}_f > 0$, and $\|\mathcal{G}_s(\mathcal{S})\hat{\mu}_a\|$ is bounded by a positive constant $\mathcal{K}_g$. In light of [29], it can be inferred that $\|e_{\mu,k}\| \leq \mathcal{L}_\mu\|e_{s,k}\|$, where $\mathcal{L}_\mu$ is a positive constant. By leveraging the comparison lemma, we are able to further deduce that

$$
\begin{aligned}
\|e_{\mu,k}\| \leq & \mathcal{L}_\mu\|e_{s,k}\| \\
\leq & \mathcal{L}_\mu\frac{\mathcal{K}_f\|\bar{\mathcal{S}}_k\| + \mathcal{K}_g}{\mathcal{K}_f}\left(e^{\mathcal{K}_f(t - \mathcal{T}_k)} - 1\right).
\end{aligned} \tag{52}
$$

When $t = \mathcal{T}_{k+1}$, based on the dynamic event-driven condition (42), one has

$$\|e_{\mu,k}\| \geq \sqrt{\frac{\frac{1}{4\gamma^4}\nabla\hat{\mathcal{P}}^{\mathsf{T}}\mathcal{G}_s\mathcal{G}_s^{\mathsf{T}}\nabla\hat{\mathcal{P}} + \frac{\mathcal{D}_p(t)}{\beta}}{\gamma^2}}. \quad (53)$$

Combining (52) and (53), we can get

$$\mathcal{T}_{k+1} - \mathcal{T}_k \geq \frac{1}{\mathcal{K}_f}\ln\left(1 + \frac{1}{\phi}\sqrt{\frac{\frac{1}{4\gamma^4}\nabla\hat{\mathcal{P}}^{\mathsf{T}}\mathcal{G}_s\mathcal{G}_s^{\mathsf{T}}\nabla\hat{\mathcal{P}} + \frac{\mathcal{D}_p(t)}{\beta}}{\gamma^2}}\right),$$

where $\phi = \mathcal{L}_\mu \frac{\mathcal{K}_f\|\bar{\mathcal{S}}_k\| + \mathcal{K}_g}{\mathcal{K}_f}$. Therefore, we can deduce that the disparity between any two triggering intervals is greater than 0, which implies that Zeno behavior will not manifest.

Remark 5: 1) Different from traditional safe control methods [16], [17], [18], which handled a single type of constraint only. This paper designs a dynamic compensator and employs a barrier function-based system transformation method, enabling the simultaneous handling of complex situations involving both state and input constraints. Furthermore, existing results [16], [17], [18], [19] are applicable only to affine nonlinear systems and rely on precise system models. However, this paper uses the dynamic compensator and the integral reinforcement learning technique to achieve optimal safe control for nonaffine nonlinear systems in a model-free manner. Therefore, the developed approach is more practical and applicable to a broader class of nonlinear systems. 2) Unlike the existing results [33], [34], [38], this paper proposes a novel dynamic event-driven mechanism within the framework of zero-sum games. Specifically, the event-driven error and the control input are regarded as two players in the zero-sum game. The control input aims to ensure system stability and minimize the performance index function, while the event-driven error tends to maximize the performance index function, potentially leading to the destabilization of the closed-loop system. By employing the ADP technique, the Nash equilibrium solution of the zero-sum game can be obtained, which ensures the stability of the closed-loop system even under the worst-case event-driven error. In fact, the worst-case event-driven error implies a larger interval between adjacent triggering times, thus minimizing the number of control policy updates and conserving computational and communication resources. Therefore, compared with traditional dynamic event-driven control methods [33], [34], [38], the proposed approach can maintain the stability of the closed-loop system while significantly reducing the frequency of control policy adjustments, which further conserves resources.

IV. SIMULATION

In this section, we will elucidate the efficacy of the ADP-based MFDEDS control method in both a Van der Pol circuit and a nonaffine system.
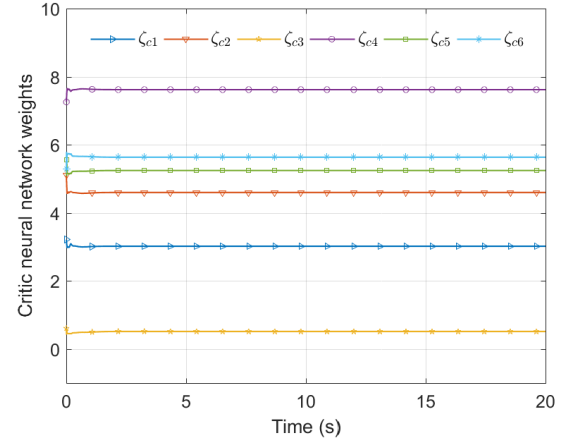
Fig. 1.   Critic neural network weights in case 1.
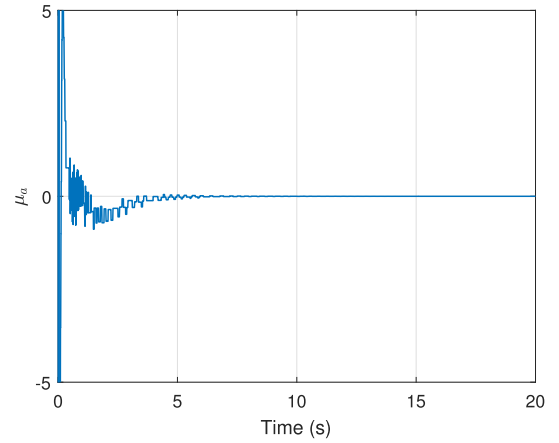
Fig. 2.   Control policy μ_a in case 1.

A. Case 1

The system model of the Van der Pol circuit system is articulated as

$$\dot{\mathcal{X}}_1 = \mathcal{X}_2,$$
$$\dot{\mathcal{X}}_2 = -2\mathcal{X}_1 + 3\mathcal{X}_2(1 - \mathcal{X}_1^2)\mu_o.$$

Next, we detail the parameter settings during the simulation process. The upper and lower bounds for the state components and the control policy of the nonaffine SCS (1) are defined as $\underline{\alpha}_{x,1} = -2$, $\underline{\alpha}_{x,2} = -2$, $\overline{\alpha}_{x,1} = 2$, $\overline{\alpha}_{x,2} = 2$, $\underline{\alpha}_\mu = -4$, and $\overline{\alpha}_\mu = 4$. The parameters in performance index function are selected as $\mathcal{M}_1 = 20\mathcal{I}_3$, $\mathcal{M}_2 = \mathcal{I}$, and $\gamma = 10$. The parameters of the dynamic event-driven condition are set to $\beta = 1$ and $\alpha = 1$. Additionally, the activation function of the critic neural network is configured as $\psi_c(\mathcal{S}) = [\mathcal{S}_1^2, \mathcal{S}_2^2, \mathcal{S}_3^2, \mathcal{S}_1\mathcal{S}_2, \mathcal{S}_1\mathcal{S}_3, \mathcal{S}_2\mathcal{S}_3]^{\mathsf{T}}$, and the learning rate is chosen as $\alpha_c = 2$.

Fig. 1 illustrates the evolution curve of the neural network weights during the training process. It is evident that, guided by the weight updating rule (41), the neural network weights ultimately converge to $\zeta_c = [3.01, 4.59, 0.51, 7.61, 5.24, 5.63]^{\mathsf{T}}$. Fig. 2 presents the control policy of the unconstrained SCS, which shows that it ultimately converges, and the staircase-like nature of the curve
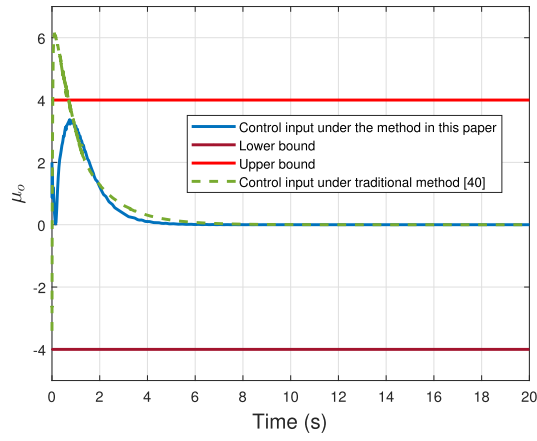
Fig. 3. Control policy μ_o in case 1.
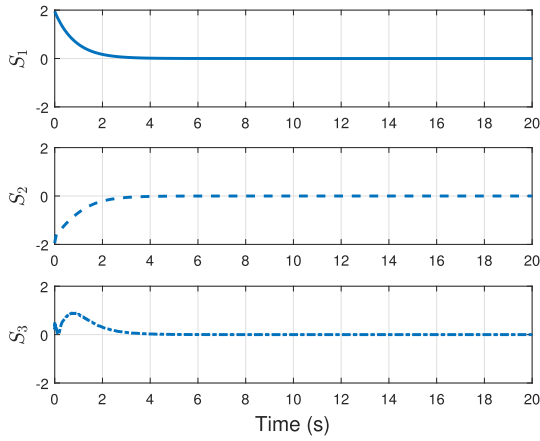
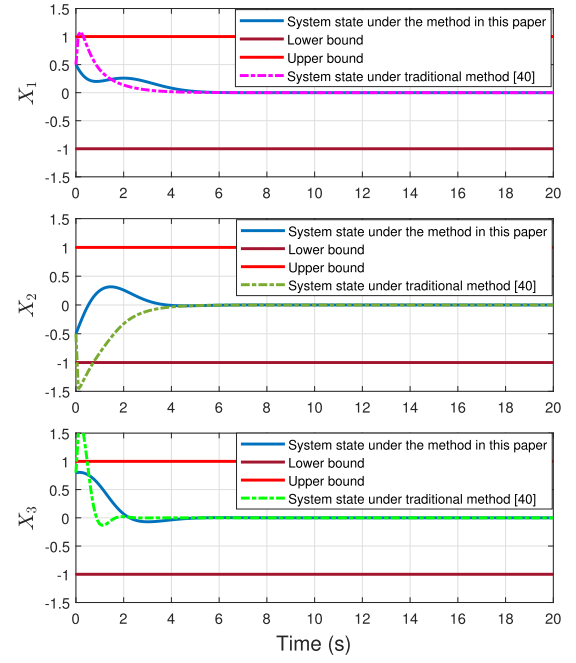Fig. 4. System states of the unconstrained SCS in case 1.

Fig. 5. System states of the nonaffine SCS in case 1.

Fig. 6. The number of control policy adjustments in case 1.

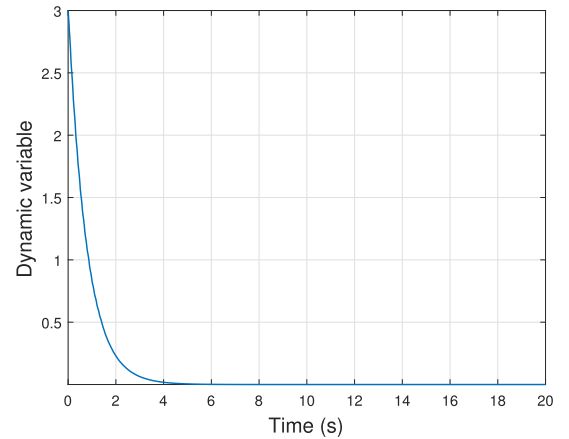Fig. 7. Dynamic variable in case 1.
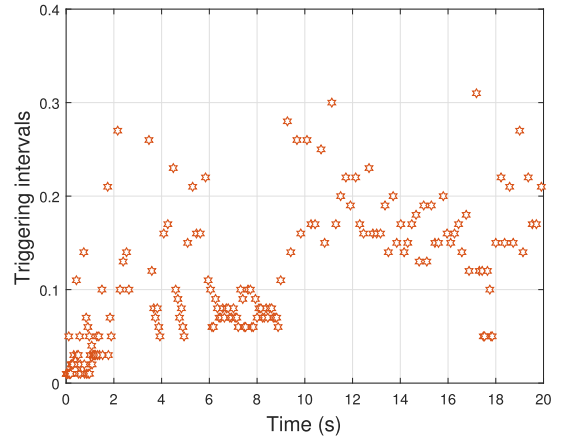
Fig. 8. Triggering intervals in case 1.

μ_a, we can derive the control policy μ_o for the nonaffine SCS, as shown in Fig. 3. It can be observed that the trajectory of the μ_o can consistently remain within the safe set. However, the traditional control method in [40] cannot guarantee that the control policy lies within the specified bounds. Figs. 4 and 5 display the system state trajectories of the unconstrained SCS and the nonaffine SCS respectively, and the results illustrate that under the influence of the control policies μ_a and μ_o, the system states eventually tend to a region close to 0,

means that the control policy remains unchanged between adjacent triggering instants. By integrating the control policy
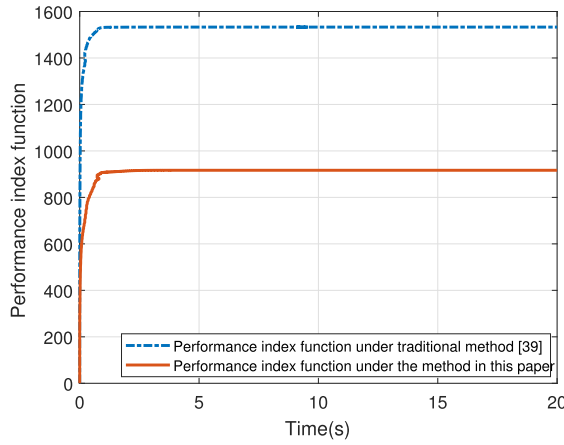
Fig. 9.   Performance index function in case 1.
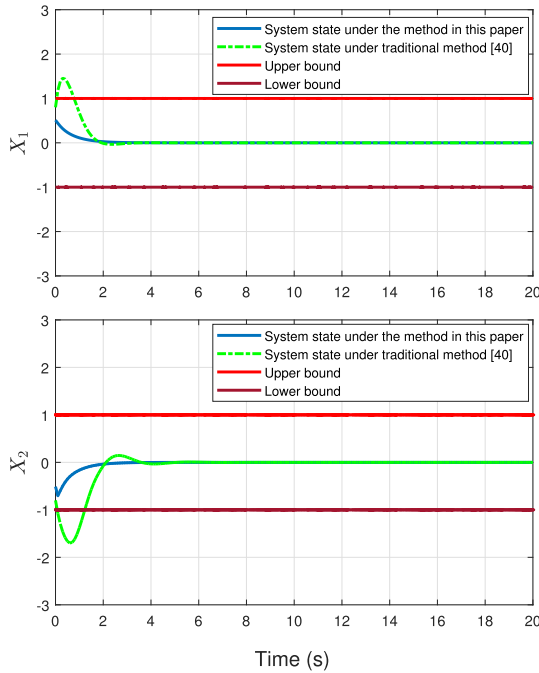
Fig. 10.   System states of the unconstrained SCS under different constrained boundaries in case 1.

which means that the stability of both systems is guaranteed. Moreover, Fig. 5 exhibits the system state trajectories under different control methods. It can be observed that the proposed method ensures that the system states remain within the predefined boundaries, whereas the system states exceed these boundaries under the traditional control method in [40]. This means that the safety of the nonaffine SCS can be guaranteed with the developed approach in this paper. Fig. 6 compares the number of control policy adjustments under time-driven (TD) [36], event-driven (ED) [10], traditional dynamic event-driven (TDED) [33], and the developed novel dynamic event-driven (DED) mechanisms. It can be observed that the novel dynamic event-driven mechanism significantly reduces the adjustment frequency of the control policy, thereby effectively saving computational and communication resources. Fig. 7 reveals the evolution curve of the dynamic variable, which remains greater than 0 and eventually converges. Fig. 8 illustrates that the time
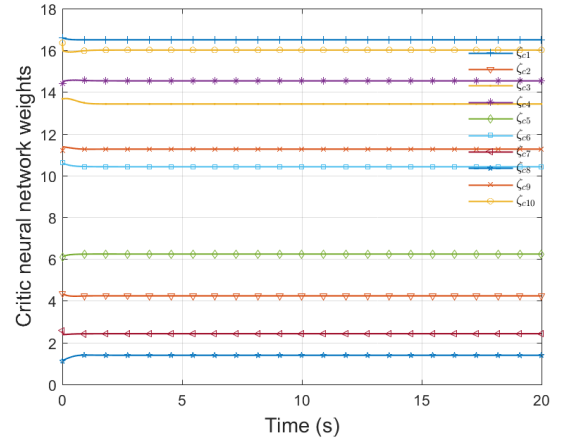
Fig. 11.   Critic neural network weights in case 2.
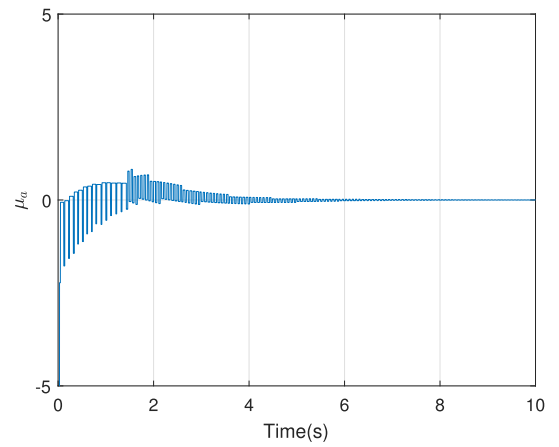
Fig. 12.   Control policy μ_a in case 2.

intervals between any two adjacent triggering moments under the novel dynamic event-driven mechanism are greater than 0, and this indicates that Zeno behavior does not occur. Fig. 9 presents the convergence values of the performance index function under different control methods. We can find that the proposed method in this paper achieves a smaller convergence value compared to the traditional control method in [39]. Therefore, the developed method can ensure the stability of the closed-loop system with lower control cost, demonstrating its optimization characteristics. Fig. 10 illustrates the system state curves when the constrained boundaries are set as $\underline{\alpha}_{x,1} = \underline{\alpha}_{x,2} = -1$ and $\overline{\alpha}_{x,1} = \overline{\alpha}_{x,2} = 1$, from which it can be concluded that the proposed method can still ensure that the system state remains within the preset boundaries.

B. Case 2

Next, we will further assess the efficacy of the ADP-based MFDEDS control approach on the following nonaffine system.

$$\dot{\mathcal{X}}_1 = -\mathcal{X}_1 + \mathcal{X}_2,$$
$$\dot{\mathcal{X}}_2 = -2\mathcal{X}_2 + \mathcal{X}_3,$$
$$\dot{\mathcal{X}}_3 = -\mathcal{X}_2 - (1 - \sin^2(\mathcal{X}_2))\mathcal{X}_3 + \sin(\mathcal{X}_2)\mu_o + \mu_o^2.$$

Initially, the simulation parameters pertinent to Case 2 are delineated in Table I. The activation function
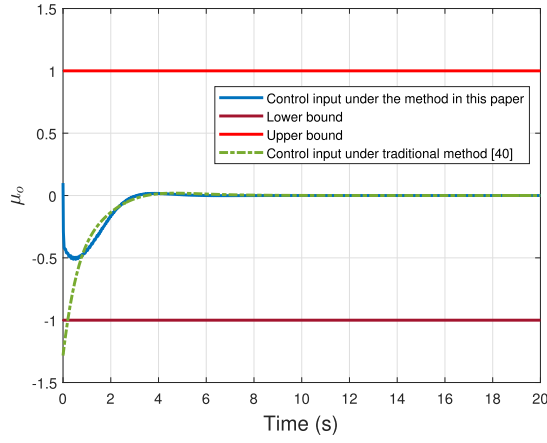
Fig. 13. Control policy μ_o in case 2.
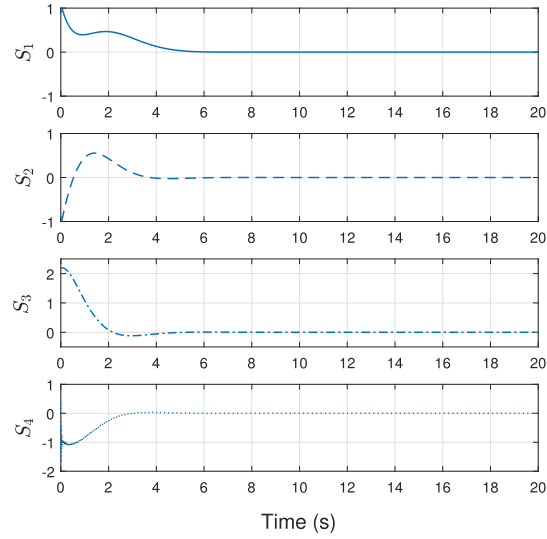
Fig. 14. System states of the unconstrained SCS in case 2.

TABLE I
SIMULATION PARAMETERS IN CASE 2

Parameter	$\underline{\alpha}_{x,1}$	$\underline{\alpha}_{x,2}$	$\underline{\alpha}_{x,3}$	$\overline{\alpha}_{x,1}$	$\overline{\alpha}_{x,2}$	$\overline{\alpha}_{x,3}$	$\underline{\alpha}_{\mu}$
Value	-1	-1	-1	1	1	1	-1
Parameter	$\overline{\alpha}_{\mu}$	$\mathcal{M}_1$	$\mathcal{M}_2$	γ	β	α	α_c
Value	1	$30\mathcal{I}_4$	$\mathcal{I}$	10	1	1	3

of the critic neural network is designated as $\varphi_c(\mathcal{S}) = [\mathcal{S}_1^2, \mathcal{S}_2^2, \mathcal{S}_3^2, \mathcal{S}_4^2, \mathcal{S}_1\mathcal{S}_2, \mathcal{S}_1\mathcal{S}_3, \mathcal{S}_1\mathcal{S}_4, \mathcal{S}_2\mathcal{S}_3, \mathcal{S}_2\mathcal{S}_4, \mathcal{S}_3\mathcal{S}_4]^{\mathsf{T}}$. Comprehensive simulation results are elucidated in Figs. 11–19. Fig. 11 elucidates the progression of the neural network weights, from which it is discernible that the ultimate convergence value is $\zeta_c = [16.50, 4.23, 13.43, 14.53, 6.24, 10.42, 2.42, 1.39, 11.27, 16.01]^{\mathsf{T}}$. Figs. 12 and 13 depict the evolution curves of the control policies of the unconstrained SCS and the nonaffine SCS, respectively. It can be seen from the experimental results that the control policy μ_a of the unconstrained SCS manifests a stepped characteristic, while the control policy μ_o of the nonaffine SCS is always kept in the safe set. Figs. 14 and 15 show the variation curves of

Fig. 15. System states of the nonaffine SCS in case 2.
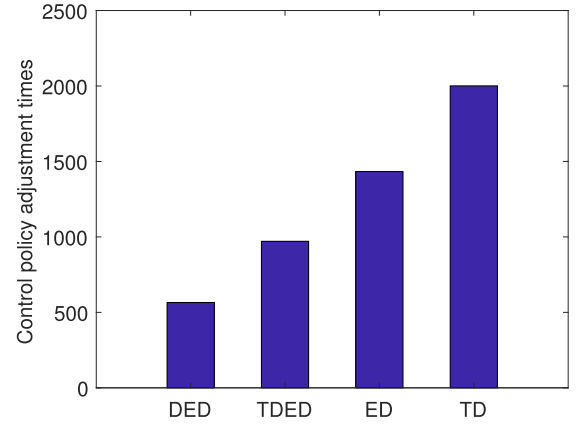
Fig. 16. The number of control policy adjustments in case 2.

Fig. 17. Dynamic variable in case 2.

the system states under two control policies μ_a and μ_o. It is clear that both of them converge to zero, and the trajectories

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12   IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS
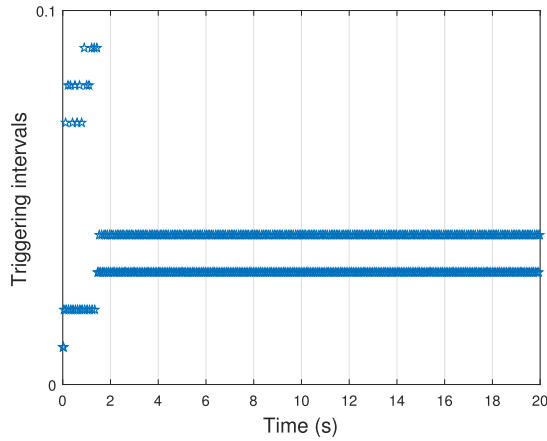
Fig. 18.   Triggering intervals in case 2.
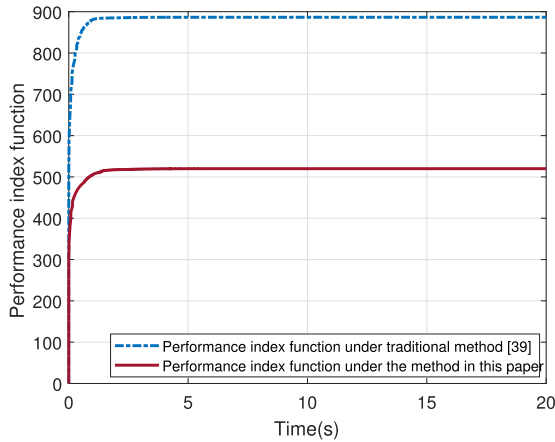
Fig. 19.   Performance index function in case 2.

of the nonaffine SCS does not exceed the boundaries of the safe set in contrast to the traditional control method in [40]. Fig. 16 shows that under different mechanisms, the adjustment times of the control policy are 567, 971, 1433, and 2000, respectively. Therefore, this indicates that the developed dynamic event-driven mechanism performs well in reducing the burden of computational and communication. Fig. 17 demonstrates that the dynamic variable remains consistently greater than zero and ultimately converges. Meanwhile, Fig. 18 illustrates all adjacent triggering time intervals, with the results indicating that these intervals are positive, which effectively mitigates the occurrence of Zeno behavior. Fig. 19 shows that compared with the existing method [39], the proposed scheme can guarantee the stability of the closed-loop system with less control cost.

Remark 6: In fact, the design of activation functions and the number of neurons in the hidden layers significantly influence the approximation performance of neural networks. However, to date, there is no unified method to guide the selection of these parameters. In this paper and most existing results [10], [15], [16], [33], a trial-and-error approach is adopted, that is, performing repeated experiments and observing the results to determine the appropriate activation function and the number of hidden layer neurons. In future research, we will carefully consider this challenge and attempt to use optimization

algorithms (such as particle swarm optimization) to find the appropriate combination.

V. CONCLUSION

This paper develops an ADP-based MFDEDS control approach for unknown nonaffine SCSs subject to state and input constraints. By establishing a dynamic compensator and implementing system transformation, the OSC problem of the nonaffine SCS with state and input constraints is converted into the optimal regulation problem of the unconstrained SCS. The designed OSC policy does not depend on the system dynamics of the nonaffine SCS, which achieves model-free near OSC control of the original system. Moreover, a novel game-based dynamic event-driven mechanism is proposed to ensure the stability of the unconstrained SCS and maximize the triggering interval, which significantly reduces the computational and communication burdens. Subsequently, a single-critic learning framework is constructed to obtain the approximately OSC policy online, and the Lyapunov's direct method is employed to demonstrate that this control policy can ensure that the SCS and the neural network weight errors are UUB. Finally, simulation results show that the designed ADP-based MFDEDS control method can ensure that the system state and the control input of nonaffine SCSs remain within their safe sets, and the implementation of the novel dynamic event-driven mechanism significantly reduces the frequency of control policy updates, thereby saving computational and communication resources. In future work, we will consider relaxing Assumption 1 to further enhance the practicality of the proposed method. Additionally, since constraints are typically time-varying in practical applications, we aim to explore novel safe control methods that can handle time-varying constraints.

REFERENCES

[1] J. Sun, J. Yang, and Z. Zeng, "Safety-critical control with control barrier function based on disturbance observer," *IEEE Trans. Autom. Control*, vol. 69, no. 7, pp. 4750–4756, Jul. 2024.

[2] K. Lu, Z. Liu, H. Yu, C. L. P. Chen, and Y. Zhang, "Inverse optimal adaptive neural control for state-constrained nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 8, pp. 10617–10628, Aug. 2024.

[3] H. Wang, W. Liu, and M. Tong, "Adaptive fuzzy fast finite-time output-feedback tracking control for switched nonlinear systems with full-state constraints," *IEEE Trans. Fuzzy Syst.*, vol. 32, no. 3, pp. 958–968, Mar. 2024.

[4] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.

[5] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[6] D. Wang, N. Gao, D. Liu, J. Li, and F. L. Lewis, "Recent progress in reinforcement learning and adaptive dynamic programming for advanced control applications," *IEEE/CAA J. Autom. Sinica*, vol. 11, no. 1, pp. 18–36, Jan. 2024.

[7] J. Li, L. Yuan, T. Chai, and F. L. Lewis, "Consensus of nonlinear multiagent systems with uncertainties using reinforcement learning based sliding mode control," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 70, no. 1, pp. 424–434, Jan. 2023.

[8] J. Sun, B. Luo, X. Xu, and C. Yang, "Boundary optimal control for parabolic distributed parameter systems with value iteration," *IEEE Trans. Cybern.*, vol. 54, no. 3, pp. 1571–1581, Mar. 2024.

[9] X. Yang, M. Xu, and Q. Wei, "Dynamic event-sampled control of interconnected nonlinear systems using reinforcement learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 1, pp. 923–937, Jan. 2024.

ZHANG et al.: MODEL-FREE GAME-BASED DYNAMIC EVENT-DRIVEN SAFETY-CRITICAL CONTROL
13

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[10] S. Xue, B. Luo, D. Liu, and Y. Gao, "Event-triggered ADP for tracking control of partially unknown constrained uncertain systems," *IEEE Trans. Cybern.*, vol. 52, no. 9, pp. 9001–9012, Sep. 2022.

[11] D. Wang, J. Wu, J. Ren, P. Xin, and J. Qiao, "Novel discounted optimal tracking design under offline and online formulations for asymmetric constrained systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 11, pp. 6886–6896, Nov. 2023.

[12] H. Xia, J. Hou, and P. Guo, "Two-level local observer-based decentralized optimal fault tolerant tracking control for unknown nonlinear interconnected systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 54, no. 3, pp. 1779–1790, Mar. 2024.

[13] C. Tan, M. Lin, H. Xia, and B. Zhao, "Fault-tolerant tracking control for nonlinear systems with multiplicative actuator faults in view of zero-sum differential games," *Neurocomputing*, vol. 567, Jan. 2024, Art. no. 127037.

[14] T. An, Y. Wang, G. Liu, Y. Li, and B. Dong, "Cooperative game-based approximate optimal control of modular robot manipulators for human–robot collaboration," *IEEE Trans. Cybern.*, vol. 53, no. 7, pp. 4691–4703, Jul. 2023.

[15] D. Lin, S. Xue, D. Liu, M. Liang, and Y. Wang, "Adaptive dynamic programming-based hierarchical decision-making of non-affine systems," *Neural Netw.*, vol. 167, pp. 331–341, Oct. 2023.

[16] B. Zhao, D. Liu, and C. Luo, "Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4330–4340, Oct. 2020.

[17] C. Qin, T. Zhu, K. Jiang, and Y. Wu, "Integral reinforcement learning-based dynamic event-triggered safety control for multiplayer Stackelberg–Nash games with time-varying state constraints," *Eng. Appl. Artif. Intell.*, vol. 133, Jul. 2024, Art. no. 108317.

[18] S. Xue, B. Luo, D. Liu, and Y. Gao, "Event-triggered integral reinforcement learning for nonzero-sum games with asymmetric input saturation," *Neural Netw.*, vol. 152, pp. 212–223, Aug. 2022.

[19] Y. Yang, D.-W. Ding, H. Xiong, Y. Yin, and D. C. Wunsch, "Online barrier-actor-critic learning for H_∞ control with full-state constraints and input saturation," *J. Franklin Inst.*, vol. 357, no. 6, pp. 3316–3344, Apr. 2020.

[20] D. Wang, H. Huang, and M. Zhao, "Model-free optimal tracking design with evolving control strategies via Q-learning," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 71, no. 7, pp. 3373–3377, Jul. 2024.

[21] M. Lin and B. Zhao, "Policy optimization adaptive dynamic programming for optimal control of input-affine discrete-time nonlinear systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 7, pp. 4339–4350, Jul. 2023.

[22] Y. Zhou, B. Luo, X. Wang, X. Xu, and L. Xiao, "RL-based adaptive optimal bipartite consensus control for nonlinear heterogeneous MASs via event-triggered state feedback," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 71, no. 9, pp. 4261–4273, Sep. 2024.

[23] R. Song, L. Liu, and B. Hu, "Aperiodic sampling artificial-actual H_∞ optimal control for interconnected constrained systems," *IEEE Trans. Autom. Sci. Eng.*, early access, Oct. 23, 2024, doi: 10.1109/TASE.2023.3324643.

[24] W. Bai, T. Li, Y. Long, and C. L. P. Chen, "Event-triggered multigradient recursive reinforcement learning tracking control for multiagent systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 1, pp. 366–379, Jan. 2023.

[25] Z. Peng, R. Luo, J. Hu, K. Shi, and B. K. Ghosh, "Distributed optimal tracking control of discrete-time multiagent systems via event-triggered reinforcement learning," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 69, no. 9, pp. 3689–3700, Sep. 2022.

[26] L. Xia, Q. Li, R. Song, and S. S. Ge, "Distributed optimized dynamic event-triggered control for unknown heterogeneous nonlinear MASs with input-constrained," *Neural Netw.*, vol. 154, pp. 1–12, Oct. 2022.

[27] B. Zhao, S. Zhang, and D. Liu, "Self-triggered approximate optimal neuro-control for nonlinear systems through adaptive dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 14, 2024, doi: 10.1109/TNNLS.2024.3362800.

[28] W. Xu, T. Wang, J. Qiu, and X. Liu, "A novel framework for game-based optimal event-triggered control of multi-input nonlinear systems," *IEEE Trans. Autom. Control*, vol. 69, no. 7, pp. 4867–4874, Jul. 2024.

[29] A. Sahoo and V. Narayanan, "Optimization of sampling intervals for tracking control of nonlinear systems: A game theoretic approach," *Neural Netw.*, vol. 114, pp. 78–90, Jun. 2019.

[30] J. Lu, Q. Wei, and F.-Y. Wang, "Parallel control for optimal tracking via adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 6, pp. 1662–1674, Nov. 2020.

[31] Q. Wei, S. Jiao, F.-Y. Wang, and Q. Dong, "Robust optimal parallel tracking control based on adaptive dynamic programming," *IEEE Trans. Cybern.*, vol. 54, no. 7, pp. 4308–4321, Jul. 2024.

[32] A. J. van der Schaft, "L_2-gain analysis of nonlinear systems and nonlinear state-feedback H_∞ control," *IEEE Trans. Autom. Control*, vol. 37, no. 6, pp. 770–784, Jun. 1992.

[33] X. Yang and D. Wang, "Reinforcement learning for robust dynamic event-driven constrained control," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 3, 2024, doi: 10.1109/TNNLS.2024.3394251.

[34] C. Mu, K. Wang, and Z. Ni, "Adaptive learning and sampled-control for nonlinear game systems using dynamic event-triggering strategy," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4437–4450, Sep. 2022.

[35] M. Lin, B. Zhao, and D. Liu, "Event-triggered robust adaptive dynamic programming for multiplayer Stackelberg–Nash games of uncertain nonlinear systems," *IEEE Trans. Cybern.*, vol. 54, no. 1, pp. 273–286, Jan. 2024.

[36] Z. Chen, K. Chen, and R. Tang, "Optimal synchronization with $\mathbb{L}^2$-gain performance: An adaptive dynamic programming approach," *Neural Netw.*, vol. 179, Nov. 2024, Art. no. 106566.

[37] Z. Zhang, K. Zhang, X. Xie, and V. Stojanovic, "ADP-based prescribed-time control for nonlinear time-varying delay systems with uncertain parameters," *IEEE Trans. Autom. Sci. Eng.*, early access, Apr. 18, 2024, doi: 10.1109/TASE.2024.3389020.

[38] X. Yang and Q. Wei, "Adaptive dynamic programming for robust event-driven tracking control of nonlinear systems with asymmetric input constraints," *IEEE Trans. Cybern.*, vol. 54, no. 11, pp. 6333–6344, Nov. 2024.

[39] D. Yao, H. Li, R. Lu, and Y. Shi, "Distributed sliding-mode tracking control of second-order nonlinear multiagent systems: An event-triggered approach," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3892–3902, Sep. 2020.

[40] J. Lin, B. Zhao, D. Liu, and Y. Wang, "Dynamic compensator-based near-optimal control for unknown nonaffine systems via integral reinforcement learning," *Neurocomputing*, vol. 564, Jan. 2024, Art. no. 126973.

Yongwei Zhang received the B.S. degree in automation from the School of Electronic and Information Engineering, Jiaying University, Meizhou, China, in 2016, and the Ph.D. degree in control science and engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2021.

He was a Post-Doctoral Fellow with the School of Automation, Guangdong University of Technology, from 2021 to 2023. He is currently a Lecturer with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou. His current research interests include adaptive dynamic programming and optimal control.

Weifeng Zhong received the Ph.D. degree in control science and engineering from Guangdong University of Technology, Guangzhou, China, in 2019. He was a Visiting Scholar with Nanyang Technological University, Singapore, in 2021, and a Visiting Student with The Hong Kong University of Science and Technology, Hong Kong, in 2016. He is currently an Associate Professor with Guangdong University of Technology. His research interests include optimization and control in smart grids, connected vehicles, and the Internet of Things.

Guoxu Zhou received the Ph.D. degree in intelligent signal and information processing from South China University of Technology, Guangzhou, China, in 2010.

He was a Research Scientist with the Brain Science Institute, RIKEN, Tokyo, Japan. He is currently a Full Professor and the Dean of the School of Automation, Guangdong University of Technology, Guangzhou. He has authored more than 70 peer-reviewed articles that have been published in prestigious journals, such as *Proceedings of the IEEE*, *IEEE Signal Processing Magazine*, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, and IEEE TRANSACTIONS ON IMAGE PROCESSING. His current research interests include tensor analysis, intelligent information processing, and artificial intelligence.

Prof. Zhou serves as an Associate Editor for IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.

Dr. Xie is a fellow of the Academy of Engineering Singapore, IFAC, and CAA. He was listed as a Highly Cited Researcher by Thomson Routers and Clarivate Analytics. He is the Editor-in-Chief of *Unmanned Systems* and served as an Editor for IET Book Series in Control and an Associate Editor for a number of journals, including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, and IEEE TRANSACTIONS ON CONTROL OF NETWORK SYSTEMS. He was an IEEE Distinguished Lecturer from January 2012 to December 2014 and the General Chair of the 62nd IEEE Conference on Decision and Control 2023. He is currently the Vice-President of the IEEE Control System Society.

Lihua Xie (Fellow, IEEE) received the Ph.D. degree in electrical engineering from The University of Newcastle, Callaghan, NSW, Australia, in 1992.

Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently the President's Chair of Control Engineering and the Director of the Center for Advanced Robotics Technology Innovation. He was the Head of the Division of Control and Instrumentation and the Co-Director of the Delta-NTU Corporate Laboratory for Cyber-Physical Systems. He held teaching appointments with the Department of Automatic Control, Nanjing University of Science and Technology, Nanjing, China, from 1986 to 1989. His research interests include robust control and estimation, networked control systems, multiagent networks, and unmanned systems. He has published ten books and numerous articles in the above areas and holds 25 patents/technical disclosures.

Shengli Xie (Fellow, IEEE) received the Ph.D. degree in control theory and applications from South China University of Technology, Guangzhou, China, in 1997. He is currently a Full Professor and the Head of the Institute of Intelligent Information Processing, Guangdong University of Technology, Guangzhou. He has co-authored two books and more than 150 research papers in refereed journals and conference proceedings. His research interests include blind signal processing, machine learning, and the Internet of Things. He was awarded the Second Prize of National Natural Science Award of China in 2009. He was awarded a Highly Cited Researcher. He is an Associate Editor of IEEE INTERNET OF THINGS JOURNAL and IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.

Reinforcement Learning-Based Distributed Robust Bipartite Consensus Control for Multispacecraft Systems With Dynamic Uncertainties

Yongwei Zhang [ID] , *Member, IEEE*, and Jun-Yi Li [ID] , *Member, IEEE*

Abstract—In this article, the reinforcement learning-based distributed robust bipartite consensus control of multispacecraft systems with dynamic uncertainties is investigated. The developed control structure includes two parts, i.e., integral sliding mode control and distributed optimal bipartite consensus control. In the first step, an integral sliding mode controller is designed for each following spacecraft to address matched uncertainties such that the dynamics of nominal spacecraft is obtained. In the second step, a novel performance index function, which contains consensus errors and their derivatives, is designed for each nominal spacecraft. As a result, the system assumption of zero equilibrium and the discount factor in performance index function are not required, which simplifies the controller design process and improves the practicability of the developed control method. Moreover, in order to solve the coupled Hamilton–Jacobi–Bellman equation of each following spacecraft, a novel policy iteration algorithm is designed and its properties are analyzed. Finally, a group of spacecraft is employed to verify the effectiveness of the present control scheme.

Index Terms—Bipartite consensus control, integral sliding mode (ISM) control, multispacecraft systems, neural networks (NNs), reinforcement learning (RL).

I. INTRODUCTION

SPACECRAFT control is an essential and interdisciplinary field of research within aerospace engineering that encompasses a range of vital technologies, including navigation, attitude control, energy management, communication, and satellite formation control. The development and advancement of these technologies are integral to ensuring the safety, stability, and effectiveness of space exploration missions. As space exploration missions increase in complexity and heterogeneity, the need for interaction and collaboration between spacecraft has grown significantly. Consequently, multispacecraft consensus control has gained interest within the aerospace research community. This research area has demonstrated practical applications, including cooperative control of spacecraft, such as the Tiangong-2 space lab and Shenzhou-11 spacecraft, NASA's drag-enhanced navigation technology, and constellation control of the Fengyun-4 A satellite, among others. It is widely acknowledged that the conventional consensus control involves mainly two types, namely, leader–follower consensus and leaderless consensus, which rely on the mutual cooperation among multiple intelligent agents to achieve a shared state or output variable. However, in practical scenarios, agents may not only have cooperative relationships, but also competitive relationships. For example, in Earth observation satellite formation systems, each satellite may possess different data collection tasks, requiring coordination among them to enhance overall mission performance. Furthermore, resource allocation and balance also need to be considered, leading to data rivalry between satellites, which calls for negotiation and competition among them. To capture this phenomenon, some scholars have proposed bipartite consensus control, where collaborating agents converge to the same state, while competitors converge to opposite states. In recent years, a number of academics have been making significant strides in the field of bipartite consensus control of multiagent systems. Liang et al. [1] investigated an asymmetric bipartite consensus problem of nonlinear multiagent systems by developing an event-triggered model-free adaptive control approach. Shahvali et al. [2] presented a novel fully distributed control scheme to address bipartite consensus control of fractional-order multiagent systems by adopting the backstepping technique and the neuro-adaptive update mechanism. Zhao et al. [3] addressed the

bipartite consensus control problem of heterogeneous nonaffine discrete-time multiagent systems by employing neural network (NN) and pseudopartial derivative techniques. Nevertheless, the abovementioned method only considers the stability of multiagent systems, neglecting the issue of control cost, and has not been actually applied to spacecraft control. As precision components are assembled on spacecraft, communication and execution capabilities are limited. Therefore, it is utmost significance to design efficient and energy-saving control strategies that can reduce control costs while ensuring that the spacecraft completes collaborative control tasks.

In recent years, reinforcement learning (RL) has emerged as a promising approach to solving control problems. RL-based control methods leverage the feedback loop between the control agent and its environment, enabling the agent to learn the optimal control strategy through trial-and-error interactions while ensuring system stability and minimizing control costs [4], [5], [6], [7], [8], [9], [10]. These methods are characterized by their adaptability, model-free operation, and good control performance, and have been deployed extensively to tackle a range of challenging problems in control theory, including optimal control [11], [12], trajectory tracking control [13], [14], robust control [15], [16], differential game [17], [18], and so on. Moreover, RL-based control approaches have been adopted to address control problems of practical systems. For example, Zhang et al. [19] developed an RL-based resilient event-triggered control approach to address the tracking control problem for rear-wheel-drive autonomous vehicles. In the area of spacecraft control, several scholars and experts considered the trajectory tracking or consensus control problems. Shi et al. [20] investigated the leader–follower spacecraft formation control problem by using RL-based event-triggered control approach. Yang et al. [21] tackled the attitude control problem for spacecraft with actuator misalignment and pointing constraints by using RL technique. Zhou et al. [22] suggested an online adaptive nonlinear control method based on heuristic dynamic programming to address the trajectory tracking control problem for spacecrafts. However, existing research on spacecraft consensus only considers cases where mutual cooperation exists, and concurrency of cooperation and competition in bipartite consensus control problems remains an open research issue in the field. In addition, spacecraft working for an extended period in outer space is exposed to numerous hostile factors, such as high and low temperatures, radiation, vacuum, and microgravity, which can severely impact the system components of the spacecraft. Therefore, the occurrence of model uncertainty is challenging to avoid, leading to an adverse effect on the control performance.

The direct acquisition of an optimal controller for uncertain systems remains a challenge in the field of control engineering. Consequently, researchers utilize robust control techniques in conjunction with optimal control methods to develop controllers that guarantee the robustness and the optimal performance of closed-loop systems. In recent years, RL-based robust control methods have been extensively investigated as a possible solution to this challenge. For example, Wang et al. [23] tackled the robust control problem of continuous-time (CT) systems.

By formulating a specific performance index function, the robust control problem can be transformed into an optimal control problem. In addition, an online policy iteration algorithm and an actor–critic framework were established as means of obtaining the approximate solution of the Hamilton–Jacobi–Bellman (HJB) equation. Afterward, Wang et al. [24] studied the robust control problem in the context of an event-triggering mechanism, and an RL-based event-triggered robust control scheme was established. Yang and He [25] addressed the event-triggered robust control problem of CT systems with mismatched uncertainties and input constraints by designing an auxiliary system. Although the abovementioned methods can deal with dynamic uncertainties effectively, they have some deficiencies. On the one hand, the existing results need to add an upper bounding function for the uncertain part into the performance index function, which increases the controller's conservatism and cannot obtain the desired metrics. On the other hand, the assumption of zero equilibrium is necessary in existing results since it guarantees that the performance index function is finite. Moreover, when dealing with the tracking problem, the control input fails to approach zero, resulting in the nonconvergence of the performance index function. Most of existing approaches add a discount factor on the performance index function, which increase the complexity of the RL algorithm and even lead to algorithm nonconvergence. In conclusion, the abovementioned approaches are suitable for single-agent systems only and do not apply to spacecrafts. Furthermore, the existing results have shortcomings in dealing with dynamic uncertainties. Due to the extensive application of spacecraft in space exploration, satellite navigation, and aerospace loading, and its susceptibility to dynamic uncertainty, investigating the distributed robust bipartite consensus (DRBC) control problem of spacecrafts is crucial.

In this article, an integral sliding mode (ISM)-based DRBC control approach for spacecrafts is presented. The innovations and contributions of this article are outlined as follows.

1) This article extends the RL-based distributed control method to spacecrafts. By designing an ISM-based DRBC control scheme, the cooperative spacecrafts in the same group can achieve consensus, while the competing spacecrafts in different groups converge to the opposite state.

2) Unlike the existing approaches [23], [24] that tackled robust control problems by introducing an upper bound function in the performance index function, this article employs ISM technique to mitigate the influence of matched uncertainties of each following spacecraft. Moreover, different from the work in [26], which required the bounded assumption of the matched uncertainties, an adaptive term is designed in ISM control laws to remove the bound assumption.

3) By developing a novel performance index function, which includes bipartite consensus errors and their derivatives for each following spacecraft, the assumption of zero equilibrium is removed and the discount factor is not required, which simplifies RL algorithm design and improves the practicability of the control method.
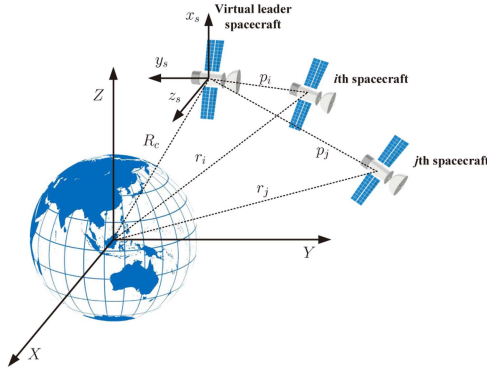
Fig. 1. Multispacecraft systems.

The rest of this article is organized as follows. In Section II, the graph theory and the problem statement are presented. In Section III, the ISM control law is developed for each following spacecraft to handle the matched uncertainties. Moreover, the distributed optimal bipartite consensus controller design of the nominal spacecraft, the novel policy iteration algorithm design and its properties analysis, and the NN implementation are provided. Section IV showcases the efficacy of the ISM-based DRBC control scheme through simulation experiments. Finally, Section V concludes this article.

II. PRELIMINARIES

A. Signed Graph Theory

Consider a signed graph denoted as $\mathscr{S}_g = (\mathscr{S}_v, \mathscr{S}_e, \mathscr{S}_a)$, where $\mathscr{S}_v = \{v_1, v_2, \ldots, v_{\mathcal{N}}\}$ is a node set, $\mathscr{S}_e = \{(v_i, v_j) | v_i, v_j \in \mathscr{S}_v\} \subseteq \mathscr{S}_v \times \mathscr{S}_v$ is an edge set, and $\mathscr{S}_a = [a_{ij}]_{\mathcal{N} \times \mathcal{N}}$ is a weighted adjacency matrix. Note that if and only if the agent i and the agent j are directly connected, then $(v_i, v_j) \in \mathscr{S}_e$. Furthermore, $a_{ij} > 0$ represents that the agent i and the agent j are cooperative, and $a_{ij} < 0$ denotes the agent i and the agent j are competitive. Define $\mathcal{N}_i = \{j : (v_i, v_j) \in \mathscr{S}_e, j \neq i\}$ is a set of neighbors of the agent i, and $\bar{\mathcal{N}}_i$ be a set of the agent i and its neighbors. The degree matrix of $\mathscr{S}_g$ is denoted as $\mathscr{D} = \mathrm{diag}\{\nu_1, \ldots, \nu_{\mathcal{N}}\}$, where $\nu_i = \sum_{j \in \mathcal{N}_i} |a_{ij}|$. The Laplacian matrix is defined as $\mathscr{L} = \mathscr{D} - \mathscr{S}_a$.

Definition 1: A signed graph is said to be structurally balanced if there exists a partition of its signed edges into two sets $\mathscr{S}_{v1}$ and $\mathscr{S}_{v2}$ such that the following conditions are satisfied.
1) $\mathscr{S}_v = \mathscr{S}_{v1} \cup \mathscr{S}_{v2}, \mathscr{S}_{v1} \cap \mathscr{S}_{v2} = \emptyset$.
2) $\forall i, j \in \mathscr{S}_{v,l}(l \in \{1, 2\}), a_{ij} > 0$.
3) $\forall i \in \mathscr{S}_{v,l}, j \in \mathscr{S}_{v,q}, l \neq q(l, q \in \{1, 2\}), a_{ij} < 0$.

B. Problem Statement

Fig. 1 portrays a schematic diagram of a multispacecraft system, where $L_e = \{X, Y, Z\}$ denotes the Earth center inertial coordinate frame and $L_c = \{x_s, y_s, z_s\}$ represents the local vertical local horizontal frame. The virtual leader spacecraft can provide the information of position and velocity for the follower spacecraft [29]. The dynamics of the ith following spacecraft is

described as

$$\dot{p}_i = \zeta_i$$

$$m_i \dot{\zeta}_i + \mathscr{C}_i \zeta_i + \mathcal{D}_i p_i + \mathcal{M}_i + \Psi_i(\dot{\zeta}_i) = u_i$$

where $p_i = [p_{i,x}, p_{i,y}, p_{i,z}] \in \mathbb{R}^3$ is the position vector of the ith spacecraft relative to the virtual leader spacecraft, $\zeta_i \in \mathbb{R}^3$ is velocity vector, m_i is the mass of the ith spacecraft, $\mathcal{M}_i = m_i \chi [R_c/r_i^3 - 1/R_c^2, 0, 0]^{\mathsf{T}}$ is the gravity vector, $\Psi_i(\dot{\zeta}_i) \in \mathbb{R}^3$ is the dynamic uncertainty, $u_i \in \mathbb{R}^3$ is the control input, and $\mathscr{C}_i \in \mathbb{R}^{3 \times 3}$ and $\mathcal{D}_i \in \mathbb{R}^{3 \times 3}$ are system matrices, which are given as

$$\mathcal{D}_i = m_i \begin{bmatrix} \chi/r_i^3 - \dot{\varphi}^2 & -\ddot{\varphi} & 0 \\ \ddot{\varphi} & \chi/r_i^3 - \dot{\varphi}^2 & 0 \\ 0 & 0 & \chi/r_i^3 \end{bmatrix}$$

$$\mathscr{C}_i = 2m_i \begin{bmatrix} 0 & -\dot{\varphi} & 0 \\ \dot{\varphi} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, n_c = \sqrt{\chi/a_c^3}$$

$$r_i = \sqrt{(R_c + p_{i,x})^2 + p_{i,y}^2 + p_{i,z}^2}$$

$$R_c = a_c(1 - \sigma_c^2)/(1 + \sigma_c \cos(\varphi))$$

$$\dot{\varphi} = n_c (1 - \sigma_c \cos(\varphi))^2 / (1 - \sigma_c^2)^{3/2} \quad (1)$$

where φ, a_c, and σ_c represent the true anomaly, semimajor axis, and eccentricity of the orbit, respectively. χ is the Earth's gravitational constant. Let $z_i = [z_{i,1}, z_{i,2}, z_{i,3}, z_{i,4}, z_{i,5}, z_{i,6}]^{\mathsf{T}} = [p_i^{\mathsf{T}}, \zeta_i^{\mathsf{T}}]^{\mathsf{T}} \in \mathbb{R}^6$, the dynamic of the ith spacecraft can be expressed as

$$\dot{z}_i = \mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)(u_i + \Psi_i(z_i)) \quad (2)$$

where

$$\mathcal{F}_i(z_i) = \begin{bmatrix} \zeta_i \\ m_i^{-1}(-\mathscr{C}_i \zeta_i - \mathcal{D}_i p_i - \mathcal{M}_i) \end{bmatrix}$$

$$\mathcal{G}_i(z_i) = \begin{bmatrix} \mathbf{0} \\ m_i^{-1} I \end{bmatrix}.$$

The dynamics model of the leader is provided as

$$\dot{z}_0 = \mathcal{F}_0(z_0) \quad (3)$$

where $z_0 \in \mathbb{R}^6$ is the state vector and $\mathcal{F}_0(\cdot) \in \mathbb{R}^6$ is a differentiable function.

Assumption 1: The time derivative of the dynamic uncertainties $\Psi_i(z_i)$ evolves slowly, i.e., $\dot{\Psi}_i(z_i) \approx 0$.

Definition 2: In the bipartite consensus control problem, the system state of the ith spacecraft satisfies

$$\lim_{t \to \infty} (z_i - s_i z_0) = 0 \quad (4)$$

where $s_i = 1$ for $i \in \mathscr{S}_{v1}$, and $s_i = -1$ for $i \in \mathscr{S}_{v2}$.

Our goal is to develop ISM-based DRBC controllers to achieve bipartite consensus among uncertain spacecrafts, that is, cooperative spacecrafts converge on a common objective, while competing spacecrafts reach to a different objective.

III. ISM-BASED DRBC CONTROL OF SPACECRAFTS

The ISM-based DRBC controller of the ith following spacecraft is developed as follows:

$$u_i(t) = \mu_i(t) + \mu_{ic}(t) \tag{5}$$

where $\mu_i(t) \in \mathbb{R}^3$ refers to a continuous distributed optimal bipartite consensus control law, while $\mu_{ic}(t) \in \mathbb{R}^3$ is a discontinuous ISM control law, which is utilized to handle the matched uncertainties of each following spacecraft.

A. ISM Controller Design

For the sake of designing the discontinuous ISM control law $\mu_{ic}(t)$, the integral sliding function is chosen as follows:

$$\mathcal{S}_i\left(z_i(t), t\right) = -\int_0^t \mathcal{P}_i(z_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i\right) d\tau + \mathcal{K}_i(z_i) - \mathcal{K}_i\left(z_i(0)\right) \tag{6}$$

where $\mathcal{K}_i(z_i) \in \mathbb{R}^3$ and $\mathcal{P}_i(z_i) = \frac{\partial \mathcal{K}_i(z_i)}{z_i} \in \mathbb{R}^{3 \times 6}$ are designed functions. It is noteworthy that the initial system state starts at sliding mode surface since that $\mathcal{S}_i(z_i(0), 0) = 0$ when $t = 0$. Therefore, the approaching condition of the sliding mode surface is not required [30]. The derivative of (6) with respect to time is calculated by

$$\dot{\mathcal{S}}_i(z_i, t) = \mathcal{P}_i(z_i)\dot{z}_i - \mathcal{P}_i(z_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i\right). \tag{7}$$

In order to guarantee that the system state remains on the sliding surface, the discontinuous ISM control law $\mu_{ic}(t)$ of the ith following spacecraft is designed as

$$\mu_{ic}(t) = -\mathcal{A}_i \mathrm{sgn}\left(\mathcal{G}_i^{\mathsf{T}}(z_i)\mathcal{P}_i^{\mathsf{T}}(z_i)\mathcal{S}_i\right) - \hat{\Psi}_i(z_i) \tag{8}$$

where $\mathcal{A}_i$ is a positive constant, $\hat{\Psi}_i(z_i)$ is the estimate of $\Psi_i(z_i)$, $\mathrm{sgn}(\mathbb{A}) = [\mathrm{sgn}(\mathbb{A}_1), \ldots, \mathrm{sgn}(\mathbb{A}_n)]$, where $\mathbb{A} = [\mathbb{A}_1, \ldots, \mathbb{A}_n]$, and $\mathrm{sgn}(\cdot)$ is a sign function.

Theorem 1: Consider the ith spacecraft with matched uncertainty (2), the integral sliding function given by (6), and Assumption 1, the discontinuous ISM control law (8) guarantees that the system state stay on the sliding mode surface with the adaptive update law

$$\dot{\hat{\Psi}}_i(z_i) = \beta_i \mathcal{G}_i^{\mathsf{T}}(z_i)\mathcal{P}_i^{\mathsf{T}}(z_i)\mathcal{S}_i \tag{9}$$

where β_i is a positive constant.

Proof: Select the Lyapunov function candidate as

$$\mathcal{L}_{T1} = \frac{1}{2}\mathcal{S}_i^{\mathsf{T}}\mathcal{S}_i + \frac{1}{2\beta_i}\tilde{\Psi}_i^{\mathsf{T}}(z_i)\tilde{\Psi}_i(z_i) \tag{10}$$

where $\tilde{\Psi}_i(z_i) = \Psi_i(z_i) - \hat{\Psi}_i(z_i)$ is the estimation error. Calculating the time derivative of (10) and adopting the system dynamics (2), it yields that

$$\dot{\mathcal{L}}_{T1} = \mathcal{S}_i^{\mathsf{T}}(\mathcal{P}_i(z_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\left(u_i + \Psi_i(z_i)\right)\right)$$
$$-\mathcal{P}_i(z_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i\right)) - \frac{1}{\beta_i}\tilde{\Psi}_i^{\mathsf{T}}(z_i)\dot{\hat{\Psi}}_i(z_i)$$
$$= \mathcal{S}_i^{\mathsf{T}}\left(\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)u_i + \mathcal{P}_i(z_i)\mathcal{G}_i(z_i)\Psi_i(z_i)\right)$$

$$-\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)\mu_i) - \frac{1}{\beta_i}\tilde{\Psi}_i^{\mathsf{T}}(z_i)\dot{\hat{\Psi}}_i(z_i)$$
$$= -\mathcal{A}_i\mathcal{S}_i^{\mathsf{T}}\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)\mathrm{sgn}\left(\mathcal{G}_i^{\mathsf{T}}(z_i)\mathcal{P}_i^{\mathsf{T}}(z_i)\mathcal{S}_i\right)$$
$$+ \mathcal{S}_i^{\mathsf{T}}\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)\left(\Psi_i(z_i) - \hat{\Psi}_i(z_i)\right)$$
$$- \frac{1}{\beta_i}\tilde{\Psi}_i^{\mathsf{T}}(z_i)\dot{\hat{\Psi}}_i(z_i)$$
$$= -\mathcal{A}_i\left|\mathcal{S}_i^{\mathsf{T}}\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)\right| + \mathcal{S}_i^{\mathsf{T}}\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)\tilde{\Psi}_i(z_i)$$
$$- \frac{1}{\beta_i}\tilde{\Psi}_i^{\mathsf{T}}(z_i)\dot{\hat{\Psi}}_i(z_i). \tag{11}$$

Letting $\mathcal{P}_i(z_i) = \mathcal{G}_i^+(z_i)$. Since $|\mathcal{S}_i^{\mathsf{T}}\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)| \geq \|\mathcal{S}_i^{\mathsf{T}}\mathcal{P}_i(z_i)\mathcal{G}_i(z_i)\|$ is held and according to (9), we further have

$$\dot{\mathcal{L}}_{T1} \leq -\|\mathcal{S}_i\|\mathcal{A}_i + \tilde{\Psi}_i^{\mathsf{T}}(z_i)\left(\mathcal{G}_i^{\mathsf{T}}(z_i)\mathcal{P}_i^{\mathsf{T}}(z_i)\mathcal{S}_i - \frac{1}{\beta_i}\dot{\hat{\Psi}}_i(z_i)\right)$$
$$\leq -\|\mathcal{S}_i\|\mathcal{A}_i. \tag{12}$$

Therefore, $\dot{\mathcal{L}}_{T1} \leq 0$ is held. This signifies that the system state remains on the sliding mode surface when applying the ISM control law (8). The proof is finished.

According to (5) and (7), and letting $\dot{\mathcal{S}}_i(z_i, t) = 0$, the equivalent control μ_{iceq} of the ith spacecraft is calculated by

$$\mu_{iceq} = -\Psi_i(z_i). \tag{13}$$

Substituting (13) into (2), the ith spacecraft without dynamic uncertainties is formulated as

$$\dot{z}_i = \mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i. \tag{14}$$

Remark 1: The adaptive updating law (9) is designed based on the Lyapunov stability principle, as illustrated in Theorem 1. To ensure that the system state remains on the sliding surface, it is required that (12) be negative. Therefore, when the adaptive updating law is formulated as in (9), (12) being negative holds. It is noted that the introduction of the adaptive updating term aims to relax the assumption of the upper bound function of uncertain term, which enhances the practicality of the control method.

B. Distributed Optimal Bipartite Consensus Controller Design

In the following, the distributed optimal bipartite consensus control law is designed for each spacecraft. The local neighborhood bipartite consensus error of the ith spacecraft is defined as

$$\delta_i = \sum_{j \in \mathcal{N}_i} |a_{ij}|\left(z_i - \mathrm{sgn}(a_{ij})z_j\right) + b_i(z_i - s_i z_0) \tag{15}$$

where $b_i \geq 0$ is a pinning gain. Thus, the dynamics of the local neighborhood bipartite consensus error is obtained by

$$\dot{\delta}_i = \sum_{j \in \mathcal{N}_i} |a_{ij}|\left(\dot{z}_i - \mathrm{sgn}(a_{ij})\dot{z}_j\right) + b_i(\dot{z}_i - s_i\dot{z}_0)$$
$$= (\nu_i + b_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i\right) - b_i s_i \mathcal{F}_0(z_0)$$

$$-\sum_{j \in \mathcal{N}_i} a_{ij} \left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j\right). \tag{16}$$

The performance index function of the ith spacecraft is designed as

$$\mathcal{J}_i(\delta_i, \dot{\delta}_i) = \int_t^\infty \mathcal{C}_i \left(\delta_i(\nu), \dot{\delta}_i(\nu)\right) d\nu$$

$$= \int_t^\infty \left(\delta_i^\mathsf{T}(\nu)\mathcal{Q}_i\delta_i(\nu) + \dot{\delta}_i^\mathsf{T}(\nu)\mathcal{R}_i\dot{\delta}_i(\nu)\right) d\nu \tag{17}$$

where $\mathcal{Q}_i \in \mathbb{R}^{6 \times 6}$ and $\mathcal{R}_i \in \mathbb{R}^{6 \times 6}$ are positive definite matrices. The Hamiltonian of the ith spacecraft is given as

$$\mathcal{H}_i \left(\delta_i, \dot{\delta}_i, \nabla \mathcal{J}_i(\delta_i, \dot{\delta}_i)\right)$$

$$= \nabla \mathcal{J}_i^\mathsf{T}(\delta_i, \dot{\delta}_i) \bigg((\nu_i + b_i) \left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i\right)$$

$$- \sum_{j \in \mathcal{N}_i} a_{ij} \left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j\right) - b_i s_i \mathcal{F}_0(z_0) \bigg)$$

$$+ \mathcal{C}_i(\delta_i, \dot{\delta}_i). \tag{18}$$

Thus, the optimal performance index function satisfies

$$\mathcal{J}_i^*(\delta_i, \dot{\delta}_i) = \min_{\mu_i \in \Re(\Omega)} \int_t^\infty \mathcal{C}_i \left(\delta_i(\nu), \dot{\delta}_i(\nu)\right) d\nu \tag{19}$$

where $\Re(\Omega)$ is a set of admissible controls. The distributed optimal bipartite consensus control law of the ith spacecraft is given by

$$\mu_i^* = -\frac{1}{2(\nu_i + b_i)} \left(\mathcal{G}_i^\mathsf{T}(z_i)\mathcal{R}_i\mathcal{G}_i(z_i)\right)^{-1}$$

$$\times \left(2\mathcal{G}_i^\mathsf{T}(z_i)\mathcal{R}_i\Gamma_i + \mathcal{G}_i^\mathsf{T}(z_i)\nabla\mathcal{J}_i(\delta_i, \dot{\delta}_i)\right) \tag{20}$$

where $\Gamma_i = (\nu_i + b_i)\mathcal{F}_i(z_i) - \sum_{j \in \mathcal{N}_i} a_{ij}(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j) - b_i s_i \mathcal{F}_0(z_0)$. According to (18) and (20), the coupled HJB equation is provided as

$$0 = \mathcal{H}_i \left(\delta_i, \dot{\delta}_i, \nabla \mathcal{J}_i^*(\delta_i, \dot{\delta}_i)\right)$$

$$= \nabla \mathcal{J}_i^{*\mathsf{T}}(\delta_i) \bigg((\nu_i + b_i) \left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^*\right)$$

$$- \sum_{j \in \mathcal{N}_i} a_{ij} \left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^*\right) - b_i s_i \mathcal{F}_0(z_0) \bigg)$$

$$+ \mathcal{C}_i \left(\delta_i, \dot{\delta}_i\right). \tag{21}$$

From (20) and (21), we can find that the distributed optimal bipartite consensus control law μ_i^* depends on the optimal performance index function $\mathcal{J}_i^*(\delta_i, \dot{\delta}_i)$. However, it is awkward to obtain the $\mathcal{J}_i^*(\delta_i, \dot{\delta}_i)$ from (21) directly. In the following, a novel policy iteration algorithm is established to cope with this issue.

Remark 2: Traditional performance index functions are typically quadratic with respect to system states and control inputs. However, in tracking control problems, control inputs

---

Algorithm 1: Novel Policy Iteration Algorithm.

Step 1: Let $k = 0$ and select initial admissible control law $\mu_i^0 \in \Re(\Omega)$. Choose a computation precision ξ.

Step 2: (**Policy evaluation**) Calculate the iterative performance index function $\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i)$ by

$$0 = \mathcal{C}_i \left(\delta_i, \dot{\delta}_i, \mu_i^{(k-1)}, \mu_{-i}^{(k-1)}\right)$$

$$+ \nabla \mathcal{J}_i^{(k)\mathsf{T}}(\delta_i, \dot{\delta}_i) \bigg((\nu_i + b_i)\mathcal{Q}_i - b_i s_i \mathcal{F}_0(z_0)$$

$$- \sum_{j \in \mathcal{N}_i} a_{ij}\mathcal{Q}_j \bigg). \tag{22}$$

where $\mathcal{Q}_i = \mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^{(k-1)}$ and $\mathcal{Q}_j = \mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k-1)}$.

Step 3: (**Policy improvement**) Update the control law by

$$\mu_i^{(k)}(\delta_i) = -\frac{1}{2(\nu_i + b_i)} \left(\mathcal{G}_i^\mathsf{T}(z_i)\mathcal{R}_i\mathcal{G}_i(z_i)\right)^{-1}$$

$$\times \left(2\mathcal{G}_i^\mathsf{T}(z_i)\mathcal{R}_i\mathcal{T}_i + \mathcal{G}_i^\mathsf{T}(z_i)\nabla\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i)\right), \tag{23}$$

where $\mathcal{T}_i = (\nu_i + b_i)\mathcal{F}_i(z_i) - \sum_{j \in \mathcal{N}_i} a_{ij}(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k-1)}) - b_i s_i \mathcal{F}_0(z_0)$.

Step 4: If $|\mathcal{J}_i^{(k+1)}(\delta_i, \dot{\delta}_i) - \mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i)| < \xi$, goto Step 5; else, let $k = k + 1$, go back to Step 2 and continue.

Step 5: Stop Algorithm.

---

do not tend toward zero. To prevent the divergence of the performance index function, a discount factor is commonly included. However, the choice of the discount factor will impact the convergence of the adaptive dynamic programming (ADP) algorithm and even the stability of the closed-loop system. In order to address this challenge, this article introduces a new performance index function that incorporates consensus errors and their derivatives, which eliminates the need of the discount factor. As a result, the process of controller design is simplified and the effectiveness of the control method is enhanced.

Remark 3: The developed ISM-based DRBC controller contains two parts, that is, the distributed optimal bipartite consensus control law μ_i and the ISM control law μ_{ic}. It is noted that for the ith nominal spacecraft (14) without dynamic uncertainties, only μ_i is required and it can guarantee that the ith nominal spacecraft achieves optimal bipartite consensus control. However, the ith spacecraft contains the dynamic uncertainty $\Psi_i(\dot{\zeta}_i)$. Therefore, in order to eliminate the effect of the dynamic uncertainty, the ISM control law μ_{ic} is developed. In fact, if the dynamic uncertainty does not exist, according to (7), one can know that the derivative of the integral sliding function $\dot{\mathcal{S}}$ is equal to 0. Since the initial value of $\mathcal{S}$ is 0, then $\mathcal{S}$ will remain at 0, which leads the ISM control law μ_{ic} being 0 as well. Therefore, the ISM

control law will not have any impact on the nominal spacecraft. Only in the presence of uncertainties, the ISM control law μ_{ic} will generate corresponding responses to counteract the effect of the dynamic uncertainty. In general, these two control laws each have their own responsibilities and will not lead to excessive control.

C. Novel Policy Iteration Algorithm

In order to obtain the optimal performance index function $\mathcal{J}_i^*(\delta_i, \dot{\delta}_i)$ and the distributed optimal bipartite consensus control law μ_i^*, inspired by the results in [11] and [27], a novel policy iteration algorithm is designed in Algorithm 1. Moreover, the properties of the novel policy iteration algorithm are analyzed in Theorems 2 and 3. It is demonstrated that iterative performance index functions exhibit a monotonic decrease and converge to the optimal value, while iterative control laws ensure the stability of the local neighborhood bipartite consensus error.

Theorem 2: Consider the ith following spacecraft (14), the iterative performance index function and the iterative distributed bipartite consensus control law are given by (22) and (23), respectively, if $\mu_i^{(k)} \in \Re(\Omega)$, then the following conditions hold.

1) $\mu_i^{(k+1)} \in \Re(\Omega)$.
2) $\mathcal{J}_i^{(k+1)}(\delta_i, \dot{\delta}_i) \leq \mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i)$.

Proof: 1) In light of $\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i) \geq 0$, the Lyapunov function candidate is chosen as

$$\mathcal{L}_{2T} = \mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i). \tag{24}$$

Taking the time derivative of $\mathcal{L}_{2T}$ and adopting the following system dynamics:

$$\dot{\delta}_i = (\nu_i + b_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^{(k+1)}\right) - b_i s_i \mathcal{F}_0(z_0)$$
$$- \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)$$

we have

$$\dot{\mathcal{J}}_i^{(k)}(\delta_i, \dot{\delta}_i)$$
$$= \nabla\mathcal{J}_i^{(k)\mathsf{T}}(\delta_i, \dot{\delta}_i)\left((\nu_i + b_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^{(k+1)}\right)\right.$$
$$\left. - b_i s_i \mathcal{F}_0(z_0) - \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)\right). \tag{25}$$

According to (20), we can get

$$\nabla\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i) = -\frac{1}{2(\nu_i + b_i)}\mathcal{R}_i\Bigg((\nu_i + b_i)\mathcal{F}_i(z_i)$$
$$- \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)$$
$$- b_i s_i \mathcal{F}_0(z_0) + (\nu_i + b_i)\mathcal{G}_i(z_i)\mu_i^{(k+1)}\Bigg). \tag{26}$$

By using (25) and (26), it holds that

$$\dot{\mathcal{J}}_i^{(k)}(\delta_i, \dot{\delta}_i)$$
$$= -\frac{1}{2(\nu_i + b_i)}\Bigg((\nu_i + b_i)\mathcal{F}_i(z_i)$$
$$- \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)$$
$$- b_i s_i \mathcal{F}_0(z_0) + (\nu_i + b_i)\mathcal{G}_i(z_i)\mu_i^{(k+1)}\Bigg)^{\mathsf{T}}\mathcal{R}_i$$
$$\times \Bigg((\nu_i + b_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^{(k+1)}\right)$$
$$- b_i s_i \mathcal{F}_0(z_0) - \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)\Bigg)$$
$$\leq 0.$$

Therefore, $\mu_i^{(k+1)}$ is an admissible control law.

2) Based on (22), it is easy to infer that if $\mu_i^{(k)}$ and $\mu_{-i}^{(k)}$ are admissible control laws, the iterative performance index function $\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i)$ satisfies

$$0 = \mathcal{C}_i\left(\delta_i, \dot{\delta}_i, \mu_i^{(k)}, \mu_{-i}^{(k)}\right)$$
$$+ \nabla\mathcal{J}_i^{(k)\mathsf{T}}(\delta_i, \dot{\delta}_i)\Bigg((\nu_i + b_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^{(k)}\right)$$
$$- b_i s_i \mathcal{F}_0(z_0) - \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)\Bigg). \tag{27}$$

Considering (23), we can get

$$\mathcal{C}_i\left(\delta_i, \dot{\delta}_i, \mu_i^{(k+1)}, \mu_{-i}^{(k)}\right)$$
$$+ \nabla\mathcal{J}_i^{(k)\mathsf{T}}(\delta_i, \dot{\delta}_i)\Bigg((\nu_i + b_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^{(k+1)}\right)$$
$$- b_i s_i \mathcal{F}_0(z_0) - \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)\Bigg)$$
$$\leq 0. \tag{28}$$

According to (28), we can get

$$\dot{\mathcal{J}}_i^{(k)\mathsf{T}}(\delta_i, \dot{\delta}_i)$$
$$= \nabla\mathcal{J}_i^{(k)\mathsf{T}}(\delta_i, \dot{\delta}_i)\Bigg((\nu_i + b_i)\left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i)\mu_i^{(k+1)}\right)$$
$$- b_i s_i \mathcal{F}_0(z_0) - \sum_{j \in \mathcal{N}_i} a_{ij}\left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j)\mu_j^{(k)}\right)\Bigg)$$
$$\leq -\mathcal{C}_i\left(\delta_i, \dot{\delta}_i, \mu_i^{(k+1)}, \mu_{-i}^{(k)}\right). \tag{29}$$

Integrate both sides of (29), it follows that:

$$\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i) \geq \mathcal{J}_i^{(k+1)}(\delta_i, \dot{\delta}_i). \tag{30}$$

The proof is concluded.

Theorem 3: Let $\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i)$ and $\mu_i^{(k)}(\delta_i)$ be obtained by (22) and (23), respectively. Then, the iterative performance index function $\mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i)$ and the iterative control law $\mu_i^{(k)}(\delta_i)$ converge to $\mathcal{J}_i^*(\delta_i, \dot{\delta}_i)$ and $\mu_i^*(\delta_i)$, respectively, as $k \to \infty$, i.e.,

$$\lim_{k \to \infty} \mathcal{J}_i^{(k)}(\delta_i, \dot{\delta}_i) = \mathcal{J}_i^*(\delta_i, \dot{\delta}_i)$$

$$\lim_{k \to \infty} \mu_i^{(k)}(\delta_i) = \mu_i^*(\delta_i).$$

Proof: The proof of Theorem 3 closely resembles that presented in [27] and [28], and hence, the details are omitted in this context.

D. NN Implementation

In the following, critic NNs are built to acquire the approximate solution of the coupled HJB equation. By utilizing the NN, the optimal performance index $\mathcal{J}_i^*(\delta_i, \dot{\delta}_i)$ is represented by

$$\mathcal{J}_i^*(\delta_i, \dot{\delta}_i) = \varpi_{ic}^{*\mathsf{T}} \varrho_{ic}(\delta_i) + \epsilon_{ic}(\delta_i) \tag{31}$$

where $\varpi_{ic}^* \in \mathbb{R}^{h_{ic}}$ is the ideal weight vector, $\varrho_{ic}(\delta_i) \in \mathbb{R}^{h_{ic}}$ is the activation function, h_{ic} is the number of hidden layer neurons, and $\epsilon_{ic}(\delta_i) \in \mathbb{R}$ is the approximation error. Then, the partial derivative of $\mathcal{J}_i^*(\delta_i, \dot{\delta}_i)$ with respect to δ_i is calculated by

$$\nabla \mathcal{J}_i^*(\delta_i, \dot{\delta}_i) = \nabla \varrho_{ic}^{\mathsf{T}}(\delta_i) \varpi_{ic}^* + \nabla \epsilon_{ic}(\delta_i). \tag{32}$$

The approximate performance index function is formulated as

$$\hat{\mathcal{J}}_i(\delta_i, \dot{\delta}_i) = \hat{\varpi}_{ic}^{\mathsf{T}} \varrho_{ic}(\delta_i) \tag{33}$$

where $\hat{\varpi}_{ic}$ is the estimate of ϖ_{ic}^*. Similarly, we can get

$$\nabla \hat{\mathcal{J}}_i(\delta_i, \dot{\delta}_i) = \nabla \varrho_{ic}^{\mathsf{T}}(\delta_i) \hat{\varpi}_{ic}. \tag{34}$$

According to (20) and (32), the distributed optimal bipartite consensus control law of the ith following spacecraft is rewritten as follows:

$$\mu_i^* = -\frac{1}{2(\nu_i + b_i)} \left(\mathcal{G}_i^{\mathsf{T}}(z_i) \mathcal{R}_i \mathcal{G}_i(z_i) \right)^{-1}$$
$$\times \left(\mathcal{G}_i^{\mathsf{T}}(z_i) \left(\nabla \varrho_{ic}^{\mathsf{T}}(\delta_i) \varpi_{ic}^* + \nabla \epsilon_{ic}(\delta_i) \right) + 2 \mathcal{G}_i^{\mathsf{T}}(z_i) \mathcal{R}_i \Gamma_i \right) \tag{35}$$

where $\Gamma_i = (\nu_i + b_i) \mathcal{F}_i(z_i) - \sum_{j \in \mathcal{N}_i} a_{ij}(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j) \mu_j) - b_i s_i \mathcal{F}_0(z_0)$. Based on (33), the approximate distributed bipartite consensus control law is provided as follows:

$$\hat{\mu}_i = -\frac{1}{2(\nu_i + b_i)} \left(\mathcal{G}_i^{\mathsf{T}}(z_i) \mathcal{R}_i \mathcal{G}_i(z_i) \right)^{-1}$$
$$\times \left(\mathcal{G}_i^{\mathsf{T}}(z_i) \nabla \varrho_{ic}^{\mathsf{T}}(\delta_i) \hat{\varpi}_{ic} + 2 \mathcal{G}_i^{\mathsf{T}}(z_i) \mathcal{R}_i \Gamma_i \right). \tag{36}$$

Combining (21) and (36), the approximate Hamiltonian is given as follows:

$$\hat{\mathcal{H}}_i \left(\delta_i, \dot{\delta}_i, \hat{\varpi}_{ic} \right)$$
$$= \mathcal{C}_i \left(\delta_i, \dot{\delta}_i \right) + \hat{\varpi}_{ic}^{\mathsf{T}} \nabla \varrho_{ic}(\delta_i) \left((\nu_i + b_i) \left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i) \hat{\mu}_i \right) \right.$$
$$\left. - \sum_{j \in \mathcal{N}_i} a_{ij} \left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j) \hat{\mu}_j \right) - b_i s_i \mathcal{F}_0(z_0) \right)$$

$$\triangleq e_{ic}. \tag{37}$$

By employing gradient descent algorithm on $E_{ic} = \frac{1}{2} e_{ic}^{\mathsf{T}} e_{ic}$, the weight tuning law is designed as follows:

$$\dot{\hat{\varpi}}_{ic} = -\alpha_c \frac{1}{(1 + \Delta_i^{\mathsf{T}} \Delta_i)^2} \left(\frac{\partial E_{ic}}{\partial \hat{\varpi}_{ic}} \right)$$
$$= -\frac{\alpha_c \Delta_i}{(1 + \Delta_i^{\mathsf{T}} \Delta_i)^2} \left(\hat{\varpi}_{ic}^{\mathsf{T}} \Delta_i + \mathcal{C}_i \left(\delta_i, \dot{\delta}_i \right) \right) \tag{38}$$

where $\alpha_c > 0$ denotes the learning rate and

$$\Delta_i = \nabla \varrho_{ic}(\delta_i) \left((\nu_i + b_i) \left(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i) \hat{\mu}_i \right) \right.$$
$$\left. - \sum_{j \in \mathcal{N}_i} a_{ij} \left(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j) \hat{\mu}_j \right) - b_i s_i \mathcal{F}_0(z_0) \right).$$

Let the critic NN weight estimation error be $\tilde{\varpi}_{ic} = \varpi_{ic}^* - \hat{\varpi}_{ic}$. Inspired by Yang and He [25], we have

$$\dot{\tilde{\varpi}}_{ic} = -\dot{\hat{\varpi}}_{ic}$$
$$= -\alpha_c \hbar_i \hbar_i^{\mathsf{T}} \tilde{\varpi}_{ic} + \frac{\alpha_c \hbar_i}{1 + \Delta_i^{\mathsf{T}} \Delta_i} \varsigma_i \tag{39}$$

where $\hbar_i = \frac{\Delta_i}{1 + \Delta_i^{\mathsf{T}} \Delta_i}$, and $\varsigma_i = -\nabla \epsilon_{ic}^{\mathsf{T}}(\delta_i)((\nu_i + b_i)(\mathcal{F}_i(z_i) + \mathcal{G}_i(z_i) \hat{\mu}_i) - \sum_{j \in \mathcal{N}_i} a_{ij}(\mathcal{F}_j(z_j) + \mathcal{G}_j(z_j) \hat{\mu}_j) - b_i s_i \mathcal{F}_0(z_0))$ is the residual error.

Theorem 4: Consider the ith spacecraft (14) and the critic NN weight tuning law provided by (38), then the critic NN weight estimation error is ensured to be uniform ultimate boundedness (UUB).

Proof: Due to space limitation, the proof of Theorem 4 is provided in the Supplementary Material.

Remark 4: Algorithm 1 provides a specific procedure of the policy iteration. By iteratively alternating between the policy evaluation and the policy improvement, we can obtain the optimal performance index function and the distributed optimal bipartite consensus control law. However, since the performance index function is unknown at each iteration, we employ the critic NN to implement Algorithm 1. The detailed process is explained as follows. In the kth policy (22), our objective is to find the performance index function $\mathcal{J}_i^{(k)}(\cdot)$ such that (22) holds. Since the $\mathcal{J}_i^{(k)}(\cdot)$ is unknown, we utilize the critic NN to approximate its value, which can be represented as $\mathcal{J}_i^{(k)}(\cdot) = \hat{\varpi}_{ic}^{\mathsf{T}} \varrho_{ic}(\delta_i)$. As the weight $\hat{\varpi}_{ic}$ is not the ideal one, the right-hand side of (22) is not equal to 0. To make the approximate weight $\hat{\varpi}_{ic}$ closer to the ideal weight, we define the right-hand side of (22) as the error function in critic NN learning, as shown in (37). Subsequently, by using the gradient descent approach, the critic NN weight updating law (38) is obtained, which guides the approximate weight toward the ideal weight. Upon obtaining the ideal weight, the right-hand side of (22) becomes 0, indicating the completion of the kth policy evaluation. Following this, the control law for the kth iteration can be obtained through policy improvement. By iterating a certain number of times, we can

ultimately obtain the optimal performance index function and the distributed optimal bipartite consensus control law.

E. Stability Analysis

This section will provide proof that the developed approximate distributed bipartite consensus control law (36) guarantees the local neighborhood bipartite consensus error to be UUB.

Assumption 2: $\mathcal{G}_i(\cdot)$, $\tilde{\varpi}_{ic}$, ϖ_{ic}^*, $\nabla\varrho_{ic}(\cdot)$, and $\nabla\epsilon_{ic}(\cdot)$ are norm-bounded, i.e.,

$$\|\mathcal{G}_i(\cdot)\| \leq \bar{\mathcal{G}}_i, \ \|\tilde{\varpi}_{ic}\| \leq \bar{\varpi}_{ic}, \ \|\varpi_{ic}^*\| \leq \bar{\varpi}_{icM}$$

$$\|\nabla\varrho_{ic}(\cdot)\| \leq \bar{\varrho}_{ic}, \ \|\nabla\epsilon_{ic}(\cdot)\| \leq \bar{\epsilon}_{ic}$$

where $\bar{\mathcal{G}}_i$, $\bar{\varpi}_{ic}$, $\bar{\varpi}_{icM}$, $\bar{\varrho}_{ic}$, and $\bar{\epsilon}_{ic}$ are positive constants.

Theorem 5: Consider the ith spacecraft (14), the approximate distributed optimal bipartite consensus control law provided by (36), the critic NN weight renovating law given by (38), and Assumption 2. Then, the bipartite consensus error is ensured to be UUB.

Proof: Due to space limitation, the proof of Theorem 5 is provided in the Supplementary Material.

Remark 5: 1) Different from the existing ADP-based control approaches [14] and [17], this article considers the bipartite consensus control for multispacecraft systems, which has a more practical application background. Moreover, this article designs a novel optimal performance index function, which contains bipartite consensus error and its derivative. Therefore, the discount factor is not required and the practicability of the control method is improved. 2) For traditional ADP-based robust control approaches [16] and [25], it is a common practice to add an upper bound function of the uncertain term into the performance index function. In other words, during the controller design process, the impact of uncertainties is taken into account in advance to derive the optimal robust controller. However, this approach requires the prior knowledge of the upper bound function of uncertainties, which increases the conservatism of the control method. To tackle this challenge, this article introduces the ISM technique to alleviate the impact of the uncertain term. Moreover, by integrating an adaptive term into the ISM control law, the assumption of the upper bound function of the uncertain term is not required. As a result, the developed ISM-based DRBC control method reduces the conservatism of the controller and improves the practicality of the control approach.

IV. SIMULATION

In this section, four spacecrafts are adopted to demonstrate the validity of the developed ISM-based DRBC control scheme. The communication topology is shown in Fig. 2, where $\mathcal{F}_i(i=1,2,3,4)$ denotes the ith follower and A_1 represents the leader. The communication topology parameters are selected as: $a_{12}=a_{13}=a_{21}=a_{24}=a_{31}=a_{34}=a_{42}=a_{43}=1$, and $b_1=1$. The system parameters of the spacecraft are chosen as: $a_c = 7178$, $\sigma_c = 0.01$, $\chi = 3.986 \times 1014\,\mathrm{m}^3/\mathrm{s}^2$, and $m_i = 100\,\mathrm{kg}$. The dynamics of the leader is given as
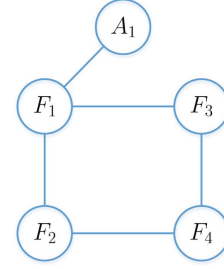
Fig. 2.   Structure of the communication topology.

follows:

$$\mathcal{F}_0 = [\cos(t), \sin(t), \cos(t), -\sin(t), \cos(t), -\sin(t)]^\mathsf{T}.$$

The dynamic uncertainties of all following spacecrafts are presented as follows:

$$\Psi_1(z_1) = \begin{bmatrix} 8\sin(z_{1,1})\cos^2(z_{1,3}) \\ 10\sin(z_{1,2})\cos^2(z_{1,4}) \end{bmatrix},$$

$$\Psi_2(z_2) = \begin{bmatrix} 5\sin(z_{2,1}) \\ 6\cos(z_{2,4}) \\ 3\sin(z_{2,2})\cos^2(z_{2,3}) \end{bmatrix},$$

$$\Psi_3(z_3) = \begin{bmatrix} 6\cos^2(z_{3,3}) \\ 3\sin(z_{3,1}) \\ 10\sin(z_{3,2})\cos^2(z_{3,3}) \end{bmatrix}, \Psi_4(z_4) = \begin{bmatrix} 6\sin^2(z_{4,1}) \\ 9\cos^2(z_{4,2}) \\ 3\sin^2(z_{4,3}) \end{bmatrix}.$$

In the first stage, an ISM controller is developed for each following spacecraft to deal with matched uncertainties. The integral sliding function is selected as (6), where $\mathcal{K}_i(z_i) = [z_{i,4}, z_{i,5}, z_{i,6}]^\mathsf{T}$ and $\mathcal{P}_i(z_i) = [0,0,0,1,0,0;0,0,0,0,1,0;0,0,0,0,0,1]$. Therefore, the ISM control law of the each following spacecraft is designed as follows:

$$\mu_{ic} = \mathcal{A}_i\tanh\left(\mathcal{G}_i^\mathsf{T}(z_i)\mathcal{P}_i^\mathsf{T}(z_i)\mathcal{S}_i/\kappa\right) - \hat{\Psi}_i(z_i) \qquad (40)$$

where $\mathcal{A}_i = 20$, $\kappa = 0.05$, and $\mathrm{sgn}(\cdot)$ is replaced by $\tanh(\cdot)$ for reducing the chattering phenomenon. According to the presented Fig. 3, the evolution curve of the ISM control law of each spacecraft can be perceptibly discerned, thereby providing an insightful glimpse into its response toward dynamic uncertainties. With the help of the ISM control law μ_{ic}, the influence of matched uncertainties can be eliminated and the nominal spacecraft system is obtained. Fig. 4 shows the variation curves of the sliding mode functions for all following spacecrafts. It can be observed that the sliding mode functions are maintained within a small neighborhood of zero, which implies that the system state can be sustained on the sliding surface.

In the following, the distributed optimal bipartite consensus controllers are designed for nominal spacecrafts. The control parameters are chosen as $\mathcal{Q}_{ii} = I_6$, $\mathcal{R}_{ii} = 0.01I_6$, and $\beta_i = 20$. The activation function of the critic NN is chosen as $\varrho_{ic}(\delta_i) = [\delta_{i,1}^2, \delta_{i,2}^2, \delta_{i,3}^2, \delta_{i,4}^2, \delta_{i,5}^2, \delta_{i,6}^2, \delta_{i,1}\delta_{i,4}, \delta_{i,2}\delta_{i,5}, \delta_{i,3}\delta_{i,6}]$.

The simulation results are shown in Figs. 5–10. Fig. 5 provides the weight updating curves. It can be observed that the weight
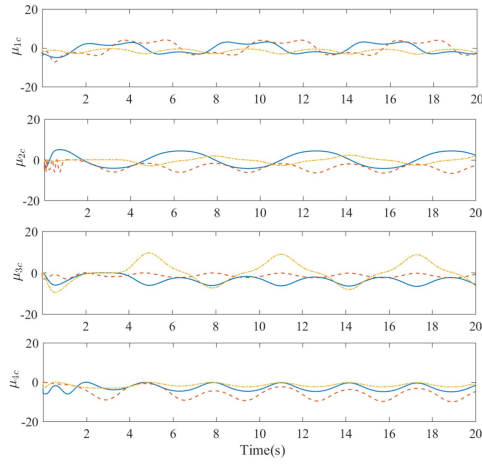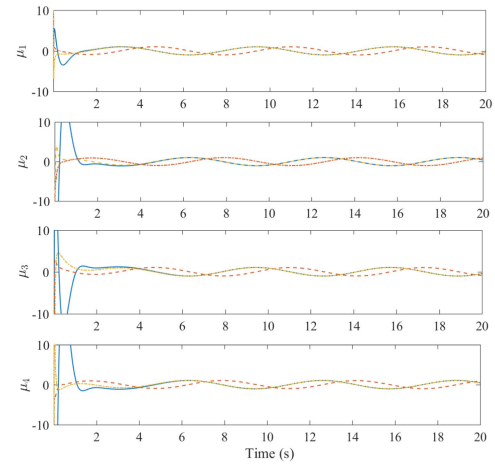
Fig. 3. ISM control laws.
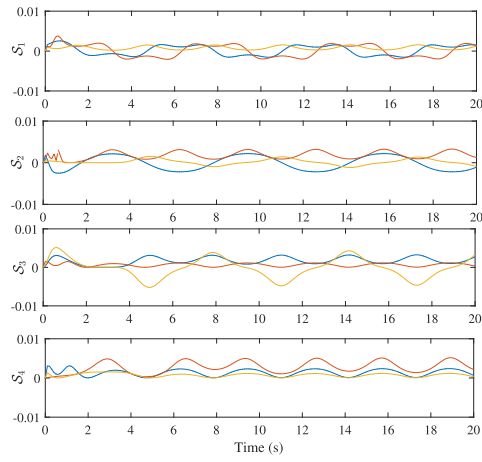
Fig. 4. Integral sliding functions.
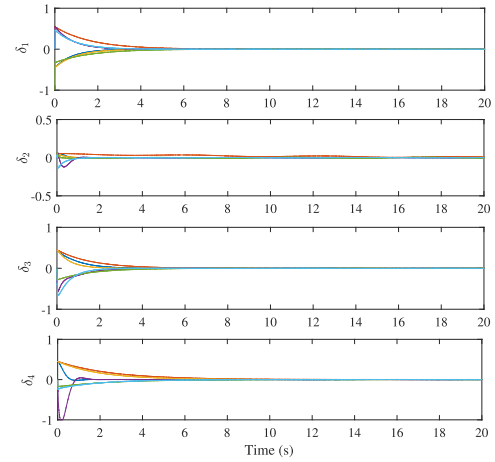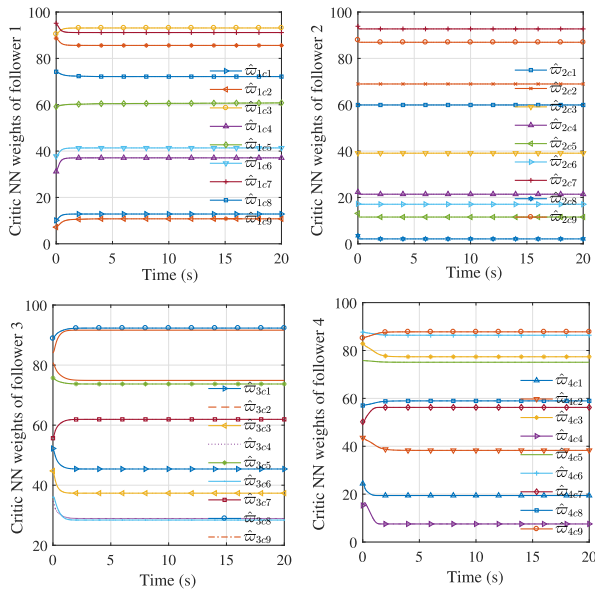
Fig. 5. Critic NN weights of all followers.

Fig. 6. Distributed optimal bipartite consensus control laws.

Fig. 7. Bipartite consensus errors.

vectors of critic NNs will converge to

$$\hat{\varpi}_{1c} = [12.84, 10.76, 93.13, 37.07, 60.79, 41.36,$$
$$91.14, 72.14, 85.59]^\mathsf{T}$$
$$\hat{\varpi}_{2c} = [59.89, 68.95, 39.09, 21.39, 11.53, 17.06,$$
$$92.71, 2.10, 86.96]^\mathsf{T}$$
$$\hat{\varpi}_{3c} = [45.39, 74.90, 37.34, 28.90, 73.70, 28.39,$$
$$61.93, 92.32, 91.62]^\mathsf{T}$$
$$\hat{\varpi}_{4c} = [19.42, 38.28, 77.37, 7.54, 75.09, 86.37,$$
$$56.23, 58.95, 87.79]^\mathsf{T}.$$

Fig. 6 shows the evolution of distributed optimal bipartite consensus control laws. Fig. 7 demonstrates that with the help of the developed control law, the bipartite consensus error of each following spacecraft can converge to a vicinity of 0. The state trajectories of the leader and the followers are illustrated in Fig. 8. It is evident that spacecrafts that belong to the same group converge to a common trajectory, while spacecrafts from different groups
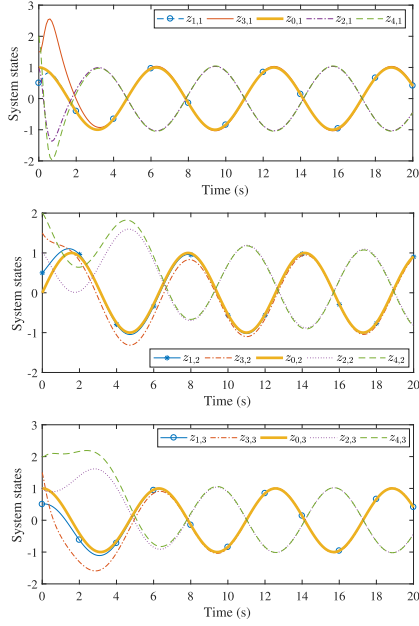
Fig. 8.   System states of all agents.
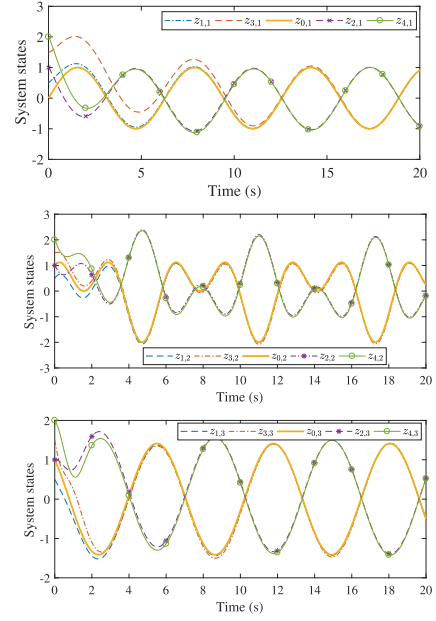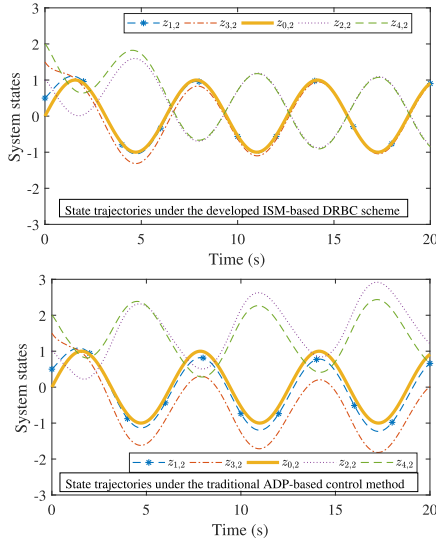
Fig. 9.   State trajectories under different control methods.

converge to opposite trajectories. Fig. 9 compares the system state trajectories under the developed ISM-based DRBC control scheme and the traditional ADP-based distributed control method in [12]. It can be observed that the traditional control method fails to separate the state trajectories of spacecrafts into two clusters. Next, we selected different leader trajectories to further validate the performance of the proposed control approach. The dynamics of the leader is selected as follows:

$$\mathcal{F}_0 = [\ \sin(t), \sin(t) + \cos(2t), \cos(t) - \sin(t),$$
$$\cos(t), \cos(t) - 2\sin(2t), -\sin(t) - \cos(t)]^\mathsf{T}.$$

Fig. 10.   System states of all agents.

Fig. 10 presents the state trajectories of all following spacecrafts. It can be observed that the spacecraft trajectories are divided into two groups, where followers 1 and 3 converge to the same group as the leader, and followers 2 and 4 converge to the group opposite to the leader trajectory. Based on the simulation results above, it can be inferred that the developed ISM-based DRBC control scheme realizes the bipartite consensus of the spacecraft cluster.

V.   CONCLUSION

In this article, an ISM-based DRBC control approach is developed for spacecrafts with matched uncertainties. To begin with, the ISM controller is designed for each following spacecraft to cope with matched uncertainties, and the dynamics of the nominal spacecraft is obtained. Subsequently, by designing a novel performance index function of each following spacecraft, the distributed optimal bipartite consensus control problem is addressed. To acquire the approximate solutions for the coupled HJB equations, a new policy iteration algorithm is introduced, and a critic-only structure is built. Theoretical analysis shows that iterative performance index functions exhibit monotonic decrease and converge to the optimal value, and the iterative control laws ensure the local neighborhood bipartite consensus error asymptotically stable. Finally, simulation results indicate the validity of the proposed ISM-based DRBC control scheme.

REFERENCES

[1] J. Liang, X. Bu, L. Cui, and Z. Hou, "Event-triggered asymmetric bipartite consensus tracking for nonlinear multi-agent systems based on model-free adaptive control," *IEEE/CAA J. Automatica Sinica*, vol. 10, no. 3, pp. 662–672, Mar. 2023.

[2] M. Shahvali, A. Azarbahram, M. Naghibi-Sistani, and J. Askari, "Bipartite consensus control for fractional-order nonlinear multi-agent systems: An output constraint approach," *Neurocomputing*, vol. 397, pp. 212–223, Jul. 2020.

[3] H. Zhao, J. Shan, L. Peng, and H. Yu, "Learning-based robust bipartite consensus control for a class of multiagent systems," *IEEE Trans. Ind. Electron.*, vol. 70, no. 4, pp. 4068–4076, Apr. 2023.

[4] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.

[5] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[6] Q. Wei, D. Liu, Y. Liu, and R. Song, "Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 2, pp. 168–176, Apr. 2017.

[7] B. Zhao, D. Liu, and C. Luo, "Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4330–4340, Oct. 2020.

[8] M. Lin, B. Zhao, and D. Liu, "Policy gradient adaptive critic designs for model-free optimal tracking control with experience replay," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 52, no. 6, pp. 3692–3703, Jun. 2022.

[9] Q. Wei, J. Lu, T. Zhou, X. Cheng, and F. Wang, "Event-triggered near-optimal control of discrete-time constrained nonlinear systems with application to a boiler-turbine system," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3926–3935, Jun. 2022.

[10] D. Wang, X. Li, M. Zhao, and J. Qiao, "Adaptive critic control design with knowledge transfer for wastewater treatment applications," *IEEE Trans. Ind. Informat.*, vol. 20, no. 2, pp. 1488–1497, Feb. 2024, doi: 10.1109/TII.2023.3278875.

[11] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 621–634, Mar. 2014.

[12] Z. Guo, H. Li, H. Ma, and W. Meng, "Distributed optimal attitude synchronization control of multiple QUAVs via adaptive dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 6, pp. 8053–8063, Jun. 2024.

[13] B. Zhao and D. Liu, "Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3054–3064, Apr. 2020.

[14] M. Ha, D. Wang, and D. Liu, "Discounted iterative adaptive critic designs with novel stability analysis for tracking control," *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 7, pp. 1262–1272, Jul. 2022.

[15] C. Mu and Y. Zhang, "Learning-based robust tracking control of quadrotor with time-varying and coupling uncertainties," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 1, pp. 259–273, Jan. 2020.

[16] M. Lin, B. Zhao, and D. Liu, "Event-triggered robust adaptive dynamic programming for multiplayer Stackelberg-Nash games of uncertain nonlinear systems," *IEEE Trans. Cybern.*, vol. 54, no. 1, pp. 273–286, Jan. 2024.

[17] S. Xue, B. Luo, and D. Liu, "Event-triggered adaptive dynamic programming for zero-sum game of partially unknown continuous-time nonlinear systems," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 9, pp. 3189–3199, Sep. 2020.

[18] J. Lu, Q. Wei, Z. Wang, T. Zhou, and F. Wang, "Event-triggered optimal control for discrete-time multi-player non-zero-sum games using parallel control," *Inf. Sci.*, vol. 584, pp. 519–535, Jan. 2022.

[19] K. Zhang, R. Su, H. Zhang, and Y. Tian, "Adaptive resilient event-triggered control design of autonomous vehicles with an iterative single critic learning framework," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 12, pp. 5502–5511, Dec. 2021.

[20] Y. Shi, Q. Hu, D. Li, and M. Lv, "Adaptive optimal tracking control for spacecraft formation flying with event-triggered input," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6418–6428, May 2023.

[21] H. Yang, Q. Hu, H. Dong, and X. Zhao, "ADP-based spacecraft attitude control under actuator misalignment and pointing constraints," *IEEE Trans. Ind. Electron.*, vol. 69, no. 9, pp. 9342–9352, Sep. 2022.

[22] Y. Zhou, E. Kampen, and Q. Chu, "Incremental model based online heuristic dynamic programming for nonlinear adaptive tracking control with partial observability," *Aerosp. Sci. Technol.*, vol. 105, Oct. 2020, Art. no. 106013.

[23] D. Wang, D. Liu, and H. Li, "Policy iteration algorithm for online design of robust control for a class of continuous-time nonlinear systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 627–632, Apr. 2014.

[24] D. Wang, C. Mu, X. Yang, and D. Liu, "Event-based constrained robust control of affine systems incorporating an adaptive critic mechanism," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 47, no. 7, pp. 1602–1612, Jul. 2017.

[25] X. Yang and H. He, "Event-triggered robust stabilization of nonlinear input-constrained systems using single network adaptive critic designs," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 50, no. 9, pp. 3145–3157, Sep. 2020.

[26] H. Zhang, Q. Qu, G. Xiao, and Y. Cui, "Optimal guaranteed cost sliding mode control for constrained-input nonlinear systems with matched and unmatched disturbances," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2112–2126, Jun. 2018.

[27] T. Wang, Y. Wang, X. Yang, and J. Yang, "Further results on optimal tracking control for nonlinear systems with nonzero equilibrium via adaptive dynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 1900–1910, Apr. 2023.

[28] D. Liu, X. Yang, and H. Li, "Adaptive optimal control for a class of continuous-time affine nonlinear systems with unknown internal dynamics," *Neural Comput. Appl.*, vol. 23, pp. 1843–1850, Dec. 2013.

[29] M. Zhuang, L. Tan, K. Li, and S. Song, "Fixed-time formation control for spacecraft with prescribed performance guarantee under input saturation," *Aerosp. Sci. Technol.*, vol. 119, Dec. 2021, Art. no. 107176.

[30] D. Yao, H. Li, R. Lu, and Y. Shi, "Distributed sliding-mode tracking control of second-order nonlinear multiagent systems: An event-triggered approach," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3892–3902, Sep. 2020.

Yongwei Zhang (Member, IEEE) received the B.S. degree in automation from the School of Electronic and Information Engineering, Jiaying University, Meizhou, China, in 2016, and the Ph.D. degree in control science and engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2021.

From 2021 to 2023, he was a Postdoctoral Fellow with the School of Automation, Guangdong University of Technology. He is currently a Lecturer with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou. His current research interests include adaptive dynamic programming and optimal control.

Jun-Yi Li (Member, IEEE) was born in Henan province, China, in 1991. He received the M.S. degree in control science and engineering from Hangzhou Dianzi University, Hangzhou, China, in 2017, and the Ph.D. degree in control science and engineering from the Guangdong University of Technology, Guangzhou, China, in 2021.

From 2019 to 2020, he was an Academic Visitor with the Department of Computer Science, Brunel University London, Uxbridge, U.K. He is currently a Lecturer with the School of Automation, Guangdong University of Technology. His research interests include complex dynamical networks and networked systems.

Event-triggered cooperative robust formation control of multi-agent systems via reinforcement learning

Yongwei Zhang[1] · Jiantao Zhang[1] · Juntao Xiong[1]

Abstract

This article develops an event-triggered cooperative robust formation control scheme for nonlinear multi-agent systems with dynamic uncertainties via reinforcement learning. By formulating a modified value function for each agent, the cooperative robust formation control problem of uncertain multi-agent systems is transformed into a cooperative optimal formation control problem of its nominal plant. To save communication and computing resources, a novel triggering condition is developed for each agent, and the controller is renovated only when an event occurs. Subsequently, the event-triggered optimal formation control law of each agent is derived by solving the coupled Hamilton-Jacobi-Bellman equation via single-critic structure. Furthermore, theoretical analysis indicates that the developed event-triggered cooperative robust formation control approach ensures the asymptotic stability of the formation error for each uncertain agent. Eventually, two simulation cases are adopted to confirm the effectiveness of the developed control approach.

1 Introduction

Cooperative control of multi-agent systems (MASs) has garnered significant research interest owing to its versatile utilization in diverse domains, for example, distributed sensor networks, aerospace systems, unmanned swarm systems, and so on. Compared with single-agent systems, MASs complete complex tasks in an efficient and robust way through the cooperation between each agent [1]. As is known to all, the formation control is one of the common and basic issues in MASs, which endeavors to prompt multiple agents to form a predetermined geometric pattern with their states or outputs. Currently, a significant number of scholars are

dedicated to addressing the challenge of formation control. Cao et al. [2] proposed a neural network (NN)-based composite dynamic surface control approach to address the fixed-time formation control problem of MASs. Pang et al. [3] addressed the time-varying formation control problem of MASs with communication delays and packet dropouts by developing a cloud-based predictive control method. Yao et al. [4] developed a sliding mode control method to handle the leader-follower formation control problem of MASs with uncertain perturbations under event-triggered mechanism. The previously mentioned researches primarily concentrate on the stability of MASs. Nevertheless, the attainment of green and high efficiency in MASs necessitates the careful consideration of control cost. To achieve this goal, it is essential to present an optimal formation control method to accomplish the cooperative task of MASs while minimizing the control cost of each agent.

It is widely acknowledged that reinforcement learning (RL) is a highly effective method for addressing the optimal control problem of nonlinear systems [5–11]. To achieve optimal solutions for consensus, formation, and containment control problems in MASs, numerous researchers have put forward RL-based cooperative control methods. At consensus control aspect, Guo et al. [12] tackled the distributed

---

Jiantao Zhang and Juntao Xiong contributed equally to this work.

✉ Juntao Xiong
  xiongjt2340@163.com

  Yongwei Zhang
  YongweiZhang@scau.edu.cn

  Jiantao Zhang
  zhangjiantao@yeah.net

[1] College of Mathematics and Informatics, South China
   Agricultural University, Guangzhou 510642, China

optimal attitude consensus control issue of unmanned aerial vehicle cluster by using RL technique. Xia et al. [13] developed an off-policy RL method to tackle the optimal synchronization issue of MASs with asymmetric input constraints. For formation control, Mu et al. [14] developed a multistep generalized policy iteration algorithm to settle the hierarchical leader-follower formation control problem of large-scale MASs. Wen et al. [15] investigated the leader-follower formation control of unknown MASs by employing a modified RL technique. For containment control, Xiao et al. [16] confronted the distributed optimal containment control problem of mobile robots by combining optimal backstepping and RL techniques. Yang et al. [17] adopted off-policy RL technique to handle the model-free optimal containment control problem of heterogeneous MASs with unknown dynamics. In general, aforementioned results develop appropriate iterative RL algorithms to acquire the approximate solutions of coupled Hamilton–Jacobi–Bellman (HJB) equations, and corresponding optimal cooperative controllers are developed to accomplish the cooperative control tasks of MASs. However, the above mentioned control methods require to update controllers at each sampling time, which consume a lot of communication and computing resources. Moreover, in MASs, each agent transmits information through a communication network, but its computing and communication capabilities are limited since agent is usually equipped with a microprocessor, such as unmanned aerial vehicles and unmanned surface vehicles, etc. To solve this problem, scholars turn to develop event-triggered cooperative control approaches for MASs, that is, agents only communicate or update controllers at necessary times to alleviate computing and communication burdens.

In recent years, several researchers have proposed event-triggered cooperative control schemes for MASs. For example, Zhao et al. [18] explored the optimal coordination control of MASs under the event-triggered framework. Chen et al. [19] presented an adaptive distributed observer-based RL algorithm to tackle the event-triggered H_∞ consensus problem. Wang et al. [20] combined integral sliding-mode and local RL techniques to address the robust optimal consensus control problem of MASs. Ren et al. [21] addressed the security distributed consensus estimation problem of nonlinear systems with deception attacks by developing an event-triggered extended Kalman filter. On the whole, all of existing researches handle the consensus control problem only, the formation control problem is not investigated. Furthermore, in practice, agents need to perform various complex tasks in harsh environments, such as unmanned aerial vehicle rescue and unmanned surface vehicle deep sea exploration. It means that the emergence of model uncertainty is inevitable. Actually, plenty of RL-based robust

control methodologies have been formulated for nonlinear systems to handle model uncertainties. Nevertheless, most of them consider single agent systems [22, 23] or large-scale systems [24, 25], the research for MASs is still in its infancy. Based on the aforementioned discussion and analysis, the paramount importance lies in effectively addressing the robust formation control problem of MASs within the event-triggered framework.

This paper introduces an innovative RL-based event-triggered cooperative robust formation (ETCRF) control scheme for uncertain MASs. The innovations and contributions of this paper are outlined in the following manner.

1. Unlike the exsiting optimal cooperative control results [12] and [13] that adopted time-triggered mechanism, this paper develops a novel triggering condition for each agent and the developed RL-based ETCRF controller is renewed at triggering instant only such that the computing and the communication resources are conserved.

2. Different from existing cooperative formation control approaches [14] and [15], which are applicable to ideal system model only, this paper considered MASs with dynamic uncertainties. Through the development of a novel value function for each agent, the ETCRF control problem is converted into an event-triggered cooperative optimal formation (ETCOF) control problem and the developed controllers guarantee all followers catch up with the leader in a specified geometric pattern even in the presence of dynamic uncertainties. Therefore, the developed RL-based ETCRF control method is more practical.

The subsequent sections of this paper are organized as follows. Section 2 introduces the graph theory and presents the problem statement. Section 3 provides a detailed explanation of the ETCRF controller design, the NN implementation, and the stability analysis. The simulation results of the RL-based ETCRF control scheme are displayed in Section 4. Finally, Section 5 provides the corresponding conclusion.

2 Preliminaries

2.1 Graph theory

Consider a MAS characterized by the presence of a solitary leader and $\mathcal{M}$ followers. Let $\mathcal{M}_i$ represents the neighbor set of the agent i and $\bar{\mathcal{M}}_i$ denotes a set that contains the agent i and its neighbors. The communication topology graph of the MAS is given by $\mathcal{T}_g = \{\mathbb{P}, \mathbb{E}, \mathbb{A}\}$, where $\mathbb{P} = \{\mathbb{P}_1, ..., \mathbb{P}_\mathcal{M}\}$ is a node set, $\mathbb{E} = \{(\mathbb{P}_i, \mathbb{P}_j) : \mathbb{P}_i, \mathbb{P}_j \in \mathbb{P}\}$ is a edge set, and $\mathbb{A} = [\alpha_{ij}]$ is a weighted adjacency matrix. Note that

$(\mathbb{P}_i, \mathbb{P}_j) \in \mathbb{E}$ if and only if the agent i and the agent j are directly connected. Furthermore, if $(\mathbb{P}_i, \mathbb{P}_j) \in \mathbb{E}$, then $\alpha_{ij} > 0$, otherwise, $\alpha_{ij} = 0$, and $\alpha_{ii} = 0$ for all $i = 1, ..., \mathcal{M}$. The degree matrix of $\mathcal{T}_g$ is defined as $\mathbb{D} = \text{diag}\{d_1, ..., d_{\mathcal{M}}\}$ with $d_i = \sum_{j \in \mathcal{M}_i} \alpha_{ij}$ and the Laplacian matrix is calculated as $\mathbb{L} = \mathbb{D} - \mathbb{A}$.

2.2 Problem statement

The system dynamics of the ith follower is expressed as

$$\dot{z}_i = \mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\big(\mu_i + \Lambda_i(z_i)\big), \tag{1}$$

where $z_i \in \mathbb{R}^{s_i}$ and $\mu_i \in \mathbb{R}^{k_i}$ are the system state and the control input of the ith follower, $\Lambda_i(z_i) \in \mathbb{R}^{k_i}$ is the matched uncertainty, and $\mathcal{A}_{f,i}(z_i) \in \mathbb{R}^{s_i}$ and $\mathcal{A}_{g,i}(z_i) \in \mathbb{R}^{s_i \times k_i}$ are nonlinear system functions.

Assumption 1 *The system functions $\mathcal{A}_{f,i}(z_i)$ and $\mathcal{A}_{g,i}(z_i)$ are Lipschitz continuous on a compact set Ω and the system (1) is stabilizable on Ω [5, 6].*

Assumption 2 *The dynamic uncertainty $\Lambda_i(z_i)$ satisfies $\|\Lambda_i(z_i)\| \leq \bar{\Lambda}_i(z_i)$, where $\bar{\Lambda}_i(z_i)$ is a known function and $\bar{\Lambda}_i(0) = 0$ [22, 23].*

The dynamics of the leader is provided as

$$\dot{z}_0 = \mathcal{A}_{f,0}(z_0), \tag{2}$$

where $z_0 \in \mathbb{R}^{s_0}$ and $\mathcal{A}_{f,0}(\cdot) \in \mathbb{R}^{s_0}$ is a differential function.

The objective of this paper is to introduce a RL-based ETCRF control method that ensures all uncertain followers accurately track the leader's trajectory within a specified formation. We will show that this objective can be achieved by addressing the ETCOF control problem of nominal MASs.

The nominal form of (1) is given as

$$\dot{z}_i = \mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i. \tag{3}$$

The formation error of the ith follower is provided as

$$\mathscr{E}_i = \sum_{j \in \mathcal{M}_i} \alpha_{ij}(z_i - \eta_i - z_j + \eta_j) + c_i(z_i - z_0 - \eta_i), \tag{4}$$

where $\eta_i \in \mathbb{R}^{s_i}$ is the formation pattern between the follower i and the leader and $c_i > 0$ is the connection coefficient between the follower i and the leader. Then, the dynamics of the formation error is calculated as

$$\dot{\mathscr{E}}_i = \sum_{j \in \mathcal{M}_i} \alpha_{ij}(\dot{z}_i - \dot{z}_j) + c_i(\dot{z}_i - \dot{z}_0)$$

$$= \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i - \mathcal{A}_{f,j}(z_j) - \mathcal{A}_{g,j}(z_j)\mu_j\big)$$

$$+ c_i\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i - \mathcal{A}_{f,0}(z_0)\big)$$

$$= (d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i\big) - c_i\mathcal{A}_{f,0}(z_0)$$

$$- \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\mu_j\big). \tag{5}$$

The novel value function of the ith follower is formulated as

$$\mathcal{J}_i(\mathscr{E}_i) = \int_t^\infty e^{-\gamma(\varsigma - t)} \mathcal{U}_i\Big(\mathscr{E}_i(\varsigma), \mu_i(\varsigma), \mu_{-i}(\varsigma)\Big) d\varsigma, \tag{6}$$

where γ is a positive constant, $\mu_{-i} = \{\mu_j : j \in \mathcal{M}_i, j \neq i\}$, and $\mathcal{U}_i(\cdot)$ is the utility function that is formulated as

$$\mathcal{U}_i(\mathscr{E}_i, \mu_i, \mu_{-i}) = \sum_{j \in \bar{\mathcal{M}}_i} \Big(\theta \bar{\Lambda}_j^2(z_j) + \mathscr{E}_j^\mathsf{T} \mathcal{Q}_j \mathscr{E}_j + \mu_j^\mathsf{T} \mathcal{R}_j \mu_j\Big),$$

where θ is a positive constant, $\mathcal{Q}_j \in \mathbb{R}^{s_j \times s_j}$ and $\mathcal{R}_j \in \mathbb{R}^{k_j \times k_j}$ are positive definite matrices. The Hamiltonian of the ith follower is given as

$$\mathcal{H}_i\big(\mathscr{E}_i, \nabla \mathcal{J}_i(\mathscr{E}_i), \mu_i, \mu_{-i}\big)$$

$$= \nabla \mathcal{J}_i^\mathsf{T}(\mathscr{E}_i)\bigg((d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i\big)$$

$$- c_i\mathcal{A}_{f,0}(z_0) - \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\mu_j\big)\bigg)$$

$$+ \mathcal{U}_i(\mathscr{E}_i, \mu_i, \mu_{-i}) - \gamma \mathcal{J}_i(\mathscr{E}_i). \tag{7}$$

The discounted optimal value function of the ith follower

$$\mathcal{J}_i^*(\mathscr{E}_i) = \min_{\mu_i \in \Re(\Omega)} \int_t^\infty e^{-\gamma(\varsigma - t)} \mathcal{U}_i\big(\mathscr{E}_i(\varsigma), \mu_i(\varsigma), \mu_{-i}(\varsigma)\big) d\varsigma$$

fulfills the following HJB equation as

$$\min_{\mu_i \in \Re(\Omega)} \mathcal{H}_i\big(\mathscr{E}_i, \nabla \mathcal{J}_i^*(\mathscr{E}_i), \mu_i, \mu_{-i}\big) = 0, \tag{8}$$

where $\Re(\Omega)$ represents the admissible control law. Subsequently, the optimal formation control law is obtained by

$$\mu_i^* = -\frac{d_i + c_i}{2} \mathcal{R}_i^{-1} \mathcal{A}_{g,i}^\mathsf{T}(z_i) \nabla \mathcal{J}_i^*(\mathscr{E}_i). \tag{9}$$

By utilizing (8) and (9), we can derive the coupled HJB equation as

$$0 = \nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\bigg((d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i^*\big)$$

$$-c_i \mathcal{A}_{f,0}(z_0) - \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\mu_j^*\big)\bigg)$$

$$+ \mathcal{U}_i\big(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*\big) - \gamma \mathcal{J}_i^*(\mathscr{E}_i). \tag{10}$$

In the following, a RL-based ETCRF control scheme is put forward to approximate the solution of the coupled HJB equation.

Remark 1 *It is worth noting that the HJB equation is a complex partial differential equation, and it is difficult to obtain its analytical solution directly. In order to solve this problem, the policy iteration algorithm is adopted to obtain the approximate solution. The algorithm pseudo-code is given as follows.*

Algorithm 1 Policy iteration algorithm.

Step 1: Let $k = 1$ and select initial admissible control law $\mu_i^0 \in \Re(\Omega)$. Choose a computation precision ξ.
Step 2: (Policy evaluation) Calculate the iterative value function $\mathcal{J}_i^{(k)}(\mathscr{E}_i)$ by

$$0 = \mathcal{U}_i\big(\mathscr{E}_i, \mu_i^{(k-1)}, \mu_{-i}^{(k-1)}\big) - \gamma \mathcal{J}_i^{(k)}(\mathscr{E}_i)$$

$$+ \nabla \mathcal{J}_i^{(k)\mathsf{T}}(\mathscr{E}_i)\Big((d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i^{(k-1)}\big)$$

$$- c_i \mathcal{A}_{f,0}(z_0) - \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\mu_j^{(k-1)}\big)\Big).$$

Step 3: (Policy improvement) Update the control law by

$$\mu_i^{(k)} = -\frac{d_i + c_i}{2}\mathcal{R}_i^{-1}\mathcal{A}_{g,i}^{\mathsf{T}}(z_i)\nabla \mathcal{J}_i^{(k)}(\mathscr{E}_i).$$

Step 4: If $|\mathcal{J}_i^{(k)}(\mathscr{E}_i) - \mathcal{J}_i^{(k-1)}(\mathscr{E}_i)| < \xi$, goto Step 5; else, let $k = k+1$, go back to Step 2 and continue.
Step 5: Stop Algorithm.

It is worth mentioning that several researchers have established theoretical analysis frameworks to show that, through continuous policy evaluation and policy improvement, the iterated value function can converge to the optimal value function and the optimal control policy is obtained after iteration completion [5, 6, 8]. Therefore, we can ultimately obtain the approximate solution of the HJB equation.

Remark 2 *This paper adopts the policy iteration, that is value-based RL algorithm, to derive the optimal value function and the optimal formation control law for MASs. In the field of control, prevalent RL algorithms include policy iteration, value iteration, and policy gradient. The former two are value-based RL algorithms, while the latter is policy-based. Typically, value-based RL algorithms are extensively applied and have proven effective in addressing classical control problems such as the optimal regulation and the trajectory*

tracking. Scholars have rigorously analyzed the convergence and the optimality of policy iteration or value iteration algorithms, ensuring the attainment of optimal control laws for closed-loop systems [5, 6]. Recently, several researchers have begun investigating the policy gradient-based RL algorithm to address the control problems of nonlinear systems. It is noteworthy that each of these algorithms has its own strengths and limitations. For example, value iteration algorithm easily determines its initial condition, but the stability of the closed-loop system is not guaranteed at each iteration. Policy iteration algorithm necessitates the admissible control law as initial condition, yet the stability of the closed-loop system is assured at each iteration. Since the stability is paramount in closed-loop systems, the policy iteration algorithm becomes the prevalent approach in control applications. In addition, the policy gradient algorithm does not require the system function information, making it suitable for closed-loop systems with unknown dynamics. In conclusion, the selection of RL algorithms should be tailored to specific problems. This paper employs the classic policy iteration algorithm to tackle the cooperative robust formation control problem for MASs and introduces an event-triggered mechanism to mitigate computational and communication burdens. In the future work, we will explore event-triggered control methods based on value iteration or policy gradient algorithms to address cooperative control problems of MASs.

Remark 3 *Compared with traditional control methods, the advantage of the RL-based control approach lies in its ability to guarantee the stability of the closed-loop system while minimizing the performance index function, thereby reducing the control cost and enhancing the control performance. Moreover, by utilizing a model-free RL algorithm, it is possible to design a controller without the need of system functions. As a result, model-free control can be achieved for closed-loop systems with unknown system dynamics. Nevertheless, the design of the RL-based control method faces the following challenges. 1) Most RL algorithms, such as value iteration and policy iteration, require a certain number of iterations to obtain the optimal control policy. As the system stability is crucial in the field of control, it is essential to establish a rigorous theoretical analysis framework to guarantee the stability of the closed-loop system during the iteration process and the optimality of the control policy upon completion of the iteration. 2) When implementing RL algorithms, NNs need to be introduced. Therefore, one of the challenges lies in designing the appropriate NN weight updating law to ensure the optimal weights are obtained, thus deriving the optimal control policy. Additionally, the NN structure, activation functions, and control parameters all impact control performance. Therefore, selecting these parameters presents another challenge.*

3 RL-based event-triggered robust formation control scheme

3.1 Event-triggered robust formation controller design

In event-triggered mechanism, the sampled state of ith follower is expressed as

$$\bar{z}_{i,\kappa} = z_i(\Gamma_\kappa), \forall t \in [\Gamma_\kappa, \Gamma_{\kappa+1}), \tag{11}$$

where Γ_κ denotes the κth sampling instant. Hence, the corresponding sampled formation error is provided as

$$\begin{aligned}\bar{\mathscr{E}}_{i,\kappa} = &\sum_{j \in \mathcal{M}_i} \alpha_{ij}(\bar{z}_{i,\kappa} - \eta_i - \bar{z}_{j,\kappa} + \eta_j) \\ &+ c_i(\bar{z}_{i,\kappa} - \bar{z}_{0,\kappa} - \eta_i), \forall t \in [\Gamma_\kappa, \Gamma_{\kappa+1}).\end{aligned} \tag{12}$$

Moreover, the event-triggered error is formulated as

$$e_{i,\kappa}(t) = \mathscr{E}_i(t) - \bar{\mathscr{E}}_{i,\kappa}, \forall t \in [\Gamma_\kappa, \Gamma_{\kappa+1}). \tag{13}$$

Based on (9), (11) and (12), the ETCOF control law of the ith follower is given as

$$\mu_i^*(\bar{\mathscr{E}}_{i,\kappa}) = -\frac{d_i + c_i}{2} \mathcal{R}_i^{-1} \mathcal{A}_{g,i}^{\mathsf{T}}(\bar{z}_{i,\kappa}) \nabla \mathcal{J}_i^*(\bar{\mathscr{E}}_{i,\kappa}). \tag{14}$$

Assumption 3 *The system function $\mathcal{A}_{g,i}(z_i)$ is norm-bounded as*

$$\|\mathcal{A}_{g,i}(z_i)\| \le \bar{g}_i, \tag{15}$$

where $\bar{g}_i$ is a positive constant [5, 6].

In the following, we will indicate that the designed ETCOF control law (14) guarantees the formation error of each follower to be asymptotically stable. This indicates that it is reasonable and effective to convert the ETCRF control problem into an ETCOF control problem.

Theorem 1 *Consider the ith follower with dynamic uncertainty (1), Assumptions 2 and 3, and the ETCOF control law provided by (14). If the triggering condition fulfills*

$$\begin{aligned}T_{i,c}(\mathscr{E}_i, \bar{\mathscr{E}}_{i,\kappa}) = &(1+\nu)\nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\Big(-c_i \mathcal{A}_{f,0}(z_0) \\ &+ (d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i^*(\bar{\mathscr{E}}_{i,\kappa})\big) \\ &- \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\mu_j^*(\bar{\mathscr{E}}_{j,\kappa})\big)\Big) \\ &+ \mathcal{U}_i(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*) + \theta \sum_{j \in \bar{\mathcal{M}}_i} \bar{\Lambda}_j^2(z_j) \\ &+ (1+\nu)\nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\nabla \mathcal{J}_i^*(\mathscr{E}_i) \\ &< 0,\end{aligned} \tag{16}$$

where $\nu > 0$ is a design parameter. Then, the formation error pertaining to the ith follower exhibits asymptotic stability.

Proof Decide on the Lyapunov function candidate as

$$\mathcal{L}_{T1} = \mathcal{J}_i^*(\mathscr{E}_i). \tag{17}$$

By taking the derivative of $\mathcal{L}_{T1}$ with respect to the solution of (1) and utilizing (16), we can derive

$$\begin{aligned}\dot{\mathcal{L}}_{T1} = &\nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\Big(-c_i \mathcal{A}_{f,0}(z_0) + (d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) \\ &+ \mathcal{A}_{g,i}(z_i)\big(\mu_i^*(\bar{\mathscr{E}}_{i,\kappa}) + \Lambda_i(z_i)\big)\big) - \sum_{j \in \mathcal{M}_i} \alpha_{ij} \\ &\Big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\big(\mu_j^*(\bar{\mathscr{E}}_{j,\kappa}) + \Lambda_j(z_j)\big)\Big)\Big) \\ = &\frac{1}{1+\nu}\Big(T_{i,c}(\mathscr{E}_i, \bar{\mathscr{E}}_{i,\kappa}) - \mathcal{U}_i(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*) \\ &- \theta \sum_{j \in \bar{\mathcal{M}}_i} \bar{\Lambda}_j^2(z_j) - (1+\nu)\nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\nabla \mathcal{J}_i^*(\mathscr{E}_i)\Big) \\ &+ (d_i + c_i)\nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\mathcal{A}_{g,i}(z_i)\Lambda_i(z_i) \\ &- (d_i + c_i)\nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\sum_{j \in \mathcal{M}_i} \alpha_{ij}\mathcal{A}_{g,j}(z_j)\Lambda_j(z_j) \\ \le &\frac{1}{1+\nu}\Big(T_{i,c}(\mathscr{E}_i, \bar{\mathscr{E}}_{i,\kappa}) - \mathcal{U}_i(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*) \\ &- \theta \sum_{j \in \bar{\mathcal{M}}_i} \bar{\Lambda}_j^2(z_j) - (1+\nu)\nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\nabla \mathcal{J}_i^*(\mathscr{E}_i)\Big) \\ &+ \nabla \mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\nabla \mathcal{J}_i^*(\mathscr{E}_i) + \frac{(d_i + c_i)}{2}\bar{g}_i^2 \bar{\Lambda}_i^2(z_i) \\ &+ \frac{\mathcal{M}_i(d_i + c_i)}{2}\sum_{j \in \mathcal{M}_i} \alpha_{ij}\bar{g}_j^2 \bar{\Lambda}_j^2(z_j) \\ \le &\frac{1}{1+\nu}\Big(T_{i,c}(\mathscr{E}_i, \bar{\mathscr{E}}_{i,\kappa}) - \mathcal{U}_i(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*) \\ &- \theta \sum_{j \in \bar{\mathcal{M}}_i} \bar{\Lambda}_j^2(z_j)\Big) + \frac{(d_i + c_i)}{2}\bar{g}_i^2 \bar{\Lambda}_i^2(z_i) \\ &+ \frac{\mathcal{M}_i(d_i + c_i)}{2}\sum_{j \in \mathcal{M}_i} \alpha_{ij}\bar{g}_j^2 \bar{\Lambda}_j^2(z_j).\end{aligned} \tag{18}$$

Let

$$\Upsilon_i = \max\left\{\frac{(d_i + c_i)\bar{g}_i^2}{2}, \frac{\mathcal{M}_i(d_i + c_i)a_{i1}\bar{g}_1^2}{2}, \\ ..., \frac{\mathcal{M}_i(d_i + c_i)a_{i\mathcal{M}_i}\bar{g}_{\mathcal{M}_i}^2}{2}\right\}.$$

Then, if θ is selected to satisfy $\frac{\theta}{1+\nu} > \Upsilon_i$, we can obtain

$$
\begin{aligned}
\dot{\mathcal{L}}_{T1} &\leq \frac{1}{1+\nu}\bigg(-\mathcal{U}_i(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*) - \theta \sum_{j \in \bar{\mathcal{M}}_i} \bar{\Lambda}_j^2(z_j) \bigg) \\
&\quad + \Upsilon_i \sum_{j \in \bar{\mathcal{M}}_i} \bar{\Lambda}_j^2(z_j) \\
&\leq -\frac{1}{1+\nu}\mathcal{U}_i(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*) < 0.
\end{aligned}
\tag{19}
$$

Therefore, with the event-triggering condition (16), the ETCOF control law (14) guarantees the asymptotic stability of the formation error for each follower. The proof has been concluded. □

Remark 4 *The design of the event-triggered condition (16) is primarily based on the Lyapunov stability principle. It is well-known that the stability of the closed-loop system can only be guaranteed when the $\dot{\mathcal{L}}_{T1}$ is less than 0. Through the proof process of Theorem 1, it can be observed that if the event-triggered condition satisfies (16), then the $\dot{\mathcal{L}}_{T1}$ is negative, which guarantees that the formation error is asymptotically stable. This constitutes the main idea behind the design of the event-triggered condition. In addition, there are certain parameter values in (16) that need to be determined by the user. These values will impact the magnitude of the triggering threshold. Generally, a larger triggering threshold leads to fewer controller updates, thus conserving computational and communication resources. However, this may risk system instability. Conversely, a smaller triggering threshold results in more frequent controller updates, which increase the computational and communication burden and make the stability of the closed-loop system easier to maintain. At present, there is no unified method to select these parameter values. Typically, iterative experiments are required for their determination.*

3.2 Neural network implementation

By leveraging the assistance of the critic NN, we can approximate the optimal value function as

$$
\mathcal{J}_i^*(\mathscr{E}_i) = \chi_{ic}^{*\mathsf{T}} \zeta_{ic}(\mathscr{E}_i) + \epsilon_{ic}(\mathscr{E}_i),
\tag{20}
$$

where $\chi_{ic}^* \in \mathbb{R}^{h_c}$ is the ideal weight vector, $\zeta_{ic}(\mathscr{E}_i) \in \mathbb{R}^{h_c}$ is the activation function, h_c is the number of hidden layer neurons, and $\epsilon_{ic}(\mathscr{E}_i) \in \mathbb{R}$ is the approximation error. Obviously, the partial derivative of $\mathcal{J}_i^*(\mathscr{E}_i)$ is given as

$$
\nabla \mathcal{J}_i^*(\mathscr{E}_i) = \nabla \zeta_{ic}^{\mathsf{T}}(\mathscr{E}_i) \chi_{ic}^* + \nabla \epsilon_{ic}(\mathscr{E}_i).
\tag{21}
$$

The approximate value function is formulated as

$$
\hat{\mathcal{J}}_i(\mathscr{E}_i) = \hat{\chi}_{ic}^{\mathsf{T}} \zeta_{ic}(\mathscr{E}_i),
\tag{22}
$$

where $\hat{\chi}_{ic}$ denotes the estimate of the optimal weight. Then, we can further get

$$
\nabla \hat{\mathcal{J}}_i(\mathscr{E}_i) = \nabla \zeta_{ic}^{\mathsf{T}}(\mathscr{E}_i) \hat{\chi}_{ic}.
\tag{23}
$$

According to (14) and (20), the ETCOF control law of the ith follower is rewritten as

$$
\mu_i^*(\bar{\mathscr{E}}_{i,\kappa}) = -\frac{d_i + c_i}{2} \mathcal{R}_i^{-1} \mathcal{A}_{g,i}^{\mathsf{T}}(\bar{z}_{i,\kappa})\big(\nabla \zeta_{ic}^{\mathsf{T}}(\bar{\mathscr{E}}_{i,\kappa})\chi_{ic}^* + \nabla \epsilon_{ic}(\bar{\mathscr{E}}_{i,\kappa})\big).
\tag{24}
$$

Combining (14) and (22), the approximate ETCOF control law is provided as

$$
\hat{\mu}_i(\bar{\mathscr{E}}_{i,\kappa}) = -\frac{d_i + c_i}{2} \mathcal{R}_i^{-1} \mathcal{A}_{g,i}^{\mathsf{T}}(\bar{z}_{i,\kappa}) \nabla \zeta_{ic}^{\mathsf{T}}(\bar{\mathscr{E}}_{i,\kappa}) \hat{\chi}_{ic}.
\tag{25}
$$

As per the equations (10) and (25), the approximate Hamiltonian is given as

$$
\begin{aligned}
\hat{\mathcal{H}}_i(\mathscr{E}_i, \hat{\chi}_{ic}) &= \hat{\chi}_{ic}^{\mathsf{T}} \nabla \zeta_{ic}(\mathscr{E}_i) \bigg(-c_i \mathcal{A}_{f,0}(z_0) \\
&\quad + (d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\mathscr{E}}_{i,\kappa})\big) \\
&\quad - \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\mathscr{E}}_{j,\kappa})\big) \bigg) \\
&\quad + \mathcal{U}_i(\mathscr{E}_i, \hat{\mu}_i, \hat{\mu}_{-i}) - \gamma \hat{\mathcal{J}}_i(\mathscr{E}_i) \triangleq \mathbf{e}_{ic}.
\end{aligned}
\tag{26}
$$

By minimizing the objective function $E_{ic} = \frac{1}{2}\mathbf{e}_{ic}^{\mathsf{T}}\mathbf{e}_{ic}$, the weight tuning law of the critic NN is designed as

$$
\begin{aligned}
\dot{\hat{\chi}}_{ic} &= -\varrho_c \frac{1}{(1 + \Psi_i^{\mathsf{T}}\Psi_i)^2}\left(\frac{\partial E_{ic}}{\partial \hat{\chi}_{ic}}\right) \\
&= -\frac{\varrho_c \Psi_i}{(1 + \Psi_i^{\mathsf{T}}\Psi_i)^2}\big(\hat{\chi}_{ic}^{\mathsf{T}}\Psi_i + \mathcal{U}_i(\mathscr{E}_i, \hat{\mu}_i, \hat{\mu}_{-i}) - \gamma \hat{\mathcal{J}}_i(\mathscr{E}_i)\big),
\end{aligned}
\tag{27}
$$

where $\varrho_c > 0$ is the learning rate and

$$
\begin{aligned}
\Psi_i &= \nabla \zeta_{ic}(\mathscr{E}_i)\bigg((d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\mathscr{E}}_{i,\kappa})\big) \\
&\quad - \sum_{j \in \mathcal{M}_i} \alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\mathscr{E}}_{j,\kappa})\big) - c_i \mathcal{A}_{f,0}(z_0)\bigg).
\end{aligned}
$$

Theorem 2 *Consider the ith nominal follower (3), the critic NN weight updating rule is provided as (27), then the critic NN weight estimation error $\tilde{\chi}_{ic} = \chi_{ic}^* - \hat{\chi}_{ic}$ is guaranteed to be uniform ultimate boundedness (UUB).*

Proof The similar proof of Theorem 2 is widely existed in [26, 27], we omit the detail proof here. □

Remark 5 *It is worth noting that each term in the triggering condition* (16) *is known and easy to compute. The reasons are explained as follows. 1) With the help of the NN, the first part $\mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)$ is expressed as* (21). *Therefore, by calculating the product of the activation function and the NN weight, the value of $\mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)$ can be obtained. 2) The second part, represented by $-c_i\mathcal{A}_{f,0}(z_0) + (d_i + c_i)\big(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\mu_i^*(\bar{\bar{\mathscr{E}}}_{i,\kappa})\big) - \sum_{j\in\mathcal{M}_i}\alpha_{ij}\big(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\mu_j^*(\bar{\bar{\mathscr{E}}}_{j,\kappa})\big)$, signifies the dynamics of the formation error. Since the control law can be calculated from* (24), *and the system functions $\mathcal{A}_{f,0}$, $\mathcal{A}_{f,i}$, and $\mathcal{A}_{g,i}$ as well as the constants c_i, d_i, and α_{ij} are all known, we can directly calculate the value of the second part. 3) The third part $\bar{\Lambda}_j(z_j)$ is the upper bound function of the uncertainty and can be chosen as $\|z_j\|$, so obtaining the value of this part is straightforward. 4) The four part $\mathcal{U}_i(\mathscr{E}_i, \mu_i^*, \mu_{-i}^*)$ is the utility function. Since $\mathscr{E}_i$, μ_i and $\bar{\Lambda}_j(z_j)$ are known, the value of this part can be readily determined. According to the above discussion, the triggering condition is easy to implement.*

3.3 Stability analysis

This section will provide proof that during the training phase of the RL-based ETCRF controller, the formation error of each follower is guaranteed to be UUB.

Assumption 4 *The system function $\mathcal{A}_{f,i}(z_i)$ is Lipschitz continuous and satisfies*

$$\|\mathcal{A}_{f,i}(z_i)\| \le L_{i,f}\|z_i\|, \tag{28}$$

where $L_{i,f}$ is a positive constant [5, 6].

Assumption 5 *$\tilde{\chi}_{ic}$, χ_{ic}^*, $\nabla\zeta_{ic}(z_i)$, and $\nabla\epsilon_{ic}(z_i)$ are norm-bounded, i.e.,*

$$\|\tilde{\chi}_{ic}\| \le \bar{\chi}_{ic}, \quad \|\chi_{ic}^*\| \le \bar{\chi}_{icM},$$
$$\|\nabla\zeta_{ic}(z_i)\| \le \bar{\zeta}_{ic}, \quad \|\nabla\epsilon_{ic}(z_i)\| \le \bar{\epsilon}_{ic}$$

where $\bar{\chi}_{ic}$, $\bar{\chi}_{icM}$, $\bar{\zeta}_{ic}$ and $\bar{\epsilon}_{ic}$ are positive constants [22, 28–31].

Theorem 3 *For the ith nominal follower* (3), *the approximate ETCOF control law provided by* (25), *the critic NN weight renovating law given by* (27), *and Assumptions 3–5. Then, the formation error is ensured to be UUB if the event-triggering condition*

$$\hat{T}_{i,c}(\mathscr{E}_i, \bar{\bar{\mathscr{E}}}_{i,\kappa})$$

$$= (1+\nu)\Bigg(\nabla\hat{\mathcal{J}}_i^{\mathsf{T}}(\mathscr{E}_i)\bigg((d_i + c_i)(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\bar{\mathscr{E}}}_{i,\kappa}))$$
$$- \sum_{j\in\mathcal{M}_i}\alpha_{ij}(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\bar{\mathscr{E}}}_{j,\kappa})) - c_i\mathcal{A}_{f,0}(z_0)\bigg)\Bigg)$$
$$+ \mathcal{U}_i(\mathscr{E}_i, \hat{\mu}_i, \hat{\mu}_{-i}) + \theta\sum_{j\in\bar{\mathcal{M}}_i}\|z_j\|^2$$
$$< 0 \tag{29}$$

holds.

Proof The selected Lyapunov function candidate is shown as

$$\mathcal{L}_{3T} = \mathcal{L}_{3T1} + \mathcal{L}_{3T2} = \mathcal{J}_i^*(\mathscr{E}_i) + \mathcal{J}_i^*(\bar{\bar{\mathscr{E}}}_{i,\kappa}). \tag{30}$$

Case 1: The event is not triggered, i.e., $t \in [\Gamma_\kappa, \Gamma_{\kappa+1})$. Calculating the time derivative of (30) and adopting system (3), we can get

$$\dot{\mathcal{L}}_{3T1} = \nabla\mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\bigg((d_i + c_i)(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\bar{\mathscr{E}}}_{i,\kappa}))$$
$$- \sum_{j\in\mathcal{M}_i}\alpha_{ij}(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\bar{\mathscr{E}}}_{j,\kappa})) - c_i\mathcal{A}_{f,0}(z_0)\bigg),$$
$$\dot{\mathcal{L}}_{3T2} = 0.$$

If the event-triggering condition (29) holds, it can be derived that

$$\dot{\mathcal{L}}_{3T1} \le \nabla\mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\bigg((d_i + c_i)(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\bar{\mathscr{E}}}_{i,\kappa}))$$
$$- \sum_{j\in\mathcal{M}_i}\alpha_{ij}(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\bar{\mathscr{E}}}_{j,\kappa}))$$
$$- c_i\mathcal{A}_{f,0}(z_0)\bigg) - \frac{1}{1+\nu}\hat{T}_{i,c}(\mathscr{E}_i, \bar{\bar{\mathscr{E}}}_{i,\kappa})$$
$$\le \nabla\mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)\bigg((d_i + c_i)(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\bar{\mathscr{E}}}_{i,\kappa}))$$
$$- \sum_{j\in\mathcal{M}_i}\alpha_{ij}(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\bar{\mathscr{E}}}_{j,\kappa}))\bigg)$$
$$- \nabla\hat{\mathcal{J}}_i^{\mathsf{T}}(\mathscr{E}_i)\bigg((d_i + c_i)(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\bar{\mathscr{E}}}_{i,\kappa}))$$
$$- \sum_{j\in\mathcal{M}_i}\alpha_{ij}(\mathcal{A}_{f,j}(z_j) + \mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\bar{\mathscr{E}}}_{j,\kappa}))\bigg)$$
$$+ \nabla\mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i)c_i\mathcal{A}_{f,0}(z_0) - \nabla\hat{\mathcal{J}}_i^{\mathsf{T}}(\mathscr{E}_i)c_i\mathcal{A}_{f,0}(z_0)$$
$$- \frac{1}{1+\nu}\mathcal{U}_i(\mathscr{E}_i, \hat{\mu}_i, \hat{\mu}_{-i}) - \frac{\theta}{1+\nu}\sum_{j\in\bar{\mathcal{M}}_i}\|z_j\|^2$$
$$\le \|\nabla\mathcal{J}_i^{*\mathsf{T}}(\mathscr{E}_i) - \nabla\hat{\mathcal{J}}_i^{\mathsf{T}}(\mathscr{E}_i)\|^2$$
$$+ \frac{1}{2}\big\|(d_i + c_i)(\mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\bar{\mathscr{E}}}_{i,\kappa}))\big\|^2$$

$$+\frac{1}{2}\left\|\sum_{j\in\mathcal{M}_i}\alpha_{ij}\big(\mathcal{A}_{f,j}(z_j)+\mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\mathscr{E}}_{j,\kappa})\big)\right\|^2$$

$$-\frac{1}{1+\nu}\mathcal{U}_i(\mathscr{E}_i,\hat{\mu}_i,\hat{\mu}_{-i})-\frac{\theta}{1+\nu}\sum_{j\in\bar{\mathcal{M}}_i}\|z_j\|^2$$

$$\leq\|\nabla\zeta_{ic}^{\mathsf{T}}(\bar{\mathscr{E}}_{i,\kappa})\chi_{ic}^*+\nabla\epsilon_{ic}(\mathscr{E}_i)-\nabla\zeta_{ic}^{\mathsf{T}}(\bar{\mathscr{E}}_{i,\kappa})\hat{\chi}_{ic}\|^2$$

$$+\frac{(d_i+c_i)^2}{2}\left\|\mathcal{A}_{f,i}(z_i)+\mathcal{A}_{g,i}(z_i)\hat{\mu}_i(\bar{\mathscr{E}}_{i,\kappa})\right\|^2$$

$$+\frac{\mathcal{M}_i}{2}\sum_{j\in\mathcal{M}_i}\alpha_{ij}^2\left\|\mathcal{A}_{f,j}(z_j)+\mathcal{A}_{g,j}(z_j)\hat{\mu}_j(\bar{\mathscr{E}}_{j,\kappa})\right\|^2$$

$$-\frac{1}{1+\nu}\mathcal{U}_i(\mathscr{E}_i,\hat{\mu}_i,\hat{\mu}_{-i})-\frac{\theta}{1+\nu}\sum_{j\in\bar{\mathcal{M}}_i}\|z_j\|^2.$$

$$(31)$$

According to Assumptions 3–5, we further obtain

$$\dot{\mathcal{L}}_{3T1}\leq\mathcal{M}_i\sum_{j\in\mathcal{M}_i}L_{j,f}^2\alpha_{ij}^2\|z_j\|^2+(d_i+c_i)^2L_{i,f}^2\|z_i\|^2$$

$$+\bar{\epsilon}_{ic}^2+\bar{\zeta}_{ic}^2\bar{\chi}_{icM}^2+(d_i+c_i)^2L_{i,f}^2\bar{g}_i^2\|\hat{\mu}_i(\bar{\mathscr{E}}_{i,\kappa})\|^2$$

$$+\mathcal{M}_i\sum_{j\in\mathcal{M}_i}\alpha_{ij}^2\bar{g}_j^2\|\hat{\mu}_j(\bar{\mathscr{E}}_{j,\kappa})\|^2-\frac{\lambda_{\min}(\mathcal{Q}_i)}{1+\nu}\|\mathscr{E}_i\|^2$$

$$-\frac{\theta}{1+\nu}\sum_{j\in\bar{\mathcal{M}}_i}\|z_j\|^2+\bar{\zeta}_{ic}^2\bar{\chi}_{ic}^2.$$

$$(32)$$

Let $\Xi_i=\max\{(d_i+c_i)^2L_{i,f}^2,\mathcal{M}_iL_{1,f}^2\alpha_{i,1}^2,...,\mathcal{M}_iL_{\mathcal{M}_i,f}^2\alpha_{i,\mathcal{M}_i}^2\}$. If we select appropriate θ to satisfy $\theta>(1+\nu)\Xi_i$, then we further get

$$\dot{\mathcal{L}}_{3T1}\leq\bar{\zeta}_{ic}^2\bar{\chi}_{ic}^2+\bar{\epsilon}_{ic}^2+\bar{\zeta}_{ic}^2\bar{\chi}_{icM}^2+(d_i+c_i)^2L_{i,f}^2\bar{g}_i^2\|\hat{\mu}_i(\bar{\mathscr{E}}_{i,\kappa})\|^2$$

$$+\mathcal{M}_i\sum_{j\in\mathcal{M}_i}\alpha_{ij}^2\bar{g}_j^2\|\hat{\mu}_j(\bar{\mathscr{E}}_{j,\kappa})\|^2-\frac{\lambda_{\min}(\mathcal{Q}_i)}{1+\nu}\|\mathscr{E}_i\|^2.$$

According to (25) and Assumptions 3 and 5, we have

$$\left\|\hat{\mu}_i(\bar{\mathscr{E}}_{i,\kappa})\right\|^2=\left\|-\frac{d_i+c_i}{2}\mathcal{R}_i^{-1}\mathcal{A}_{g,i}^{\mathsf{T}}(\bar{z}_{i,\kappa})\nabla\zeta_{ic}^{\mathsf{T}}(\bar{\mathscr{E}}_{i,\kappa})\hat{\chi}_{ic}\right\|^2$$

$$\leq\frac{(d_i+c_i)^2}{4}\|\mathcal{R}_i^{-1}\|^2\bar{g}_i^2\bar{\zeta}_{ic}^2\bar{\chi}_{ic}^2\triangleq\Phi_i.\qquad(33)$$

Based on (33), $\dot{\mathcal{L}}_{3T1}$ can be deduced as

$$\dot{\mathcal{L}}_{3T1}\leq\bar{\zeta}_{ic}^2\bar{\chi}_{ic}^2+\bar{\epsilon}_{ic}^2+\bar{\zeta}_{ic}^2\bar{\chi}_{icM}^2+(d_i+c_i)^2L_{i,f}^2\bar{g}_i^2\Phi_i$$

$$+\mathcal{M}_i\sum_{j\in\mathcal{M}_i}\alpha_{ij}^2\bar{g}_j^2\Phi_j-\frac{\lambda_{\min}(\mathcal{Q}_i)}{1+\nu}\|\mathscr{E}_i\|^2$$

$$\leq-\frac{\rho^2\lambda_{\min}(\mathcal{Q}_i)}{1+\nu}\|\mathscr{E}_i\|^2+\frac{(\rho^2-1)\lambda_{\min}(\mathcal{Q}_i)}{1+\nu}\|\mathscr{E}_i\|^2+\Theta,$$

$$(34)$$

where $0<\rho<1$ and $\Theta=\bar{\zeta}_{ic}^2\bar{\chi}_{ic}^2+\bar{\epsilon}_{ic}^2+\bar{\zeta}_{ic}^2\bar{\chi}_{icM}^2+(d_i+c_i)^2L_{i,f}^2\bar{g}_i^2\Phi_i+\mathcal{M}_i\sum_{j\in\mathcal{M}_i}\alpha_{ij}^2\bar{g}_j^2\Phi_j$. Hence, $\dot{\mathcal{L}}_{3T}<0$ if the formation error $\mathscr{E}_i$ lies outside the compact set

$$\Omega_{\mathscr{E}_i}=\left\{\mathscr{E}_i:\|\mathscr{E}_i\|\leq\sqrt{\frac{\Theta(1+\nu)}{(1-\rho^2)\lambda_{\min}(\mathcal{Q}_i)}}\right\}.\qquad(35)$$

Case 2: The event is triggered, i.e., $\forall t=\Gamma_{\kappa+1}$. According to (30), we can get

$$\Delta\mathcal{L}_{3T}(t)=\Delta\mathcal{L}_{3T1}(t)+\Delta\mathcal{L}_{3T2}(t).\qquad(36)$$

Based on the result in *Case 1*, it is evident that $\dot{\mathcal{L}}_{3T}<0$ for all $t\in[\Gamma_\kappa,\Gamma_{\kappa+1})$. Therefore, we further have

$$\Delta\mathcal{L}_{3T1}(t)=\mathcal{J}_i^*(\bar{\mathscr{E}}_{i,\kappa+1})-\mathcal{J}_i^*\big(\mathscr{E}(\Gamma_{\kappa+1}^-)\big)\leq0,$$

$$\Delta\mathcal{L}_{3T2}(t)=\mathcal{J}_i^*(\bar{\mathscr{E}}_{i,\kappa+1})-\mathcal{J}_i^*(\bar{\mathscr{E}}_{i,\kappa})\leq-\iota\big(\|e_{i,\kappa+1}(\Gamma_\kappa)\|\big),$$

where $\iota(\cdot)$ is a class-$\mathcal{K}$ function and $e_{i,\kappa+1}(\Gamma_\kappa)=\bar{\mathscr{E}}_{i,\kappa+1}-\bar{\mathscr{E}}_{i,\kappa}$. According to the above discussion, we can conclude that $\dot{\mathcal{L}}_{3T}<0$ is still satisfied when the event occurs. The proof is completed. □

Theorem 4 *Consider the ith nominal follower* (3), *the approximate ETCOF control law given by* (25), *the event-triggering condition given by* (29), *and* Assumptions 3–5. *Then, the minimal intersampling time satisfies* $\Delta\Gamma_{\min}>0$.

Proof According to the existing results [18] and [32], it is easy to obtain the conclusion of Theorem 4, so the detailed proof is omitted. □

Remark 6 *The difficulty of designing the RL-based ETCRF controller is mainly reflected in the following two aspects. Firstly, this paper considers the MAS with dynamic uncertainties. In order to ensure that each following agent can catch up with the leader trajectory with the specified formation in the presence of uncertain terms, the designed controller is required to possess robustness. Therefore, a novel value function which contains an upper bound function of the uncertainty, is designed. The objective is to proactively account for the impact of uncertainties during the controller design process, thus obtaining a robust formation controller. Second, in order to save computing and communication resources, this paper designs the robust formation controller under the event triggered mechanism, that is, the controller is updated only when the triggering condition is not met. To this end, it is crucial to design a reasonable*

triggering condition, which can save computing and communication resources and maintain the desired formation between each agent. In this paper, a new triggering condition is designed by using the Lyapunov stability principle. Theoretical analysis shows that the developed RL-based ETCRF controller ensures the formation error to be asymptotically stable under the designed triggering condition.

4 Simulation

In this section, three simulation cases are applied to showcase the availability of the presented RL-based ETCRF control approach.

4.1 Case 1

This case selects a MAS with one leader and three followers. The communication network is illustrated in Fig. 1, with L_0 acting as the leader and F_i ($i = 1, 2, 3$) serving as the respective follower. The parameters of communication topology are given as $\alpha_{12} = 1$, $\alpha_{21} = 1$, $\alpha_{13} = 1$, $\alpha_{31} = 1$, $\alpha_{23} = 1$, $\alpha_{32} = 1$, and $c_1 = 1$. The dynamics of the ith follower is given as

$$\dot{z}_i = \mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\big(\mu_i + \Lambda_i(z_i)\big), \tag{37}$$

where

$$\mathcal{A}_{f,1}(z_1) = \begin{bmatrix} z_{1,2} \\ \mathcal{F}_1 \end{bmatrix}, \mathcal{A}_{f,2}(z_2) = \begin{bmatrix} z_{2,2} \\ \mathcal{F}_2 \end{bmatrix}, \mathcal{A}_{f,3}(z_3) = \begin{bmatrix} z_{3,2} \\ \mathcal{F}_3 \end{bmatrix},$$

$$\mathcal{A}_{g,1}(z_1) = \begin{bmatrix} 0 \\ 2z_{1,1} \end{bmatrix}, \mathcal{A}_{g,2}(z_2) = \begin{bmatrix} 0 \\ \cos(2z_{2,1}^2) + 2 \end{bmatrix},$$

$$\mathcal{A}_{g,3}(z_3) = \begin{bmatrix} 0 \\ \sin(4z_{3,1}^2) + 2 \end{bmatrix},$$
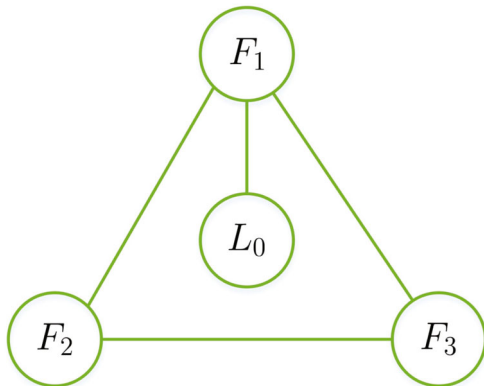
Fig. 1 The structure of communication topology
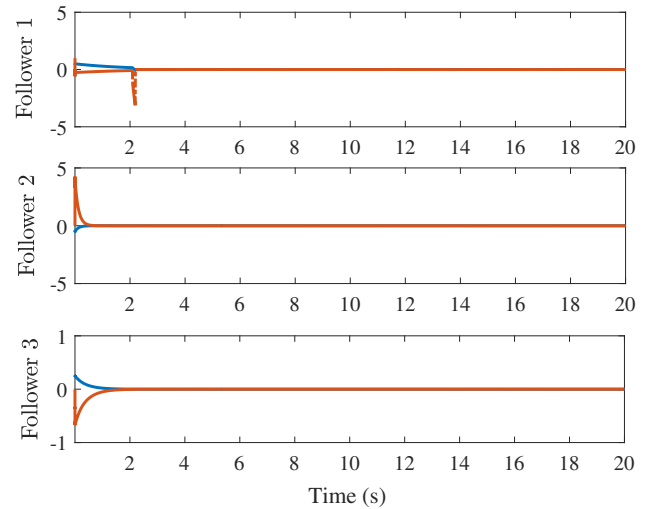
Fig. 2 Formation errors of all followers in case 1

$$\mathcal{F}_1 = -0.5z_{1,1} - 0.5z_{1,2}(1 + (\cos(2z_{1,1}) + 2)^2),$$
$$\mathcal{F}_2 = -z_{2,1} - 0.5z_{2,2} + 0.5z_{2,1}^2 z_{2,2},$$
$$\mathcal{F}_3 = -0.5z_{3,1} - z_{3,2} + z_{3,1}^2 z_{3,2} - 0.25z_{3,2}(\cos(2z_{3,1})$$
$$+ 2)^2, +0.25z_{i,2}(\sin(4z_{i,2}^2) + 2)^2,$$
$$\Lambda_1(z_1) = 2z_{1,1}\cos^3(z_{1,1})\cos^2(z_{1,2}), \Lambda_2(z_2)$$
$$= 2z_{2,1}\sin^6(z_{2,1})\sin^2(z_{2,2}),$$
$$\Lambda_3(z_3) = 2z_{3,1}\cos^3(z_{3,1})\sin^5(z_{3,2}), z_i = [z_{i,1}, z_{i,2}]^\mathsf{T},$$
$$i = 1, 2, 3.$$

The dynamics of the leader is chosen as

$$\dot{z}_0 = \begin{bmatrix} \sin(t) \\ \cos(t) \end{bmatrix}.$$
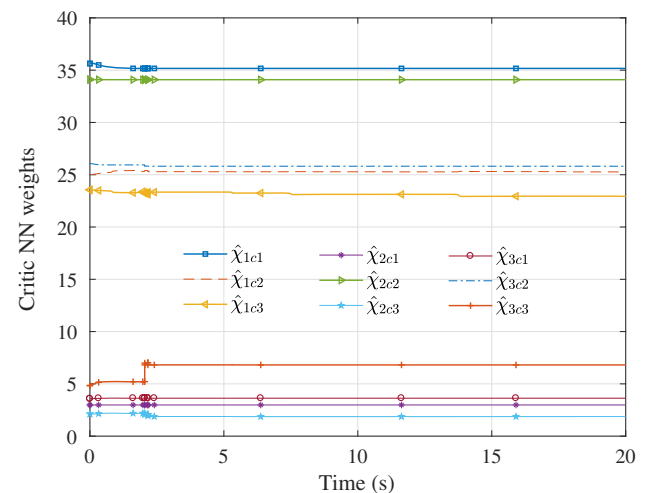
Fig. 3 Critic NN weights in case 1
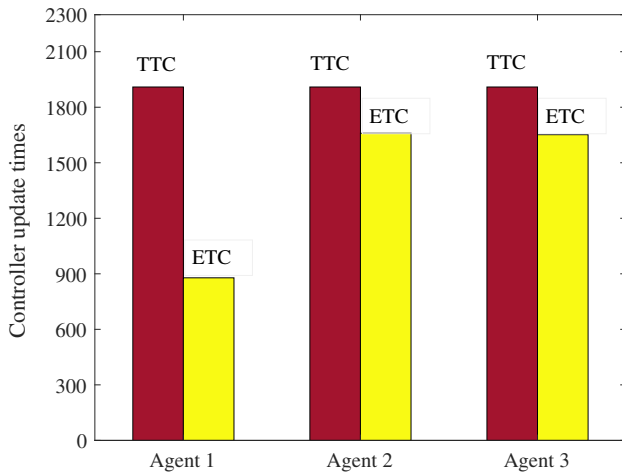
Fig. 4 The controller update times of all followers in case 1

Fig. 6 System state curves of all followers in case 1

The formation patterns of all followers are chosen as $\eta_1 = [0.5; 0]$, $\eta_2 = [1; 0]$, and $\eta_3 = [1.5; 0]$. The parameters in value function are determined as $\mathcal{Q}_i = I_2$, $\mathcal{R}_i = 0.1I$, $\gamma = 0.5$ and $\theta = 10$, respectively. The activation function of each follower is designed as $\zeta_{ic} = [\mathscr{E}_{i,1}^2, \mathscr{E}_{i,2}^2, \mathscr{E}_{i,1}\mathscr{E}_{i,2}]$. The learning rate is picked as $\varrho_c = 2$. The parameter in triggering condition is selected as $\nu = 0.5$.

Simulation results of this case are exhibited in Figs. 2, 3, 4, 5 and 6. Figure 2 displays that the formation error of each nominal follower will converge to zero, which implies that the desired formation between each follower can be maintained. The critic NN weight evolution curves are provided in Fig. 3. It is apparent that the weight vectors of critic NNs will converge to $\hat{\chi}_{1c} = [35.16, 25.29, 22.90]^T$, $\hat{\chi}_{2c} = [2.98, 34.08, 1.87]^T$, and $\hat{\chi}_{3c} = [3.62, 25.81, 6.81]^T$, respectively. Figure 4 compares the controller update times of each follower under the time-triggered control (TTC) mechanism and the event-triggered control (ETC) mechanism. It is clear that the ETC mechanism can reduce the controller update times and the computing and communication

burdens of each follower is alleviated. Figure 5 presents the triggering time curves for each follower, from which it can be observed that under the event-triggered framework, the triggering times are not evenly spaced. Figure 6 reveals the state curve of each follower with dynamic uncertainty. We can conclude that the expected formation between each uncertain follower can be guaranteed with the presented RL-based ETCRF control method. Figure 7 depicts the system state trajectories under different control methods. It can be observed that due to the influence of uncertainties, the system state trajectories under traditional formation control method deviate from the ideal trajectory.

4.2 Case 2

In this case, three spacecrafts are adopted to demonstrate the validity of the proposed RL-based ETCRF control approach. According to [33], the system dynamics of the ith spacecraft
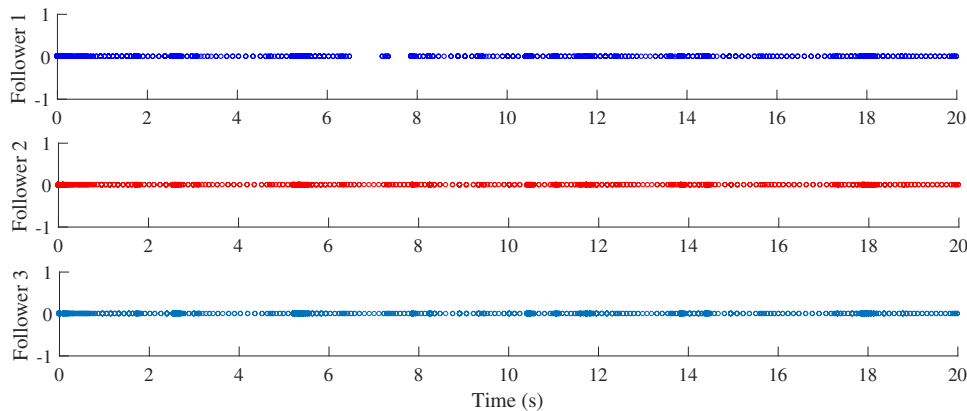
Fig. 5 Triggering times in case 1
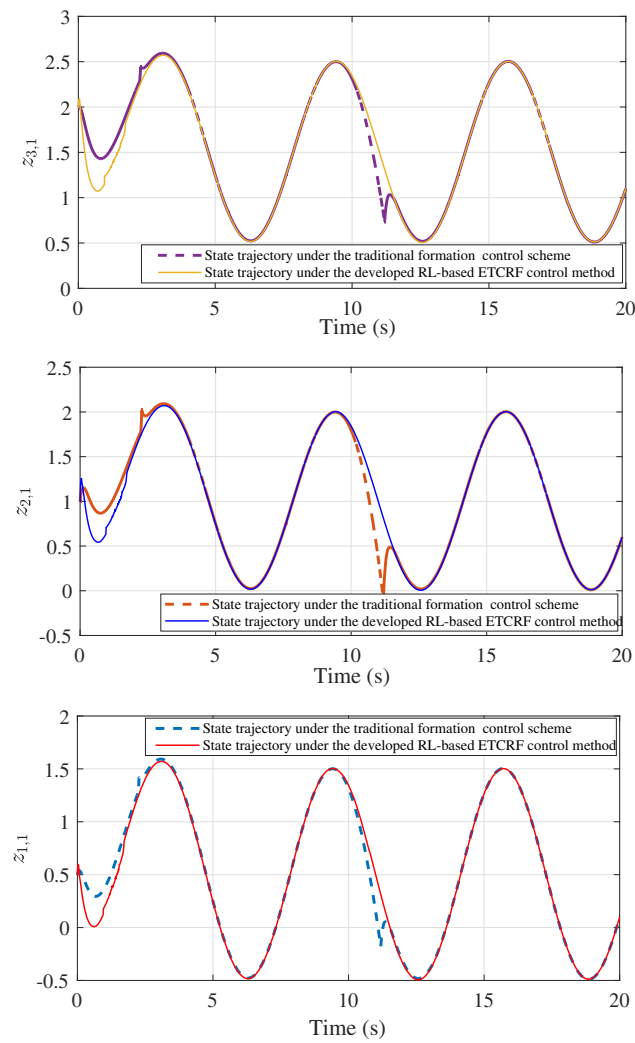
Fig. 7 System state curves of all uncertain agents in case 1

is provided as

$$\dot{\mathcal{P}}_i = \mathcal{V}_i,$$
$$m_i \dot{\mathcal{V}}_i + \mathcal{C}_i \mathcal{V}_i + \mathcal{D}_i \mathcal{P}_i + \mathcal{N}_i = \mu_i + \Lambda_i, \tag{38}$$

Table 1 Parameters of the spacecraft

Parameter	φ	a_c	σ_c
Description	True anomaly	Semimajor axis	Orbit eccentricity
Value	$\varphi(0) = 0$	7178	0.01
Parameter	ϖ	m_i	
Description	Earth's gravitational	Mass of the spacecraft	
Value	$3.986 \times 10^{14} m^3/s^2$	100kg	
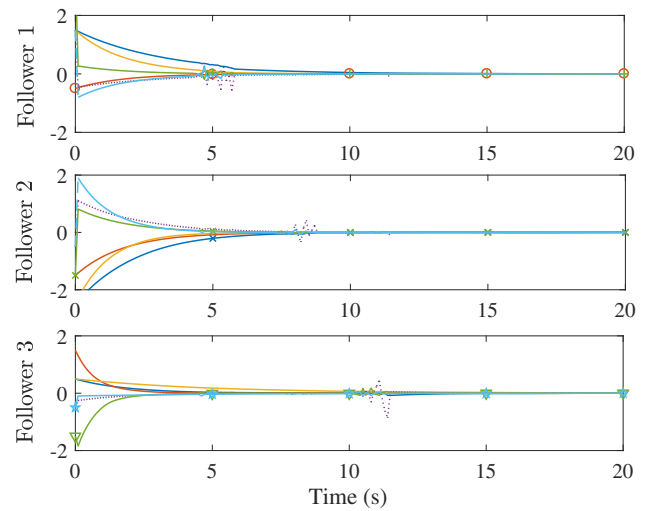
Fig. 8 Formation errors of all followers in case 2
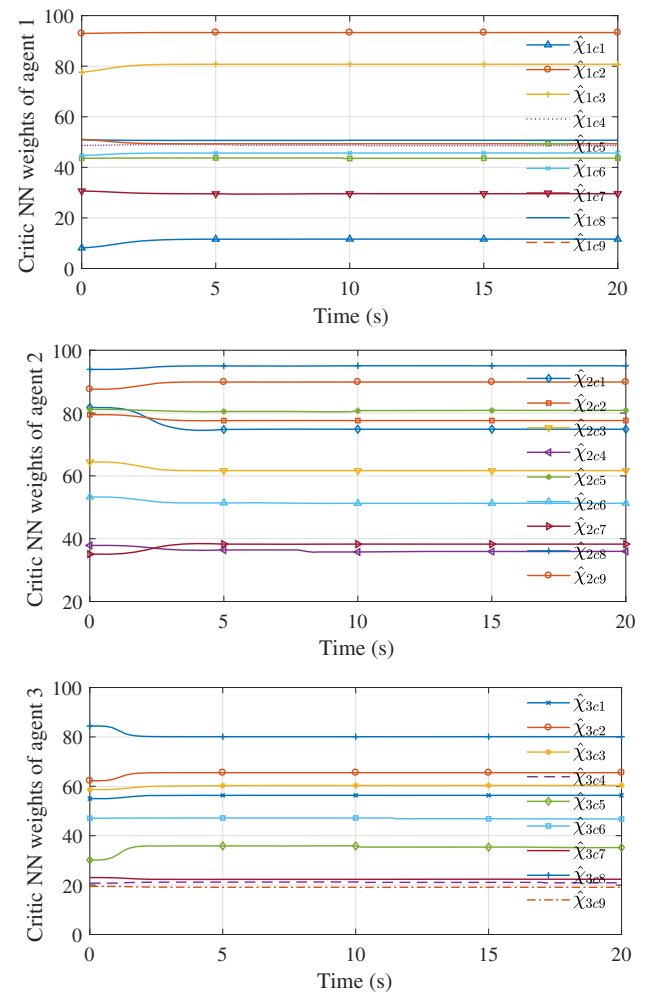
Fig. 9 Critic NN weights in case 2
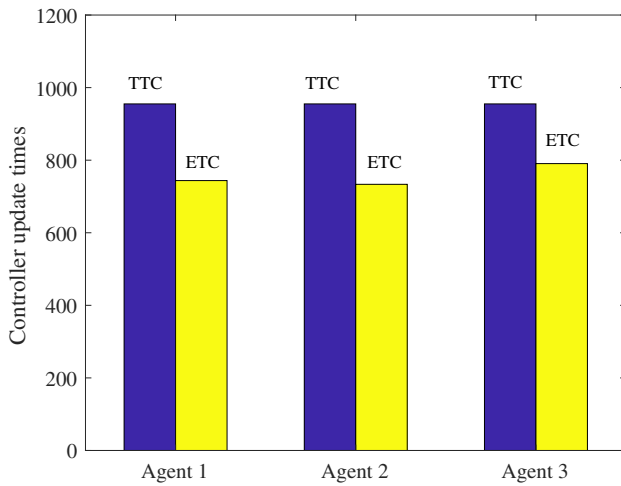
Fig. 10 Controller update times of all spacecrafts in case 2

where $\mathcal{P}_i = [\mathcal{P}_{i,1}, \mathcal{P}_{i,2}, \mathcal{P}_{i,3}]^\mathsf{T} \in \mathbb{R}^3$ and $\mathcal{V}_i \in \mathbb{R}^3$ denote the position and velocity vectors of the ith spacecraft, respectively, μ_i is the control input, Λ_i is the dynamic uncertainty, $\mathcal{C}_i$, $\mathcal{D}_i$ and $\mathcal{N}_i$ are system functions, which provide as

$$\mathcal{C}_i = 2m_i \begin{bmatrix} 0 & -\dot{\varphi} & 0 \\ \dot{\varphi} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

$$\mathcal{D}_i = m_i \begin{bmatrix} \varpi/r_i^3 - \dot{\varphi}^2 & -\ddot{\varphi} & 0 \\ \ddot{\varphi} & \varpi/r_i^3 - \dot{\varphi}^2 & 0 \\ 0 & 0 & \varpi/r_i^3 \end{bmatrix},$$

$$\mathcal{N}_i = m_i \varpi \begin{bmatrix} R_c/r_i^3 - 1/R_c^2 \\ 0 \\ 0 \end{bmatrix},$$

$$r_i = \sqrt{(R_c + \mathcal{P}_{i,1})^2 + \mathcal{P}_{i,2}^2 + \mathcal{P}_{i,3}^2},$$

$$R_c = a_c(1 - \sigma_c^2)/(1 + \sigma_c\cos(\varphi)),$$

$$\dot{\varphi} = n_c(1 - \sigma_c\cos(\varphi))^2/(1 - \sigma_c^2)^{3/2}, n_c = \sqrt{\varpi/a_c^3}.$$

The physical interpretations and values of the corresponding variables are given in Table 1.

Let $z_i = [\mathcal{P}_i^\mathsf{T}, \dot{\mathcal{P}}_i^\mathsf{T}]^\mathsf{T} = [z_{i,1}, z_{i,2}, z_{i,3}, z_{i,4}, z_{i,5}, z_{i,6}]^\mathsf{T}$. Then, the system dynamics of the ith spacecraft is reformulated as

$$z_i = \mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)(\mu_i + \Lambda_i), \tag{39}$$

where

$$\Lambda_1 = \begin{bmatrix} z_{1,1} \\ \cos^5(z_{1,1}) \\ \sin^3(z_{1,5}) \end{bmatrix}, \Lambda_2 = \begin{bmatrix} z_{2,1} \\ \sin^6(z_{2,2}) \\ \sin^2(z_{2,3}) \end{bmatrix},$$

$$\Lambda_3 = \begin{bmatrix} z_{3,1} \\ \cos^3(z_{3,3}) \\ \sin^6(z_{3,1}) \end{bmatrix},$$

$$\mathcal{A}_{f,i}(z_i) = \begin{bmatrix} \mathcal{V}_i \\ -\frac{1}{m_i}(\mathcal{C}_i\mathcal{V}_i + \mathcal{D}_i\mathcal{P}_i + \mathcal{N}_i) \end{bmatrix},$$

$$\mathcal{A}_{g,i}(z_i) = \frac{1}{m_i} \begin{bmatrix} \mathbf{0}_{3\times3} \\ \mathbf{I}_3 \end{bmatrix}.$$

The dynamics of the leader is given as

$$\dot{z}_0 = \mathcal{A}_{f,0}(z_0),$$

where

$$\mathcal{A}_{f,0}(z_0) = [\cos(t), \sin(t), \cos(t), -\sin(t), \cos(t), -\sin(t)]^\mathsf{T}.$$
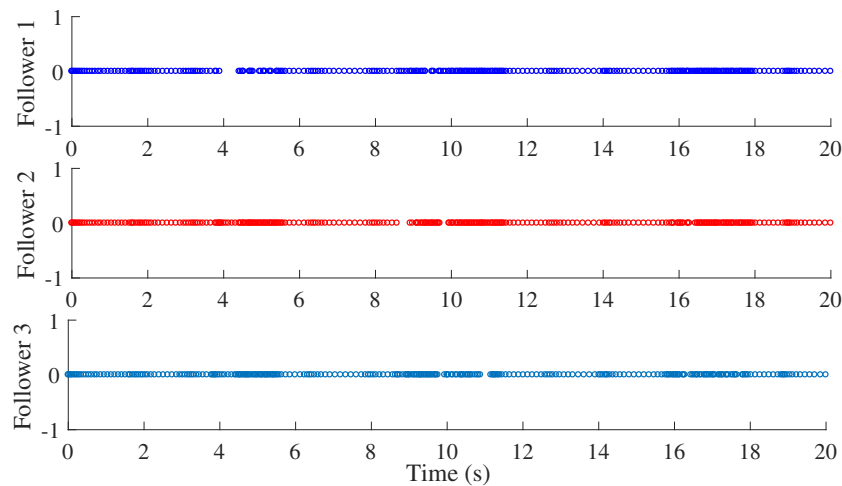
Fig. 11 Triggering times in case 2

The communication topology of this case is selected the same as in case 1. The formation patterns of all followers are chosen as $\eta_1 = [1; 1; 1; 0; 0; 0]$, $\eta_2 = [2; 2; 2; 0; 0; 0]$, and $\eta_3 = [1.5; 1.5; 1.5; 0; 0; 0]$. The activation function of the ith following spacecraft is designed as $\zeta_{ic} = [\mathscr{E}_{i,1}^2, \mathscr{E}_{i,2}^2, \mathscr{E}_{i,3}^2, \mathscr{E}_{i,4}^2, \mathscr{E}_{i,5}^2, \mathscr{E}_{i,6}^2, \mathscr{E}_{i,1}\mathscr{E}_{i,4}, \mathscr{E}_{i,2}\mathscr{E}_{i,5}, \mathscr{E}_{i,3}\mathscr{E}_{i,6}]$. The control parameters of this case are given as: $\mathcal{Q}_i = I_6$, $\mathcal{R}_i = 0.1 I_3$, $\gamma = 0.5$, $\theta = 20$, $\varrho_c = 0.5$, $\nu = 0.8$.

Figures 8, 9, 10, 11 and 12 display the simulation results of this case. The convergence of the formation error of each following spacecraft is shown in Fig. 8. Figure 9 reveals the critic NN weight updating curves. It is readily apparent that the weight vectors of the critic NN will converge to $\hat{\chi}_{1c} = [11.61, 93.29, 80.76, 48.61, 43.63, 45.67, 29.56, 50.70, 49.31]^T$, $\hat{\chi}_{2c} = [74.85, 77.61, 61.67, 35.97, 80.84, 51.28, 38.27, 95.04, 89.88]^T$, and $\hat{\chi}_{3c} = [56.33, 65.51, 60.34, 20.93, 35.19, 46.78, 22.37, 80.09, 19.12]^T$, respectively. The controller update times of the TTC and ETC approaches are provided in Fig. 10. Obviously, the ETC method requires less controller update and saves communication and computing resources. Figure 11 further illustrates the triggering time of each following spacecraft. It can be seen that under the event-triggered framework, the controller for each following spacecraft is not continuously updated. Figure 12 demonstrates that the proposed RL-based ETCRF control approach maintains the expected formation of the spacecraft cluster with dynamic uncertainty. Figure 13 compares the spacecraft state trajectories under different control methods, from which we can see that the developed RL-based ETCRF control approach method can effectively resist the influence of uncertainties.

4.3 Case 3

This case adopts three robotic arms to further validate the effectiveness of the developed RL-based ETCRF control method. The dynamics of the ith robotic arm is given as

$$\dot{z}_i = \mathcal{A}_{f,i}(z_i) + \mathcal{A}_{g,i}(z_i)\big(\mu_i + \Lambda_i(z_i)\big), \tag{40}$$

where

$$\mathcal{A}_{f,i}(z_i) = \begin{bmatrix} z_{i,2} \\ -4.905\sin(z_{i,1}) - 0.2 z_{i,2} \end{bmatrix}, \quad \mathcal{A}_{g,i}(z_1) = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix},$$
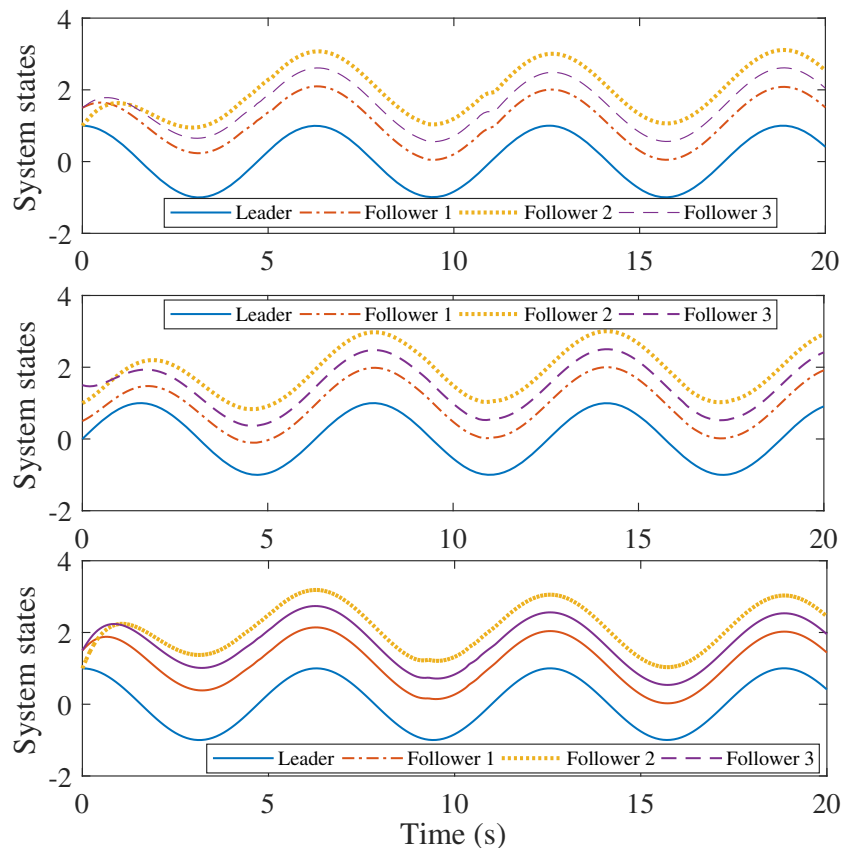
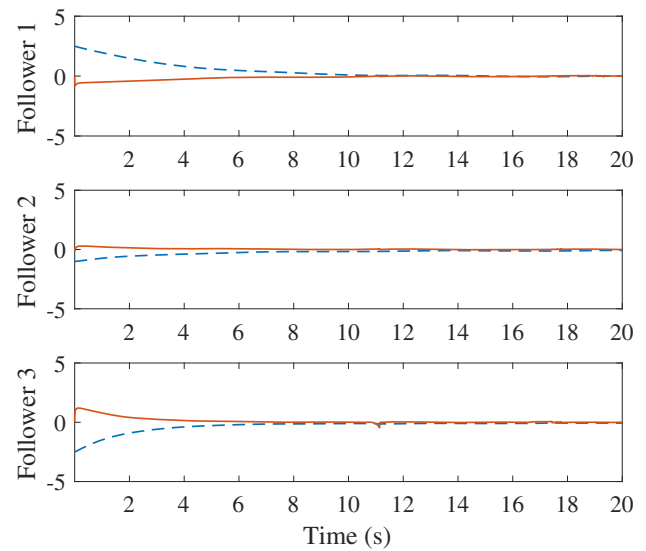**Fig. 12** System state curves of all uncertain agents in case 2

Fig. 14 Formation errors of all followers in case 3

each robotic arm can be achieved. Figure 15 displays the changes of critic NN weights, which ultimately converge to $\hat{\chi}_{1c} = [64.91, 42.81, 82.14]^{\mathsf{T}}$, $\hat{\chi}_{2c} = [21.93, 49.54, 91.65]^{\mathsf{T}}$, and $\hat{\chi}_{3c} = [70.47, 77.97, 77.10]^{\mathsf{T}}$, respectively. Figure 16 compares the controller update times between the ETC method and the TTC method. It is evident that the ETC method requires fewer controller updates, thus alleviating computing and communication burdens. Figure 17 reveals the moments of event occurrences, indicating that they are not evenly spaced. Only when the triggering condition is not met will it be recorded as a triggering moment. Figure 18 demonstrates that even in the presence of system uncertainties, each robotic arm can maintain the specified formation

Fig. 13 System states under different control approaches in case 2

$$\Lambda_1(z_1) = \cos(z_{1,1})\sin^2(z_{1,2}), \Lambda_2(z_2) = z_{2,1}\sin^3(z_{2,1}),$$
$$\Lambda_3(z_3) = z_{3,1}\cos^2(z_{3,1})\sin^5(z_{3,2}), z_i = [z_{i,1}, z_{i,2}]^{\mathsf{T}},$$
$$i = 1, 2, 3.$$

The dynamics of the leader, the communication topology, and the formation patterns of all followers are the same as in Case 1. The simulation parameters are selected as $\mathcal{Q}_i = I_2$, $\mathcal{R}_i = 0.01I$, $\gamma = 0.3$, $\varrho_c = 3$, $\theta = 15$ and $\nu = 0.5$. The activation function of each robotic arm is designed as $\zeta_{ic} = [\mathscr{E}_{i,1}^2, \mathscr{E}_{i,2}^2, \mathscr{E}_{i,1}\mathscr{E}_{i,2}]$.

Figure 14 shows that the formation errors can eventually converge to near zero. Therefore, the desired formation of

Fig. 15 Critic NN weights in case 3

Fig. 16 Controller update times of all robotic arms in case 3

Fig. 18 System state curves of all uncertain agents in case 3

and track the leader's trajectory. Figure 19 demonstrates that compared with the traditional method, the developed RL-based ETCRF control approach exhibits robustness and has the ability to eliminate the influence of uncertainties.

Remark 7 *a) Equations (1)-(10), which appear in the problem statement section, primarily describe the system dynamics, the formation error and its dynamics, the value function, the optimal formation control law, and the HJB equation. b) Equations (11)-(19), appearing in the section of event-triggered robust formation controller design, provide the definition of formation error and the expression of optimal formation control law under the event-triggered mechanism. In addition, it includes the detailed derivation process of Theorem 1. c) Equations (20)-(27) describe the expressions of the value function, the formation control law, the Hamiltonian, and the NN weight updating law under the NN framework. b) Equations (28)-(36) present a detailed derivation of Theorem 3. e) Equations (37)-(40) provide the expressions of closed-loop systems in simulation section.*

5 Conclusion

In this paper, the RL-based ETCRF control method is presented for MASs with matched uncertainties. The ETCRF control problem is transformed to a ETCOF control problem by developing a discounted value function for each agent. For the sake of reducing computing and communication burdens, the developed robust formation controller of each agent is improved when the novel event-triggering condition is contravened only. Afterwards, a single critic structure is established to acquire the approximate solution of the coupled HJB equation of each agent. Additionally, through theoretical analysis, it is revealed that the proposed RL-based ETCRF control method is capable of preserving the prescribed formation between each agent. In the end, two simulation cases are presented to confirm the effectiveness of the introduced RL-based ETCRF control approach. Due to the unique advantages of different types of RL algorithms, in our future work, we will propose event-triggered control

Fig. 17 Triggering times in case 3
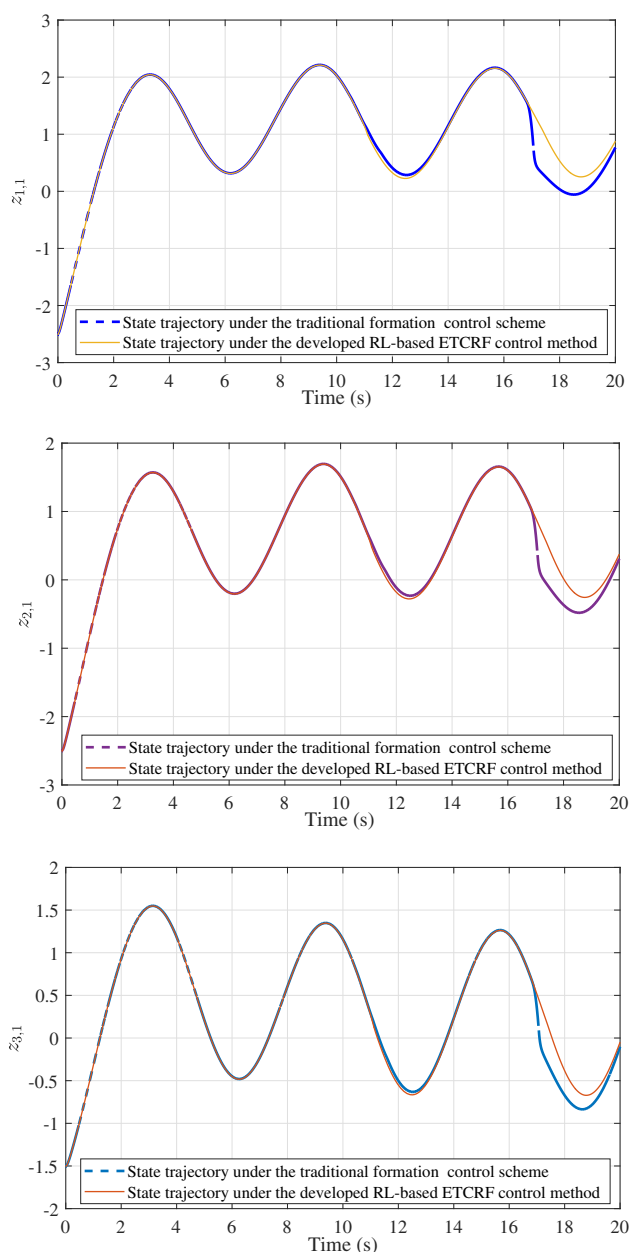
Fig. 19 System states under different control appraoches in case 3

methods based on value iteration or policy gradient to address cooperative control problems of MASs. Additionally, we will apply the RL-based ETCRF control approach to solve formation control problems of practical unmanned autonomous systems such as drones or ships.

Data Availability The authors can confirm that all relevant data are included in the article.

Declarations

Conflicts of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethical and informed consent for data used This paper does not involve data collection, human subjects or sensitive personal information, therefore ethical approval and informed consent were not required.

References

1. Liu Y, Chi R, Li H, Wang L, Lin N (2023) HiTL-based adaptive fuzzy tracking control of MASs: A distributed fixed-time strategy. Sci China Technol Sci 66:2907–2916
2. Cao L, Cheng Z, Liu Y, Li H (2022) Event-based adaptive NN fixed-time cooperative formation for multiagent systems. IEEE Trans Neural Netw Learn Syst Early Access. https://doi.org/10.1109/TNNLS.2022.3210269
3. Pang Z, Zheng C, Li C, Liu G, Han Q (2022) Cloud-based time-varying formation predictive control of multi-agent systems with random communication constraints and quantized signals. IEEE Trans Circ Syst II: Express Briefs 69(3):1282–1286
4. Yao D, Li H, Lu R, Shi Y (2022) Event-based distributed sliding mode formation control of multi-agent systems and its applications to robot manipulators. Inf Sci 614:87–103
5. Liu D, Wei Q, Wang D, Yang X, Li H (2017) Adaptive Dynamic Programming With Applications in Optimal Control. Springer, Cham, Switzerland, p 2017
6. Liu D, Xue S, Zhao B, Luo B, Wei Q (2021) Adaptive dynamic programming for control: A survey and recent advances. IEEE Trans Syst Man Cybern: Syst 51(1):142–160
7. Qin C, Zhang Z, Shang Z, Zhang J, Zhang D (2023) Adaptive optimal safety tracking control for multiplayer mixed zero-sum games of continuous-time systems. Appl Intell 53:17460–17475
8. Liu D, Wei Q (2014) Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. IEEE Trans Neural Netw Learn Syst 25(3):621–634
9. Wei Q, Liu D, Liu Y, Song R (2017) Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming. IEEE/CAA J Autom Sin 4(2):168–176
10. Su H, Luan X, Zhang H, Liang X, Yang J, Wang J (2024) Decentralized optimal control of large-scale partially unknown nonlinear mismatched interconnected systems based on dynamic event-triggered control. Neurocomputing 568:127013
11. Qin C, Qiao X, Wang J, Zhang D, Hou Y, Hu S (2024) Barrier-critic adaptive robust control of nonzero-sum differential games for uncertain nonlinear systems with state constraints. IEEE Trans Syst Man Cybern: Syst 54(1):50–63
12. Guo Z, Li H, Ma H, Meng W (2022) Distributed optimal attitude synchronization control of multiple QUAVs via adaptive dynamic programming. IEEE Trans Neural Netw Learn Syst Early Access. https://doi.org/10.1109/TNNLS.2022.3224029

13. Xia L, Li Q, Song S, Modares H (2022) Optimal synchronization control of heterogeneous asymmetric input-constrained unknown nonlinear MASs via reinforcement learning. IEEE/CAA J Autom Sin 9(3):520–532

14. Mu C, Peng J, Sun C (2023) Hierarchical multiagent formation control scheme via actor-critic learning. IEEE Trans Neural Netw Learn Syst 34(11):8764–8777

15. Wen G, Chen C, Li B (2020) Optimized formation control using simplified reinforcement learning for a class of multiagent systems with unknown dynamics. IEEE Trans Ind Electron 67(9):7879–7888

16. Xiao W, Zhou Q, Liu Y, Li H, Lu R (2022) Distributed reinforcement learning containment control for multiple nonholonomic mobile robots. IEEE Trans Circ Syst I: Reg Papers 69(2):896–907

17. Yang Y, Modares H, Wunsch D, Yin Y (2019) Optimal containment control of unknown heterogeneous systems with active leaders. IEEE Trans Control Syst Technol 27(3):1228–1236

18. Zhao W, Zhang H (2019) Distributed optimal coordination control for nonlinear multi-agent systems using event-triggered adaptive dynamic programming method. ISA Trans 91:184–195

19. Chen Z, Chen K, Chen S, Zhang Y (2022) Event-triggered H_∞ consensus for uncertain nonlinear systems using integral sliding mode based adaptive dynamic programming. Neural Netw 156:258–270

20. Wang J, Zhang Z, Tian B, Zong Q (2024) Event-based robust optimal consensus control for nonlinear multiagent system with local adaptive dynamic programming. IEEE Trans Neural Netw Learn Syst 35(1):1073–1086

21. Ren H, Cheng Z, Qin J, Lu R (2023) Deception attacks on event-triggered distributed consensus estimation for nonlinear systems. Automatica 154:111100

22. Wang D, Mu C, Yang X, Liu D (2017) Event-based constrained robust control of affine systems incorporating an adaptive critic mechanism. IEEE Trans Syst Man Cybern: Syst 47(7):1602-1612

23. Zhang Q, Zhao D, Wang D (2018) Event-based robust control for uncertain nonlinear systems using adaptive dynamic programming. IEEE Trans Neural Netw Learn Syst 29(1):37–50

24. Huo Y, Wang D, Li M, Qiao J (2024) Decentralized event-triggered asymmetric constrained control through adaptive critic designs for nonlinear interconnected systems. IEEE Trans Syst Man Cybern: Syst 54(1):391–402

25. Yang X, Zeng Z, Gao Z (2022) Decentralized neurocontroller design with critic learning for nonlinear-interconnected systems. IEEE Trans Cybern 52(11):11672–11685

26. Zhao B, Liu D (2020) Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming. IEEE Trans Ind Electron 67(4):3054–3064

27. Xue S, Luo B, Liu D (2020) Event-triggered adaptive dynamic programming for zero-sum game of partially unknown continuous-time nonlinear systems. IEEE Trans Syst Man Cybern Syst 50(9):3189–3199

28. Zhang Y, Zhao B, Liu D, Zhang S (2022) Event-triggered control of discrete-time zero-sum games via deterministic policy gradient adaptive dynamic programming. IEEE Trans Syst Man Cybern Syst 52(8):4823–4835

29. Yang X, He H (2020) Event-triggered robust stabilization of nonlinear input-constrained systems using single network adaptive critic designs. IEEE Trans Syst Man Cybern: Syst 50(9):3145–3157

30. Lin M, Zhao B, Liu D (2024) Event-triggered robust adaptive dynamic programming for multiplayer stackelberg-nash games of uncertain nonlinear systems. IEEE Trans Cybern 54(1):273–286

31. Zhao B, Liu D, Luo C (2020) Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints. IEEE Trans Neural Netw Learn Syst 31(10):4330–4340

32. Luo B, Yang Y, Liu D, Wu H (2020) Event-triggered optimal control with performance guarantees using adaptive dynamic programming. IEEE Trans Neural Netw Learn Syst 31(1):76–88

33. Wei C, Wu X, Xiao B, Wu J, Zhang C (2022) Adaptive leader-following performance guaranteed formation control for multiple spacecraft with collision avoidance and connectivity assurance. Aerospace Sci Technol 120:107266

Contents lists available at ScienceDirect

European Journal of Control

journal homepage: www.sciencedirect.com/journal/european-journal-of-control

Integral sliding mode-based event-triggered optimal fault tolerant tracking control of continuous-time nonlinear systems☆

Yongwei Zhang [a], Shunchao Zhang [b],*

[a] College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China
[b] School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou 510521, China

ARTICLE INFO

ABSTRACT

In this paper, the integral sliding mode-based event-triggered optimal fault tolerant tracking control of continuous-time nonlinear systems is investigated via adaptive dynamic programming. The developed control scheme consists of two parts, i.e., integral sliding mode control and event-triggered optimal tracking control. For the first part, an integral sliding mode controller is designed to eliminate the affect of actuator fault and the dynamics of nominal nonlinear systems is obtained. For the second part, a novel quadratic cost function with respect to the tracking error and its dynamics is developed such that the feedforward control law or the discount factor is not required, which reduces the complexity of the control method and guarantees the tracking performance. Moreover, a critic-only structure is established to obtain the solution of tracking Hamilton–Jacobi–Bellman equation. It should be noted that the optimal tracking control law is updated only at triggering moments in order to preserve computing and communication resources. Finally, the effectiveness of the present approach is demonstrated through simulation examples of a robotic arm system and a Van der Pol circuit system.

1. Introduction

In real-world control systems, such as those encountered in unmanned aerial vehicles (UAVs), mobile robots, and unmanned surface vehicles (USVs), simultaneous tracking of a predetermined trajectory and optimization of performance indicators poses significant challenges. Consequently, the optimal tracking control (OTC) problem emerges, necessitating the resolution of the intricate Hamilton–Jacobi–Bellman (HJB) equation. Due to the presence of a nonlinear partial derivative term, obtaining an analytical solution of the HJB equation is intractable. Thankfully, adaptive dynamic programming (ADP), initially introduced by Werbos (Werbos, 1992) as a self-learning optimization algorithm, integrates neural networks (NNs) to obtain an approximate solution of the HJB equation in a forward-in-time manner, which effectively addresses the "curse of dimensionality" in dynamic programming. In the early stage, scholars mainly developed classic ADP algorithms such as value iteration, policy iteration, Q-learning, etc., and provided proofs of their convergence and optimality (Liu, Wei, Wang, Yang, & Li, 2017; Liu, Xue, Zhao, Luo, & Wei, 2021; Luo, Huang, & Liu, 2021; Yanez & Souza, 2022; Zhao, Zhang, & Liu, 2023). With the continuous development of ADP, it is generally employed to address

various control problems, including OTC (Carolis & Saccon, 2020; Lu, Wei, Liu, Zhou, & Wang, 2022; Modares & Lewis, 2014; Mu, Zhang, Gao, & Sun, 2020; Tang, Luo, & Liao, 2023), robust stabilization (Gao, Jiang, & Davari, 2019; Yang, Guo, Xiong, Ding, Yin, & Wunsch, 2019; Yang & He, 2020), fault-tolerant control (Fan & Yang, 2016; Liu, Zhao, & Liu, 2020; Yang, Li, Xie, & Zhang, 2020; Zhang, Yuan, & Guo, 2021; Zhang, Zhao, Liu, & Zhang, 2022; Zhao, Liu, & Li, 2017; Zhao, Wang, Xu, Zong, & Zhao, 2023), and so on. In general, by employing the ADP algorithm and the NN approximator, an approximate optimal cost function can be obtained, thereby obtaining an approximate optimal control policy.

Regarding the OTC problem, the aforementioned ADP-based approaches can be categorized into two groups. The first approach designs a tracking controller that consists of a feedforward and a feedback component. The feedforward component ensures tracking performance, while the feedback component stabilizes the tracking error dynamics and optimizes system performance indicators. The second approach transforms the OTC problem into an optimal control problem by augmenting the original system with a command generator. Nevertheless, both methods possess certain limitations. In the first approach, the

design of the feedforward component explicitly demands accurate system functions and an invertible control input matrix. In the second approach, convergence is ensured by supplementing the cost function with an additional discount factor, but this factor can recursively impact the optimization metrics and the convergence of the ADP algorithm. To overcome above-mentioned difficulties, Li, Ding, Lewis, and Chai (2021) and Wang, Wang, Yang, and Yang (2023) addressed the OTC problem by designing novel quadratic performance index functions in regard to the tracking error and the tracking error dynamics. Nevertheless, existing results consider ideal scenario only. In the practical system, the existence of actuator fault is inevitable due to the control system works in hostile environment for a long time.

Over the past decades, a mass of researchers investigated the fault-tolerant problem by adopting ADP technique in a near optimal manner (Liu et al., 2020; Stojanovic, 2023; Zhao et al., 2017). Liu et al. (2020) addressed the fault-tolerant tracking control problem by developing an NN-based fault observer to compensate the control input online. Zhao et al. (2017) developed an ADP-based fault-tolerant control approach for nonlinear systems with actuator faults. Stojanovic (2023) combined ADP and fault compensation techniques to design an approximated optimal fault-tolerant control approach by adopting real-time input/output data. Among the aforementioned results, the majority of them employ an observer to obtain the fault information, which is then incorporated directly into the cost function or optimal control policy. Regrettably, the achievement of desired performance metrics can be notably arduous, and the use of such observational techniques can often amplify the conservatism of the controlling mechanism. It is worth mentioning that sliding mode control method has been regarded as an effective approach for handling faults or external disturbances, and has been extensively utilized in various types of control systems, including ordinary differential systems (Fan & Yang, 2016), partial differential systems (Zhang, Song, Song, & Stojanovic, 2023), and fractional-order systems (Mathiyalagan & Sangeetha, 2020). In recent years, many scholars have integrated integral sliding mode control (ISM) technique and ADP technique to achieve optimal fault-tolerant control. The main idea is to design a composite law, which contains an ISM law to eliminate the actuator fault and an optimal law to stablize the sliding mode dynamics/nominal systems (Fan & Yang, 2016; Tajrishi & Kalat, 2024; Yang et al., 2020). This method is less conservative, because there is no need to add fault information into the cost function such that the expected performance of control systems can be guaranteed as far as possible. Despite these efforts, it is noteworthy that the current ISM-based optimal fault-tolerant control approaches have only addressed bias faults and considered regulation problems exclusively. As such, the optimal fault-tolerant tracking (OFTT) control with more generalized fault behavior remains largely unexplored. Moreover, existing methods are time-triggered control methods, where controllers are updated periodically, leading to heavy computational and communication burdens. It is well-known that event-triggered control approach can economize computing and communication resources due to the control signal update at triggering moment only. Consequently, ADP-based event-triggered control schemes have been put forward to cope with the optimal regulation (Djordjevic, Tao, Song, He, Gao, & Stojanovic, 2023; Vamvoudakis, 2014; Yang et al., 2023), trajectory tracking (Lu et al., 2022; Peng, Yan, Huang, Cheng, Shi, & Ghosh, 2023), robust stabilization (Lin, Zhao, & Liu, 2024; Yang & He, 2020), differential game (Lin, Xue, Liu, Liang, & Wang, 2023; Wang, Hu, Zhao, & Qiao, 2023; Xue, Luo, & Liu, 2020) of nonlinear systems. For example, Djordjevic et al. (2023) developed an ADP-based event-triggered data driven controller for hydraulic servo actuators with unknown dynamics. Lu et al. (2022) addressed the event-triggered optimal parallel tracking control problem of discrete-time nonlinear systems. Lin et al. (2024) presented an event-triggered robust ADP algorithm to cope with the multiplayer stackelberg games of uncertain nonlinear systems. Xue et al. (2020) investigated the zero-sum game problem of partially unknown nonlinear systems by

developing an ADP-based event-triggered control scheme. Nevertheless, to our best knowledge, the OFTT control problem under event-triggered framework is rarely studied yet, which inspires our research.

To deal with the aforementioned limitations, an ISM-based event-triggered optimal fault-tolerant tracking (ETOFTT) control approach for continuous-time (CT) nonlinear systems is proposed. The primary contributions and novelties of this research are outlined below.

(1) This paper integrates ADP and ISM techniques to deal with the OFTT control problem for CT nonlinear systems subject to general actuator fault. On the one hand, with the help of ISM technique and a novel adaptive law, both the bias fault and the loss of effectiveness fault can be eliminated effectively. On the other hand, ADP-based OFTT control scheme completes the trajectory tracking assignment in an optimal manner.
(2) Different from the existing ADP-based OTC control approaches (Lu et al., 2022; Mu et al., 2020) which need to design a discount cost function or develop a tracking controller contains the feedforward part and the feedback part, this article designs a quadratic cost function in regard to the tracking error and its dynamics such that the inadequacies of existing methods can be improved. Moreover, unlike the existing approaches (Ha, Wang, & Liu, 2022; Li et al., 2021; Wang, Ren, & Ha, 2023; Wang, Wang, Yang, & Yang, 2023) which considered ideal fault-free systems only, this paper considers actuator fault such that the practicability of the control method is improved.

The subsequent section of this paper is arranged as follows. In Section 2, we introduce the problem statement. Section 3 is dedicated to the development of ISM control laws to address actuator fault. Additionally, we provide details on the design of an event-triggered optimal tracking controller, the NN implementation, and the analysis of the closed-loop system's stability. In Section 4, simulations are employed to substantiate the theoretical findings, and Section 5 encompasses the conclusions.

2. Problem statement

Consider a category of nonlinear systems provided by

$$\dot{S}(t) = \mathcal{I}_f\big(S(t)\big) + \mathcal{I}_g\big(S(t)\big)\mu_o(t), \tag{1}$$

where $S(t) \in \mathbb{R}^n$ is the system state, $\mu_o(t) \in \mathbb{R}^m$ is the output of the actuator, and $\mathcal{I}_f\big(S(t)\big) \in \mathbb{R}^n$ and $\mathcal{I}_g\big(S(t)\big) \in \mathbb{R}^{n\times m}$ are nonlinear system functions.

Assumption 1. The system functions $\mathcal{I}_f(S)$ and $\mathcal{I}_g(S)$ are Lipschitz continuous over a compact set Ω, and the controllability of system (1) holds within Ω.

For the OTC problem, it is imperative to develop an optimal controller to guarantee the system state $S(t)$ follows the predesigned trajectory $\pi(t)$ whose dynamics is given by

$$\dot{\pi}(t) = \psi_d\big(\pi(t)\big), \tag{2}$$

where $\psi_d(\cdot)$ is a continuously differentiable function with $\psi_d(0) = 0$.

In practice, actuator fault is inevitable, and a mathematical model illustrating the actuator fault is provided by

$$\mu_o(t) = \sigma\mu(t) + \mu_f(t), t > t_f \tag{3}$$

where $0 < \sigma < 1$ denotes the unknown loss of effectiveness fault, $\mu(t)$ represents the control input signal, $\mu_f(t)$ means the unknown bias fault, and t_f indicates the time when the fault occurred.

The primary objective of this paper is to introduce an ISM-based ETOFTT control method capable of ensuring the system state closely follows a pre-specified trajectory even in the presence of actuator

fault. To realize this objective, the ISM-based ETOFTT controller is formulated by

$$\mu(t) = \mu_c(t) + \hat{\theta}\mu_n(t), \tag{4}$$

where $\mu_n(t) \in \mathbb{R}^m$ is an event-triggered OTC law which used to guarantee the tracking performance of the nominal plant, $\mu_c(t) \in \mathbb{R}^m$ is a discontinuous ISM control law which adopted to deal with the actuator fault, $\theta = \frac{1}{\sigma}$ and $\hat{\theta}$ is the estimate of θ.

Assumption 2. The unknown bias fault $\mu_f(t)$ and the inverse of the loss of effectiveness fault θ are norm-bounded, i.e., $\|\mu_f(t)\| \le \bar{\mu}_f$ and $\|\theta\| \le \bar{\theta}$, where $\bar{\mu}_f$ and $\bar{\theta}$ are positive constants.

3. Integral sliding mode-based event-triggered optimal fault tolerant tracking control approach

3.1. ISM controller design

With the objective of acquiring the discontinuous control law $\mu_c(t)$, the integral sliding function is specified as

$$\mathcal{P}(S(t), t) = -\int_0^t \mathcal{M}(S(\tau))\Big(\mathcal{I}_f(S(\tau)) + \mathcal{I}_g(S(\tau))\mu_n\Big)d\tau + \mathcal{K}(S) - \mathcal{K}(S(0)), \tag{5}$$

where $\mathcal{K}(S) \in \mathbb{R}^m$ and $\mathcal{M}(S) = \frac{\partial \mathcal{K}(S)}{S} \in \mathbb{R}^{m \times n}$ are user-designed functions. The time derivative of (5) is calculated as

$$\dot{\mathcal{P}}(S, t) = \mathcal{P}(S)\dot{S} - \mathcal{P}(S)\big(\mathcal{I}_f(S) + \mathcal{I}_g(S)\mu_n\big). \tag{6}$$

Letting $\dot{\mathcal{P}}(S, t) = 0$, the equivalent control is calculated as

$$\mu_{ceq} = -\frac{1}{\sigma}\mu_n - \frac{1}{\sigma}\mu_f. \tag{7}$$

Substituting (7) into (1), the nominal system dynamic is

$$\dot{S} = \mathcal{I}_f(S) + \mathcal{I}_g(S)\mu_n. \tag{8}$$

Consider the equivalent control is unavailable, we turn to design the following ISM control law $\mu_c(t)$ to guarantee the system state stays on the sliding surface

$$\mu_c(t) = -\mathcal{A}\,\mathrm{sgn}\big(\mathcal{I}_g^{\mathsf{T}}(S)\mathcal{M}^{\mathsf{T}}(S)\mathcal{P}(S)\big), \tag{9}$$

where $\mathcal{A}$ is a positive constant, $\mathrm{sgn}(\mathbb{A}) = [\mathrm{sgn}(\mathbb{A}_1), \dots, \mathrm{sgn}(\mathbb{A}_n)]$ with $\mathbb{A} = [\mathbb{A}_1, \dots, \mathbb{A}_n]$ and $\mathrm{sgn}(\cdot)$ is a sign function.

Theorem 1. *For the nonlinear system subject to actuator fault (1), the integral sliding mode function designed by (5), and Assumptions 1 and 2, if the renovating law of $\hat{\theta}$ is formulated as*

$$\dot{\hat{\theta}} = \frac{1}{\gamma}\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_n, \tag{10}$$

where $\gamma > 0$ is an user-defined parameter, then the system state stays on the sliding surface using the discontinuous ISM control law (9).

Proof. We design a Lyapunov function candidate as

$$\mathcal{L}_{T1} = \frac{1}{2\sigma}\mathcal{P}^{\mathsf{T}}(S)\mathcal{P}(S) + \frac{\gamma}{2}\tilde{\theta}^{\mathsf{T}}\tilde{\theta}, \tag{11}$$

where $\tilde{\theta} = \theta - \hat{\theta}$. Calculating the time derivative of (11) and considering system dynamics (1) yields

$$\dot{\mathcal{L}}_{T1} = \frac{1}{\sigma}\mathcal{P}^{\mathsf{T}}(S)\Big(\mathcal{M}(S)\big(\mathcal{I}_f(S) + \mathcal{I}_g(S)(\sigma\mu + \mu_f)\big)$$
$$\quad - \mathcal{M}(S)\big(\mathcal{I}_f(S) + \mathcal{I}_g(S)\mu_n\big)\Big) - \gamma\tilde{\theta}^{\mathsf{T}}\dot{\hat{\theta}}$$
$$= \mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu + \frac{1}{\sigma}\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_f$$
$$\quad - \frac{1}{\sigma}\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_n - \gamma\tilde{\theta}^{\mathsf{T}}\dot{\hat{\theta}}$$
$$= \mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_c - \mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_n\tilde{\theta}$$

$$+ \theta\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_f - \gamma\tilde{\theta}^{\mathsf{T}}\dot{\hat{\theta}}$$
$$= -\mathcal{A}\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mathrm{sgn}\big(\mathcal{I}_g^{\mathsf{T}}(S)\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\big)$$
$$\quad - \tilde{\theta}^{\mathsf{T}}\big(\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_n + \gamma\dot{\hat{\theta}}\big)$$
$$\quad + \theta\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\mu_f. \tag{12}$$

Based on Assumption 2 and the adaptive update law (10), we further have

$$\dot{\mathcal{L}}_{T1} \le -\mathcal{A}\|\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\| + \bar{\theta}\|\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\|\bar{\mu}_f$$
$$\le -\|\mathcal{P}^{\mathsf{T}}(S)\mathcal{M}(S)\mathcal{I}_g(S)\|(\mathcal{A} - \bar{\theta}\bar{\mu}_f). \tag{13}$$

Therefore, if $\mathcal{A} > \bar{\theta}\bar{\mu}_f$ is held, then $\dot{\mathcal{L}}_{T1} \le 0$. This implies that the ISM control law (9) ensures the system state sustain on the sliding surface. This accomplishes the proof.

3.2. Event-triggered optimal tracking controller design

Next, an event-triggered optimal tracking controller is developed for nominal system (8) to ensure the system state pursues the predesigned trajectory. The tracking error is expressed as

$$v(t) = S(t) - \pi(t). \tag{14}$$

The cost function of the system (8) is defined as

$$\mathcal{J}(v) = \int_t^\infty C\big(v(v), \dot{v}(v)\big)dv, \tag{15}$$

where $C(v, \dot{v}) = v^{\mathsf{T}}Qv + \dot{v}^{\mathsf{T}}\mathcal{R}\dot{v}$ is the utility function, $Q \in \mathbb{R}^{n \times n}$ and $\mathcal{R} \in \mathbb{R}^{m \times m}$ are positive definite matrices. The Hamiltonian is described as

$$\mathcal{H}\big(v, \mu_n, \nabla\mathcal{J}(v)\big) = \nabla\mathcal{J}^{\mathsf{T}}(v)\big(\mathcal{I}_f(S) + \mathcal{I}_g(S)\mu_n - \psi_d(\pi)\big) + C(v, \dot{v}). \tag{16}$$

The optimal cost function satisfies

$$\mathcal{J}^*(v) = \min_{\mu_n \in \Re(\Omega)} \int_t^\infty C\big(v(v), \dot{v}(v)\big)dv, \tag{17}$$

where $\Re(\Omega)$ represents the admissible control sets. Thus, the OTC law is obtained by

$$\mu_n^*(v) = \big(\mathcal{I}_g^{\mathsf{T}}(S)\mathcal{R}\mathcal{I}_g(S)\big)^{-1}\Big(-\mathcal{I}_g^{\mathsf{T}}(S)\mathcal{R}\mathcal{I}_f(S) + \mathcal{I}_g^{\mathsf{T}}(S)\mathcal{R}\psi_d(\pi) - \frac{1}{2}\mathcal{I}_g^{\mathsf{T}}(S)\nabla\mathcal{J}(v)\Big). \tag{18}$$

Combining to (16) and (18), the HJB equation is provided as

$$0 = \nabla\mathcal{J}^{*\mathsf{T}}(v)\big(\mathcal{I}_f(S) + \mathcal{I}_g(S)\mu_n^* - \psi_d(\pi)\big) + C(v, \dot{v}). \tag{19}$$

Traditional ADP-based control approaches obtain the optimal cost function in an iterative manner. However, there consume abundant computing and communication resources since control laws update at each sampling moment. To conquer this problem, an event-triggered framework is established and the system state and the predesigned trajectory at the triggering moment are expressed as

$$\bar{S}_k(t) = S(P_k), P_k \le t < P_{k+1} \tag{20}$$

$$\bar{\pi}_k(t) = \pi(P_k), P_k \le t < P_{k+1} \tag{21}$$

where P_k the kth triggering instant. Hence, the homologous tracking error is determined by

$$\bar{v}_k(t) = \bar{S}_k(t) - \bar{\pi}_k(t), P_k \le t < P_{k+1} \tag{22}$$

The next triggering instant is calculated as

$$P_{k+1} = \inf\{t > P_k : \|\mathcal{B}_k(t)\|^2 \ge \alpha_1(\|v(t)\|^2)\}, \tag{23}$$

where $\mathcal{B}_k(t) = \bar{v}_k(t) - v(t)$ is the measurement error and α_1 is a $\mathcal{K}_\infty$ function. According to (18), (20), (21) and (22), the event-triggered OTC law is provided as

$$\mu_n^*(\bar{v}_k) = \left(\mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\mathcal{I}_g(\bar{S}_k)\right)^{-1}\left(-\mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\mathcal{I}_f(\bar{S}_k) + \mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\psi_d(\bar{\pi}_k)\right.$$
$$\left. - \frac{1}{2}\mathcal{I}_g^\top(\bar{S}_k)\nabla\mathcal{J}^*(\bar{v}_k)\right). \tag{24}$$

Assumption 3. The OTC law is Lipschitz continuous in regard to $\mathcal{B}_k(t)$, that is,

$$\left\|\mu_n^*(v(t)) - \mu^*(\bar{v}_k(t))\right\| \le L_\mu\|\mathcal{B}_k(t)\|, \tag{25}$$

where L_μ is a positive constant.

Assumption 4. The partial derivative of optimal cost function $\mathcal{J}^*(v)$ in regard to v satisfies,

$$\|\nabla\mathcal{J}^*(v)\| \le c_1\|v\|, \tag{26}$$

where c_1 is a positive constant.

Assumption 5. The system function $\mathcal{I}_g(S)$ is norm-bounded, i.e.,

$$0 < \|\mathcal{I}_g(S)\| \le \bar{\mathcal{I}}_g, \tag{27}$$

where $\bar{\mathcal{I}}_g$ is a positive constant.

Theorem 2. *Given the nominal nonlinear system* (8)*, the event-triggered OTC law provided by* (24)*, and Assumptions* 3–5*, if the triggering condition*

$$\|\mathcal{B}_k(t)\|^2 \le \frac{2(1-\rho_1^2)\lambda_{\min}(Q) - c_1^2\bar{\mathcal{I}}_g^2}{L_\mu^2}\|v\|^2 \tag{28}$$

and the following inequation

$$2(1-\rho_1^2)\lambda_{\min}(Q) > c_1^2\bar{\mathcal{I}}_g^2 \tag{29}$$

hold, where $0 < \rho_1 < 1$*, then the asymptotic stability of the tracking error is assured, indicating that the system state effectively follows the predetermined trajectory.*

Proof. The Lyapunov function candidate is formulated as

$$\mathcal{L}_{T2} = \mathcal{J}^*(v). \tag{30}$$

Based on (8), the time derivative of (30) is calculated by

$$\dot{\mathcal{L}}_{T2} = \nabla\mathcal{J}^{*\top}(v)\left(\mathcal{I}_f(S) + \mathcal{I}_g(S)\mu_n^*(\bar{v}_k) - \psi_d(\pi)\right). \tag{31}$$

According to (19) and Assumptions 3–5, it holds that

$$\dot{\mathcal{L}}_{T2} = -C(v,\dot{v}) + \nabla\mathcal{J}^{*\top}(v)\mathcal{I}_g(S)\left(\mu_n^*(\bar{v}_k) - \mu_n^*(v)\right)$$
$$\le -v^\top Q v + \frac{1}{2}\|\nabla\mathcal{J}^{*\top}(v)\mathcal{I}_g(S)\|^2 + \frac{1}{2}\|\mu_n^*(\bar{v}_k) - \mu_n^*(v)\|^2$$
$$\le -\rho_1^2\lambda_{\min}(Q)\|v\|^2 + (\rho_1^2-1)\lambda_{\min}(Q)\|v\|^2 + \frac{1}{2}L_\mu^2\|\mathcal{B}_k(t)\|^2 + \frac{c_1^2\bar{\mathcal{I}}_g^2}{2}\|v\|^2. \tag{32}$$

As a result, it can be inferred that $\dot{\mathcal{L}}_{T2} < 0$ when the triggering condition (28) is fulfilled, implying the asymptotic stability of the tracking error.

3.3. Neural network implementation

In this part, the critic NN is adopted to approximate the optimal cost function $\mathcal{J}^*(v)$, which is shown as

$$\mathcal{J}^*(v) = \varpi_c^{*\top}\zeta_c(v) + \epsilon_c(v), \tag{33}$$

where $\varpi_c^* \in \mathbb{R}^{h_c}$ is the optimal weight vector, $\zeta_c(v) \in \mathbb{R}^{h_c}$ is the activation function, h_c is the number of hidden layer neurons, and $\epsilon_c(v) \in \mathbb{R}$ is the approximation error. Then, we derive the partial derivative of $\mathcal{J}^*(v)$ with respect to e as

$$\nabla\mathcal{J}^*(v) = \nabla\zeta_c^\top(v)\varpi_c^* + \nabla\epsilon_c(v). \tag{34}$$

The estimated cost function is represented as

$$\hat{\mathcal{J}}(v) = \hat{\varpi}_c^\top\zeta_c(v), \tag{35}$$

where $\hat{\varpi}_c$ is the estimate of ϖ_c^*. Similarly, we can obtain

$$\nabla\hat{\mathcal{J}}(v) = \nabla\zeta_c^\top(v)\hat{\varpi}_c. \tag{36}$$

Consequently, the event-triggered OTC law is rewritten as

$$\mu_n^*(\bar{v}_k) = \left(\mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\mathcal{I}_g(\bar{S}_k)\right)^{-1}\left(\mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\psi_d(\bar{\pi}_k)\right.$$
$$\left. - \frac{1}{2}\mathcal{I}_g^\top(\bar{S}_k)\left(\nabla\zeta_c^\top(v)\varpi_c^* + \nabla\epsilon_c(v)\right) - \mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\mathcal{I}_f(\bar{S}_k)\right) \tag{37}$$

Hereafter, we can obtain the event-triggered approximate OTC law as

$$\hat{\mu}_n(\bar{v}_k) = \left(\mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\mathcal{I}_g(\bar{S}_k)\right)^{-1}\left(-\mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\mathcal{I}_f(\bar{S}_k) + \mathcal{I}_g^\top(\bar{S}_k)\mathcal{R}\psi_d(\bar{\pi}_k)\right.$$
$$\left. - \frac{1}{2}\mathcal{I}_g^\top(\bar{S}_k)\nabla\zeta_c^\top(v)\hat{\varpi}_c\right). \tag{38}$$

The approximate Hamiltonian is given as

$$\mathcal{H}\left(v,\hat{\mu}_n,\hat{\varpi}_c\right) = \hat{\varpi}_c^\top\nabla\zeta_c(v)\left(\mathcal{I}_f(S) + \mathcal{I}_g(S)\hat{\mu}_n - \psi_d(\pi)\right) + C(v,\dot{v}) \triangleq e_c. \tag{39}$$

By utilizing the gradient descent method, the critic NN weight is updated as

$$\dot{\hat{\varpi}}_c = -\alpha_c\frac{1}{(1+\Delta^\top\Delta)^2}\left(\frac{\partial E_c}{\partial\hat{\varpi}_c}\right)$$
$$= -\frac{\alpha_c\Delta}{(1+\Delta^\top\Delta)^2}\left(\hat{\varpi}_c^\top\Delta + C(v,\dot{v})\right), \tag{40}$$

where $\alpha_c > 0$ is the learning rate and $\Delta = \nabla\zeta_c(v)\left(\mathcal{I}_f(S) + \mathcal{I}_g(S)\hat{\mu}_n\right)$.

Lemma 1. *Given the nominal nonlinear system* (8)*, the critic NN weight tuning rule* (40) *ensures that the estimation error of the critic NN weights* $\tilde{\varpi}_c$ *will be uniformly ultimately bounded.*

Proof. The detailed proof of Lemma 1 is available in Chen, Chen, Chen, and Zhang (2022), Vamvoudakis (2014), Xue et al. (2020), and thus is skipped here.

3.4. Stability analysis

Assumption 6. $\tilde{\varpi}_c$, ϖ_c^*, $\nabla\zeta_c(v)$, and $\nabla\epsilon_c(v)$ satisfy

$$\|\tilde{\varpi}_c\| \le \bar{\varpi}_c, \ \|\varpi_c^*\| \le \bar{\varpi}_{cm}, \ \|\nabla\zeta_c(v)\| \le \bar{\xi}_c, \ \|\nabla\epsilon_c(v)\| \le \bar{\epsilon}_c$$

where $\bar{\varpi}_c$, $\bar{\varpi}_{cm}$, $\bar{\varpi}_c$, and $\bar{\epsilon}_c$ are positive constants.

Theorem 3. *For the nominal nonlinear system* (8)*, the event-triggered approximate OTC law provided by* (38)*, and Assumptions* 3–6*, if the triggering condition satisfies*

$$\|\mathcal{B}_k(t)\|^2 \le \frac{(1-\rho_2^2)\|v\|^2}{L_\mu^2}, \tag{41}$$

where $0 < \rho_2 < 1$*, then the tracking error is insured to be uniform ultimate boundedness.*

Proof. The entire process of proving is split into two components, and the construction of the Lyapunov function candidate is as follows.

$$\mathcal{L}_{T3} = \mathcal{J}^*(v) + \mathcal{J}^*(\bar{v}_k). \tag{42}$$

Part 1: The event is not triggered, i.e., $t \in [P_k, P_{k+1})$.
By calculating the time derivative of (42), one can obtain

$$\dot{\mathcal{L}}_{T3} = \nabla\mathcal{J}^{*\top}(v)\left(\mathcal{I}_f(S) + \mathcal{I}_g(S)\hat{\mu}_n(\bar{v}_k) - \psi_d(\pi)\right)$$
$$= -C(v,\dot{v}) + \nabla\mathcal{J}^{*\top}(v)\mathcal{I}_g(S)\left(\hat{\mu}_n(\bar{v}_k) - \mu_n^*(v)\right)$$
$$\le -v^\top Q v + \frac{1}{2}\|\nabla\mathcal{J}^{*\top}(v)\mathcal{I}_g(S)\|^2 + \frac{1}{2}\|\hat{\mu}_n(\bar{v}_k) - \mu_n^*(v)\|^2. \tag{43}$$

According to Assumption 3, (37) and (38), the last part of (43) can be derived as

$$\frac{1}{2}\|\hat{\mu}_n(\bar{v}_k) - \mu_n^*(v)\|^2 \le \|\hat{\mu}_n(\bar{v}_k) - \mu_n^*(\bar{v}_k)\|^2 + \|\mu_n^*(\bar{v}_k) - \mu_n^*(v)\|^2$$

$$\le L_\mu^2 \|\mathcal{B}_k(t)\|^2 + \|\hat{\mu}_n(\bar{v}_k) - \mu_n^*(\bar{v}_k)\|^2$$

$$\le \left\| -\frac{1}{2}\left(\mathcal{I}_g^{\mathsf{T}}(\bar{S}_k)\mathcal{R}\mathcal{I}_g(\bar{S}_k)\right)^{-1}\mathcal{I}_g^{\mathsf{T}}(\bar{S}_k)\nabla\zeta_c^{\mathsf{T}}(v)\tilde{\varpi}_c \right.$$

$$\left. -\frac{1}{2}\left(g^{\mathsf{T}}(\bar{S}_k)\mathcal{R}\mathcal{I}_g(\bar{S}_k)\right)^{-1}\mathcal{I}_g^{\mathsf{T}}(\bar{S}_k)\nabla\epsilon_c(\bar{v}_k) \right\|^2$$

$$+ L_\mu^2 \|\mathcal{B}_k(t)\|^2. \tag{44}$$

Based on Assumption 5, we assume that $\left(\mathcal{I}_g^{\mathsf{T}}(\bar{S}_k)\mathcal{R}\mathcal{I}_g(\bar{S}_k)\right)^{-1}$ satisfies $\left\|\left(\mathcal{I}_g^{\mathsf{T}}(\bar{S}_k)\mathcal{R}\mathcal{I}_g(\bar{S}_k)\right)^{-1}\right\| \le \lambda_1$, where λ_1 is a positive constant. Therefore, we further have

$$\frac{1}{2}\|\hat{\mu}_n(\bar{v}_k) - \mu_n^*(v)\|^2 \le \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\zeta}_c^2\bar{\varpi}_c^2 + \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\epsilon}_c^2 + L_\mu^2\|\mathcal{B}_k(t)\|^2. \tag{45}$$

Substituting (45) into (43), we can obtain

$$\dot{\mathcal{L}}_{T3} \le -v^{\mathsf{T}}Qv + \frac{1}{2}\|\nabla\mathcal{J}^{*\mathsf{T}}(v)g(S)\|^2 + \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\zeta}_c^2\bar{\varpi}_c^2$$

$$+ \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\epsilon}_c^2 + L_\mu^2\|\mathcal{B}_k(t)\|^2$$

$$\le -\rho_2^2\lambda_{\min}(Q)\|v\|^2 + (\rho_2^2 - 1)\lambda_{\min}(Q)\|v\|^2 + L_\mu^2\|\mathcal{B}_k(t)\|^2$$

$$+ \frac{1}{2}\bar{I}_g^2\bar{\zeta}_c^2\bar{\varpi}_c^2 + \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\zeta}_c^2\bar{\varpi}_c^2 + \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\epsilon}_c^2.$$

Letting $\Theta_1 = \frac{1}{2}\bar{I}_g^2\bar{\zeta}_c^2\bar{\varpi}_c^2 + \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\zeta}_c^2\bar{\varpi}_c^2 + \frac{1}{2}\lambda_1^2\bar{I}_g^2\bar{\epsilon}_c^2$. Therefore, if the system state v is located outside the following compact set, $\dot{\mathcal{L}}_{T4}$ will be negative.

$$\Omega_v = \left\{ v : \|v\| \le \sqrt{\frac{\Theta_1}{(1 - \rho_2^2)\lambda_{\min}(Q)}} \right\}. \tag{46}$$

Part 2: The event is triggered, i.e., $\forall t = P_{k+1}$. According to (42), one can get

$$\Delta\mathcal{L}_{T3}(t) = \Delta\mathcal{L}_{T3,1}(t) + \Delta\mathcal{L}_{T3,2}(t). \tag{47}$$

Based on the result in *Case 1*, we can conclude that $\dot{\mathcal{L}}_{T3} < 0$ for every $t \in [P_k, P_{k+1})$. Therefore, we further get

$$\Delta\mathcal{L}_{T3,1}(t) = \mathcal{J}^*(\bar{v}_{k+1}) - \mathcal{J}^*(v(P_{k+1}^-)) \le 0,$$

$$\Delta\mathcal{L}_{T3,2}(t) = \mathcal{J}^*(\bar{v}_{k+1}) - \mathcal{J}^*(\bar{v}_k) \le -\iota(\|\mathcal{B}_{k+1}(P_k)\|),$$

where $\iota(\cdot)$ is a class-$\mathcal{K}$ function, and $\mathcal{E}_{k+1}(P_k) = \bar{v}_{k+1} - v_k$. Based on the aforementioned analysis, it is evident that $\dot{\mathcal{L}}_{T3} < 0$ holds at the triggering instants. This completes the proof.

Remark 1. This paper develops an ISM-based ETOFTT control method for CT nonlinear systems with general actuator fault. The advantages of the proposed control scheme are outlined as follows. (1) Different from traditional ADP-based OTC control approaches (Lu et al., 2022; Mu et al., 2020) which need to design a discount cost function, this paper develops a novel cost function without discount factor. As a result, the problem of the discount factor affecting the system stability is avoided. Therefore, the controller design process is simplified and the practicability of the control method is improved. In addition, unlike existing results (Ha et al., 2022; Li et al., 2021; Liu et al., 2020) that developed time-triggered OTC methods, the proposed OTC law is updated only at triggering moments. Hence, the computational and communication burden is reduced. (2) Unlike existing optimal fault tolerant control methods (Liu et al., 2020; Zhao et al., 2017) addressed bias faults only, this paper considers general actuator faults which contains the loss of effectiveness fault and the bias fault. Therefore, the developed ISM-based ETOFTT control method is more suitable in practical scenarios. Moreover, by designing the ISM control law and the adaptive updating law, the effects of both the loss of effectiveness fault and the bias fault can be eliminated.
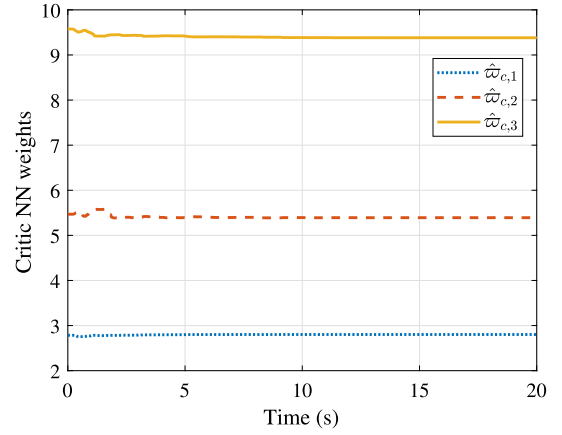

Fig. 1. Critic NN weights.

Remark 2. In this paper, we introduce a novel cost function to characterize the system's performance index. Traditional cost functions are typically quadratic in terms of system states and control inputs. However, for tracking control problems, control inputs do not approach zero. In order to prevent the cost function from diverging, a discount factor is usually incorporated. Nonetheless, the selection of the discount factor will impact the convergence of the ADP algorithm and even the stability of the closed-loop system. This paper proposes a novel cost function which contains tracking errors and their derivatives and the need of the discount factor is avoided. Therefore, the controller design process is simplified and the practicability of the control method is improved.

4. Simulation

4.1. Example 1

In this part, we employ a robotic arm as a means of showcasing the applicability of the ISM-based ETOFTT control approach. The dynamics of the robotic arm are depicted as

$$\dot{S}_1 = S_2$$
$$\dot{S}_2 = -4.905\sin(S_1) - 0.2S_2 + 0.1\mu_o \tag{48}$$

Due to actuator fault, the actual control input become

$$\mu_o(t) = \sigma\mu(t) + \mu_f(t), \tag{49}$$

where $\sigma = 0.5$ and

$$\mu_f(t) = \begin{cases} \sin^2(t), & t < 5 \\ 0, & t \ge 5 \end{cases}$$

The dynamics of the predesigned trajectory is chosen as

$$\dot{\pi}(t) = \begin{bmatrix} 3\cos(3t) \\ -9\sin(3t) \end{bmatrix}. \tag{50}$$

Initially, an ISM controller is proposed with the purpose of mitigating the effects caused by actuator fault. Let the user-defined functions in (5) be $\mathcal{K}(S) = S_2$ and $\mathcal{M}(S) = [0, 1]$, the positive constants in (9) and (10) be $\mathcal{A} = 30$ and $\gamma = 0.01$. Under the ISM control law (9), the actuator fault can be compensated and the nominal robotic arm system is obtained. Next, the event-triggered optimal tracking controller is developed for the nominal robotic arm system to assure the system state tracks the predesigned trajectory. The parameters in (15) are set as $Q = I_2$ and $\mathcal{R} = I$, the activation function of the critic NN is designed as $\zeta_c(v) = [v_1^2, v_1v_2, v_2^2]^{\mathsf{T}}$, and the parameters in (41) are picked as $\rho_2 = 0.1$ and $L_\mu = 10$.
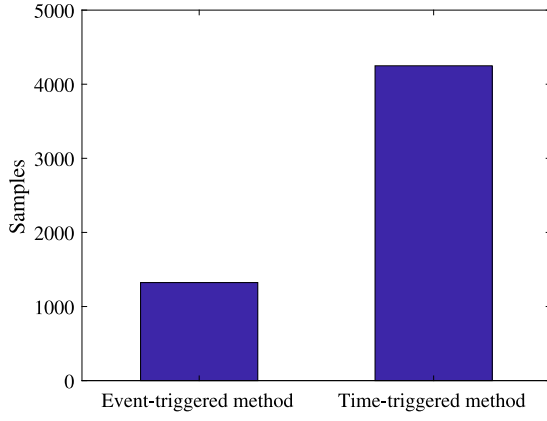
Fig. 2. Controller update time.
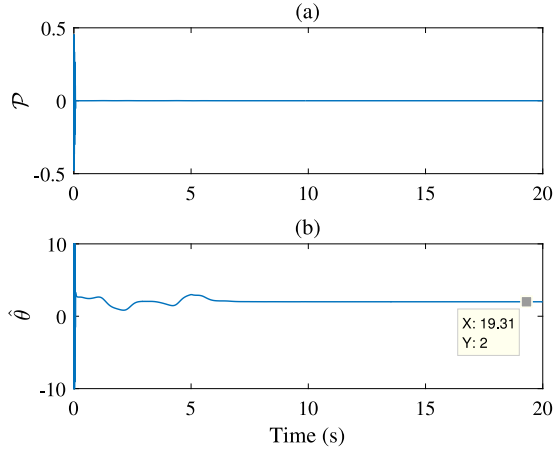

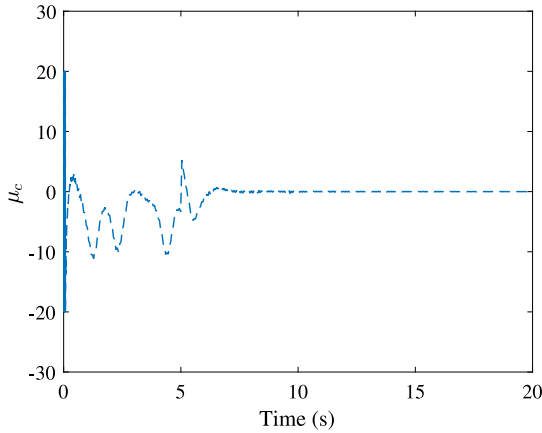**Fig. 3.** (a) Sliding mode function. (b) Adaptive term.

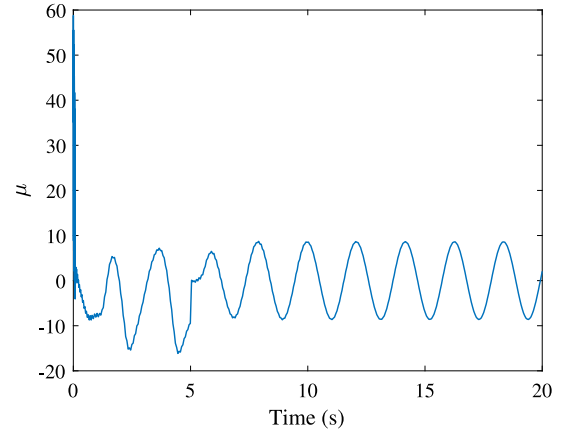**Fig. 4.** ISM control law.

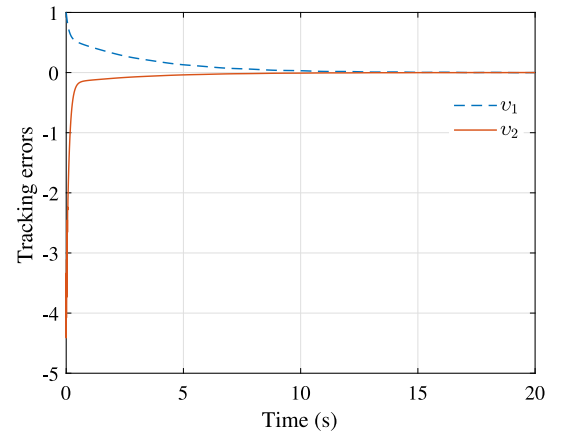

Fig. 5. ISM-based ETOFTT control law.
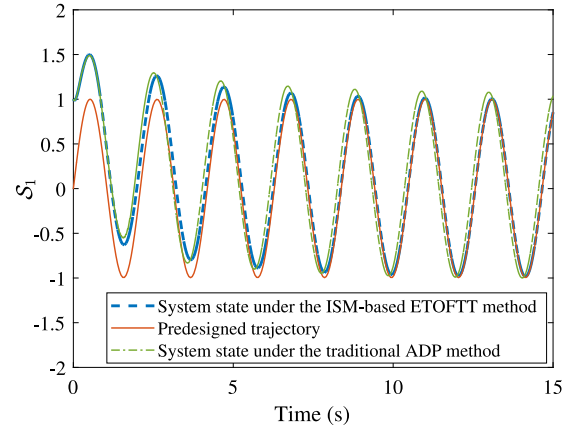

Fig. 6. Trajectories errors.

Fig. 7. Trajectories tracking.

The simulation validation outcomes are presented in Figs. 1–5. Fig. 1 depicts the weight change curves, showing that the critic NN weight vector arrives to $\hat{\varpi}_c = [2.80, 5.39, 9.38]^\mathsf{T}$. Fig. 2 presents a comparison of the amount of updates required by the event-triggered controller and the time-triggered controller. The results indicate that the event-triggered controller necessitates only 1323 updates, while the time-triggered controller in Zhao et al. (2017) demands 4249 updates, underscoring the former's superiority in terms of computational and communication resource conservation. In Fig. 3(a), the progression of

the sliding mode function is shown, revealing that the sliding mode function remains in a tiny zone of zero, indicating that the robotic arm system state remains on the sliding mode surface. Fig. 3(b) displays the evolution of the estimate of the loss of effectiveness fault. The curve of the ISM control law is illustrated in Fig. 4, demonstrating the capability of the proposed ISM controller to offer responses in dealing with actuator fault. Fig. 5 displays the changing curve of the ISM-based ETOFTT control law. The tracking error curves is provided in Fig. 6, where we can conclude that the system state catches up with
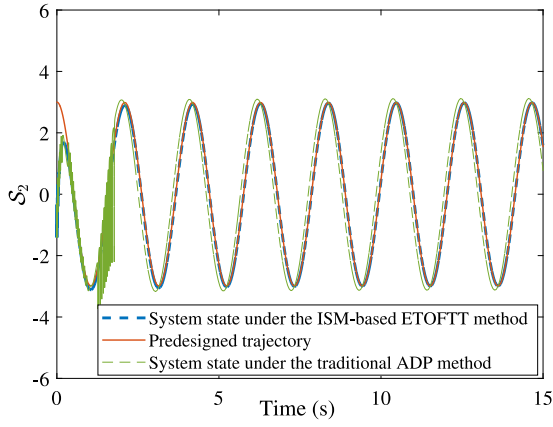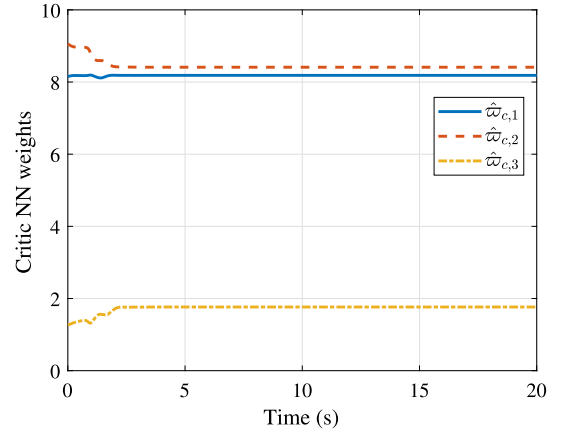
Fig. 8. Trajectories tracking.

Fig. 9. Cost function.
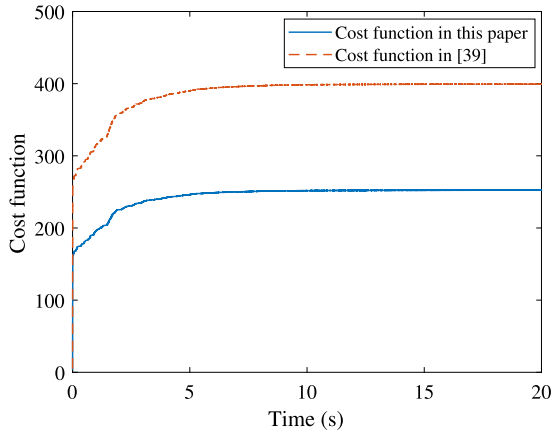

Fig. 10. Critic NN weights.

Fig. 11. Controller update time.

Fig. 12. (a) Sliding mode function. (b) Adaptive term.
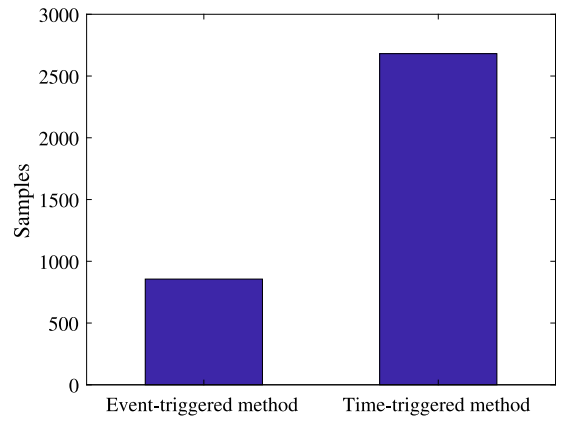
the predesigned trajectory within 10 s In Figs. 7 and 8, we compared the tracking performance between the ISM-based ETOFTT control approach and the traditional ADP method. It is clear that the ISM-based ETOFTT control approach ensures the system state to catch up with the desired trajectory, whereas the traditional ADP method fails to track the desired trajectory due to actuator faults. Fig. 9 compares the cost functions between the ISM-based ETOFTT control method and the traditional method in Pan, Yang, Pan, and Yu (2018), indicating that the convergence value of the cost function in the developed method is smaller than the traditional one. This implies that the ISM-based ETOFTT control method incurs lower control costs and the proposed control method exhibits optimized performance. On the whole, the ISM-based ETOFTT controller guarantees the tracking performance of the robotic arm subject to actuator fault and the developed ISM-based ETOFTT control scheme is effective.

4.2. Example 2

Next, we will further confirm the efficacy of the ISM-based ETOFTT control method by employing a Van der Pol circuit system with the following dynamics

$$\dot{S}_1 = S_2$$
$$\dot{S}_2 = -2S_1 + 3(1 - S_1^2)S_2 + \mu_o \tag{51}$$

As a consequence of the actuator fault, the genuine control input is formulated as

$$\mu_o(t) = \sigma\mu(t) + \mu_f(t), \tag{52}$$

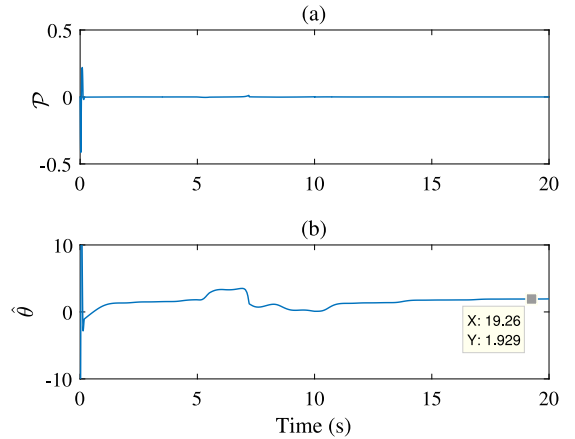where $\sigma = 0.5$ and

$$\mu_f(t) = \begin{cases} 0, & t < 5 \\ 10\sin(t)\cos(t), & 5 \leq t < 10 \\ 0, & t \geq 10 \end{cases}$$

The dynamics of the predetermined trajectory is selected as

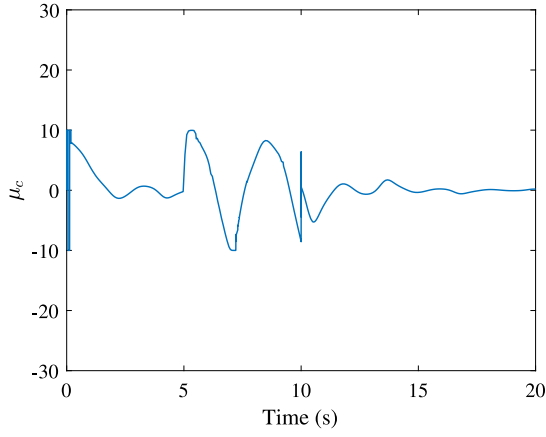$$\dot{\pi}(t) = \begin{bmatrix} -0.5\sin(t) + 0.6\cos(3t) \\ -0.5\cos(t) - 1.8\sin(3t) \end{bmatrix}. \tag{53}$$
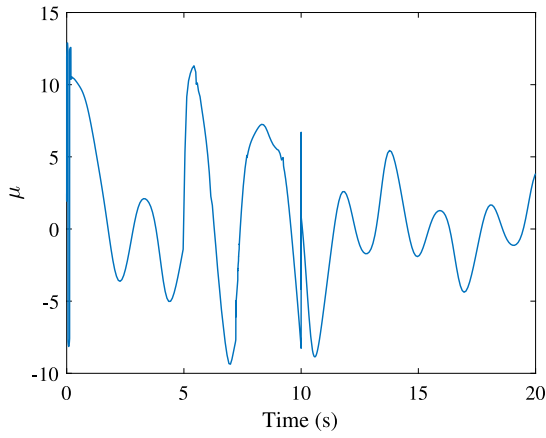
Fig. 13. ISM control law.

Fig. 16. Trajectories tracking.

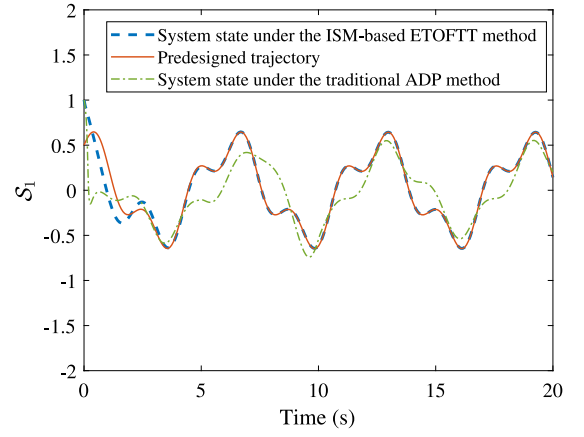**Fig. 14.** ISM-based ETOFTT control law.

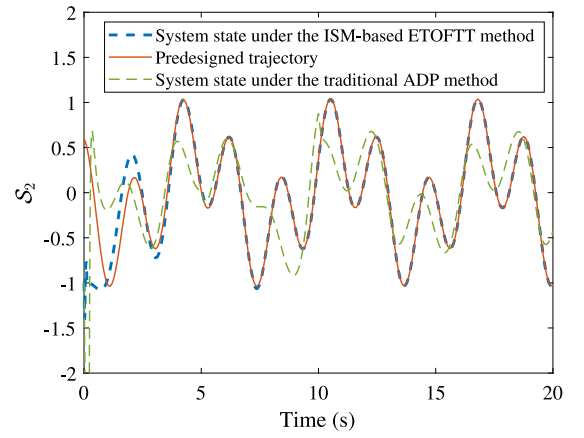

Fig. 17. Trajectories tracking.
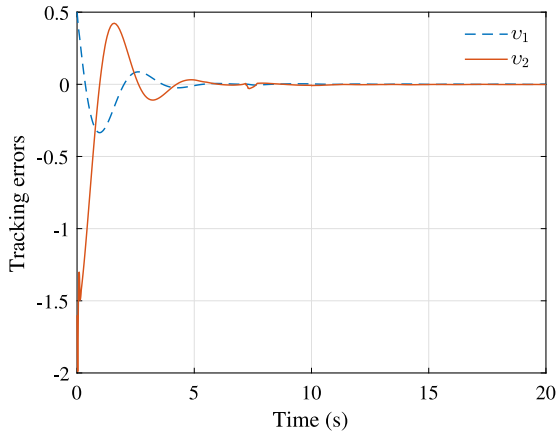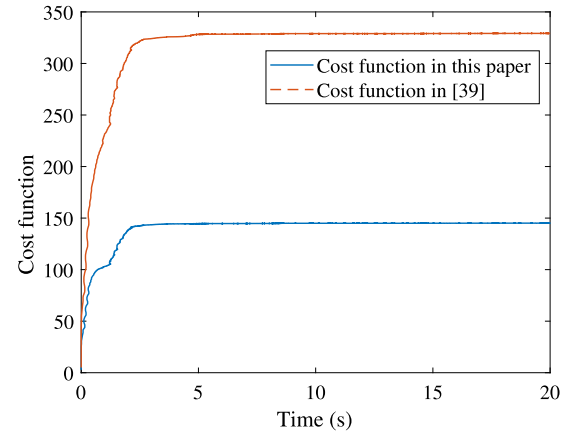

Fig. 15. Trajectories errors.

Fig. 18. Cost function.

Firstly, an ISM controller is developed to mitigate the impact of actuator fault. The user-defined functions in (5) are selected as $\mathcal{K}(S) = S_2$ and $\mathcal{M}(S) = [0, 1]$, respectively. The positive constants in (9) and (10) are assigned as $\mathcal{A} = 50$ and $\gamma = 0.001$. By employing the ISM control law (9), it becomes feasible to effectively compensate the actuator fault, thereby leading to the acquisition of the nominal Van der Pol circuits system. Furthermore, an event-triggered optimal tracking controller is established for the nominal Van der Pol circuits system to assure that the system state accurately catch up the predetermined

trajectory. The parameters in (15) are picked as $\mathcal{Q} = I_2$ and $\mathcal{R} = I$. The activation function of the critic NN is designed as $\zeta_c(v) = [v_1^2, v_1 v_2, v_2^2]^{\mathsf{T}}$. Additionally, the parameters in (41) are selected as $\rho_2 = 0.1$ and $L_\mu = 10$.

Fig. 10 illustrates the weight change curves, demonstrating the convergence of the critic NN weight vector to $\hat{\varpi}_c = [8.17, 8.58, 1.62]^{\mathsf{T}}$. In Fig. 11, a comparison is shown between the number of updates required by the event-triggered controller and the time-triggered controller. The results demonstrate that the event-triggered controller only

requires 856 updates, whereas the time-triggered controller requires 2681 updates. This highlights the superiority of the event-triggered controller in terms of conserving computational and communication resources. Fig. 12(a) illustrates the progression of the sliding mode function, demonstrating that it consistently resides within a narrow region surrounding zero. This indicates that the system state of the Van der Pol circuits system remains confined to the sliding mode surface. Fig. 12(b) showcases the development of the estimated loss of effectiveness fault. Fig. 13 illustrates the curve of the ISM control law. It can be observed that when the fault occurs, the ISM control law will provide corresponding compensatory responses to counteract the impact of the actuator fault. Fig. 14 depicts the overall ISM-based ETOFTT control law curve. Fig. 15 provides the tracking error curves, from which it can be observe that the tracking error converges after 10 s. This indicates that the system state successfully catches up with the predetermined trajectory. Figs. 16 and 17 depict the system state and desired trajectory curves under the developed ISM-based ETOFTT control method and the traditional ADP method. It is evident that the proposed ISM-based ETOFTT control method can ensure that the system state catches up with the ideal trajectory in the presence of actuator faults. In contrast, the traditional ADP method is unable to track the desired trajectory due to the impact of actuator faults. Fig. 18 illustrates that the control cost of the ISM-based ETOFTT control method is smaller than that of the traditional one, which means that this approach can achieve fault-tolerant tracking control in an efficient and energy-saving way. In summary, the ISM-based ETOFTT controller ensures the tracking performance of the Van der Pol circuits system even with the actuator fault, thus confirming the efficacy of the proposed ISM-based ETOFTT control approach.

Remark 3. In this paper, we evaluate the effectiveness of the proposed ISM-based ETOFTT control method through the tracking error, the number of controller update, and the convergence value of the cost function. The details are illustrated as follows. (1) In simulation, we adopt tracking error to evaluate the trajectory tracking performance. As shown in Figs. 6 and 15, the tracking error converges to a small region around zero, indicating that the system state closely follows the desired trajectory. This means that the proposed control method can ensure that the system state tracks the desired trajectory. (2) In order to highlight the advantages of event-triggered mechanism, we compared the number of controller update time between the time-triggered control method and the event-triggered control method. It can be observed from Figs. 2 and 11 that the event-triggered control method requires fewer controller updates compared to the time-triggered one. As each controller update process consumes computational and communication resources, the event-triggered control method can alleviate the computational and communication burden of the closed-loop system. (3) To demonstrate the optimization of the developed control methods, we compared the convergence values of the cost functions between the ISM-based ETOFTT control method and traditional control method in Pan et al. (2018). It can be observed from Figs. 9 and 18 that the convergence value of the cost function in the ISM-based ETOFTT control method is smaller than the traditional one. Generally, a smaller convergence value of the cost function indicates lower control costs and lower energy consumption for the closed-loop system. Therefore, the simulation results indicate that the ISM-based ETOFTT control method can reduce control costs and achieve fault-tolerant tracking control in an efficient and energy-saving way.

Remark 4. In fact, the parameter $\mathcal{A}$ in the ISM control law, the parameters $\mathcal{Q}$ and $\mathcal{R}$ in the cost function, and the parameter L_μ in the triggering condition will affect the control performance. For example, (1) In the ISM control law, if $\mathcal{A}$ is too small and does not satisfy $\mathcal{A} > \bar{\theta}\bar{\mu}_f$, the designed ISM control law cannot guarantee that the system state remains on the sliding surface, thus the impact of actuator

faults cannot be eliminated effectively. If $\mathcal{A}$ is too large, it may cause the closed-loop system oscillations or even instability. (2) In the cost function, the $\mathcal{Q}$ value is adopted to limit tracking errors and ensure the speed of system response. The $\mathcal{R}$ value is used to limit the amplitude and the smoothness of the control input, ensuring the stability of the closed-loop system. Additionally, it affect the energy consumption during the control process. (3) The parameter L_μ in the triggering condition affects the controller update frequency. If L_μ is too large, the triggering threshold will be small, leading to a higher controller update frequency, which brings a large amount of computation. However, if L_μ is too small, the controller update frequency will decrease but the system state may not track the desired trajectory. In general, the above-mentioned parameters will affect the control performance of the developed ISM-based ETOFTT control method. However, there is currently no unified method to select the values of these parameters. Researchers usually obtain suitable parameter values based on experience and continuous tuning. In this paper, we select these parameter values by "trial and error" with repetitive simulations.

5. Conclusion

In this article, an ISM-based ETOFTT approach is presented for CT nonlinear systems subject to actuator fault. At first, an ISM control law is designed to cope with actuator fault and acquire the nominal nonlinear system plant. Subsequently, a novel cost function is developed to satisfactorily tackle the OTC problem. To acquire an approximate solution for the tracking HJB equation, a critic-only framework is utilized. Moreover, the developed control law is updated aperiodically to conserve computing and communication resources. Theoretical analysis demonstrates that the ISM-based ETOFTT controller assures asymptotic stability of the tracking error. Finally, simulation outcomes certify the validity of the presented ISM-based ETOFTT scheme. In practice, multi-agent systems are widely encountered, such as drone swarms and autonomous driving systems. Since agents operate in complex environments, the occurrence of actuator faults is inevitable. However, compared with single-agent systems, the fault-tolerant control of multi-agent systems is more intricate due to the fact that the fault of each agent can propagate to affect other agents through the communication network. Therefore, in future work, we will further integrate ISM and ADP techniques to address the fault-tolerant control problem of multi-agent systems.

CRediT authorship contribution statement

Yongwei Zhang: Methodology, Writing – original draft, Writing – review & editing. **Shunchao Zhang:** Conceptualization, Investigation, Writing – review & editing.

Declaration of competing interest

The authors have no competing interests to declare that are relevant to the content of this article.

References

Carolis, C., & Saccon, A. (2020). On linear quadratic optimal control for time-varying multimodal linear systems with time-triggered jumps. *IEEE Control Systems Letters, 4*(1), 217–222.

Chen, Z., Chen, K., Chen, S., & Zhang, Y. (2022). Event-triggered consensus for uncertain nonlinear systems using integral sliding mode based adaptive dynamic programming. *Neural Networks, 156*, 258–270.

Djordjevic, V., Tao, H., Song, X., He, S., Gao, W., & Stojanovic, V. (2023). Data-driven control of hydraulic servo actuator: An event-triggered adaptive dynamic programming approach. *Mathematical Biosciences and Engineering, 20*(5), 8561–8582.

Fan, Q., & Yang, G. (2016). Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances. *IEEE Transactions on Neural Networks and Learning Systems, 27*(1), 165–177.

Gao, W., Jiang, Y., & Davari, M. (2019). Data-driven cooperative output regulation of multi-agent systems via robust adaptive dynamic programming. *IEEE Transactions on Circuits and Systems II: Express Briefs, 66*(3), 447–451.

Ha, M., Wang, D., & Liu, D. (2022). Discounted iterative adaptive critic designs with novel stability analysis for tracking control. *IEEE/CAA Journal of Automatica Sinica, 9*(7), 1262–1272.

Li, C., Ding, J., Lewis, F., & Chai, T. (2021). A novel adaptive dynamic programming based on tracking error for nonlinear discrete-time systems. *Automatica, 129,* Article 109687.

Lin, D., Xue, S., Liu, D., Liang, M., & Wang, Y. (2023). Adaptive dynamic programming-based hierarchical decision-making of non-affine systems. *Neural Networks, 167,* 331–341.

Lin, M., Zhao, B., & Liu, D. (2024). Event-triggered robust adaptive dynamic programming for multiplayer stackelberg-nash games of uncertain nonlinear systems. *IEEE Transactions on Cybernetics, 54*(1), 273–286.

Liu, D., Wei, Q., Wang, D., Yang, X., & Li, H. (2017). *Adaptive dynamic programming with applications in optimal control.* Cham, Switzerland: Springer.

Liu, D., Xue, S., Zhao, B., Luo, B., & Wei, Q. (2021). Adaptive dynamic programming for control: a survey and recent advances. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 51*(1), 142–160.

Liu, X., Zhao, B., & Liu, D. (2020). Fault tolerant tracking control for nonlinear systems with actuator failures through particle swarm optimization-based adaptive dynamic programming. *Applied Soft Computing, 97,* Article 106766.

Lu, J., Wei, Q., Liu, Y., Zhou, T., & Wang, F. (2022). Event-triggered optimal parallel tracking control for discrete-time nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 52*(6), 3772–3784.

Luo, B., Huang, T., & Liu, D. (2021). Periodic event-triggered suboptimal control with sampling period and performance analysis. *IEEE Transactions on Cybernetics, 51*(3), 1253–1261.

Mathiyalagan, K., & Sangeetha, G. (2020). Second-order sliding mode control for nonlinear fractional-order systems. *Applied Mathematics and Computation, 383*(15), Article 125264.

Modares, H., & Lewis, F. (2014). Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. *Automatica, 50*(7), 1780–1792.

Mu, C., Zhang, Y., Gao, Z., & Sun, C. (2020). ADP-based robust tracking control for a class of nonlinear systems with unmatched uncertainties. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50*(11), 4056–4067.

Pan, Y., Yang, C., Pan, L., & Yu, H. (2018). Integral sliding mode control: performance, modification, and improvement. *IEEE Transactions on Industrial Informatics, 14,* 3087–3096.

Peng, Z., Yan, W., Huang, R., Cheng, H., Shi, K., & Ghosh, B. (2023). Event-triggered learning robust tracking control of robotic systems with unknown uncertainties. *IEEE Transactions on Circuits and Systems II: Express Briefs, 70*(7), 2540–2544.

Stojanovic, V. (2023). Fault-tolerant control of a hydraulic servo actuator via adaptive dynamic programming. *Mathematical Modelling and Control, 3*(3), 181–191.

Tajrishi, M., & Kalat, A. (2024). Fast finite time fractional-order robust-adaptive sliding mode control of nonlinear systems with unknown dynamics. *Journal of Computational and Applied Mathematics, 438,* Article 115554.

Tang, R., Luo, B., & Liao, Y. (2023). Adaptive dynamic programming based composite control for profile tracking with multiple constraints. *Neurocomputing, 557,* Article 126711.

Vamvoudakis, K. (2014). Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems. *IEEE/CAA Journal of Automatica Sinica, 1*(3), 282–293.

Wang, D., Hu, L., Zhao, M., & Qiao, J. (2023). Dual event-triggered constrained control through adaptive critic for discrete-time zero-sum games. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 53*(3), 1584–1595.

Wang, D., Ren, J., & Ha, M. (2023). Discounted linear Q-learning control with novel tracking cost and its stability. *Information Sciences, 626,* 339–353.

Wang, T., Wang, Y., Yang, X., & Yang, J. (2023). Further results on optimal tracking control for nonlinear systems with nonzero equilibrium via adaptive dynamic programming. *IEEE Transactions on Neural Networks and Learning Systems, 34*(4), 1900–1910.

Werbos, P. (1992). Approximate dynamic programming for real-time control and neural modeling. In D. A. White, & D. A. Sofge (Eds.), *Handbook of intelligent control: neural, fuzzy, and adaptive approaches.* New York, NY, USA: Van Nostrand Reinhold, ch. 13.

Xue, S., Luo, B., & Liu, D. (2020). Event-triggered adaptive dynamic programming for zero-sum game of partially unknown continuous-time nonlinear systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50*(9), 3189–3199.

Yanez, W., & Souza, F. (2022). On the effect of probing noise in optimal control LQR via Q-learning using adaptive filtering algorithms. *European Journal of Control, 65,* Article 100633.

Yang, Y., Fan, X., Gao, W., Yue, W., Liu, A., Geng, S., et al. (2023). Event-triggered output feedback control for a class of nonlinear systems via disturbance observer and adaptive dynamic programming. *IEEE Transactions on Fuzzy Systems, 31*(9), 3148–3160.

Yang, Y., Guo, Z., Xiong, H., Ding, D., Yin, Y., & Wunsch, D. (2019). Data-driven robust control of discrete-time uncertain linear systems via off-policy reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems, 30*(12), 3735–3747.

Yang, X., & He, H. (2020). Event-triggered robust stabilization of nonlinear input-constrained systems using single network adaptive critic designs. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50*(9), 3145–3157.

Yang, D., Li, T., Xie, X., & Zhang, H. (2020). Event-triggered integral sliding-mode control for nonlinear constrained-input systems with disturbances via adaptive dynamic programming. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 50*(11), 4086–4096.

Zhang, Q., Song, X., Song, S., & Stojanovic, V. (2023). Finite-time sliding mode control for singularly perturbed PDE systems. *Journal of the Franklin Institute, 360*(2), 841–861.

Zhang, P., Yuan, Y., & Guo, L. (2021). Fault-tolerant optimal control for discrete-time nonlinear system subjected to input saturation: A dynamic event-triggered approach. *IEEE Transactions on Cybernetics, 51*(6), 2956–2968.

Zhang, Y., Zhao, B., Liu, D., & Zhang, S. (2022). Distributed fault tolerant consensus control of nonlinear multiagent systems via adaptive dynamic programming, *IEEE Transactions on Neural Networks and Learning Systems,* early access. http://dx.doi.org/10.1109/TNNLS.2022.3217774.

Zhao, B., Liu, D., & Li, Y. (2017). Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems. *Information Sciences, 384,* 21–33.

Zhao, Y., Wang, H., Xu, N., Zong, G., & Zhao, X. (2023). Reinforcement learning-based decentralized fault tolerant control for constrained interconnected nonlinear systems. *Chaos, Solitons Fractals, 167,* Article 113034.

Zhao, B., Zhang, Y., & Liu, D. (2023). Adaptive dynamic programming-based cooperative motion/force control for modular reconfigurable manipulators: A joint task assignment approach. *IEEE Transactions on Neural Networks and Learning Systems, 34*(12), 10944–10954.

Distributed Fault Tolerant Consensus Control of Nonlinear Multiagent Systems via Adaptive Dynamic Programming

Yongwei Zhang, *Member, IEEE*, Bo Zhao, *Senior Member, IEEE*,
Derong Liu, *Fellow, IEEE*, and Shunchao Zhang

Abstract—This article develops a distributed fault-tolerant consensus control (DFTCC) approach for multiagent systems by using adaptive dynamic programming. By establishing a local fault observer, the potential actuator faults of each agent are estimated. Subsequently, the DFTCC problem is transformed into an optimal consensus control problem by designing a novel local value function for each agent which contains the estimated fault, the consensus errors, and the control laws of the local agent and its neighbors. In order to solve the coupled Hamilton–Jacobi–Bellman equation of each agent, a critic-only structure is established to obtain the approximate local optimal consensus control law of each agent. Moreover, by using Lyapunov's direct method, it is proven that the approximate local optimal consensus control law guarantees the uniform ultimate boundedness of the consensus error of all agents, which means that all following agents with potential actuator faults synchronize to the leader. Finally, two simulation examples are provided to validate the effectiveness of the present DFTCC scheme.

Index Terms—Adaptive dynamic programming, fault-tolerant control, multiagent systems, optimal consensus control.

I. INTRODUCTION

DISTRIBUTED coordination in multiagent systems (MASs) has received tremendous attention due to its broad applications in multirobot systems [1], distributed sensor networks [2], battery management [3], spring-mass systems [4], and unmanned air vehicles [5], etc. For such systems, cooperative control policies are required to drive all agents to achieve a unified goal. It is well-known that system consensus or synchronization is one of the most significant tasks for MASs. The consensus control can be divided into two categories, i.e., the leaderless consensus and the leader-follower consensus. In recent years, many researchers have paid attention to developing consensus control schemes for MASs. In [6], a robust adaptive fault-tolerant control (FTC) scheme was developed to address the leaderless consensus of MASs with uncertain nonidentical dynamics and actuator faults. In [7], a distributed sliding-mode controller was developed for second-order MASs to achieve leader-follower consensus. In [8], the distributed resilience consensus problem of MASs with actuator faults was investigated by designing an adaptive controller. It is noted that all the above-mentioned results only guarantee the stability of the consensus error. However, the control efficiency such as the energy consumption and the production cost which are important indicators in practical applications is not taken into account. Consequently, optimal consensus control receives wide attention and it aims to develop distributed control protocols based on the local information of each agent and its neighbors, such that all agents achieve synchronization and optimal performance. It is worth pointing out that the game theory provides an ideal perspective to solve the optimal consensus control problem of MASs [9], [10], [11], [12]. Under the game theory, each agent chooses a local optimal control policy to minimize its performance index. However, obtaining the analytic solution is intractable by solving the coupled Hamilton–Jacobi–Bellman (HJB) equation due to its high nonlinearities [13], [14], [15], [16], [17].

For the sake of the "curse of dimensionality" in solving the coupled HJB equation by dynamic programming, adaptive dynamic programming (ADP) has been extensively investigated [18], [19], [20], [21], [22], [23], [24]. In recent years, several researchers have developed optimal consensus control schemes for MASs by using the ADP technique. For linear systems, in [25], an off-policy reinforcement learning (RL) method was proposed to address optimal consensus control problems. Under the framework of graphical games, a local performance index function in terms of the tracking error

and the control input was designed for each agent, and the off-policy RL method was developed to obtain the approximate solution of the coupled HJB equation. In [26], the optimal consensus control problem of discrete-time (DT) linear MASs was addressed by using the RL method. By designing a discounted performance index function, the coupled DT HJB equation was solved by an actor-critic framework. For nonlinear systems, in [27], an augmented system was established and the solutions of the corresponding coupled HJB equations were obtained by using the policy iteration (PI) algorithm. In [28], a Q-function-based PI algorithm was developed to realize the model-free distributed optimal consensus control of DT nonlinear MASs.

It is worth pointing out that the aforementioned works have not considered actuator faults. As industrial systems become complex and large-scale, the occurrence of actuator faults is inevitable. Once the faults occur, the control performance will be degraded or even become unstable [29]. To ensure reliability and control accuracy, many ADP-based FTC methods have been developed recently. In [30], a PI-based online fault compensation control method was developed for continuous-time (CT) nonlinear systems. In order to handle actuator faults, the FTC was constructed by an online fault compensation term and a PI-based optimal regulation term. In [31], the FTC problem of CT nonlinear systems was tackled by developing an ADP-based control method. The actuator fault was estimated by a fault observer, and a novel performance index function was designed to transform the FTC problem into an optimal control problem. In [32], a sliding-mode FTC method was proposed to deal with actuator faults of nonlinear systems by using the RL method. In [33], the FTC problem of complex unknown dynamical networks was considered. The static feedback gain which was obtained by the iterative ADP algorithm was employed to compensate for the actuator faults.

The above-mentioned works only considered single agent systems, but multiple agents are required to fulfill complex practical tasks, such as industrial manufacturing [34], attitude alignment of space crafts [35], and disaster relief [36]. However, the occurrence of actuator faults is inevitable in systems with large-scale, distributed, and autonomous agents. As aforementioned, the ADP-based FTC has shown proper advantages, i.e., optimality, adaptivity, and learning ability. Therefore, it is reasonable to tackle the fault-tolerant consensus control (FTCC) problem of MASs by using the ADP technique. Compared with a single-agent system, the FTCC problem of MASs is more intractable because agents in MASs are interconnected through mechanical interconnections or communication networks. It implies that a fault that occurs in one agent may affect other interconnected agents or even destroy the control performance of the whole system. Moreover, the interconnection of MASs leads to complications of fault detection and low fault estimation accuracy. Thus, the major challenges lie in that: 1) how to obtain the accurate fault information of each agent and 2) how to design a fault-tolerant controller to compensate for the actuator faults. From the above discussion, it is urgent to develop a distributed fault-tolerant consensus control (DFTCC) method to guarantee the stability of MASs with unpredictable faults. However, this problem has been rarely investigated, which motivates our research.

In this article, an ADP-based DFTCC approach is developed for MASs with potential actuator faults. The contributions and novelties of this work are summarized as follows.

1) Different from existing approaches [30], [31] which addressed the FTC problem for single-agent systems, this article develops an ADP-based DFTCC approach for MASs. It guarantees the consensus error of each agent to be uniformly ultimately bounded (UUB), which means that all agents agree upon the leader state even if faults exist.

2) A local fault observer is established to estimate the potential actuator faults of each agent. Subsequently, the FTCC problem is converted to an optimal consensus control problem by designing a local value function that contains the estimated fault, the local consensus error, and the control inputs of the agent and its neighbors.

3) Compared with traditional ADP-based control approaches to address simple HJB equations, a distributed PI algorithm is developed to solve coupled HJB equations that reflect the interconnections among each agent and its neighbors. Moreover, the critic neural network (NN) is adopted to obtain the approximate local optimal consensus control law of each agent.

The rest of this article is organized as follows. In Section II, the FTCC problem for MASs is formulated. In Section III, a local fault observer and a critic NN are established for each agent, and the ADP-based DFTCC approach is developed. Moreover, the stability of the closed-loop system is provided. In Section IV, the effectiveness of the developed method is verified by two simulation examples. In Section V, a brief conclusion is given.

II. Preliminaries

A. Graph Theory

Consider the undirected communication topology graph denoted by $\Pi_g = \{\mathcal{P}, \xi, \mathcal{A}\}$, where $\mathcal{P} = \{p_1, \ldots, p_N\}$ is a set of nodes, $\xi = \{(p_i, p_j): p_i, p_j \in \mathcal{P}\}$ is a set of edges, and $\mathcal{A} = [a_{ij}]$ is a weighted adjacency matrix. If and only if agent i and agent j are directly connected, then $(p_i, p_j) \in \xi$. Moreover, $a_{ij} > 0$ if $(p_i, p_j) \in \xi$, $a_{ij} = 0$ if $(p_i, p_j) \notin \xi$, and $a_{ii} = 0$ for all $i = 1, \ldots, N$. Let $N_i = \{j: (p_i, p_j) \in \xi, j \neq i\}$ be a set of neighbors of the agent i, $\bar{N}_i$ be a set of agent i and its neighbors, $\mathcal{D} = \text{diag}\{d_1, \ldots, d_N\}$ with $d_i = \sum_{j \in N_i} a_{ij}$ be the degree matrix of Π_g, and $\mathcal{L} = \mathcal{D} - \mathcal{A} = [l_{ij}]$ be the Laplacian matrix with $l_{ij} = -a_{ij}$ and $l_{ii} = \sum_{j=1}^{N_i} a_{ij}$.

B. Problem Formulation

Consider the nonlinear MASs with one leader and N followers. The system dynamics of each following agent is described as

$$\dot{x}_i = \bar{f}_i(x_i) + \mathcal{G}_i(x_i)(u_i - u_{if}), \quad i = 1, 2, \ldots, N \quad (1)$$

where $x_i \in \mathbb{R}^{n_i}$ is the system states of the agent i, $u_i \in \mathbb{R}^{m_i}$ is the control input of the agent i, $u_{if} \in \mathbb{R}^{m_i}$ is the actuator fault,

and $\bar{f}_i(x_i) \in \mathbb{R}^{n_i}$ and $\mathcal{G}_i(x_i) \in \mathbb{R}^{n_i \times m_i}$ are known nonlinear system functions.

Assumption 1: The system functions $\bar{f}_i(x_i)$ and $\mathcal{G}_i(x_i)$ are Lipschitz continuous on a compact set Ω containing the origin with $f(0) = 0$, and the system (1) is stabilizable on Ω.

The system dynamics of the leader is given by

$$\dot{x}_0 = f_0(x_0) \tag{2}$$

where $x_0 \in \mathbb{R}^{n_0}$, and $f_0(x_0) \in \mathbb{R}^{n_0}$ is a differentiable function. The local neighborhood consensus error of the agent i is defined as

$$\mathcal{E}_i = \sum_{j \in N_i} a_{ij}(x_i - x_j) + c_i(x_i - x_0) \tag{3}$$

where $c_i \geq 0$ is the pinning gain. Thus, the dynamics of the local neighborhood consensus error can be obtained by differentiating (3) as

$$
\begin{aligned}
\dot{\mathcal{E}}_i &= \sum_{j \in N_i} a_{ij}(\dot{x}_i - \dot{x}_j) + c_i(\dot{x}_i - \dot{x}_0) \\
&= \sum_{j \in N_i} a_{ij}\big(\bar{f}_i(x_i) + \mathcal{G}_i(x_i)(u_i - u_{if}) \\
&\qquad - \bar{f}_j(x_j) - \mathcal{G}_j(x_j)(u_j - u_{jf})\big) \\
&\qquad + c_i\big(\bar{f}_i(x_i) + \mathcal{G}_i(x_i)(u_i - u_{if}) - f_0(x_0)\big) \\
&= \sum_{j \in N_i} a_{ij}\big(\bar{f}_i(x_i) + \mathcal{G}_i(x_i)(u_i - u_{if})\big) \\
&\qquad + c_i\big(\bar{f}_i(x_i) + \mathcal{G}_i(x_i)(u_i - u_{if})\big) \\
&\qquad - c_i f_0(x_0) - \sum_{j \in N_i} a_{ij} f_0(x_0) + \sum_{j \in N_i} a_{ij} f_0(x_0) \\
&\qquad - \sum_{j \in N_i} a_{ij}\big(\bar{f}_j(x_j) + \mathcal{G}_j(x_j)(u_j - u_{jf})\big) \\
&= \left(\sum_{j \in N_i} a_{ij} + c_i\right)\big(\bar{f}_i(x_i) + \mathcal{G}_i(x_i)(u_i - u_{if}) - f_0(x_0)\big) \\
&\qquad - \sum_{j \in N_i}(a_{ij} + b_{ij})\big(\bar{f}_j(x_j) + \mathcal{G}_j(x_j)(u_j - u_{jf}) - f_0(x_0)\big) \\
&= (l_{ii} + b_{ii})\big(\mathcal{F}_i + \mathcal{G}_i(x_i)(u_i - u_{if})\big) \\
&\qquad + \sum_{j \in N_i}(l_{ij} + b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)(u_j - u_{jf})\big) \\
&= \sum_{j \in \bar{N}_i}(l_{ij} + b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)(u_j - u_{jf})\big) \tag{4}
\end{aligned}
$$

where $\mathcal{F}_i = \bar{f}_i(x_i) - f_0(x_0)$, $b_{ii} = c_i$ and $b_{ij} = 0$ if $j \in N_i$.

In order to solve the FTCC problem of system (1), we need to obtain a set of feedback control policies $u_1(x), \ldots, u_N(x)$, such that all the followers synchronize with the leader even if the fault occurs. Next, we will show that the FTCC problem of system (1) is transformed into an optimal consensus control problem by designing a modified local value function.

The nominal system corresponding to system (1) without actuator faults is expressed by

$$\dot{x}_i = \bar{f}_i(x_i) + \mathcal{G}_i(x_i)u_i. \tag{5}$$

Then, the dynamics of the local neighborhood consensus error without actuator faults is given by

$$\dot{\mathcal{E}}_i = \sum_{j \in \bar{N}_i}(l_{ij} + b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)u_j\big). \tag{6}$$

The local value function of the agent i is defined as

$$\mathcal{V}_i(\mathcal{E}_i) = \int_t^\infty \big(\rho \hat{u}_{if}^\mathsf{T} \hat{u}_{if} + \mathcal{C}_i(\mathcal{E}_i, u_i, u_{(-i)})\big) d\tau$$

where ρ is a positive constant, $\hat{u}_{if} \in \mathbb{R}^{n_i}$ is the estimation of u_{if}, and the utility function $\mathcal{C}_i(\cdot)$ is designed as

$$\mathcal{C}_i(\mathcal{E}_i, u_i, u_{(-i)}) = \mathcal{E}_i^\mathsf{T} Q_{ii} \mathcal{E}_i + u_i^\mathsf{T} R_{ii} u_i + \sum_{j \in N_i} u_j^\mathsf{T} R_{ij} u_j \tag{7}$$

where $u_{(-i)} = \{u_j | j \in N_i\}$ are the control inputs of the neighbors of the agent i, $Q_{ii} \in \mathbb{R}^{n_i \times n_i}$, $R_{ii} \in \mathbb{R}^{m_i \times m_i}$ and $R_{ij} \in \mathbb{R}^{m_j \times m_j}$ are positive definite matrices.

Assume $\mathcal{V}_i(\mathcal{E}_i) \in \mathcal{Z}^1$, where $\mathcal{Z}^1$ is a space on Ω of functions with continuous first derivative. Then, the Hamiltonian of the agent i is defined as

$$
\begin{aligned}
&\mathcal{H}_i\big(\mathcal{E}_i, \nabla \mathcal{V}_i(\mathcal{E}_i), u_i, u_{(-i)}\big) \\
&= \rho \hat{u}_{if}^\mathsf{T} \hat{u}_{if} + \mathcal{C}_i(\mathcal{E}_i, u_i, u_{(-i)}) \\
&\quad + \nabla \mathcal{V}_i^\mathsf{T}(\mathcal{E}_i)\left(\sum_{j \in \bar{N}_i}(l_{ij} + b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)u_j\big)\right).
\end{aligned}
$$

The local optimal value function of the agent i

$$
\begin{aligned}
&\mathcal{V}_i^*(\mathcal{E}_i) \\
&= \min_{u_i \in \Re(\Omega)} \int_t^\infty \left(\rho \hat{u}_{if}^\mathsf{T} \hat{u}_{if} + \mathcal{C}_i\big(\mathcal{E}_i(\tau), u_i(\tau), u_{(-i)}(\tau)\big)\right) d\tau
\end{aligned}
$$

satisfies the HJB equation as

$$\min_{u_i \in \Re(\Omega)} \mathcal{H}_i\big(\mathcal{E}_i, \nabla \mathcal{V}_i^*(\mathcal{E}_i), u_i, u_{(-i)}\big) = 0 \tag{8}$$

where $\Re(\Omega)$ is a set of admissible controls. Then, the local optimal consensus control law is derived by

$$u_i^* = -\frac{d_i + c_i}{2} R_{ii}^{-1} \mathcal{G}_i^\mathsf{T}(x_i) \nabla \mathcal{V}_i^*(\mathcal{E}_i). \tag{9}$$

Based on (8) and (9), we can obtain

$$
\begin{aligned}
0 &= \rho \hat{u}_{if}^\mathsf{T} \hat{u}_{if} + \mathcal{C}_i\big(\mathcal{E}_i, u_i^*, u_{(-i)}^*\big) \\
&\quad + \nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i)\left(\sum_{j \in \bar{N}_i}(l_{ij} + b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)u_j^*\big)\right). \tag{10}
\end{aligned}
$$

Noticing that (10) is the HJB equation, which is difficult to solve due to its high nonlinearities [38], [39], [44], [45]. In the next section, the ADP technique is adopted to overcome this bottleneck.

Remark 1: In the optimal control community, the value function represents the objective to be optimized. For the nominal system, a value function is a quadratic form with respect to the system state and the control input. In this article, the fault estimation which is obtained by the fault observer is added to the value function to compensate for the actuator fault of each agent in real-time. Therefore, the influence of

the actuator fault can be eliminated. It is worth mentioning that the controller design process not only compensates for the actuator fault but also optimizes the control performance. Therefore, the DFTCC is achieved in an optimal manner.

III. Fault Tolerant Controller Design

A. Problem Transformation

In this section, by using the Lyapunov stability theorem, we demonstrate that the local optimal consensus control law (9) guarantees the UUB of the local neighborhood consensus error of each agent even suffering from actuator faults, i.e., all following agents with potential actuator faults synchronize with the leader. It means that the FTCC problem of system (1) is transformed into an optimal consensus control problem, and the local optimal consensus control law (9) is thus the solution to the FTCC problem. Before proving, some assumptions which were used in [30], [31] and [40] are provided.

Assumption 2: The actuator fault u_{if} is norm-bounded, that is,

$$\|u_{if}\| \le \bar{u}_{if} \tag{11}$$

where $\bar{u}_{if}$ is a positive constant.

Assumption 3: The system function $\mathcal{G}_j(x_j)$ is norm-bounded, that is,

$$\|\mathcal{G}_j(x_j)\| \le \bar{\mathcal{G}}_j \tag{12}$$

where $\bar{\mathcal{G}}_j$ is a positive constant.

Theorem 1: Considering the nonlinear MASs with the leader (2) and followers (1), the dynamics of the local neighborhood consensus error given by (4), the local optimal consensus control law given by (9), and the Assumptions 2 and 3, the local neighborhood consensus error of each following agent with potential actuator faults is guaranteed to be UUB if the following inequality:

$$\chi^2 \lambda_{\min}(Q_{ii})\|\mathcal{E}_i\|^2 > \left\|\nabla\mathcal{V}_i^*(\mathcal{E}_i)\right\|^2 \tag{13}$$

holds, where $0 < \chi < 1$.

Proof: Select a Lyapunov function candidate as

$$\mathcal{L} = \mathcal{V}_i^*(\mathcal{E}_i). \tag{14}$$

Taking the time derivative of (14) along with the local neighborhood consensus error (3), we have

$$\dot{\mathcal{L}} = \nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)\big(u_j^* - u_{jf}\big)\big)$$

$$= \nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)u_j^*\big)$$

$$- \nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}. \tag{15}$$

According to (10), we can obtain

$$\nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)u_j^*\big)$$

$$= -\mathcal{C}_i\big(\mathcal{E}_i, u_i^*, u_{(-i)}^*\big) - \rho\hat{u}_{if}^{\mathsf{T}}\hat{u}_{if}. \tag{16}$$

Combining (15) with (16), we can get

$$\dot{\mathcal{L}} = -\rho\hat{u}_{if}^{\mathsf{T}}\hat{u}_{if} - \mathcal{E}_i^{\mathsf{T}}Q_{ii}\mathcal{E}_i - u_i^{*\mathsf{T}}R_{ii}u_i^* - \sum_{j\in N_i}u_j^{*\mathsf{T}}R_{ij}u_j^*$$

$$- \nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}$$

$$\le -\rho\hat{u}_{if}^{\mathsf{T}}\hat{u}_{if} - \mathcal{E}_i^{\mathsf{T}}Q_{ii}\mathcal{E}_i$$

$$- \nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}$$

$$\le -\rho\hat{u}_{if}^{\mathsf{T}}\hat{u}_{if} - \mathcal{E}_i^{\mathsf{T}}Q_{ii}\mathcal{E}_i - \frac{1}{2}\nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i)\nabla\mathcal{V}_i^*(\mathcal{E}_i)$$

$$+ \frac{1}{2}\left(\nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) - \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}\right)^{\mathsf{T}}$$

$$\times \left(\nabla\mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) - \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}\right)$$

$$- \frac{1}{2}\left(\sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}\right)^{\mathsf{T}}$$

$$\times \left(\sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}\right)$$

$$\le -\rho\hat{u}_{if}^{\mathsf{T}}\hat{u}_{if} - \mathcal{E}_i^{\mathsf{T}}Q_{ii}\mathcal{E}_i + \left\|\nabla\mathcal{V}_i^*(\mathcal{E}_i)\right\|^2$$

$$+ \left\|\sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})\mathcal{G}_j(x_j)u_{jf}\right\|^2$$

$$\le -\chi^2\lambda_{\min}(Q_{ii})\|\mathcal{E}_i\|^2 + (\chi^2-1)\lambda_{\min}(Q_{ii})\|\mathcal{E}_i\|^2$$

$$- \rho\hat{u}_{if}^{\mathsf{T}}\hat{u}_{if} + \left\|\nabla\mathcal{V}_i^*(\mathcal{E}_i)\right\|^2$$

$$+ \bar{N}_i \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})^2\bar{\mathcal{G}}_j^2\bar{u}_{jf}^2. \tag{17}$$

If the inequality (13) holds, we further have

$$\dot{\mathcal{L}} \le (\chi^2-1)\lambda_{\min}(Q_{ii})\|\mathcal{E}_i\|^2 + \lambda_1 \tag{18}$$

where $\lambda_1 = \bar{N}_i \sum_{j\in\bar{N}_i}(l_{ij}+b_{ij})^2\bar{\mathcal{G}}_j^2\bar{u}_{jf}^2$. Hence, $\dot{\mathcal{L}} < 0$ if $\mathcal{E}_i$ lies outside the compact set

$$\Omega_{\mathcal{E}_i} = \left\{\mathcal{E}_i: \|\mathcal{E}_i\| \le \sqrt{\frac{\lambda_1}{(1-\chi^2)\lambda_{\min}(Q_{ii})}}\right\}. \tag{19}$$

It means that the local optimal consensus control law given by (9) can guarantee the local neighborhood consensus error of each agent with potential actuator faults to be UUB, i.e., (9) is the solution to the FTCC problem. Thus, the problem transformation is reasonable. This completes the proof.

B. Fault Observer Design

In this section, a local fault observer is designed to estimate unknown potential actuator faults. Consider the agent i with actuator faults (1), a local fault observer is designed as

$$\dot{\hat{x}}_i = \bar{f}(\hat{x}_i) + \mathcal{G}_i(\hat{x}_i)(u_i - \hat{u}_{if}) + L_1(x_i - \hat{x}_i) \tag{20}$$

where $\hat{x}_i$ is the observation of the system state x_i, L_1 is the positive definite observer gain matrix, and $\hat{u}_{if}$ is the estimated actuator faults which is updated by

$$\dot{\hat{u}}_{if} = -L_2 \mathcal{G}_i^{\mathsf{T}}(\hat{x}_i) e_{io} \tag{21}$$

where L_2 is a positive definite matrix, and $e_{io} = x_i - \hat{x}_i$ is the state observation error of the agent i. According to (1) and (20), we can get

$$\dot{e}_{io} = \tilde{f}(x_i) + \tilde{\mathcal{G}}_i(x_i)u_i - \mathcal{G}_i(x_i)u_{if} + \mathcal{G}_i(\hat{x}_i)\hat{u}_{if} - L_1 e_{io} \tag{22}$$

where $\tilde{f}(x_i) = \bar{f}(x_i) - \bar{f}(\hat{x}_i)$ and $\tilde{\mathcal{G}}_i(x_i) = \mathcal{G}_i(x_i) - \mathcal{G}_i(\hat{x}_i)$.

Lemma 1: Considering the agent i with actuator faults (1) and the local fault observer given by (20), the fault observation error can be guaranteed to be UUB with the adaptive updating law (21).

Proof: The proof of Lemma 1 has been provided in [31], so the detail is omitted here.

C. Neural Network Implementation

In this section, critic NNs are adopted to approximate the solutions of HJB equations. According to the universal approximation property of NNs, the local optimal value function of the agent i can be expressed as

$$\mathcal{V}_i^*(\mathcal{E}_i) = W_{ic}^{*\mathsf{T}} \varphi_{ic}(\mathcal{E}_i) + \varepsilon_{ic}(\mathcal{E}_i) \tag{23}$$

where $W_{ic}^* \in \mathbb{R}^{h_{ic}}$ is the ideal weight vector, $\varphi_{ic}(\mathcal{E}_i) \in \mathbb{R}^{h_{ic}}$ is the activation function, h_{ic} is the number of hidden neurons, and $\varepsilon_{ic}(\mathcal{E}_i) \in \mathbb{R}$ is the reconstruction error. Then, the partial derivative of $\mathcal{V}_i^*(\mathcal{E}_i)$ with respect to $\mathcal{E}_i$ is given by

$$\nabla \mathcal{V}_i^*(\mathcal{E}_i) = \nabla \varphi_{ic}^{\mathsf{T}}(\mathcal{E}_i) W_{ic}^* + \nabla \varepsilon_{ic}^{\mathsf{T}}(\mathcal{E}_i). \tag{24}$$

The approximate local value function is defined as

$$\hat{\mathcal{V}}_i(\mathcal{E}_i) = \hat{W}_{ic}^{\mathsf{T}} \varphi_{ic}(\mathcal{E}_i) \tag{25}$$

where $\hat{W}_{ic} \in \mathbb{R}^{h_{ic}}$ is the estimate of W_{ic}. Similarly, we have

$$\nabla \hat{\mathcal{V}}_i(\mathcal{E}_i) = \nabla \varphi_{ic}^{\mathsf{T}}(\mathcal{E}_i) \hat{W}_{ic}. \tag{26}$$

According to (9) and (24), the local optimal consensus control law of the agent i is expressed as

$$u_i^* = -\frac{d_i + c_i}{2} R_{ii}^{-1} \mathcal{G}_i^{\mathsf{T}}(x_i) \left(\nabla \varphi_{ic}^{\mathsf{T}}(\mathcal{E}_i) W_{ic}^* + \nabla \varepsilon_{ic}^{\mathsf{T}}(\mathcal{E}_i) \right). \tag{27}$$

Then, the approximate local consensus control law of the agent i is given by

$$\hat{u}_i = -\frac{d_i + c_i}{2} R_{ii}^{-1} \mathcal{G}_i^{\mathsf{T}}(x_i) \nabla \varphi_{ic}^{\mathsf{T}}(\mathcal{E}_i) \hat{W}_{ic}. \tag{28}$$

Based on (10) and (28), the approximate Hamiltonian is

$$\begin{aligned}
&\hat{\mathcal{H}}_i(\mathcal{E}_i, \hat{W}_{ic}, \hat{u}_i, \hat{u}_{(-i)}) \\
&= \mathcal{E}_i^{\mathsf{T}} Q_{ii} \mathcal{E}_i + \hat{u}_i^{\mathsf{T}} R_{ii} \hat{u}_i + \sum_{j \in N_i} \hat{u}_j^{\mathsf{T}} R_{ij} \hat{u}_j \\
&\quad + \hat{W}_{ic}^{\mathsf{T}} \underbrace{\nabla \varphi_{ic}(\mathcal{E}_i) \left(\sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})(\mathcal{F}_j + \mathcal{G}_j(x_j)\hat{u}_j) \right)}_{\Gamma_i} \\
&\triangleq e_{ic}.
\end{aligned} \tag{29}$$

Let $\tilde{W}_{ic} = W_{ic} - \hat{W}_{ic}$ be the weight estimation error. The gradient descent algorithm is employed to minimize the target function $E_{ic} = (1/2)e_{ic}^{\mathsf{T}}e_{ic}$. Hence, the critic NN weight updating rule is given by

$$\dot{\hat{W}}_{ic} = -\alpha_c \frac{1}{\left(1 + \Gamma_i^{\mathsf{T}}\Gamma_i\right)^2} \left(\frac{\partial E_{ic}}{\partial \hat{W}_{ic}} \right) = -\frac{\alpha_c e_{ic} \Gamma_i}{\left(1 + \Gamma_i^{\mathsf{T}}\Gamma_i\right)^2} \tag{30}$$

where $\alpha_c > 0$ is the learning rate.

Theorem 2: Consider the agent i with potential actuator faults (1), if the critic NN weights are updated by (30), then the weight approximation error $\tilde{W}_{ic}$ can be guaranteed to be UUB.

Proof: The proof of Theorem 2 is similar to that in [31], [40], [41], [42], and [43], so it is omitted here.

According to the above discussion, the structure of the ADP-based DFTCC approach is displayed in Fig. 1.

Remark 2: In this article, the critic-only structure is established to obtain the approximate value function, rather than the actor-critic structure. That is because: 1) according to (28), we know that the local consensus control law relies on the value function. Once the approximate value function is obtained by using the critic NN, then the approximate local consensus control law can be obtained via (28). Therefore, only critic NN is employed and 2) in fact, the critic-only structure is widely used in existing ADP-based control schemes [31], [32], [40]. Compared with the actor-critic structure, the critic-only structure has a low control complexity, which is beneficial to practical applications.

Remark 3: As we all know, existing control architectures can be divided into three categories, i.e., centralized control, distributed control, and decentralized control. The centralized control approach requires the overall system information, while the distributed control uses the information of the local agent and its neighbors, and the decentralized control needs the information of the local agent only. Thus, in this article, the developed ADP-based DFTCC approach is a distributed one since the states of the local agent and its neighbors are the only required information.

D. Stability Analysis

In this section, we will prove that the approximate local consensus control law (28) can guarantee the local neighborhood consensus error of each following agent with potential actuator faults to be UUB. Before the stability analysis, the following assumption which is common in ADP literature [41] is provided.

Assumption 4: $\nabla \varphi_{ic}(\mathcal{E}_i)$, $\nabla \varepsilon_{ic}(\mathcal{E}_i)$, $\tilde{W}_{ic}$ and W_{ic}^* are norm-bounded, that is,

$$\|\nabla \varphi_{ic}(\mathcal{E}_i)\| \le \bar{\varphi}_{ic}, \quad \|\nabla \varepsilon_{ic}(\mathcal{E}_i)\| \le \bar{\varepsilon}_{ic}$$
$$\|\tilde{W}_{ic}\| \le \bar{W}_{ic}, \quad \|W_{ic}^*\| \le \bar{W}_{ic}^*$$

where $\bar{\varphi}_{ic}$, $\bar{\varepsilon}_{ic}$, $\bar{W}_{ic}$ and $\bar{W}_{ic}^*$ are positive constants.

Theorem 3: Consider the following agent i with potential actuator faults given by (1), the dynamics of the local neighborhood consensus error of each agent (6), the critic NN weights updated by (30), and the Assumptions 2–4. Then, the approximate local consensus control law (28) can guarantee
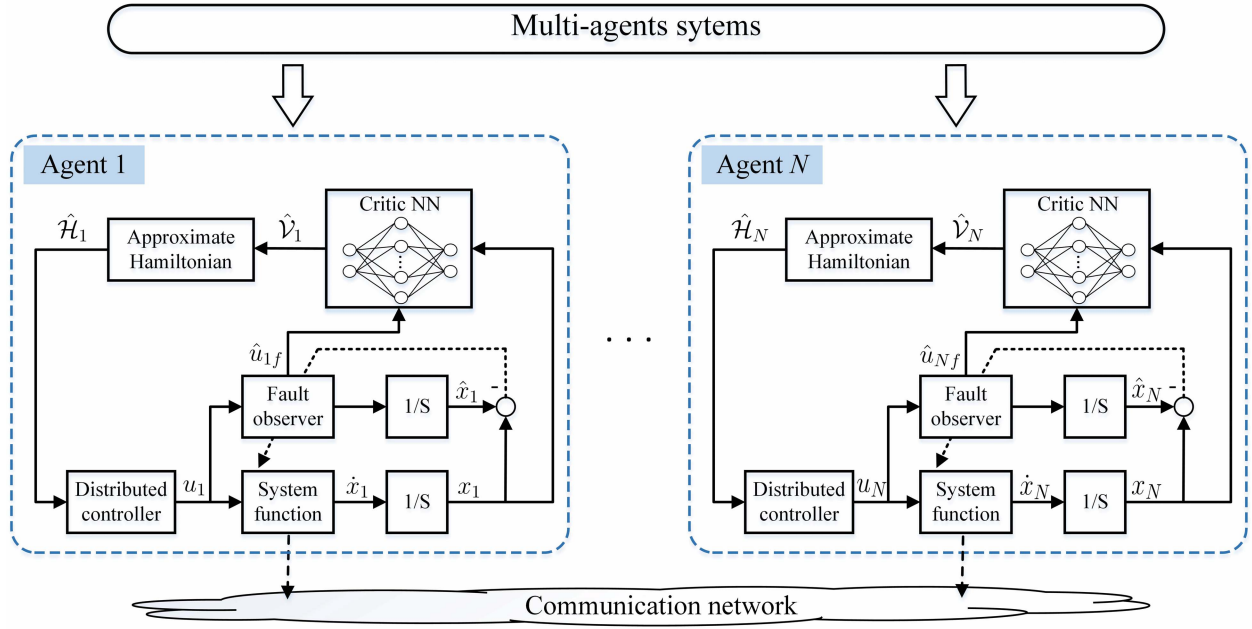
Fig. 1.   Structure of the ADP-based DFTCC approach.

the local neighborhood consensus error of each following agent with potential actuator faults to be UUB.

Proof: Select a Lyapunov function candidate as

$$\mathcal{L}_1 = \mathcal{V}_i^*(\mathcal{E}_i). \tag{31}$$

Taking the time derivative of (31) along the local neighborhood consensus error (3), we have

$$\dot{\mathcal{L}}_1 = \nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)(\hat{u}_j - u_{jf})\big)$$

$$= \nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\big(\mathcal{F}_j + \mathcal{G}_j(x_j)\hat{u}_j\big)$$

$$- \nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\mathcal{G}_j(x_j)u_{jf}. \tag{32}$$

According to (10), we can obtain

$$\nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\mathcal{F}_j$$

$$= -\mathcal{C}_i\big(\mathcal{E}_i, u_i^*, u_{(-i)}^*\big) - \nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\mathcal{G}_j(x_j)u_j^*. \tag{33}$$

Let $\Psi = [\mathcal{G}_1(x_1), \ldots, \mathcal{G}_{\bar{N}_i}(x_{\bar{N}_i})]^{\mathsf{T}}[\mathcal{G}_1(x_1), \ldots, \mathcal{G}_{\bar{N}_i}(x_{\bar{N}_i})]$ and $\lambda_{\max}(\Psi)$ be the maximum eigenvalue of Ψ. Based on (33) and using Young's inequality $\|A + B\|^2 \leq 2\|A\|^2 + 2\|B\|^2$, we can obtain (34), as shown at the bottom of the the page. According to (27) and (28), we can get

$$\|\hat{u}_j - u_j^* - u_{jf}\|^2$$

$$= \left\| -\frac{d_j + c_j}{2} R_{jj}^{-1} \mathcal{G}_j^{\mathsf{T}}(x_j) \nabla \varphi_{jc}^{\mathsf{T}}(\mathcal{E}_j)\hat{W}_{jc} \right.$$

$$\dot{\mathcal{L}}_1 = -\mathcal{C}_i\big(\mathcal{E}_i, u_i^*, u_{(-i)}^*\big) + \nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\mathcal{G}_j(x_j)\big(\hat{u}_j - u_j^* - u_{jf}\big)$$

$$\leq -\mathcal{C}_i\big(\mathcal{E}_i, u_i^*, u_{(-i)}^*\big) + \frac{1}{2}\nabla \mathcal{V}_i^{*\mathsf{T}}(\mathcal{E}_i)\nabla \mathcal{V}_i(\mathcal{E}_i)$$

$$+ \frac{1}{2}\begin{bmatrix} (l_{i1} + b_{i1})(\hat{u}_1 - u_1^* - u_{1f}) \\ \vdots \\ (l_{i\bar{N}_i} + b_{i\bar{N}_i})(\hat{u}_{\bar{N}_i} - u_{\bar{N}_i}^* - u_{\bar{N}_i f}) \end{bmatrix}^{\mathsf{T}} \Psi \begin{bmatrix} (l_{i1} + b_{i1})(\hat{u}_1 - u_1^* - u_{1f}) \\ \vdots \\ (l_{i\bar{N}_i} + b_{i\bar{N}_i})(\hat{u}_{\bar{N}_i} - u_{\bar{N}_i}^* - u_{\bar{N}_i f}) \end{bmatrix}$$

$$\leq -\mathcal{C}_i\big(\mathcal{E}_i, u_i^*, u_{(-i)}^*\big) + \frac{1}{2}\|\nabla \varphi_{ic}^{\mathsf{T}}(x)W_{ic}^* + \nabla \varepsilon_{ic}(x)\|^2 + \frac{1}{2}\lambda_{\max}(\Psi) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\|\hat{u}_j - u_j^* - u_{jf}\|^2$$

$$\leq -\mathcal{C}_i\big(\mathcal{E}_i, u_i^*, u_{(-i)}^*\big) + \bar{\varphi}_{ic}^2 \bar{W}_{ic}^{*2} + \bar{\varepsilon}_{ic}^2 + \frac{1}{2}\lambda_{\max}(\Psi) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})\|\hat{u}_j - u_j^* - u_{jf}\|^2 \tag{34}$$

$$+ \frac{d_j + c_j}{2} R_{jj}^{-1} \mathcal{G}_j^{\mathsf{T}}(x_j) \nabla \varphi_{jc}^{\mathsf{T}}(\mathcal{E}_j) W_{jc}$$

$$+ \frac{d_j + c_j}{2} R_{jj}^{-1} \mathcal{G}_j^{\mathsf{T}}(x_j) \nabla \varepsilon_{jc}^{\mathsf{T}}(\mathcal{E}_j) - u_{jf} \Big\|^2$$

$$= \Big\| \frac{d_j + c_j}{2} R_{jj}^{-1} \mathcal{G}_j^{\mathsf{T}}(x_j) \nabla \varphi_{jc}^{\mathsf{T}}(\mathcal{E}_j) \tilde{W}_{jc}$$

$$+ \frac{d_j + c_j}{2} R_{jj}^{-1} \mathcal{G}_j^{\mathsf{T}}(x_j) \nabla \varepsilon_{jc}^{\mathsf{T}}(\mathcal{E}_j) - u_{jf} \Big\|^2$$

$$\leq \Big\| (d_j + c_j) R_{jj}^{-1} \mathcal{G}_j^{\mathsf{T}}(x_j) \nabla \varphi_{jc}^{\mathsf{T}}(\mathcal{E}_j) \tilde{W}_{jc} \Big\|^2$$

$$+ \Big\| (d_j + c_j) R_{jj}^{-1} \mathcal{G}_j^{\mathsf{T}}(x_j) \nabla \varepsilon_{jc}^{\mathsf{T}}(\mathcal{E}_j) \Big\|^2 + 2\|u_{jf}\|^2$$

$$\leq (d_j + c_j)^2 R_{jj}^{-2} \bar{\mathcal{G}}_j^2 \bar{\varphi}_{jc}^2 \bar{W}_{jc}^2 + (d_j + c_j)^2 R_{jj}^{-2} \bar{\mathcal{G}}_j^2 \bar{\varepsilon}_{jc}^2$$

$$+ 2\bar{u}_{jf}^2. \tag{35}$$

Substituting (35) into (34), we have

$$\dot{\mathcal{L}}_1 \leq -\mathcal{C}_i(\mathcal{E}_i, u_i^*, u_{(-i)}^*) + \bar{\varphi}_{ic}^2 \bar{W}_{ic}^{*2} + \bar{\varepsilon}_{ic}^2$$

$$+ \frac{1}{2} \lambda_{\max}(\Psi) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})$$

$$\times \Big((d_j + c_j)^2 R_{jj}^{-2} \bar{\mathcal{G}}_j^2 \bar{\varphi}_{jc}^2 \bar{W}_{jc}^2$$

$$+ (d_j + c_j)^2 R_{jj}^{-2} \bar{\mathcal{G}}_j^2 \bar{\varepsilon}_{jc}^2 + 2\bar{u}_{jf}^2 \Big)$$

$$\leq -\mathcal{E}_i^{\mathsf{T}} Q_{ii} \mathcal{E}_i + \lambda_2$$

$$\leq -\zeta^2 \lambda_{\min}(Q_{ii}) \|\mathcal{E}_i\|^2 - (1 - \zeta^2) \lambda_{\min}(Q_{ii}) \|\mathcal{E}_i\|^2 + \lambda_2$$

where

$$\lambda_2 = \bar{\varphi}_{ic}^2 \bar{W}_{ic}^{*2} + \bar{\varepsilon}_{ic}^2$$

$$+ \frac{1}{2} \lambda_{\max}(\Psi) \sum_{j \in \bar{N}_i} (l_{ij} + b_{ij})$$

$$\times \Big((d_j + c_j)^2 R_{jj}^{-2} \bar{\mathcal{G}}_j^2 \bar{\varphi}_{jc}^2 \bar{W}_{jc}^2$$

$$+ (d_j + c_j)^2 R_{jj}^{-2} \bar{\mathcal{G}}_j^2 \bar{\varepsilon}_{jc}^2 + 2\bar{u}_{jf}^2 \Big).$$

Therefore, $\dot{\mathcal{L}}_1 < 0$ if $\mathcal{E}_i$ lies outside the compact set

$$\Omega_{\mathcal{E}_i} = \left\{ \mathcal{E}_i : \|\mathcal{E}_i\| \leq \sqrt{\frac{\lambda_2}{(1 - \zeta^2) \lambda_{\min}(Q_{ii})}} \right\}. \tag{36}$$

It means that the approximate local consensus control law (28) guarantees the UUB of the local neighborhood consensus error of each agent with potential faults. The proof is completed.

Remark 4: It is noted that W_{ic}^* and ε_{ic} are the optimal weight vector and the reconstruction error of the critic NN, respectively. In fact, after the critic NN is successfully trained, the obtained optimal weight vector and the approximate error cannot be infinite. Therefore, it is reasonable to assume that they are norm-bounded. Moreover, according to Theorem 2, the weight approximation error $\tilde{W}_{ic}$ is guaranteed to be UUB, so it is reasonable to assume that it is also norm-bounded.

Remark 5: Different from existing results [27] and [28] which tackled the consensus control problem of MASs only, this article further considers actuator faults and develops a
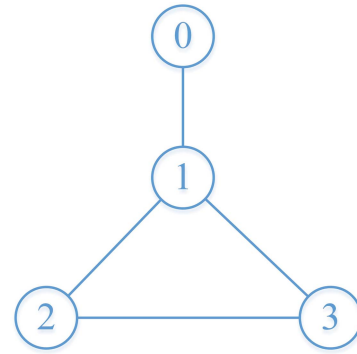

Fig. 2. Structure of communication topology.

TABLE I
PARAMETERS OF THE COMMUNICATION TOPOLOGY

Parameter	c_1	c_1	c_2	c_3	a_{12}	a_{13}	a_{21}	a_{23}	a_{31}	a_{32}
Value	1	0	0	0	0.1	0.1	0.5	0.5	0.4	0.4

DFTCC approach via ADP. In order to eliminate the influence of actuator fault, a fault observer is designed for each following agent to obtain the fault information, and a local value function is designed, which reflects the estimated fault, the local consensus error, and the control inputs of the agent and its neighbors.

Remark 6: In recent years, several FTC approaches have been developed for MASs [6], [7], [8]. However, existing results only guarantee the consensus of MASs with actuator faults, but the control performance such as the energy consumption and the production cost which are important in practice are not taken into account. In this article, the FTC problem of MASs is investigated via ADP. The developed ADP-based DFTCC approach not only guarantees the stability of MASs with actuator faults but also optimizes the control performance.

IV. NUMERICAL SIMULATION

In this section, simulation examples are adopted to verify the effectiveness of the developed ADP-based DFTCC scheme.

A. Example 1

Consider a MAS consisting of one leader and three followers, the communication topology is displayed in Fig. 2, and the corresponding parameters are provided in Table I. The dynamics of each following agent with potential actuator faults is provided as

$$x_i = \bar{f}_i(x_i) + \mathcal{G}_i(x_i)(u_i - u_{if}) \tag{37}$$

where

$$\bar{f}_i(x_i) = \begin{bmatrix} x_{i,2} \\ \mathcal{F}_1 \end{bmatrix}, \quad \mathcal{G}_i(x_i) = \begin{bmatrix} 0 \\ \cos(2x_{i,1}) + 2 \end{bmatrix}$$

$$\mathcal{F}_1 = -0.5x_{i,1} - x_{i,2} + x_{i,1}^2 x_{i,2} - 0.25x_{i,2}\big(\cos(2x_{i,1}) + 2\big)^2$$

$$+ 0.25x_{i,2}\big(\sin^2(4x_{i,2}^2 + 2)\big)$$

$x_i = [x_{i,1}, x_{i,2}]^{\mathsf{T}}$ is the system state of the agent i, and the actuator fault $u_{if} \in \mathbb{R}$ is given as

$$u_{1f} = \begin{cases} \cos(5t/2\pi) + \sin(t), & 10s \leq t \leq 20s \\ 0, & \text{others} \end{cases}$$
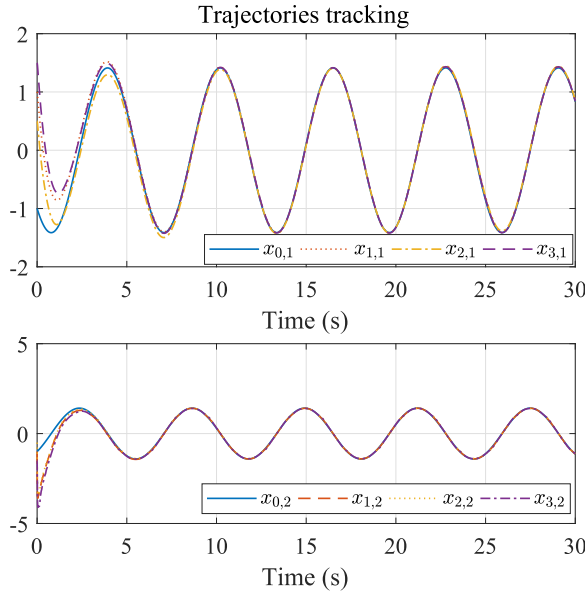
Fig. 3.   Trajectories tracking of Example 1.

Fig. 4.   Consensus errors of Example 1.

$$u_{2f} = 0, \quad 0 \le t \le 30s$$

$$u_{3f} = \begin{cases} 10\sin(t/2\pi) - \cos(2t), & 20s \le t \le 25s \\ 0, & \text{others.} \end{cases}$$

We suppose that agents 1 and 3 are faulty, and agent 2 is fault-free.

The reference dynamics of the leader is provided as

$$r(t) = \begin{bmatrix} -\cos(t) - \sin(t) \\ \sin(t) - \cos(t) \end{bmatrix}. \tag{38}$$

In the simulation, the initial states of the following agents are chosen as $x_1 = [1, -1]^{\mathsf{T}}$, $x_2 = [0.5, -0.5]^{\mathsf{T}}$ and $x_3 = [1.5, -1.5]^{\mathsf{T}}$, respectively, and the initial states of the observers are selected as $\hat{x}_1 = [2, -2]^{\mathsf{T}}$, $\hat{x}_2 = [0.5, -0.5]^{\mathsf{T}}$ and $\hat{x}_3 = [1.5, -1.5]^{\mathsf{T}}$. Let $Q_{ii} = 200I_2$, $R_{ii} = 0.1I$, $R_{ij} = 0.01I$, $\rho = 1$, $L_2 = 100$, $L_1 = 300$. Let the activation function of the critic NN be $\varphi_{ic} = [\mathcal{E}_{i,1}^2, \mathcal{E}_{i,2}\mathcal{E}_{i,1}, \mathcal{E}_{i,2}^2]$, and the learning rate of the critic NN be $\alpha_c = 1$.

Simulation results are displayed in Figs. 3–8. In Fig. 3, we can observe that the trajectories of all following agents catch up with the leader within 10 s. Fig. 4 illustrates the consensus errors of each following agent converge to a small region of zero. The fault estimations of all following agent are given in Fig. 5, where we can find that the actuator faults can be estimated by fault observers precisely. Fig. 6 shows the critic NN weight vectors will converge to $\hat{W}_1 = [4.13, 8.33, 8.92]^{\mathsf{T}}$, $\hat{W}_2 = [20.00, 8.79, 15.42]^{\mathsf{T}}$ and $\hat{W}_3 = [12.16, 8.32, 4.70]^{\mathsf{T}}$, respectively. Fig. 7 displays the control inputs of all following agents. It is clear that the controllers provide quick compensation after faults occur. Therefore, the leader and all followers can maintain consensus in the presence of actuator faults.

In order to verify the trajectory tracking performance of the developed ADP-based DFTCC approach, a different reference
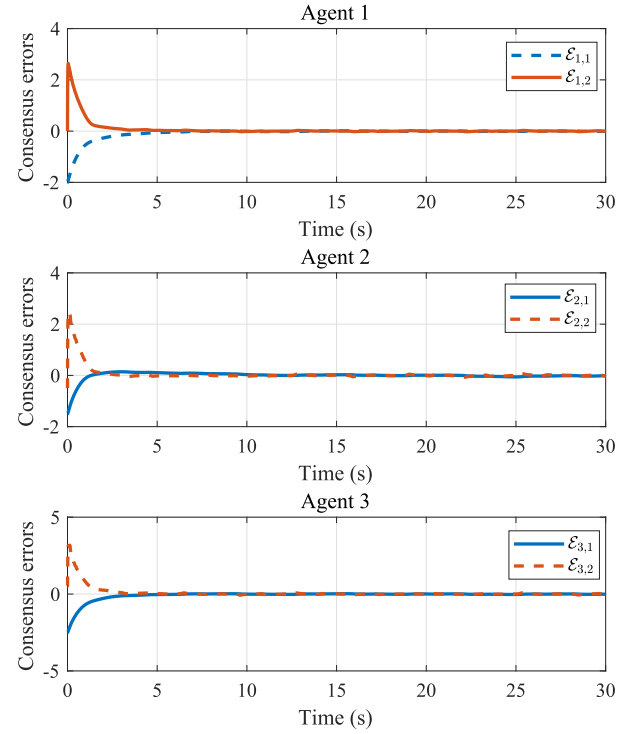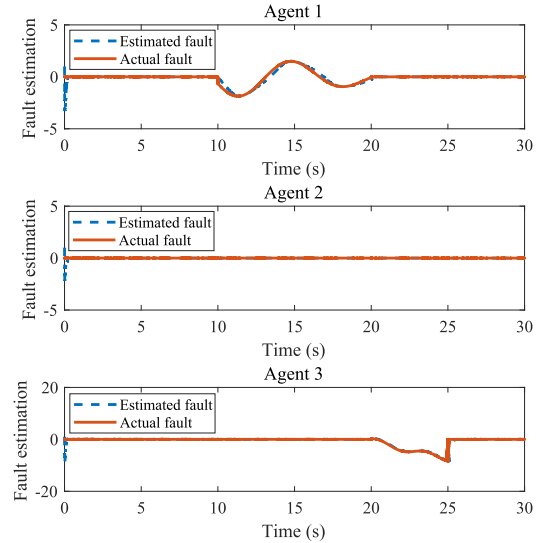

Fig. 5.   Fault estimation of each agent of Example 1.

dynamics of the leader is selected as

$$\dot{x}_0 = \begin{bmatrix} -0.5x_{0,1} - x_{0,2}\cos(x_{0,1}) \\ \sin(x_{0,1}) - x_{0,2} \end{bmatrix} \tag{39}$$

where $x_0 = [x_{0,1}, x_{0,2}]^{\mathsf{T}}$ is the leader's state vector. From Fig. 8, we can find that the trajectories of the leader and all followers can achieve consensus.

B. Example 2

To further verify the effectiveness of the developed ADP-based DFTCC approach, three 2-DOF modular
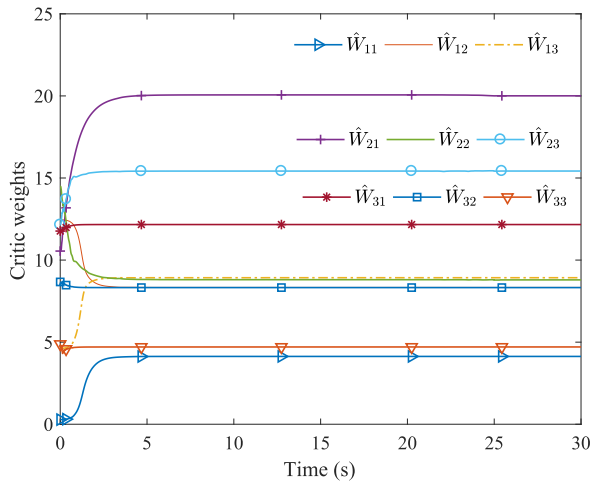
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG et al.: DFTCC OF NONLINEAR MULTIAGENT SYSTEMS

9

Fig. 6. Critic weights of Example 1.

Fig. 7. Control inputs of Example 1.

reconfigurable robots are adopted. The communication topology is the same as those of Example 1. The dynamics of each following agent with potential actuator faults is given as

$$M_i(q_i)\ddot{q}_i + C_i(q_i, \dot{q}_i)\dot{q}_i + G_i(q_i) = u_i - u_{if} \qquad (40)$$

where $q_i = [q_{i1}, q_{i2}] \in \mathbb{R}^2$ is the joint displacements of the agent i, $M_i(q_i) \in \mathbb{R}^{2\times 2}$ is the inertia matrix, $C_i(q_i, \dot{q}_i) \in \mathbb{R}^{2\times 2}$ is the Coriolis and centripetal force, $G_i(q_i) \in \mathbb{R}^2$ is the gravity term, and $u_{if} \in \mathbb{R}^2$ is the unknown actuator faults. The system functions and the actuator faults of each agent are provided at
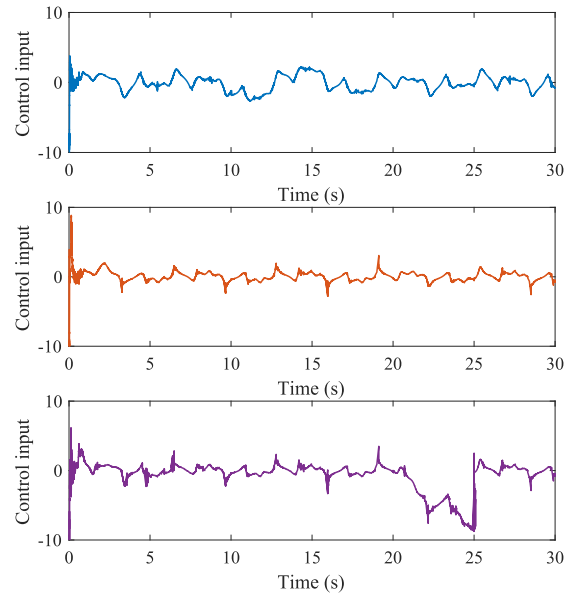
the bottom of the page. In this case, we assume that agents 1 and 2 are faulty, and agent 3 is fault-free.

The reference trajectory of the leader is selected as

$$r(t) = \begin{bmatrix} 0.6\cos(3t) - 0.4\sin(4t) \\ 0.6\cos(2t) - 0.2\sin(t) \end{bmatrix}. \qquad (41)$$

Let $Q_{ii} = 150I_4$, $R_{ii} = 0.1I_2$, $R_{ij} = 0.01I_2$, $\rho = 1$, $L_2 = 100$, the local fault observer gain be $L_1 = 100$, the initial states

Agent 1:

$$M_1(q_1) = \begin{bmatrix} 0.17 - 0.1166\cos^2(q_{12}) & -0.06\cos(q_{12}) \\ -0.06\cos(q_{12}) & 0.1233 \end{bmatrix}, \quad C_1(q_1, \dot{q}_1) = \begin{bmatrix} 0.1166\sin(2q_{12})\dot{q}_{12} & 0.06\sin(q_{12})\dot{q}_{12} \\ 0.06\sin(q_{12})\dot{q}_{12} - 0.0583\sin(q_{12})\dot{q}_{11} & 0.06\sin(q_{12})\dot{q}_{11} \end{bmatrix}$$

$$\mathcal{G}_1(q_1) = \begin{bmatrix} 0 \\ -5.88\cos(q_{12}) \end{bmatrix}$$

Agent 2:

$$M_2(q_2) = \begin{bmatrix} 0.17 - 0.1166\cos^2(q_{22}) & -0.06\cos(q_{22}) \\ -0.06\cos(q_{22}) & 0.1233 \end{bmatrix}, \quad C_2(q_2, \dot{q}_2) = \begin{bmatrix} 0.1166\sin(2q_{22})\dot{q}_{22} & 0.06\sin(q_{22})\dot{q}_{22} \\ 0.06\sin(q_{22})\dot{q}_{22} - 0.0583\sin(q_{22})\dot{q}_{21} & 0.06\sin(q_{22})\dot{q}_{21} \end{bmatrix}$$

$$\mathcal{G}_2(q_2) = \begin{bmatrix} -5.88\cos(q_{21})\sin(q_{22}) + 3.92\sin(q_{21}) \\ -5.88\cos(q_{22}) \end{bmatrix}$$

Agent 3:

$$M_3(q_3) = \begin{bmatrix} 0.36\cos(q_{32}) - 0.6066 & -0.18\cos(q_{32}) + 0.1233 \\ -0.18\cos(q_{32}) + 0.1233 & 0.1233 \end{bmatrix}, \quad C_3(q_3, \dot{q}_3) = \begin{bmatrix} -0.36\sin(q_{32})\dot{q}_{32} & -0.18\sin(q_{32})\dot{q}_{32} \\ 0.18\sin(q_{32})(\dot{q}_{31} - \dot{q}_{32}) & 0.18\sin(q_{32})\dot{q}_{31} \end{bmatrix}$$

$$\mathcal{G}_3(q_3) = \begin{bmatrix} -5.88\sin(q_{31} + q_{32}) - 17.64\sin(q_{31}) \\ -5.88\sin(q_{31} + q_{32}) \end{bmatrix}$$

$$u_{1f} = \begin{cases} [1 + \sin(3t/2\pi) + \cos(3t); 1 + \cos(t/\pi) + \sin(3t)], & 10s \le t \le 20s \\ [0; 0], & \text{others} \end{cases}$$

$$u_{2f} = \begin{cases} [3 + 5\sin(3t/2\pi) - \sin(2t); 2 + 5\cos(5t/2\pi) + \sin(2t)], & 20s \le t \le 25s \\ [0; 0], & \text{others} \end{cases}$$
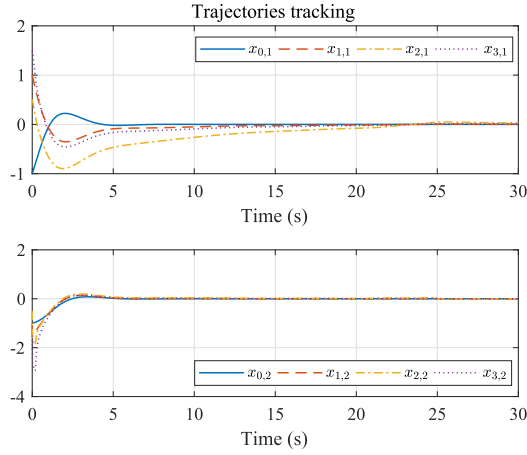
$$u_{3f} = [0; 0], \quad 0 \le t \le 30s$$

Fig. 8.   Consensus under convergence trajectories.
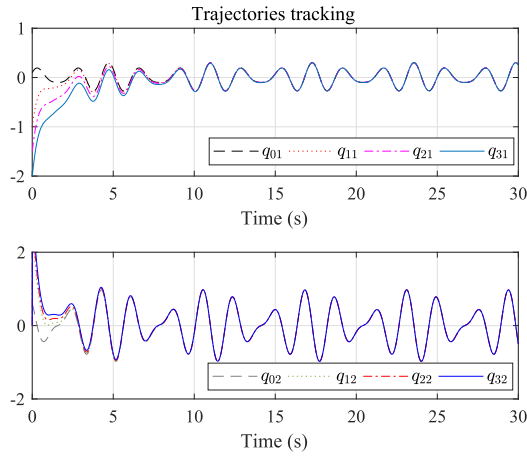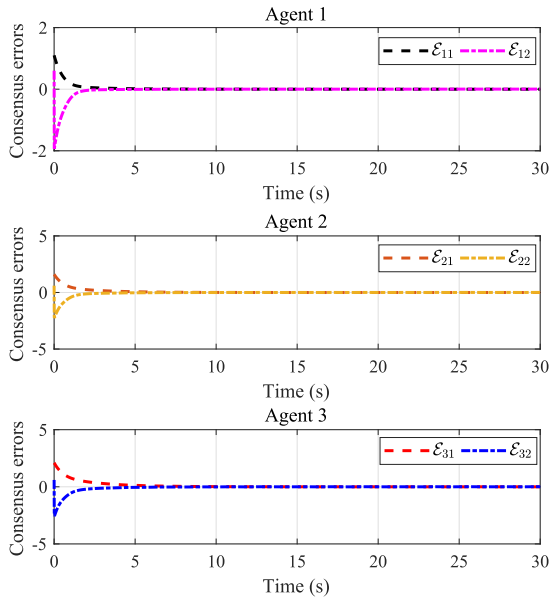
Fig. 9.   Trajectories tracking of Example 2.

Fig. 10.   Consensus errors of Example 2.

Fig. 11.   Critic weights of Example 2.
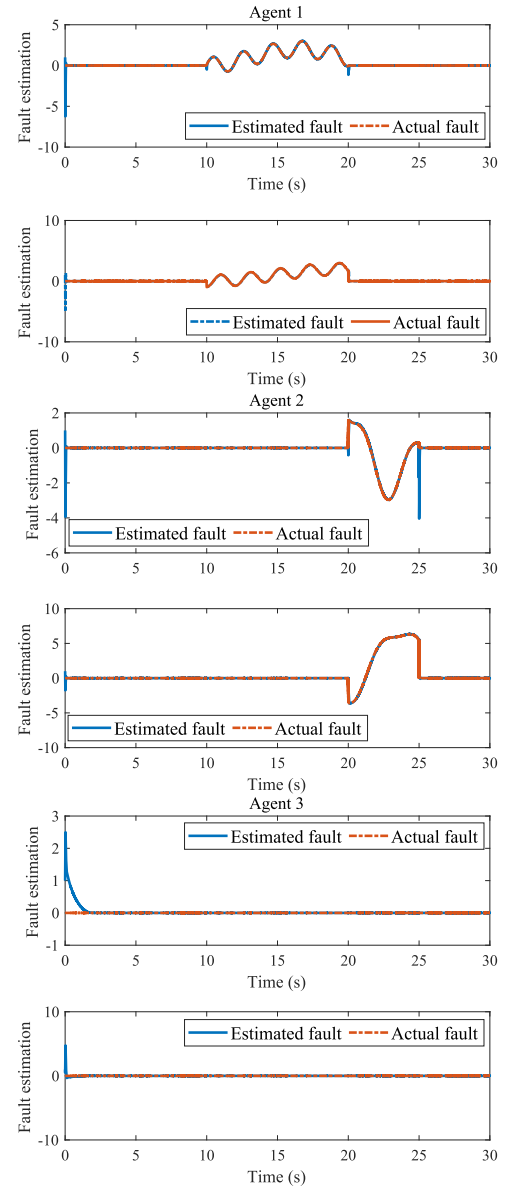
Fig. 12.   Fault estimation of each agent of Example 2.

of the agents be $x_1 = [-1, 1, 0, 0]^\mathsf{T}$, $x_2 = [-1.5, 1.5, 0, 0]^\mathsf{T}$ and $x_3 = [-2, 2, 0, 0]^\mathsf{T}$, the initial states of the observers be $\hat{x}_1 = \hat{x}_2 = \hat{x}_3 = [2, 1.5, -1, -0.5]^\mathsf{T}$, the activation function

of the critic NN be $\varphi_{ic} = [\mathcal{E}_{i,1}^2, \mathcal{E}_{i,1}\mathcal{E}_{i,3}, \mathcal{E}_{i,3}^2, \mathcal{E}_{i,2}^2, \mathcal{E}_{i,2}\mathcal{E}_{i,4}, \mathcal{E}_{i,4}^2]$, and the learning rate of the critic NN be $\alpha_c = 0.8$.

Simulation results are provided in Figs. 9–13. The trajectories of the leader and all following agents are given in

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

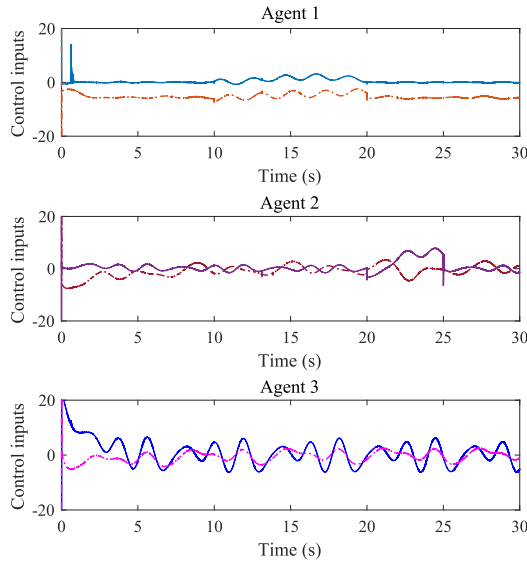ZHANG et al.: DFTCC OF NONLINEAR MULTIAGENT SYSTEMS

11

Fig. 13.   Control inputs of Example 2.

Fig. 9. We can find that all following agents can follow the leader within 10 s. Fig. 10 shows the consensus errors of all agents, where we can observe that the consensus errors converge to small region of zero as time increases. The weight evolution curves of the critic NNs are shown in Fig. 11, which converge to $\hat{W}_1 = [53.99, 107.88, 19.28, 25.46, 33.15, 67.50]^{\mathsf{T}}$, $\hat{W}_2 = [69.48, 5.08, 25.51, 22.23, 34.21, 92.50]^{\mathsf{T}}$ and $\hat{W}_3 = [57.18, 62.52, 77.67, 23.37, 38.54, 56.15]^{\mathsf{T}}$, respectively. Fig. 12 displays the fault estimation of each agent. We can observe that the local fault observers can estimate actuator faults accurately. The control inputs of all agents are illustrated in Fig. 13. It is found that when faults occur, the controllers present a quick response to resist faults so that the trajectory of the leader can be tracked even if the faults occur. From the above results and analysis, we conclude that all following agents with potential faults can successfully follow the leader. It means that the developed ADP-based DFTCC approach is effective.

Remark 7: In fact, the ADP-based DFTCC is developed based on the critic NN, which means that if the critic NN approximates the optimal value function successfully, the control input is obtained by calculating (28). Therefore, the implementation complexity of the controller mainly depends on the acquisition of the optimal weights of the critic NN. Once the optimal critic NN weights are obtained, the developed ADP-based distributed fault-tolerant controller can be employed on MASs directly. It is noted that the activation function, the learning rate and the initial weights all affect the training process of the critic NN, how to select them is a challenging problem. However, there is no guiding way to select them and researchers usually choose them by "trial and error" with repetitive simulations. Therefore, the selection of them is not unique and they can be selected to be different as long as the simulation results are satisfactory.

V. Conclusion

In this article, the FTCC problem of the MASs is addressed by proposing an ADP-based DFTCC approach. To begin with,

a local fault observer is designed to estimate the potential actuator faults of each agent. Then, the FTCC problem is converted to an optimal consensus control problem by designing a novel local value function that contains the estimated faults, the local consensus errors, and the control laws of the local agent and its neighbors. A critic-only structure is adopted to solve the coupled HJB equation. Moreover, the Lyapunov-based stability analysis demonstrates that the consensus errors are UUB. Finally, the effectiveness of the developed ADP-based DFTCC approach is verified by two simulation examples. The main contribution of this article lies in designing an ADP-based fault-tolerant controller for each agent such that all following agents with potential actuator faults can still follow the leader. The related future work is given as follows.

1) Assumption 4 is necessary for guaranteeing the stability of the local neighborhood consensus error of each agent, and we will try to relax it in our future work.
2) Since the convergence rate of the ADP algorithm is a challenging problem and significant in practical applications, we will try to investigate it in our future work.

References

[1] S. Li, R. Kong, and Y. Guo, "Cooperative distributed source seeking by multiple robots: Algorithms and experiments," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 6, pp. 1810–1820, Dec. 2014.

[2] J. Chen, X. Cao, P. Cheng, Y. Xiao, and Y. Sun, "Distributed collaborative control for industrial automation with wireless sensor and actuator networks," *IEEE Trans. Ind. Electron.*, vol. 57, no. 12, pp. 4219–4230, Dec. 2010.

[3] Q. Wei, D. Liu, G. Shi, and Y. Liu, "Multibattery optimal coordination control for home energy management systems via distributed iterative adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 62, no. 7, pp. 4203–4214, Jul. 2015.

[4] H. Yan, F. Qian, H. Zhang, F. Yang, and G. Guo, "H_∞ fault detection for networked mechanical spring-mass systems with incomplete information," *IEEE Trans. Ind. Electron.*, vol. 63, no. 9, pp. 5622–5631, Sep. 2016.

[5] R. W. Beard, T. W. McLain, M. A. Goodrich, and E. P. Anderson, "Coordinated target assignment and intercept for unmanned air vehicles," *IEEE Trans. Robot. Autom.*, vol. 18, no. 6, pp. 911–922, Dec. 2002.

[6] Y. Wang, Y. Song, and F. L. Lewis, "Robust adaptive fault-tolerant control of multi-agent systems with uncertain nonidentical dynamics and undetectable actuation failures," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3978–3988, Jun. 2015.

[7] D. Yao, H. Li, R. Lu, and Y. Shi, "Distributed sliding-mode tracking control of second-order nonlinear multiagent systems: An event-triggered approach," *IEEE Trans. Cybern.*, vol. 50, no. 9, pp. 3892–3902, Sep. 2020.

[8] C. Chen, K. Xie, F. L. Lewis, S. Xie, and A. Davoudi, "Fully distributed resilience for adaptive exponential synchronization of heterogeneous multiagent systems against actuator faults," *IEEE Trans. Autom. Control*, vol. 64, no. 8, pp. 3347–3354, Aug. 2019.

[9] M. I. Abouheaf, F. L. Lewis, K. G. Vamvoudakis, S. Haesaert, and R. Babuska, "Multi-agent discrete-time graphical games and reinforcement learning solutions," *Automatica*, vol. 50, no. 12, pp. 3038–3053, 2014.

[10] Q. Wei, D. Liu, and F. L. Lewis, "Optimal distributed synchronization control for continuous-time heterogeneous multi-agent differential graphical games," *Inf. Sci.*, vol. 317, pp. 96–113, Oct. 2015.

[11] H. Zhang, J. Zhang, G.-H. Yang, and Y. Luo, "Leader-based optimal coordination control for the consensus problem of multiagent differential games via fuzzy adaptive dynamic programming," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 1, pp. 152–163, Feb. 2015.

[12] W. Xiao, L. Cao, H. Li, and R. Lu, "Observer-based adaptive consensus control for nonlinear multi-agent systems with time-delay," *Sci. China Inf. Sci.*, vol. 63, no. 3, pp. 185–201, Mar. 2020.

[13] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.

[14] Q. Wei, D. Liu, Y. Liu, and R. Song, "Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 2, pp. 168–176, Apr. 2017.

[15] D. Liu, Y. Xu, Q. Wei, and X. Liu, "Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 36–46, Jan. 2018.

[16] D. Liu and Q. Wei, "Finite-approximation-error-based optimal control approach for discrete-time nonlinear systems," *IEEE Trans. Cybern.*, vol. 43, no. 2, pp. 779–789, Apr. 2013.

[17] Y. Yang, H. Zhu, Q. Zhang, B. Zhao, Z. Li, and D. C. Wunsch, "Sparse online kernelized actor-critic learning in reproducing kernel Hilbert space," *Artif. Intell. Rev.*, vol. 55, no. 1, pp. 23–58, Jan. 2022, doi: 10.1007/S10462-021-10045-9.

[18] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York, NY, USA: Van Nostrand Reinhold, 1992, ch. 13.

[19] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 3, pp. 621–634, Mar. 2014.

[20] D. Liu, D. Wang, and H. Li, "Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 2, pp. 418–428, Feb. 2014.

[21] D. Liu, X. Yang, D. Wang, and Q. Wei, "Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1372–1385, Jul. 2015.

[22] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[23] B. Zhao, Y. Zhang, and D. Liu, "Adaptive dynamic programming-based cooperative motion/force control for modular reconfigurable manipulators: A joint task assignment approach," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, May 11, 2022, doi: 10.1109/TNNLS.2022.3171828.

[24] Y. Zhang, B. Zhao, D. Liu, and S. Zhang, "Adaptive dynamic programming-based event-triggered robust control for multiplayer nonzero-sum games with unknown dynamics," *IEEE Trans. Cybern.*, early access, Jun. 10, 2022, doi: 10.1109/TCYB.2022.3175650.

[25] J. Li, H. Modares, T. Chai, F. L. Lewis, and L. Xie, "Off-policy reinforcement learning for synchronization in multiagent graphical games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2434–2445, Oct. 2017.

[26] H. Zhang, H. Jiang, Y. Luo, and G. Xiao, "Data-driven optimal consensus control for discrete-time multi-agent systems with unknown dynamics using reinforcement learning method," *IEEE Trans. Ind. Electron.*, vol. 64, no. 5, pp. 4091–4100, May 2017.

[27] J. Zhang, H. Zhang, and T. Feng, "Distributed optimal consensus control for nonlinear multiagent system with unknown dynamic," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3339–3348, Aug. 2018.

[28] W. Wang, X. Chen, H. Fu, and M. Wu, "Model-free distributed consensus control based on actor–critic framework for discrete-time nonlinear multiagent systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4123–4134, Nov. 2020.

[29] H. Yang et al., "Fault-tolerant cooperative control of multiagent systems: A survey of trends and methodologies," *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 4–17, Jan. 2020.

[30] B. Zhao, D. Liu, and Y. Li, "Online fault compensation control based on policy iteration algorithm for a class of affine non-linear systems with actuator failures," *IET Control Theory Appl.*, vol. 10, no. 15, pp. 1816–1823, Jun. 2016.

[31] B. Zhao, D. Liu, and Y. Li, "Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems," *Inf. Sci.*, vol. 384, pp. 21–33, Apr. 2017.

[32] H. Zhang, Y. Liang, H. Su, and C. Liu, "Event-driven guaranteed cost control design for nonlinear systems with actuator faults via reinforcement learning algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4135–4150, Nov. 2020.

[33] X. Xiao and X.-J. Li, "Adaptive dynamic programming method-based synchronisation control of a class of complex dynamical networks with unknown dynamics and actuator faults," *IET Control Theory Appl.*, vol. 12, no. 2, pp. 291–298, Jan. 2018.

[34] R. Konda, H. M. La, and J. Zhang, "Decentralized function approximated Q-learning in multi-robot systems for predator avoidance," *IEEE Robot. Autom. Lett.*, vol. 5, no. 4, pp. 6342–6349, Oct. 2020.

[35] W. Ren and R. W. Beard, "Decentralized scheme for spacecraft formation flying via the virtual structure approach," *J. Guid., Control, Dyn.*, vol. 27, no. 1, pp. 73–82, 2004.

[36] Z. Su, M. Dai, Q. Xu, and R. Li, "A novel resource allocation scheme with unmanned aerial vehicles in disaster relief networks," in *Proc. IEEE/CIC Int. Conf. Commun. China (ICCC)*, Chongqing, China, Aug. 2020, pp. 102–106.

[37] B. Zhao and D. Liu, "Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3054–3064, Apr. 2020.

[38] B. Zhao, D. Liu, and C. Alippi, "Sliding-mode surface-based approximate optimal control for uncertain nonlinear systems with asymptotically stable critic structure," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 2858–2869, Jun. 2021.

[39] B. Zhao, D. Liu, and C. Luo, "Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4330–4340, Oct. 2020.

[40] X. Yang and H. He, "Event-driven H_∞-constrained control using adaptive critic learning," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 4860–4872, Oct. 2021, doi: 10.1109/TCYB.2020.2972748.

[41] D. Wang, C. Mu, D. Liu, and H. Ma, "On mixed data and event driven design for adaptive-critic-based nonlinear H_∞ control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 993–1005, Apr. 2018.

[42] H. Zhang, K. Zhang, G. Xiao, and H. Jiang, "Robust optimal control scheme for unknown constrained-input nonlinear systems via a plug-n-play event-sampled critic-only algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 9, pp. 3169–3180, Sep. 2020.

[43] K. G. Vamvoudakis, "Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems," *IEEE/CAA J. Autom. Sinica*, vol. 1, no. 3, pp. 282–293, Jul. 2014.

[44] Y. Zhang, B. Zhao, D. Liu, and S. Zhang, "Event-triggered control of discrete-time zero-sum games via deterministic policy gradient adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 8, pp. 4823–4835, Aug. 2022.

[45] M. Lin, B. Zhao, and D. Liu, "Policy gradient adaptive critic designs for model-free optimal tracking control with experience replay," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 6, pp. 3692–3703, Jun. 2022.

Yongwei Zhang (Member, IEEE) received the B.S. degree in automation from the School of Electronic and Information Engineering, Jiaying University, Meizhou, China, in 2016, and the Ph.D. degree in control science and engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2021.

He is currently a Post-Doctoral Fellow with the Guangdong University of Technology. His current research interests include adaptive dynamic programming, optimal control, and multiagent systems.

Bo Zhao (Senior Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control science and engineering from Jilin University, Changchun, China, in 2009 and 2014, respectively.

He was a Post-Doctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2014 to 2017. Then, he joined the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, from 2017 to 2018. He is currently an Associate Professor with the School of Systems Science, Beijing Normal University, Beijing. He was selected as the Beijing Normal University Tang Scholar in 2021. He has authored or coauthored over 120 journal and conference papers and authorized four patents. His research interests include adaptive dynamic programming, robot control, fault diagnosis and tolerant control, optimal control, and artificial intelligence-based control.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG et al.: DFTCC OF NONLINEAR MULTIAGENT SYSTEMS 13

Dr. Zhao is a member of Asian-Pacific Neural Network Society (APNNS) and the Chinese Association for Artificial Intelligence (CAAI). He received the Best Paper Award of the 2022 IEEE 11th Data Driven Control and Learning Systems Conference. He is the Secretary General of Adaptive Dynamic Programming and Reinforcement Learning Technical Committee of Chinese Automation Association (CAA). He serves as an Associate Editor for the IEEE Transactions on Systems, Man and Cybernetics: Systems and *Neurocomputing*, an Early Career Advisory Board Member of IEEE/CAA Journal of Automatica Sinica, and a Guest Editor for *Complex and Intelligent Systems*, and also served many academic conferences.

Derong Liu (Fellow, IEEE) received the B.S. degree in mechanical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1982, the M.S. degree in automatic control theory and applications from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1987, and the Ph.D. degree in electrical engineering from University of Notre Dame, Notre Dame, IN, USA, in 1994.

He was a Product Design Engineer with the China North Industries Corporation, Jilin, China, from 1982 to 1984. He was an Instructor with the Graduate School of Chinese Academy of Sciences, Beijing, from 1987 to 1990. He was a Staff Fellow with the General Motors Research and Development Center, from 1993 to 1995. He was an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, from 1995 to 1999. He joined the University of Illinois at Chicago, Chicago, IL, USA, in 1999, and became a Full Professor of electrical and computer engineering and computer Science in 2006. He was selected for the 100 Talents Program by the Chinese Academy of Sciences in 2008, and he served as the Associate Director for the State Key Laboratory of Management and Control for Complex Systems at the Institute of Automation, from 2010 to 2016. He is currently a Full Professor with the School of Automation, Guangdong University of Technology. He has published 13 books and 260 articles in international journals.

Dr. Liu is a fellow of the International Neural Network Society and the International Association for Pattern Recognition. He is a member of Academia Europaea (The Academy of Europe). He was elected three times an AdCom member of the IEEE Computational Intelligence Society in 2006, 2015, and 2022, respectively. He was the Editor-in-Chief of IEEE Transactions on Neural Networks and Learning Systems, from 2010 to 2015. He was elected twice as a Distinguished Lecturer of the IEEE Computational Intelligence Society in 2012 and 2016, respectively. He served as a member for the Council of International Federation of Automatic Control from 2014 to 2017. He served as the President for Asia Pacific Neural Network Society in 2018.

He was a General Chair of the 2014 IEEE World Congress on Computational Intelligence, the 2016 World Congress on Intelligent Control and Automation, and the 2017 International Conference on Neural Information Processing. He received the Faculty Early Career Development Award from the National Science Foundation in 1999, the University Scholar Award from University of Illinois from 2006 to 2009, the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008, and the Outstanding Achievement Award from Asia Pacific Neural Network Assembly in 2014. He received the International Neural Network Society's Gabor Award in 2018, the IEEE Transactions on Neural Networks and Learning Systems Outstanding Paper Award in 2018; the IEEE Systems, Man and Cybernetics Society Andrew P. Sage Best Transactions Paper Award in 2018; the IEEE/CCA J. Automatica Sinica Hsue-Shen Tsien Paper Award in 2019. He is recipient of the IEEE CIS Neural Network Pioneer Award in 2022. He has been named a highly cited researcher consecutively for five years from 2017 to 2021 by Clarivate. He was a plenary/keynote speaker at 32 international conferences. He is currently the Editor-in-Chief of *Artificial Intelligence Review*, the Deputy Editor-in-Chief of the IEEE/CAA Journal of Automatica Sinica, the Deputy Editor-in-Chief of the *CAAI Transactions on Artificial Intelligence*, and the Chair of the IEEE Guangzhou Section.
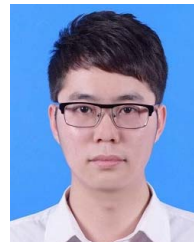
Shunchao Zhang received the B.S. degree in measurement and control technology and instruments from the School of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan, China, in 2016, the M.S. degree in control engineering and the Ph.D. degree in control science and engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2019 and 2022, respectively, where he is currently pursuing the Ph.D. degree in control science and engineering with the School of Automation.

He is currently a Lecturer with the School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou. His current research interests include optimal control and adaptive dynamic programming.

Adaptive Dynamic Programming-Based Event-Triggered Robust Control for Multiplayer Nonzero-Sum Games With Unknown Dynamics

Yongwei Zhang , *Member, IEEE*, Bo Zhao , *Senior Member, IEEE*,
Derong Liu , *Fellow, IEEE*, and Shunchao Zhang

Abstract—In this article, the event-triggered robust control of unknown multiplayer nonlinear systems with constrained inputs and uncertainties is investigated by using adaptive dynamic programming. To relax the requirement of system dynamics, a neural network-based identifier is constructed by using the system input-output data. Subsequently, by designing a nonquadratic value function, which contains the bounded functions, the system states, and the control inputs of all players, the event-triggered robust stabilization problem is converted into an event-triggered constrained optimal control problem. To obtain the approximate solution of the event-triggered Hamilton–Jacobi (HJ) equation, a critic network for each player is established with a novel weight updating law to relax the persistence of excitation condition based on the experience replay technique. Furthermore, according to the Lyapunov stability theorem, the present event-triggered robust optimal control ensures the multiplayer system to be uniformly ultimately bounded. Finally, two simulation examples are employed to show the effectiveness of the present method.

Index Terms—Adaptive dynamic programming (ADP), event-triggered control (ETC), multiplayer nonzero-sum games (MNSG), neural networks (NNs), robust control.

I. Introduction

WITH the rapid development of modern industries, control systems are becoming more and more complex

and uncertain, which may severely degrade the control performance or even lead to system's instability. Much effort has been devoted to the robustness of such systems. In recent decades, the robustness has attracted extensive attention in designing optimal control systems. In [1], the robust adaptive dynamic programming (ADP), which combined backstepping, robust redesign, and small-gain technique, was proposed to solve optimal control problems for uncertain nonlinear systems.

For the purpose of designing optimal controllers for nonlinear systems, it is required to solve the Hamilton–Jacobi–Bellman (HJB) equation, which is difficult or impossible due to its high nonlinearities. Fortunately, ADP, which was put forward by Werbos, is an effective approach to deal with this difficulty [2]–[8]. During the past decade, extensive ADP-based literature has been reported to address varieties of control problems of discrete-time (DT) systems [9]–[11] and continuous-time (CT) systems [12], [13] with trajectory tracking [14], [15], input constraints [16], fault tolerance [17], and so on. For robust stabilization problems, several ADP-based approaches were also developed. In [19], the robust controller for CT nonlinear systems with input constraints was designed by using reinforcement-learning (RL)/ADP approach. A suitable value function was selected to cope with the constrained input and the matched perturbation. For systems with unmatched disturbances, in [18], a robust control approach was proposed for nonlinear affine systems via ADP. By constructing an auxiliary system and a modified value function, the robust stabilization problem was converted into an optimal regulation problem. In [20], a robust control approach was proposed to deal with the general uncertainties. Through system transformation, an optimal controller of the nominal system was designed to stabilize the original system. Moreover, the uniform ultimate boundedness (UUB) was analyzed for both the nominal plant and the original uncertain system. However, the aforementioned robust control strategies were proposed based on the time-triggered mechanism with heavy computational and communication burden since they require data transmission at every sampling instant.

Different from time-triggered control schemes, the event-triggered control (ETC) updates system states and executes with a proper triggering condition; thus, it requires less sampling instants and less computational and communication resources [21], [22]. Thus, the ADP-based ETC (ADPETC)

has become a hot research topic recently. In [25], the event-triggered neurodynamic programming with an actor-critic (AC) framework was applied to acquire an optimal controller for DT nonlinear systems. In [23], the decentralized trajectory tracking control problem for modular reconfigurable robots was investigated by using the event-triggered method. In [16], a robust ETC (RETC) strategy for uncertain nonlinear CT systems with constrained input was proposed by using an adaptive critic structure. In [24], a decentralized ETC approach was presented for CT nonlinear interconnected systems. In addition, ADPETC methods were applied to network control systems [26] and power systems [27]. It should be noticed that the controllers developed in aforementioned works were updated at the triggering moments, hence, the computational and communication burden are alleviated.

In the aforementioned literature, single controllers with centralized control structure are employed to drive the systems that are not sufficiently large scale and complex. However, for complex and large-scale systems, such as communication networks, power systems, and networked control systems, multiple controllers, which can be regarded as multiplayers, are necessarily required to perform satisfactory control performance [28], [29]. The optimal control problem of multiplayer systems can be regarded as a nonzero-sum game, which presents both competitive and cooperative relationships, and its objective is to generate a set of control policies to approach a Nash equilibrium, which not only minimize the value function for each player but also guarantee the system to be stable. In [30], the game theory and the optimal control theory were integrated to address the DT multiplayer nonzero-sum game (MNSG) problem with the AC structure. The policy iteration algorithm was adopted to acquire a series of control laws to minimize the value function for each player. For CT MNSG problems, in [31], an ADPETC scheme was developed. The value function and the control strategy of each player were approximated by a citric neural network (NN).

It is worth pointing out that most of the existing results are developed for nominal multiplayer systems only. However, multiplayer systems are large scale, which indicates that their mathematical models are difficult to establish. Even if the mathematical model is obtained, dynamical uncertainties inevitably exist. Moreover, due to the physical characteristics of the actuator, the amplitude of its input or output is usually limited to result in a decreased execution ability. Consequently, the control performance may be reduced and even the system stability may get compromised. To the best of our knowledge, input constraints and system uncertainties have not been considered simultaneously in previous works. In addition, the existing time-triggered control approaches require plenty of computational and communication resources since the controllers are updated at every sampling instant. Based on the above discussion, it is urgent to investigate the robust control problem of multiplayer systems with input constraints and dynamical uncertainties based on the event-triggered mechanism. The main challenges are as follows: 1) a suitable value function needs to be constructed for each player to deal with input constraints and dynamical uncertainties; 2) the coupled Hamilton–Jacobi (HJ) equation is necessary but difficult to

solve to obtain the Nash equilibrium; and 3) to reduce the computational and communication burdens, an event-triggering condition, which is suitable for multiple controllers, needs to be designed. These motivate our research.

To tackle this problem, an ADP-based RETC approach is presented. To begin with, an NN-based identifier is established to estimate the unknown system dynamics. Then, critic NNs are constructed to approximate the solution of HJ equation of each player, and the experience replay (ER) technique is adopted to remove the persistence of excitation (PE) condition. Moreover, a novel event-triggering condition is derived based on Lyapunov's direct method. Consequently, the developed ADP-based RETC approach reduces the computational and communication burden in contrast to time-triggered ADP-based methods. The novelties and contributions of this article are presented as follows.

1) Different from existing methods [16], [33], which addressed the robust control problem for nonlinear systems with single controller only, this article develops an ADP-based RETC approach for multiplayer nonlinear systems. By system transformation and designing a modified nonquadratic value function for each player, the robust stabilization problem is converted to a constrained optimal control problem.

2) Unlike existing results [29], [35], which developed time-triggered control methods for multiplayer systems, this article investigates the MNSG problem with the event-triggered mechanism. It reduces the computational and communication burden in two aspects, that is: a) the developed robust controllers are updated at triggering moments and b) the value function of each player is approximated by critic NN.

3) An NN-based identifier is established to estimate the unknown system dynamics by adopting the measured system data. Moreover, the ER method, which removes the PE condition, is employed to design novel weight updating laws. Furthermore, under a new event-triggering condition, the critic NN weight estimate error dynamics and the multiplayer system are both guaranteed to be UUB.

The remainder of this article is organized as follows. In Section II, the problem statement is presented. In Section III, the unknown dynamics is reconstructed by the NN-based identifier, and the ADP-based RETC method is designed. Then, the NN implementation and the stability analysis are given. Section IV provides simulation results of the developed control method. In Section V, corresponding conclusions are given.

II. PROBLEM STATEMENT

Consider the unknown multiplayer CT nonlinear systems with uncertainties as

$$\dot{x}(t) = \mathcal{F}(x(t)) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x(t))\big(u_j(t) + \Lambda_j(x(t))\big) \qquad (1)$$

where $x(t) \in \mathbb{R}^n$ is the system state, $u_j(t) = [u_{j1}, u_{j2}, \ldots, u_{jm_j}]^{\mathsf{T}} \in \mathbb{R}^{m_j}$ is the control input of the jth player and satisfies $|u_{j\kappa}| \leq \bar{u}_{j\kappa}, \kappa = 1, 2, \ldots, m_j$, where

$\bar{u}_{j\kappa} > 0$ is the bound, $\Lambda_j(x(t)) \in \mathbb{R}^{m_j}$ represents the norm-bounded uncertainty, $\|\Lambda_j(x)\| \leq \eta_{jM}(x)$ with $\eta_{jM}(x)$ is a known positive function, $\eta_{jM}(0) = 0$, and $\mathcal{N}$ is the number of controllers. $\mathcal{F}(x) \in \mathbb{R}^n$ and $\mathcal{G}_j(x) \in \mathbb{R}^{n \times m_j}$ are unknown nonlinear system functions. Assuming that $\mathcal{F}(x)$ and $\mathcal{G}_j(x)$ are Lipschitz continuous on a compact set Ω with $\mathcal{F}(0) = 0$.

The nominal system corresponding to system (1) can be given as

$$\dot{x}(t) = \mathcal{F}(x(t)) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x(t))u_j(t). \tag{2}$$

Denote $u_{-i} = \{u_j : j = 1, 2, \ldots, \mathcal{N}, j \neq i\}$ as the supplementary set of player i. The value function for each player is defined as

$$\mathbb{V}_i(x) = \int_t^\infty \left(\xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x(v)) + \mathcal{C}_i(x(v), u_i(v), u_{-i}(v)) \right) dv$$
$$i = 1, 2, \ldots, \mathcal{N} \tag{3}$$

where ξ is a positive constant, $\mathcal{C}_i(x(v), u_i(v), u_{-i}(v)) = x^\mathsf{T} \mathcal{Q}_i x + \mathcal{W}_i(\mathcal{U}_\mathcal{N}) \geq 0$ is the utility function, where $\mathcal{Q}_i \in \mathbb{R}^{n \times n}$ is a symmetric positive-definite matrix and $\mathcal{U}_\mathcal{N} = [u_1, \ldots, u_\mathcal{N}]$. In order to cope with the optimal control for systems with constraints, inspired by [16] and [39], $\mathcal{W}_i(\mathcal{U}_\mathcal{N})$ is chosen as

$$\mathcal{W}_i(\mathcal{U}_\mathcal{N}) = 2 \sum_{j=1}^{\mathcal{N}} \int_0^{u_j} \delta^{-\mathsf{T}}\left(\Psi_j^{-1} s\right) ds \tag{4}$$

where $\delta(\cdot) \in \mathbb{R}^{m_j}$ is a monotonic odd function satisfying $|\delta_q(\cdot)| < 1$, $q = 1, 2, \ldots, m_j$ with $\delta^{-1}(\cdot) = [\delta_1^{-1}(\cdot), \ldots, \delta_{m_j}^{-1}(\cdot)]^\mathsf{T}$, and $\Psi_j = \text{diag}\{\bar{u}_{j1}, \ldots, \bar{u}_{jm_j}\}$ is a diagonal matrix that contains all the bounds. In this article, $\delta_q(\cdot)$ is selected as the hyperbolic tangent function, that is, $\delta_q(\cdot) = \tanh(\cdot)$.

Definition 1 ([28], [30], [36]): An $\mathcal{N}$-tuple of admissible policies $\{u_1^*, \ldots, u_\mathcal{N}^*\}$ is called the Nash equilibrium for the $\mathcal{N}$-player nonzero-sum game, if for any u_i and $1 \leq i \leq \mathcal{N}$, the following inequality is satisfied:

$$\mathbb{V}_i\left(u_1^*, \ldots, u_i^*, \ldots, u_\mathcal{N}^*\right) \leq \mathbb{V}_i\left(u_1^*, \ldots, u_i, \ldots, u_\mathcal{N}^*\right). \tag{5}$$

Denote $\Re(\Omega)$ as a set of admissible control. Assuming (3) is continuously differentiable, the Hamiltonian of system (2) is defined as

$$\mathcal{H}_i(x, \nabla\mathbb{V}_i(x), \mathcal{U}_\mathcal{N}) = \nabla\mathbb{V}_i^\mathsf{T}(x)\left(\mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x)u_j\right)$$
$$+ \xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x) + \mathcal{C}_i(x, u_i, u_{-i}). \tag{6}$$

Thus, the optimal value function

$$\mathbb{V}_i^*(x) = \min_{u_i \in \Re(\Omega)} \int_t^\infty \left(\xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x(v)) + \mathcal{C}_i(x(v), u_i(v), u_{-i}(v)) \right) dv \tag{7}$$

satisfies the HJ equation

$$\min_{u_i \in \Re(\Omega)} \mathcal{H}_i(x, \nabla\mathbb{V}_i^*(x), \mathcal{U}_\mathcal{N}) = 0. \tag{8}$$

Then, the optimal control law for the ith player can be obtained by differentiating (6) with respect to u_i as

$$u_i^*(x) = \arg\min_{u_i \in \Re(\Omega)} \mathcal{H}_i(x, \nabla\mathbb{V}_i^*(x), \mathcal{U}_\mathcal{N})$$
$$= -\Psi_i \delta\left(\frac{1}{2}\mathcal{G}_i^\mathsf{T}(x)\nabla\mathbb{V}_i^*(x)\right). \tag{9}$$

According to (6) and (9), we can obtain

$$0 = \mathcal{H}_i(x, \nabla\mathbb{V}_i^*(x), \mathcal{U}_\mathcal{N}^*)$$
$$= \nabla\mathbb{V}_i^{*\mathsf{T}}(x)\left(\mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x)u_j^*\right)$$
$$+ \xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x) + x^\mathsf{T}\mathcal{Q}_i x + \mathcal{W}_i(\mathcal{U}_\mathcal{N}^*). \tag{10}$$

It is noticed that the time-triggered HJ equation (10) is solved with huge amount of transmitted data and results in enormous computational and communication burden in existing ADP-based control methods. In the next section, the ADP-based RETC approach is developed to overcome these shortcomings.

Remark 1: Inspired by [16] and [39], the nonquadratic form with a monotonic odd function $\tanh(\cdot)$ was adopted to deal with the constrained input. It is noticed that existing approaches [16] and [39] have considered single controller only; however, this article aims to develop multiple robust controllers for multiplayer systems with dynamical uncertainties. Therefore, the designed value function contains the bound functions, the system states, and the control inputs of all players, which reflects the uncertainties, the regulation, and the control simultaneously. By using this modified value function, the robust stabilization problem is converted to a constrained optimal control problem.

Remark 2: Different from zero-sum games and full cooperative games, the MNSG reflects both competitive and cooperative relationships, that is, all players have their individual control goal and a common goal [28], [31], [40], [41]. For MNSG, it aims at obtaining a set of control policies to find the Nash equilibrium, which not only minimize the value function for each player but also guarantee the system to be stable. On the one hand, the competition illustrates that each player expects to achieve their own target, that is, minimize individual value function (3). However, since the value function of one player contains the control laws of other players, the realization of optimal control goal of one player will affect the control performance of others. Therefore, there exists competitive relationship among players. On the other hand, the cooperation lies in that the control policies generated by all players make contribution to stabilize the entire system. It implies that each player cooperates with others to achieve the same goal.

III. Event-Triggered Robust Controller Design

A. Event-Triggered Robust Controller Design

The sequence of triggering instants is defined as $\{S_\vartheta\}_{\vartheta=0}^{\infty}$, where S_ϑ is the ϑth sampling instant. Under the event-triggered framework, for $\forall t \in [S_\vartheta, S_{\vartheta+1}), \vartheta \in \mathbb{N}$, the event-triggered error $e_\vartheta(t)$ is defined as

$$e_\vartheta(t) = \bar{x}_\vartheta(t) - x(t) \tag{11}$$

where $\bar{x}_\vartheta(t) = x(S_\vartheta)$ is the sampled state. The event-triggered optimal control law is given as

$$u_i^*(\bar{x}_\vartheta, t) = u_i^*(\bar{x}_\vartheta)$$
$$= -\Psi_i \delta\left(\frac{1}{2}\mathcal{G}_i^\mathsf{T}(\bar{x}_\vartheta)\nabla\mathbb{V}_i^*(\bar{x}_\vartheta)\right), t \in [S_\vartheta, S_{\vartheta+1}). \tag{12}$$

According to (10), the event-triggered HJ equation is defined as

$$\mathcal{H}_i\big(x, \nabla\mathbb{V}_i^*(x), \mathcal{U}_\mathcal{N}^*(\bar{x}_\vartheta)\big)$$
$$= \nabla\mathbb{V}_i^{*\mathsf{T}}(x)\left(\mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)u_j^*(\bar{x}_\vartheta)\right)$$
$$+ \xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) + x^\mathsf{T}\mathcal{Q}_i x + \mathcal{W}_i\big(\mathcal{U}_\mathcal{N}^*(\bar{x}_\vartheta)\big) = 0 \tag{13}$$

where $\mathcal{U}_\mathcal{N}^*(\bar{x}_\vartheta) = [u_1^*(\bar{x}_\vartheta), \ldots, u_\mathcal{N}^*(\bar{x}_\vartheta)]$.

Assumption 1 [38]: $\mathcal{G}_j(x)$ and $\nabla\mathbb{V}_i^*(x)$ are norm-bounded, that is,

$$\left\|\mathcal{G}_j(x)\right\| \le \bar{\mathcal{G}}_j, \left\|\nabla\mathbb{V}_i^*(x)\right\| \le \bar{\mathbb{V}}_i \tag{14}$$

where $\bar{\mathcal{G}}_j$ and $\bar{\mathbb{V}}_i$ are positive constants.

Assumption 2 [21], [22]: The control law is Lipschitz continuous, that is

$$\left\|u_i^*(x) - u_i^*(\bar{x}_\vartheta)\right\|^2 \le \mathcal{L}_u\|e_\vartheta(t)\|^2 \tag{15}$$

where $\mathcal{L}_u$ is a positive constant.

Theorem 1: Consider the unknown multiplayer nonlinear system (1) with its nominal form (2), and the event-triggered optimal control law (12). The multiplayer system (1) is guaranteed to be UUB only if the following triggering condition:

$$\|e_\vartheta(t)\|^2 \le \frac{(1-\tau^2)\lambda_{\min}(\mathcal{Q}_i)\|x\|^2 + \big(\xi - \bar{\mathcal{G}}^2\big)\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x)}{\bar{\mathcal{G}}^2\mathcal{L}_u\mathcal{N}}$$
$$\triangleq \|e_T\|^2 \tag{16}$$

and the inequality $\xi > \bar{\mathcal{G}}^2$ hold, where $0 < \tau < 1$ and $\bar{\mathcal{G}} > 0$ are design parameters.

Proof: Select the Lyapunov function candidate as

$$\mathcal{L}_{T1} = \mathbb{V}_i^*(x). \tag{17}$$

Taking the time derivative of (17) along with the solutions of (1), we can obtain

$$\dot{\mathcal{L}}_{T1} = \nabla\mathbb{V}_i^{*\mathsf{T}}(x)\left(\mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)u_j^*(\bar{x}_\vartheta) + \sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)\Lambda_j(x)\right). \tag{18}$$

According to (10), we can obtain

$$\nabla\mathbb{V}_i^{*\mathsf{T}}(x)\mathcal{F}(x) = -\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x - \mathcal{W}_i(\mathcal{U}_\mathcal{N}^*)$$
$$- \nabla\mathbb{V}_i^{*\mathsf{T}}(x)\sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)u_j^*(x).$$

$$\dot{\mathcal{L}}_{T1} = -\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x - \mathcal{W}_i\big(\mathcal{U}_\mathcal{N}^*\big) + \nabla\mathbb{V}_i^{*\mathsf{T}}(x)\sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)\big(u_j^*(\bar{x}_\vartheta) + \Lambda_j(x) - u_j^*(x)\big)$$

$$= -\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x - \mathcal{W}_i\big(\mathcal{U}_\mathcal{N}^*\big) + \frac{1}{2}\left\|\nabla\mathbb{V}_i^*(x)\right\|^2 + \frac{1}{2}\left\|\sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)\big(u_j^*(\bar{x}_\vartheta) + \Lambda_j(x) - u_j^*(x)\big)\right\|^2$$

$$\le -\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x + \frac{1}{2}\bar{\mathbb{V}}_i^2$$

$$+ \frac{1}{2}\begin{bmatrix} u_1^*(\bar{x}_\vartheta) + \Lambda_1(x) - u_1^*(x) \\ \vdots \\ u_\mathcal{N}^*(\bar{x}_\vartheta) + \Lambda_\mathcal{N}(x) - u_\mathcal{N}^*(x) \end{bmatrix}^\mathsf{T} [\mathcal{G}_1(x), \ldots, \mathcal{G}_\mathcal{N}(x)]^\mathsf{T}[\mathcal{G}_1(x), \ldots, \mathcal{G}_\mathcal{N}(x)]\begin{bmatrix} u_1^*(\bar{x}_\vartheta) + \Lambda_1(x) - u_1^*(x) \\ \vdots \\ u_\mathcal{N}^*(\bar{x}_\vartheta) + \Lambda_\mathcal{N}(x) - u_\mathcal{N}^*(x) \end{bmatrix}$$

$$\le -\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x + \frac{1}{2}\bar{\mathbb{V}}_i^2 + \frac{1}{2}\|\mathcal{G}\|^2\sum_{j=1}^{\mathcal{N}}\left\|u_j^*(\bar{x}_\vartheta) + \Lambda_j(x) - u_j^*(x)\right\|^2$$

$$\le -\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x + \frac{1}{2}\bar{\mathbb{V}}_i^2 + \bar{\mathcal{G}}^2\sum_{j=1}^{\mathcal{N}}\left\|u_j^*(\bar{x}_\vartheta) - u_j^*(x)\right\|^2 + \bar{\mathcal{G}}^2\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x)$$

$$\le -(1-\tau^2)\lambda_{\min}(\mathcal{Q}_i)\|x\|^2 + \bar{\mathcal{G}}^2\mathcal{L}_u\mathcal{N}\|e_\vartheta(t)\|^2 + \big(\bar{\mathcal{G}}^2 - \xi\big)\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - \tau^2\lambda_{\min}(\mathcal{Q}_i)\|x\|^2 + \frac{1}{2}\bar{\mathbb{V}}_i^2 \tag{19}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG *et al.*: ADP-BASED EVENT-TRIGGERED ROBUST CONTROL FOR MNSGs

5

Define $\mathcal{G} = [\mathcal{G}_1(x), \ldots, \mathcal{G}_{\mathcal{N}}(x)]^{\mathsf{T}}[\mathcal{G}_1(x), \ldots, \mathcal{G}_{\mathcal{N}}(x)]$ and it is assumed to be norm-bounded as $\|\mathcal{G}\| \leq \bar{\mathcal{G}}$. Then, (19) is obtained. If the triggering condition (16) holds and $\xi > \bar{\mathcal{G}}^2$ is satisfied, then (19), as shown at the bottom of the previous page becomes

$$\dot{\mathcal{L}}_{T1} \leq -\tau^2 \lambda_{\min}(\mathcal{Q}_i)\|x\|^2 + \frac{1}{2}\bar{\mathbb{V}}_i^2. \tag{20}$$

Thus, $\dot{\mathcal{L}}_{T1} \leq 0$ as long as the state x lies outside the compact set

$$\Omega_x = \left\{ x : \|x\| \leq \sqrt{\frac{\bar{\mathbb{V}}_i^2}{2\tau^2 \lambda_{\min}(\mathcal{Q}_i)}} \right\}. \tag{21}$$

■

B. System Identification via NN

In this section, an NN is adopted to estimate the unknown multiplayer system dynamics by utilizing measured input/output data. Inspired by [34], system (2) is reconstructed as

$$\dot{x} = \Pi^{\mathsf{T}} x + \omega_f^{\mathsf{T}} \sigma_f(x) + \sum_{j=1}^{\mathcal{N}} \omega_{\mathcal{G}j}^{\mathsf{T}} \sigma_{\mathcal{G}j}(x) u_j + \varepsilon(t) \tag{22}$$

where $\Pi \in \mathbb{R}^{n \times n}$, $\omega_f \in \mathbb{R}^{n \times n}$, and $\omega_{\mathcal{G}j} \in \mathbb{R}^{n \times n}$ are optimal NN weight matrices, $\sigma_f(x) \in \mathbb{R}^n$ represents the activation function, and $\sigma_{\mathcal{G}j}(x) = [\sigma_{\mathcal{G}j1}(\psi_1^{\mathsf{T}} x), \sigma_{\mathcal{G}j2}(\psi_2^{\mathsf{T}} x), \ldots, \sigma_{\mathcal{G}jn}(\psi_n^{\mathsf{T}} x)]^{\mathsf{T}} \in \mathbb{R}^{n \times m_j}$ is a matrix function with momotonically increasing function $\sigma_{\mathcal{G}jl}(\cdot) \in \mathbb{R}^{m_j}$ and constant matrix $\psi_l \in \mathbb{R}^{n \times m_j}$, $l = 1, 2, \ldots, n$. $\varepsilon(t) \in \mathbb{R}^n$ is the reconstruction error. Then, the NN-based identifier is

$$\dot{\hat{x}} = \hat{\Pi}^{\mathsf{T}}(t)\hat{x} + \hat{\omega}_f^{\mathsf{T}}(t)\sigma_f(\hat{x}) + \sum_{j=1}^{\mathcal{N}} \hat{\omega}_{\mathcal{G}j}^{\mathsf{T}}(t)\sigma_{\mathcal{G}j}(\hat{x})u_j + \upsilon \tag{23}$$

where $\hat{\Pi}(t) \in \mathbb{R}^{n \times n}$, $\hat{\omega}_f(t) \in \mathbb{R}^{n \times n}$, and $\hat{\omega}_{\mathcal{G}j}(t) \in \mathbb{R}^{n \times n}$ are the estimates of the corresponding weight matrices, $\upsilon \in \mathbb{R}^n$ is defined as $\upsilon = \theta\tilde{x}$, $\theta > 0$ is a design parameter, and $\tilde{x} \triangleq x - \hat{x}$ is the identification error. Then, the identification error dynamics is

$$\dot{\tilde{x}} = \Pi^{\mathsf{T}}\tilde{x} + \tilde{\Pi}^{\mathsf{T}}(t)\hat{x} + \tilde{\omega}_f^{\mathsf{T}}(t)\sigma_f(\hat{x})$$
$$+ \sum_{j=1}^{\mathcal{N}} \left(\tilde{\omega}_{\mathcal{G}j}^{\mathsf{T}}(t)\sigma_{\mathcal{G}j}(\hat{x}) + \omega_{\mathcal{G}j}^{\mathsf{T}}\left(\sigma_{\mathcal{G}j}(x) - \sigma_{\mathcal{G}j}(\hat{x})\right)\right)u_j$$
$$+ \omega_f^{\mathsf{T}}\left(\sigma_f(x) - \sigma_f(\hat{x})\right) + \varepsilon(t) - \theta\tilde{x} \tag{24}$$

where $\tilde{\Pi}(t) = \Pi - \hat{\Pi}(t)$, $\tilde{\omega}_f(t) = \omega_f - \hat{\omega}_f(t)$, and $\tilde{\omega}_{\mathcal{G}j}(t) = \omega_{\mathcal{G}j} - \hat{\omega}_{\mathcal{G}j}(t)$.

Assumption 3: The NN reconstruction error $\varepsilon(t)$ satisfies

$$\varepsilon^{\mathsf{T}}(t)\varepsilon(t) \leq \rho_1 \tilde{x}^{\mathsf{T}}\tilde{x} \tag{25}$$

where ρ_1 is a positive constant.

Assumption 4: The optimal NN weights are norm-bounded, that is

$$\|\omega_f\| \leq \bar{\omega}_f, \|\omega_{\mathcal{G}j}\| \leq \bar{\omega}_{\mathcal{G}j} \tag{26}$$

where $\bar{\omega}_f$ and $\bar{\omega}_{\mathcal{G}j}$ are positive constants.

Theorem 2: Consider the nominal system (2), Assumptions 3 and 4, and the reconstructed system dynamics (23). The identification error $\tilde{x}$ will reach zero if NN weights $\hat{\Pi}(t)$, $\hat{\omega}_f(t)$, and $\hat{\omega}_{\mathcal{G}j}(t)$ are updated by

$$\dot{\hat{\Pi}}(t) = \Xi_a \hat{x}\tilde{x}^{\mathsf{T}} \tag{27}$$
$$\dot{\hat{\omega}}_f(t) = \Xi_f \sigma_f(\hat{x})\tilde{x}^{\mathsf{T}} \tag{28}$$
$$\dot{\hat{\omega}}_{\mathcal{G}j}(t) = \Xi_{\mathcal{G}j} \sigma_{\mathcal{G}j}(\hat{x})u_j\tilde{x}^{\mathsf{T}} \tag{29}$$

where Ξ_a, Ξ_f, and $\Xi_{\mathcal{G}j}$ are symmetric positive-definite matrices.

Proof: The proof of Theorem 2 has been provided in [34, Th. 3.1], so the detail is omitted here.

According to Theorem 2, system (2) is expressed by

$$\dot{x} = \hat{\Pi}^{\mathsf{T}}x + \hat{\omega}_f \sigma_f(x) + \sum_{j=1}^{\mathcal{N}} \hat{\omega}_{\mathcal{G}j}^{\mathsf{T}}\sigma_{\mathcal{G}j}(x)u_j \tag{30}$$

where $\hat{\Pi}$, $\hat{\omega}_f$, and $\hat{\omega}_{\mathcal{G}j}$ are converged values of corresponding weights. Then, from (2) and (30), we have

$$\mathcal{F}(x) = \hat{\Pi}^{\mathsf{T}}x + \hat{\omega}_f \sigma_f(x), \quad \mathcal{G}_j(x) = \hat{\omega}_{\mathcal{G}j}^{\mathsf{T}}\sigma_{\mathcal{G}j}(x). \tag{31}$$

■

Remark 3: The designed NN-based identifier learns the unknown system dynamics offline. Compared with the online observer, more historical system data are available and used in the offline learning and the data usage efficiency is improved. Moreover, the identification error of the NN-based identifier will be asymptotically stable, rather than UUB [33].

C. Neural Network Implementation

In this section, critic NNs are constructed to approximate the solutions of event-triggered HJ equations. The optimal value function $\mathbb{V}_i^*(x)$ is approximated by NN as

$$\mathbb{V}_i^*(x) = \omega_{ci}^{\mathsf{T}}\sigma_c(x) + \varepsilon_{ci}(x) \tag{32}$$

where $\omega_{ci} \in \mathbb{R}^{\mathcal{L}_c}$ is the ideal weight vector, $\sigma_c(x) \in \mathbb{R}^{\mathcal{L}_c}$ is the activation function, $\mathcal{L}_c$ is the number of hidden layer neurons, and $\varepsilon_{ci}(x) \in \mathbb{R}$ is the approximation error. Then, according to (32), we have

$$\nabla\mathbb{V}_i^*(x) = \nabla\sigma_c^{\mathsf{T}}(x)\omega_{ci} + \nabla\varepsilon_{ci}(x). \tag{33}$$

The approximate value function is formulated as

$$\hat{\mathbb{V}}_i(x) = \hat{\omega}_{ci}^{\mathsf{T}}\sigma_c(x) \tag{34}$$

where $\hat{\omega}_{ci}$ is the estimate of ω_{ci}. Similarly, we have

$$\nabla\hat{\mathbb{V}}_i(x) = \nabla\sigma_c^{\mathsf{T}}(x)\hat{\omega}_{ci}. \tag{35}$$

Based on (12) and (33), the event-triggered optimal control law is

$$u_i(\bar{x}_\vartheta) = -\Psi_i\delta\left(\frac{1}{2}\mathcal{G}_i^{\mathsf{T}}(\bar{x}_\vartheta)\left(\nabla\sigma_c^{\mathsf{T}}(\bar{x}_\vartheta)\omega_{ci} + \nabla\varepsilon_{ci}(\bar{x}_\vartheta)\right)\right).$$

Based on (35), we have

$$\hat{u}_i(\bar{x}_\vartheta) = -\Psi_i\delta\left(\frac{1}{2}\mathcal{G}_i^{\mathsf{T}}(\bar{x}_\vartheta)\nabla\sigma_c^{\mathsf{T}}(\bar{x}_\vartheta)\hat{\omega}_{ci}\right). \tag{36}$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6

IEEE TRANSACTIONS ON CYBERNETICS

According to (6) and (36), the approximate Hamiltonian is

$$\hat{\mathcal{H}}_i\left(x, \hat{\mathcal{U}}_{\mathcal{N}}(\bar{x}_\vartheta), \hat{\omega}_{ci}\right)$$

$$= \xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x) + x^{\mathsf{T}} \mathcal{Q}_i x + \mathcal{W}_i\left(\hat{\mathcal{U}}_{\mathcal{N}}(\bar{x}_\vartheta)\right)$$

$$+ \underbrace{\hat{\omega}_{ci}^{\mathsf{T}} \nabla \sigma_c(x) \left(\mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x) \hat{u}_j(\bar{x}_\vartheta)\right)}_{\Upsilon_i} \triangleq e_{ci} \quad (37)$$

where $\hat{\mathcal{U}}_{\mathcal{N}}(\bar{x}_\vartheta) = [\hat{u}_1(\bar{x}_\vartheta), \ldots, \hat{u}_n(\bar{x}_\vartheta)]$. To derive the updating law of the critic NN, the objective function $E_{ci} = \frac{1}{2} e_{ci}^{\mathsf{T}} e_{ci}$ is minimized by the gradient descent algorithm as

$$\dot{\hat{\omega}}_{ci} = -\alpha_c \frac{1}{(1 + \Upsilon_i^{\mathsf{T}} \Upsilon_i)^2} \left(\frac{\partial E_{ci}}{\partial \hat{\omega}_{ci}}\right)$$

$$= -\frac{\alpha_c \Upsilon_i}{(1 + \Upsilon_i^{\mathsf{T}} \Upsilon_i)^2} \left(\hat{\omega}_{ci}^{\mathsf{T}} \Upsilon_i + \xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x) + x^{\mathsf{T}} \mathcal{Q}_i x \right.$$

$$\left. + \mathcal{W}_i\left(\hat{\mathcal{U}}_{\mathcal{N}}(\bar{x}_\vartheta)\right)\right) \quad (38)$$

where $\alpha_c > 0$ is the learning rate. For the purpose of relaxing the PE condition, ER technique is adopted. Inspired by [24], by using the historical system data, a new critic NN weight tuning rule is designed by

$$\dot{\hat{\omega}}_{ci} = -\frac{\alpha_c \Upsilon_i}{(1 + \Upsilon_i^{\mathsf{T}} \Upsilon_i)^2} \left(\hat{\omega}_{ci}^{\mathsf{T}} \Upsilon_i + \xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x) \right.$$

$$\left. + x^{\mathsf{T}} \mathcal{Q}_i x + \mathcal{W}_i\left(\hat{\mathcal{U}}_{\mathcal{N}}(\bar{x}_\vartheta)\right)\right)$$

$$- \sum_{d=1}^{\mathcal{N}_{\mathcal{D}}} \frac{\alpha_c \Upsilon_{id}}{(1 + \Upsilon_{id}^{\mathsf{T}} \Upsilon_{id})^2} \left(\hat{\omega}_{ci}^{\mathsf{T}} \Upsilon_{id} + \xi \sum_{j=1}^{\mathcal{N}} \eta_{jM}^2(x(t_d))\right.$$

$$\left. + x^{\mathsf{T}}(t_d) \mathcal{Q}_i x(t_d) + \mathcal{W}_i\left(\hat{\mathcal{U}}_{\mathcal{N}}(\bar{x}_\vartheta)\right)\right)$$

$$t_d \in [S_\vartheta, S_{\vartheta+1}) \quad (39)$$

where $d \in \{1, 2, \ldots, \mathcal{N}_{\mathcal{D}}\}$ is the index of stored data, and $\Upsilon_{id} = \nabla \sigma_c(x(t_d))(\mathcal{F}(x(t_d)) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x(t_d)) \hat{u}_j(\bar{x}_\vartheta))$.

Define the weight estimation error as $\tilde{\omega}_{ci} = \omega_{ci} - \hat{\omega}_{ci}$ and recall $\dot{\tilde{\omega}}_{ci} = -\dot{\hat{\omega}}_{ci}$. Inspired by [33], we have

$$\dot{\tilde{\omega}}_{ci} = -\alpha_c \left(\Phi_i \Phi_i^{\mathsf{T}} + \sum_{d=1}^{\mathcal{N}_{\mathcal{D}}} \Phi_{id} \Phi_{id}^{\mathsf{T}}\right) \tilde{\omega}_{ci} + \frac{\alpha_c \Phi_i}{1 + \Upsilon_i^{\mathsf{T}} \Upsilon_i} \varepsilon_{Hi}$$

$$+ \sum_{d=1}^{\mathcal{N}_{\mathcal{D}}} \frac{\alpha_c \Phi_{id}}{1 + \Upsilon_{id}^{\mathsf{T}} \Upsilon_{id}} \varepsilon_{Hid}, \quad t_d \in [S_\vartheta, S_{\vartheta+1}) \quad (40)$$

where

$$\varepsilon_{Hi} = -\nabla \varepsilon_{ci}^{\mathsf{T}}(x) \left(\mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x) \hat{u}_j(\bar{x}_\vartheta)\right)$$
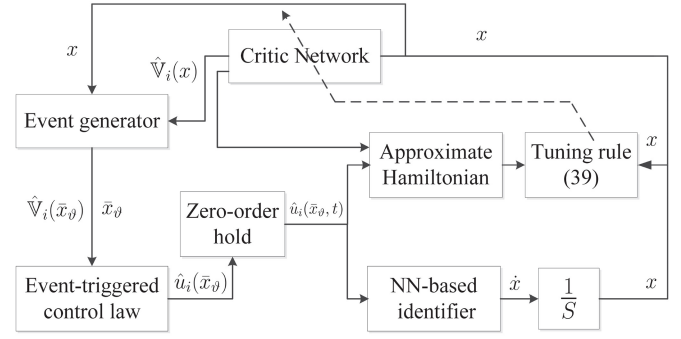

Fig. 1. Structure of the ADP-Based RETC method.

$$\varepsilon_{Hid} = -\nabla \varepsilon_{ci}^{\mathsf{T}}(x(t_d)) \left(\mathcal{F}(x(t_d)) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x(t_d)) \hat{u}_j(\bar{x}_\vartheta)\right)$$

$$\Phi_{id} = \frac{\Upsilon_{id}}{1 + \Upsilon_{id}^{\mathsf{T}} \Upsilon_{id}}$$

and $\Phi_i = (\Upsilon_i / [1 + \Upsilon_i^{\mathsf{T}} \Upsilon_i])$.

Remark 4: In the present ADP-based RETC method, the ER technique uses historical system data to update the network weights. Compared with traditional methods to relax PE condition, the ER technique is convenient to implement in practice since the historical system data are easy collected during the learning process.

Remark 5: Let $\Sigma = [\sigma_c(x(t_1)), \ldots, \sigma_c(x(t_{\mathcal{N}_{\mathcal{D}}}))]$ be the historical data matrix. According to [36], the matrix Σ requires to include enough linearly independent elements such that $\text{rank}(\Sigma) = \mathcal{L}_c$.

Remark 6: According to (36), we find that the designed event-triggered optimal controller of each player relies on the overall system state. Therefore, the developed ADP-based RETC approach is a kind of centralized control method.

Based on the above discussion, the structure of the ADP-based RETC method is shown in Fig. 1.

D. Stability Analysis

In this section, the stability of the multiplayer system (2) is demonstrated by using the Lyapunov stability theorem. Before proving, some assumptions, which were used in [16], [32], [34], and [42]–[44], are provided.

Assumption 5 [32], [34], [43], [44]: ω_{ci}, $\nabla \sigma_c(x)$, $\nabla \varepsilon_{ci}(x)$, and ε_{Hi} are norm-bounded, that is

$$\|\omega_{ci}\| \leq \bar{\omega}_{ci}, \quad \|\nabla \sigma_c(x)\| \leq \bar{\sigma}_c$$

$$\|\nabla \varepsilon_{ci}(x)\| \leq \bar{\varepsilon}_{ci}, \quad \|\varepsilon_{Hi}\| \leq \bar{\varepsilon}_{Hi}$$

where $\bar{\omega}_{ci}$, $\bar{\sigma}_c$, $\bar{\varepsilon}_{ci}$, and $\bar{\varepsilon}_{Hi}$ are positive constants.

Assumption 6 [16], [42]: $\delta(\cdot)$ is Lipschitz continuous, that is

$$\|\delta(\zeta_1) - \delta(\zeta_2)\| \leq \mathcal{L}_\delta \|\zeta_1 - \zeta_2\|$$

where $\mathcal{L}_\delta$ is a positive constant, and ζ_1 and ζ_2 are m-dimensional vectors

Theorem 3: Consider the unknown multiplayer nonlinear system (1) with the nominal form (2), the event-triggered approximate optimal control law (36), and the critic NN weight

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG *et al.*: ADP-BASED EVENT-TRIGGERED ROBUST CONTROL FOR MNSGs

7

updating law (39). Then, both the multiplayer system (2) and the critic NN weight estimation error $\tilde{\omega}_{ci}$ are UUB when the event-triggering condition

$$\|e_\vartheta(t)\|^2 \leq \frac{\sum_{i=1}^{\mathcal{N}}\left(\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) + (1-\tau^2)\lambda_{\min}(\mathcal{Q}_i)\|x\|^2\right)}{\frac{1}{2}\mathcal{N}\bar{\mathcal{G}}^2\left(1+\frac{1}{\beta^2}\right)\mathcal{L}_u^2}$$

$$\triangleq \|e_T\|^2 \tag{41}$$

and the inequality

$$\alpha_c > \frac{(1+\beta^2)\bar{\mathcal{G}}^2\|\Psi_i\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_i^2\bar{\sigma}_c^2}{2\lambda_{i\min}\left(\Phi_i\Phi_i^\mathsf{T} + \sum_{d=1}^{\mathcal{N}_\mathcal{D}}\Phi_{id}\Phi_{id}^\mathsf{T}\right)} \tag{42}$$

hold, where $\beta > 0$ is a design parameter.

Proof: Select the Lyapunov function candidate as

$$\mathcal{L} = \mathcal{L}_1 + \mathcal{L}_2 + \mathcal{L}_3$$
$$= \sum_{i=1}^{\mathcal{N}}\mathbb{V}_i^*(x) + \sum_{i=1}^{\mathcal{N}}\mathbb{V}_i^*(\bar{x}_\vartheta) + \frac{1}{2}\sum_{i=1}^{\mathcal{N}}\tilde{\omega}_{ci}^\mathsf{T}\tilde{\omega}_{ci}. \tag{43}$$

Case 1 (Event Is Not Triggered): Taking the time derivative of (43) and utilizing system (2), we have

$$\dot{\mathcal{L}}_1 = \sum_{i=1}^{\mathcal{N}}\left(\nabla\mathbb{V}_i^{*\mathsf{T}}(x)\left(\mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)\hat{u}_j(\bar{x}_\vartheta)\right)\right) \tag{44}$$

$$\dot{\mathcal{L}}_2 = 0 \tag{45}$$

$$\dot{\mathcal{L}}_3 = \sum_{i=1}^{\mathcal{N}}\left(-\alpha_c\tilde{\omega}_{ci}^\mathsf{T}\left(\Phi_i\Phi_i^\mathsf{T} + \sum_{d=1}^{\mathcal{N}_\mathcal{D}}\Phi_{id}\Phi_{id}^\mathsf{T}\right)\tilde{\omega}_{ci}\right.$$
$$\left. + \frac{\alpha_c\tilde{\omega}_{ci}^\mathsf{T}\Phi_i}{1+\Upsilon_i^\mathsf{T}\Upsilon_i}\varepsilon_{Hi} + \sum_{d=1}^{\mathcal{N}_\mathcal{D}}\frac{\alpha_c\tilde{\omega}_{ci}^\mathsf{T}\Phi_{id}}{1+\Upsilon_{id}^\mathsf{T}\Upsilon_{id}}\varepsilon_{Hid}\right). \tag{46}$$

According to (10), we can obtain (47), as shown at the bottom of the page. By using Young's inequality, we further obtain

$$\left\|\hat{u}_j(\bar{x}_\vartheta) - u_j^*(x)\right\|^2$$
$$= \left\|\left(\hat{u}_j(\bar{x}_\vartheta) - u_j^*(\bar{x}_\vartheta)\right) + \left(u_j^*(\bar{x}_\vartheta) - u_j^*(x)\right)\right\|^2$$
$$\leq (1+\beta^2)\left\|\hat{u}_j(\bar{x}_\vartheta) - u_j^*(\bar{x}_\vartheta)\right\|^2$$
$$+ \left(1+\frac{1}{\beta^2}\right)\left\|u_j^*(\bar{x}_\vartheta) - u_j^*(x)\right\|^2$$
$$\leq \frac{1}{2}(1+\beta^2)\|\Psi_j\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_j^2\left(\bar{\sigma}_c^2\|\tilde{\omega}_{cj}\|^2 + \bar{\varepsilon}_{cj}^2\right)$$
$$+ \left(1+\frac{1}{\beta^2}\right)\mathcal{L}_u^2\|e_\vartheta\|^2 \tag{48}$$

where

$$\left\|\hat{u}_j(\bar{x}_\vartheta) - u_j^*(\bar{x}_\vartheta)\right\|^2$$
$$= \left\|\Psi_j\delta\left(\frac{1}{2}\mathcal{G}_j^\mathsf{T}(\bar{x}_\vartheta)\left(\nabla\sigma_c^\mathsf{T}(\bar{x}_\vartheta)\omega_{cj} + \nabla\varepsilon_{cj}(\bar{x}_\vartheta)\right)\right)\right.$$
$$\left. - \Psi_j\delta\left(\frac{1}{2}\mathcal{G}_j^\mathsf{T}(\bar{x}_\vartheta)\nabla\sigma_c^\mathsf{T}(\bar{x}_\vartheta)\hat{\omega}_{cj}\right)\right\|^2$$
$$\leq \frac{1}{4}\|\Psi_j\|^2\mathcal{L}_\delta^2\|\mathcal{G}_j^\mathsf{T}(\bar{x}_\vartheta)\left(\nabla\sigma_c^\mathsf{T}(\bar{x}_\vartheta)\omega_{cj} + \nabla\varepsilon_{cj}(\bar{x}_\vartheta)\right)$$
$$- \mathcal{G}_j^\mathsf{T}(\bar{x}_\vartheta)\nabla\sigma_c^\mathsf{T}(\bar{x}_\vartheta)\hat{\omega}_{cj}\|^2$$
$$\leq \frac{1}{4}\|\Psi_j\|^2\mathcal{L}_\delta^2\left\|\mathcal{G}_j^\mathsf{T}(\bar{x}_\vartheta)\right\|^2\left\|\nabla\sigma_c^\mathsf{T}(\bar{x}_\vartheta)\tilde{\omega}_{cj} + \nabla\varepsilon_{cj}(\bar{x}_\vartheta)\right\|^2$$
$$\leq \frac{1}{2}\|\Psi_j\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_j^2\left(\bar{\sigma}_c^2\|\tilde{\omega}_{cj}\|^2 + \bar{\varepsilon}_{cj}^2\right). \tag{49}$$

According to (47) and (48), we have

$$\dot{\mathcal{L}}_1 \leq \sum_{i=1}^{\mathcal{N}}\left(-\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x\right) + \mathcal{N}\bar{\sigma}_c^2\sum_{i=1}^{\mathcal{N}}\bar{\omega}_{ci}^2$$

$$\dot{\mathcal{L}}_1 = \sum_{i=1}^{\mathcal{N}}\left(-\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x - \mathcal{W}_i(\mathcal{U}_\mathcal{N}^*)\right) - \sum_{i=1}^{\mathcal{N}}\nabla\mathbb{V}_i^{*\mathsf{T}}(x)\sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)u_j^*(x) + \sum_{i=1}^{\mathcal{N}}\nabla\mathbb{V}_i^{*\mathsf{T}}(x)\sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)\hat{u}_j(\bar{x}_\vartheta)$$

$$= \sum_{i=1}^{\mathcal{N}}\left(-\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x - \mathcal{W}_i(\mathcal{U}_\mathcal{N}^*)\right) + \left[\nabla\mathbb{V}_1^*(x) + \cdots + \nabla\mathbb{V}_\mathcal{N}^*(x)\right]^\mathsf{T}\sum_{j=1}^{\mathcal{N}}\mathcal{G}_j(x)\left(\hat{u}_j(\bar{x}_\vartheta) - u_j^*(x)\right)$$

$$\leq \sum_{i=1}^{\mathcal{N}}\left(-\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x - \mathcal{W}_i(\mathcal{U}_\mathcal{N}^*)\right) + \frac{1}{2}\left[\nabla\mathbb{V}_1^*(x) + \cdots + \nabla\mathbb{V}_\mathcal{N}^*(x)\right]^\mathsf{T}\left[\nabla\mathbb{V}_1^*(x) + \cdots + \nabla\mathbb{V}_\mathcal{N}^*(x)\right]$$

$$+ \frac{1}{2}\begin{bmatrix}\hat{u}_1(\bar{x}_\vartheta) - u_1^*(x)\\ \vdots \\ \hat{u}_\mathcal{N}(\bar{x}_\vartheta) - u_\mathcal{N}^*(x)\end{bmatrix}^\mathsf{T}[\mathcal{G}_1(x), \ldots, \mathcal{G}_\mathcal{N}(x)]^\mathsf{T}[\mathcal{G}_1(x), \ldots, \mathcal{G}_\mathcal{N}(x)]\begin{bmatrix}\hat{u}_1(\bar{x}_\vartheta) - u_1^*(x)\\ \vdots \\ \hat{u}_\mathcal{N}(\bar{x}_\vartheta) - u_\mathcal{N}^*(x)\end{bmatrix}$$

$$\leq \sum_{i=1}^{\mathcal{N}}\left(-\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x\right) + \frac{\mathcal{N}}{2}\sum_{i=1}^{\mathcal{N}}\left\|\nabla\sigma_c^\mathsf{T}(x)\omega_{ci} + \nabla\varepsilon_{ci}(x)\right\|^2 + \frac{1}{2}\bar{\mathcal{G}}^2\sum_{j=1}^{\mathcal{N}}\left\|\hat{u}_j(\bar{x}_\vartheta) - u_j^*(x)\right\|^2$$

$$\leq \sum_{i=1}^{\mathcal{N}}\left(-\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_i x\right) + \mathcal{N}\bar{\sigma}_c^2\sum_{i=1}^{\mathcal{N}}\bar{\omega}_{ci}^2 + \mathcal{N}\sum_{i=1}^{\mathcal{N}}\bar{\varepsilon}_{ci}^2 + \frac{1}{2}\bar{\mathcal{G}}^2\sum_{j=1}^{\mathcal{N}}\left\|\hat{u}_j(\bar{x}_\vartheta) - u_j^*(x)\right\|^2 \tag{47}$$

$$+ \frac{1}{4}\left(1+\beta^2\right)\bar{\mathcal{G}}^2\mathcal{L}_\delta^2\sum_{j=1}^{\mathcal{N}}\|\Psi_j\|^2\bar{\mathcal{G}}_j^2\left(\bar{\sigma}_c^2\|\tilde{\omega}_{cj}\|^2 + \bar{\varepsilon}_{cj}^2\right)$$

$$+ \mathcal{N}\sum_{i=1}^{\mathcal{N}}\bar{\varepsilon}_{ci}^2 + \frac{1}{2}\mathcal{N}\bar{\mathcal{G}}^2\left(1+\frac{1}{\beta^2}\right)\mathcal{L}_u^2\|e_\vartheta\|^2. \qquad (50)$$

Apply the inequality $A^\mathsf{T}B \le A^\mathsf{T}A/2 + B^\mathsf{T}B/2$ to the last two terms of (46), we have

$$\sum_{i=1}^{\mathcal{N}}\frac{\alpha_c\tilde{\omega}_{ci}^\mathsf{T}\Phi_i\varepsilon_{Hi}}{1+\Upsilon_i^\mathsf{T}\Upsilon_i} \le \sum_{i=1}^{\mathcal{N}}\frac{\alpha_c}{2}\tilde{\omega}_{ci}^\mathsf{T}\Phi_i\Phi_i^\mathsf{T}\tilde{\omega}_{ci}$$
$$+ \sum_{i=1}^{\mathcal{N}}\frac{\alpha_c}{2}\varepsilon_{Hi}^\mathsf{T}\varepsilon_{Hi} \qquad (51)$$

$$\sum_{i=1}^{\mathcal{N}}\sum_{d=1}^{\mathcal{N}_\mathcal{D}}\frac{\alpha_c\tilde{\omega}_{ci}^\mathsf{T}\Phi_{id}}{1+\Upsilon_i^\mathsf{T}\Upsilon_i}\varepsilon_{Hid} \le \frac{\alpha_c}{2}\sum_{i=1}^{\mathcal{N}}\sum_{d=1}^{\mathcal{N}_\mathcal{D}}\tilde{\omega}_{ci}^\mathsf{T}\Phi_{id}\Phi_{id}^\mathsf{T}\tilde{\omega}_{ci}$$
$$+ \frac{\alpha_c}{2}\sum_{i=1}^{\mathcal{N}}\sum_{d=1}^{\mathcal{N}_\mathcal{D}}\varepsilon_{Hid}^\mathsf{T}\varepsilon_{Hid}. \qquad (52)$$

Then, according to (51) and (52), we can obtain

$$\dot{\mathcal{L}}_3 \le -\frac{\alpha_c}{2}\sum_{i=1}^{\mathcal{N}}\lambda_{i\min}\left(\Phi_i\Phi_i^\mathsf{T} + \sum_{d=1}^{\mathcal{N}_\mathcal{D}}\Phi_{id}\Phi_{id}^\mathsf{T}\right)\|\tilde{\omega}_{ci}\|^2$$
$$+ \frac{\alpha_c(\mathcal{N}_\mathcal{D}+1)}{2}\sum_{i=1}^{\mathcal{N}}\bar{\varepsilon}_{Hi}^2. \qquad (53)$$

Combining (45), (50), and (53), we obtain

$$\dot{\mathcal{L}} \le \sum_{i=1}^{\mathcal{N}}\left(-\xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x) - x^\mathsf{T}\mathcal{Q}_ix\right) + \mathcal{N}\bar{\sigma}_c^2\sum_{i=1}^{\mathcal{N}}\bar{\omega}_{ci}^2$$
$$+ \mathcal{N}\sum_{i=1}^{\mathcal{N}}\bar{\varepsilon}_{ci}^2 + \frac{1}{2}\mathcal{N}\bar{\mathcal{G}}^2\left(1+\frac{1}{\beta^2}\right)\mathcal{L}_u^2\|e_\vartheta\|^2$$
$$+ \frac{1}{4}\left(1+\beta^2\right)\bar{\mathcal{G}}^2\sum_{j=1}^{\mathcal{N}}\|\Psi_j\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_j^2\bar{\sigma}_c^2\|\tilde{\omega}_{cj}\|^2$$
$$+ \frac{1}{4}\left(1+\beta^2\right)\bar{\mathcal{G}}^2\sum_{j=1}^{\mathcal{N}}\|\Psi_j\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_j^2\bar{\varepsilon}_{cj}^2$$
$$- \frac{\alpha_c}{2}\sum_{i=1}^{\mathcal{N}}\lambda_{i\min}\left(\Phi_i\Phi_i^\mathsf{T} + \sum_{d=1}^{\mathcal{N}_\mathcal{D}}\Phi_{id}\Phi_{id}^\mathsf{T}\right)\|\tilde{\omega}_{ci}\|^2$$
$$+ \frac{\alpha_c(\mathcal{N}_\mathcal{D}+1)}{2}\sum_{i=1}^{\mathcal{N}}\bar{\varepsilon}_{Hi}^2. \qquad (54)$$

Let

$$\lambda_1 = \frac{1}{4}\left(1+\beta^2\right)\bar{\mathcal{G}}^2\sum_{j=1}^{\mathcal{N}}\|\Psi_j\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_j^2\bar{\varepsilon}_{cj}^2$$
$$+ \mathcal{N}\bar{\sigma}_c^2\sum_{i=1}^{\mathcal{N}}\bar{\omega}_{ci}^2 + \frac{\alpha_c(\mathcal{N}_\mathcal{D}+1)}{2}\sum_{i=1}^{\mathcal{N}}\bar{\varepsilon}_{Hi}^2 + \mathcal{N}\sum_{i=1}^{\mathcal{N}}\bar{\varepsilon}_{ci}^2.$$

Then, (54) becomes

$$\dot{\mathcal{L}} \le \sum_{i=1}^{\mathcal{N}}\left(-\left(1-\tau^2\right)\lambda_{\min}(\mathcal{Q}_i)\|x\|^2 - \xi\sum_{j=1}^{\mathcal{N}}\eta_{jM}^2(x)\right)$$
$$+ \frac{1}{2}\mathcal{N}\bar{\mathcal{G}}^2\left(1+\frac{1}{\beta^2}\right)\mathcal{L}_u^2\|e_\vartheta\|^2 - \sum_{i=1}^{\mathcal{N}}\tau^2\lambda_{\min}(\mathcal{Q}_i)\|x\|^2$$
$$- \sum_{i=1}^{\mathcal{N}}\left(\frac{1}{2}\alpha_c\lambda_{i\min}\left(\Phi_i\Phi_i^\mathsf{T} + \sum_{d=1}^{\mathcal{N}_\mathcal{D}}\Phi_{id}\Phi_{id}^\mathsf{T}\right)\right.$$
$$\left. - \frac{1}{4}\left(1+\beta^2\right)\bar{\mathcal{G}}^2\|\Psi_i\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_i^2\bar{\sigma}_c^2\right)\|\tilde{\omega}_{ci}\|^2 + \lambda_1.$$
$$(55)$$

Hence, under conditions (41) and (42), $\dot{\mathcal{L}} < 0$ if $\tilde{\omega}_{ci}$ or x lies outside the compact set (56), as shown at the bottom of the page or

$$\Omega_x = \left\{x : \|x\| \le \sqrt{\frac{\lambda_1}{\tau^2\sum_{i=1}^{\mathcal{N}}\lambda_{\min}(\mathcal{Q}_i)}}\right\}. \qquad (57)$$

Case 2 (Event Is Triggered): According to (43), we can obtain

$$\Delta\mathcal{L}(t) = \Delta\mathcal{L}_1(t) + \Delta\mathcal{L}_2(t) + \Delta\mathcal{L}_3(t). \qquad (58)$$

From case 1, we have $\dot{\mathcal{L}} < 0$ for all $t \in [S_\vartheta, S_{\vartheta+1})$. Then, we further obtain

$$\Delta\mathcal{L}_1(t) = \sum_{i=1}^{\mathcal{N}}\mathbb{V}_i^*(\bar{x}_{\vartheta+1}) - \sum_{i=1}^{\mathcal{N}}\mathbb{V}_i^*\left(x\left(S_{\vartheta+1}^-\right)\right) \le 0$$
$$\Delta\mathcal{L}_2(t) = \sum_{i=1}^{\mathcal{N}}\mathbb{V}_i^*(\bar{x}_{\vartheta+1}) - \sum_{i=1}^{\mathcal{N}}\mathbb{V}_i^*(\bar{x}_\vartheta) \le -\varpi\left(\|e_{\vartheta+1}(S_\vartheta)\|\right)$$
$$\Delta\mathcal{L}_3(t) = \frac{1}{2}\left(\sum_{i=1}^{\mathcal{N}}\tilde{\omega}_{ci}^\mathsf{T}(\bar{x}_{\vartheta+1})\tilde{\omega}_{ci}(\bar{x}_{\vartheta+1})\right.$$
$$\left. - \sum_{i=1}^{\mathcal{N}}\tilde{\omega}_{ci}^\mathsf{T}\left(x\left(S_{\vartheta+1}^-\right)\right)\tilde{\omega}_{ci}\left(x\left(S_{\vartheta+1}^-\right)\right)\right) \le 0$$

where $\varpi(\cdot)$ is a class-$\mathcal{K}$ function, and $e_{\vartheta+1}(S_\vartheta) = \bar{x}_{\vartheta+1} - \bar{x}_\vartheta$. From the above analysis, we know that $\dot{\mathcal{L}} < 0$ is satisfied at the triggering instant. The proof is completed. ∎

$$\Omega_{\tilde{\omega}_{ci}} = \left\{\tilde{\omega}_{ci} : \|\tilde{\omega}_{ci}\| \le \sqrt{\frac{4\lambda_1}{2\alpha_c\lambda_{i\min}\left(\Phi_i\Phi_i^\mathsf{T} + \sum_{d=1}^{\mathcal{N}_\mathcal{D}}\Phi_{id}\Phi_{id}^\mathsf{T}\right) - \left(1+\beta^2\right)\bar{\mathcal{G}}^2\|\Psi_i\|^2\mathcal{L}_\delta^2\bar{\mathcal{G}}_i^2\bar{\sigma}_c^2}}\right\} \qquad (56)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG et al.: ADP-BASED EVENT-TRIGGERED ROBUST CONTROL FOR MNSGs

9

E. Exclusion of the Zeno Behavior

In this section, we will demonstrate that the Zeno behavior will not occur.

Assumption 7: $\mathcal{F}(x)$ satisfies

$$\|\mathcal{F}(x)\| \le \mathcal{Z}_f \|x\| \tag{59}$$

where $\mathcal{Z}_f$ is a positive constant.

Theorem 4: Considering the unknown multiplayer nonlinear system (1) with the nominal form (2) and the event-triggered approximate optimal control law (36), the minimal intersampling time $\Delta t_{\min}$ has a lower bound given by

$$\Delta t_{\min} \ge \frac{1}{\mathcal{Z}_f} \ln\left(1 + \frac{\|e_\vartheta\|}{\mathcal{B}}\right) > 0 \tag{60}$$

where $\mathcal{Z}_f$ and $\mathcal{B}$ are positive constants.

Proof: Taking the time derivative of the event-triggered error $e_\vartheta(t)$, we can obtain

$$\dot{e}_\vartheta(t) = \dot{x}(t) - \dot{\bar{x}}_\vartheta(t) \equiv \dot{x}(t), t \in [S_\vartheta, S_{\vartheta+1}). \tag{61}$$

Substituting (2) into (61), we have

$$
\begin{aligned}
\|\dot{e}_\vartheta\| &= \|\dot{x}\| \\
&= \left\| \mathcal{F}(x) + \sum_{j=1}^{\mathcal{N}} \mathcal{G}_j(x)\hat{u}_j(\bar{x}_\vartheta) \right\| \\
&\le \mathcal{Z}_f \|x\| + \mathcal{Z}_{gh} \\
&\le \mathcal{Z}_f \|\bar{x}_\vartheta + e_\vartheta\| + \mathcal{Z}_{gh} \\
&\le \mathcal{Z}_f \|e_\vartheta\| + \mathcal{Z}_f \|\bar{x}_\vartheta\| + \mathcal{Z}_{gh}
\end{aligned} \tag{62}
$$

where $\mathcal{Z}_{gh} = \sum_{j=1}^{\mathcal{N}} \bar{\mathcal{G}}_j \bar{u}_j$. According to [23, Th. 4], we have

$$\|e_\vartheta\| \le \frac{\mathcal{Z}_f \|\bar{x}_\vartheta\| + \mathcal{Z}_{gh}}{\mathcal{Z}_f} \left(e^{\mathcal{Z}_f(t-S_\vartheta)-1}\right) \tag{63}$$

for all $t \in [S_\vartheta, S_{\vartheta+1})$. According to (63), it indicates that the ϑth intersampling time satisfies

$$S_{\vartheta+1} - S_\vartheta \ge \frac{1}{\mathcal{Z}_f} \ln\left(1 + \frac{\|e_\vartheta\|}{\mathcal{B}}\right) > 0$$

where $\mathcal{B} = \frac{\mathcal{Z}_f \|\bar{x}_\vartheta\| + \mathcal{Z}_{gh}}{\mathcal{Z}_f}$. It means that $\Delta t_{\min} = \min\{S_{\vartheta+1} - S_\vartheta\} > 0$ in (60). This ends the proof. ∎

Remark 7: It is noticed that the event-triggering condition (16) in Theorem 1 and the event-triggering condition (41) in Theorem 3 are different. The event-triggering condition (16) is developed for original system (1), which shows the multiplayer system (1) is guaranteed to be UUB under this condition. However, the event-triggering condition (41) is designed for the nominal system (2), which proves that the multiplayer system (2) and the critic NN weight estimation errors are both UUB if the condition (41) holds.

IV. SIMULATION

In this section, two simulation examples are employed to verify the effectiveness of the present ADP-based RETC scheme.
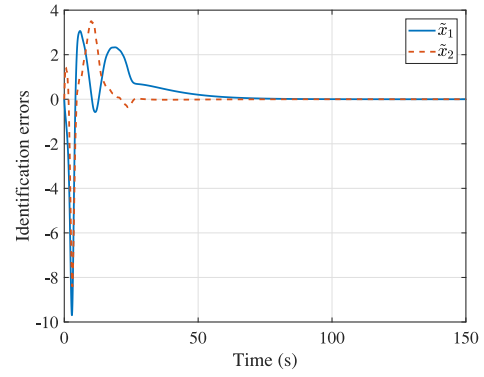

Fig. 2. Curves of identification errors.

A. Example 1

Considering the uncertain CT nonlinear system as

$$
\begin{aligned}
\dot{x} = \mathcal{F}(x) &+ \mathcal{G}_1(x)(u_1 + p_1 x_1 \sin(x_2)) \\
&+ \mathcal{G}_2(x)(u_2 + p_2 x_1 \cos(x_2))
\end{aligned} \tag{64}
$$

where

$$
\mathcal{F}(x) = \begin{bmatrix} x_2 \\ -0.5x_1 - 0.5x_2(1 + (\cos(2x_1) + 2)^2) \end{bmatrix}
$$

$$
\mathcal{G}_1(x) = \begin{bmatrix} 0 \\ \cos(2x_1) + 2 \end{bmatrix}, \mathcal{G}_2(x) = \begin{bmatrix} 0 \\ \sin(4x_1^2) + 2 \end{bmatrix}
$$

$x = [x_1, x_2]^\mathsf{T}$, p_1 and p_2 are unknown parameters and chosen randomly within $[-(\sqrt{2}/2), (\sqrt{2}/2)]$. For simplicity, we choose $p_1 = p_2 = 1$, $\xi = 10$, and $\eta_{1M}(x) = \eta_{2M}(x) = \|x\|$ in this simulation. Then, the nominal system of (64) is given as $\dot{x} = \mathcal{F}(x) + \mathcal{G}_1(x)u_1 + \mathcal{G}_2(x)u_2$. Assume the control input constraints be $\bar{u}_1 = 0.3$ and $\bar{u}_2 = 0.2$.

The initial weights of NN-based identifier $\hat{\Pi}(t)$, $\hat{\omega}_f(t)$, $\hat{\omega}_{\mathcal{G}1}(t)$, and $\hat{\omega}_{\mathcal{G}2}(t)$ are chosen randomly within $[-1, 1]$. The activation functions $\sigma_f(\cdot)$ and $\sigma_{\mathcal{G}il}(\cdot)$ are chosen as $\tanh(\cdot)$, respectively. ψ_l is selected randomly within $[-1, 1]$ and remains unchanged. Other parameters of NN-based identifer are selected as $\Xi_a = \Xi_f = \Xi_{\mathcal{G}1} = \Xi_{\mathcal{G}2} = [0.01, 0.001; 0.001, 0.01]$ and $\theta = 1$. The activation function and the learning rate of critic NNs are selected as $\sigma_c(x) = [x_1^2, x_1x_2, x_2^2]^\mathsf{T}$ and $\alpha_c = 10$, respectively. The initial weights of the critic NN are selected randomly within $[-1, 1]$. Let $x_0 = [1, -1]^\mathsf{T}$, $\mathcal{Q}_1 = 25I_2$, $\mathcal{Q}_2 = 30I_2$, $\beta = 1$, $\mathcal{L}_u = 1.4$, and $\mathcal{N}_\mathcal{D} = 8$.

Simulation results are depicted in Figs. 2–9. Fig. 2 illustrates that the identification errors of the NN-based identifer converge to equilibrium after $t = 75$ s. It means that the NN-based identifier can identify the unknown dynamics of nominal system successfully. In Fig. 3(a), we can observe that nominal system states reach a small region of zero after 30 s. The ETC curves are displayed in Fig. 3(b). As illustrated in Fig. 3(b), the control inputs are piecewise signals and satisfy $u_i < \bar{u}_i$ ($i = 1, 2$), which means that control input signals vary within the control constraints. The weight updating curves of critic NNs are displayed in Fig. 4, which converge to $\hat{\omega}_{c1} = [-0.466, -0.456, 0.478]^\mathsf{T}$ and

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10   IEEE TRANSACTIONS ON CYBERNETICS
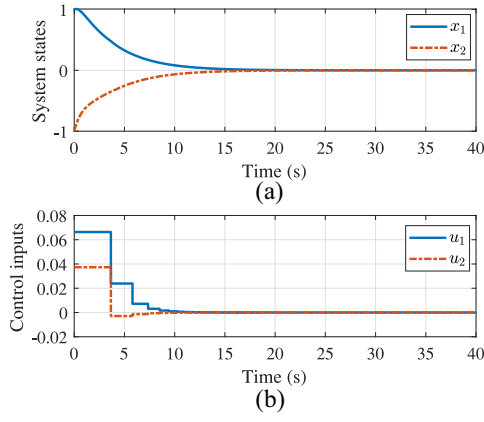

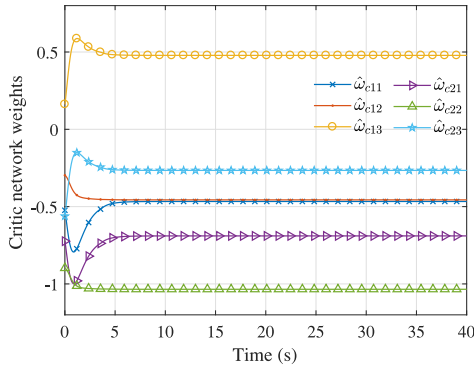Fig. 3.   (a) Trajectories of nominal system states x_1 and x_2. (b) Trajectories of control inputs.
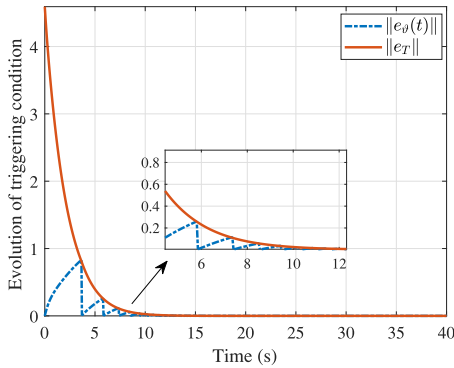

Fig. 4.   Trajectories of critic network weights.

Fig. 5.   Evolution of triggering condition.

Fig. 6.   Sampling period.

Fig. 7.   Samples.

Fig. 8.   State trajectory of multiplayer system.

$\hat{\omega}_{c2} = [-0.689, -1.034, -0.265]^\mathsf{T}$, respectively. The evolution of the triggering condition is displayed in Fig. 5, where we can see that the event-triggered error $e_\vartheta(t)$ and the threshold e_T converge to a small region of equilibrium when time increases. Fig. 6 shows the sampling period of the ETC law. We can find that the sampling periods are multiples of 0.05 s. Comparison results on the numbers of samples between the ADP-based RETC method and the time-triggered one are shown in Fig. 7. It is clear that the event-triggered controller only updates 455 times, but the time-triggered controller requires 800 times. Hence, the ADP-based RETC method reduces the computational and communication burden. Fig. 8 shows that the ADP-based RETC method can ensure the system (64) to be
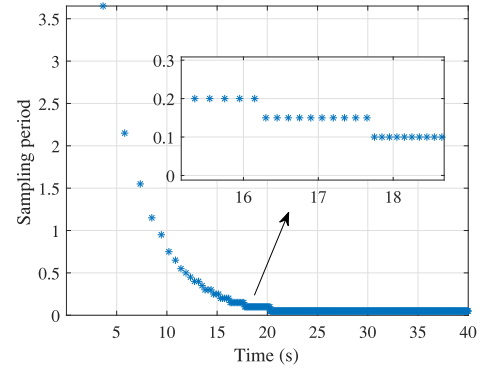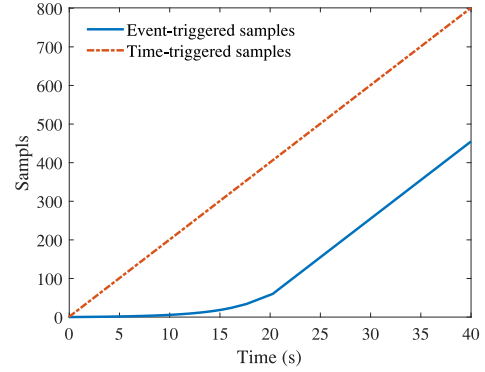
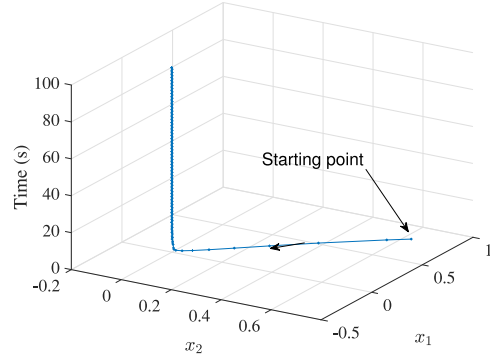stable. Fig. 9 displays the value functions of two players. We can find that the value functions converge to different values, which means that two players achieve their individual objectives.

Remark 8: The activation functions of all players are selected as $\sigma_c(x) = [x_1^2, x_1 x_2, x_2^2]^\mathsf{T}$. In fact, selecting activation function is a challenging issue since it affects the control performance directly. In this article, we select $\sigma_c(x)$ by "trial and error" with repeated simulations.

Remark 9: The design parameter $\mathcal{L}_u$ affects event-triggered controller design. On the one hand, a large $\mathcal{L}_u$ will lead to a small triggering threshold e_T; thus, more frequent control updating and computation to maintain the control performance and system stability. On the other hand, if the selected $\mathcal{L}_u$
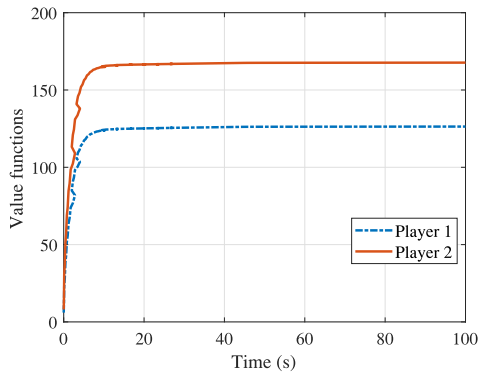
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ZHANG et al.: ADP-BASED EVENT-TRIGGERED ROBUST CONTROL FOR MNSGs

11

Fig. 9.   Evolution of value functions.

Fig. 10.   Curves of identification errors.

is too small, the controllers will update less frequently but may lead to system instability. Therefore, we need to select an appropriate $\mathcal{L}_u$ by "trial and error" with repeated simulations to guarantee the system to be stable and to reduce the computational and communication burden.

B. Example 2

Consider the torsional pendulum system [37] expressed by

$$\frac{d\chi}{dt} = \mathcal{W} + \chi u_1$$
$$\mathcal{J}\frac{d\mathcal{W}}{dt} = u_2 - \mathcal{M}gl_p\sin(\chi) - f_d\frac{d\chi}{dt} + \mathcal{W}u_3$$

where χ is the angle, $\mathcal{W}$ is the angular velocity, $\mathcal{M} = 1/3$ kg is the mass, $l_p = 2/3$ m is the length of the pendulum bar, $\mathcal{J} = (4/3)\mathcal{M}l_p^2$ kg·m² is the rotary inertia, $f_d = 0.2$ is the frictional factor, and $g = 9.8$ m/s² is the gravity acceleration. By replacing χ and $\mathcal{W}$ with x_1 and x_2, the torsional pendulum system with uncertainties is expressed by

$$\dot{x} = \mathcal{F}(x) + \mathcal{G}_1(x)\left(u_1 + p_1x_1\sin^5(x_2)\cos^2(x_1)\right)$$
$$+ \ \mathcal{G}_2(x)\left(u_2 + p_2x_2\cos^5(x_1)\sin^2(x_2)\right)$$
$$+ \ \mathcal{G}_3(x)\left(u_3 + p_3x_1\cos^5(x_2)\sin^2(x_2)\right) \qquad (65)$$

where

$$\mathcal{F}(x) = \begin{bmatrix} x_2 \\ -\frac{\mathcal{M}gl_p}{\mathcal{J}}\sin(x_1) - \frac{f_d}{\mathcal{J}}x_2 \end{bmatrix}, \mathcal{G}_1 = \begin{bmatrix} x_1 \\ -\frac{f_d}{\mathcal{J}}x_1 \end{bmatrix}$$

$$\mathcal{G}_2 = \begin{bmatrix} 0 \\ \frac{1}{\mathcal{J}} \end{bmatrix}, \mathcal{G}_3 = \begin{bmatrix} 0 \\ \frac{x_2}{\mathcal{J}} \end{bmatrix}$$

p_1, p_2, and p_3 are selected randomly within $[-2, 2]$. According to (65), the nominal system is presented as $\dot{x} = \mathcal{F}(x) + \mathcal{G}_1(x)u_1 + \mathcal{G}_2(x)u_2 + \mathcal{G}_3(x)u_3$. In this simulation, let $\xi = 10$, $\mathcal{Q}_1 = I_2$, $\mathcal{Q}_2 = 2I_2$, $\mathcal{Q}_3 = 1.5I_2$, $\beta = 1$, $\mathcal{L}_u = 1$, $\mathcal{N}_D = 8$, $p_1 = p_2 = p_3 = 1$, $\eta_{1M}(x) = \eta_{2M}(x) = \eta_{3M}(x) = \|x\|$, and $\bar{u}_1 = 0.8, \bar{u}_2 = 0.6, \bar{u}_3 = 0.5$. The initial weights of the NN-based identifier are selected randomly within $[-1, 1]$, the activation functions $\sigma_f(\cdot)$ and $\sigma_{Gil}(\cdot)$ are the same as those of Example 1, $\Xi_a = \Xi_f = \Xi_{G1} = \Xi_{G2} = \Xi_{G3} = [0.01, 0.001; 0.001, 0.01]$ and $\theta = 0.5$.

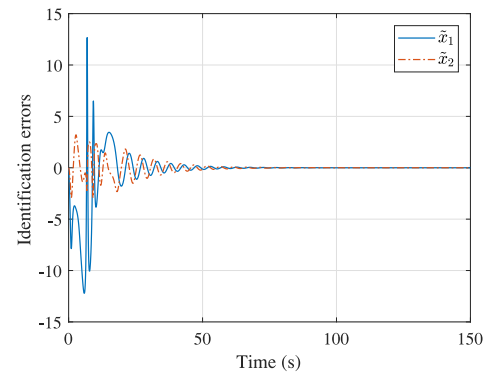Simulation results are depicted in Figs. 10–17. The identification errors are displayed in Fig. 10, where we find that
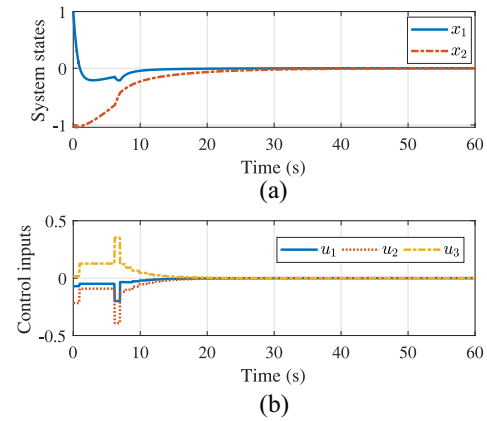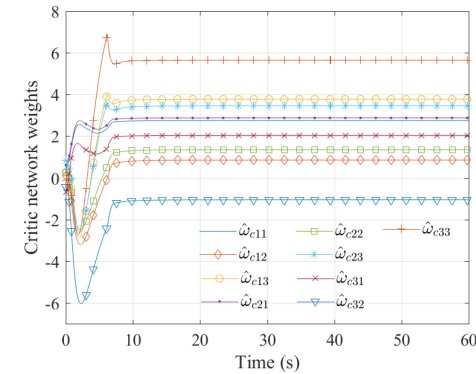

Fig. 11.   (a) Trajectories of nominal system states x_1 and x_2. (b) Trajectories of control inputs.

Fig. 12.   Trajectories of critic network weights.

identification errors reach zero after 80 s. Then, the ADP-based RETC approach is applied to the system (65). Fig. 11(a) shows the nominal system states converge to a small region of zero with the developed control inputs displayed in Fig. 11(b). We can observe that control inputs are piecewise signals and satisfy $u_i < \bar{u}_i$ ($i = 1, 2, 3$), which means that the control laws are limited within the constraints. Fig. 12 reveals that the critic NN weights converge to $\hat{\omega}_{c1} = [2.745, 0.847, 3.774]^\mathsf{T}$, $\hat{\omega}_{c2} = [2.872, 1.342, 3.453]^\mathsf{T}$, and $\hat{\omega}_{c3} = [2.031, -1.060, 5.645]^\mathsf{T}$, respectively. Fig. 13 shows that the event-triggered error $e_\vartheta(t)$ and the threshold e_T converge to a small region of equilibrium after $t = 30$ s. The sampling period is illustrated in Fig.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

12

IEEE TRANSACTIONS ON CYBERNETICS
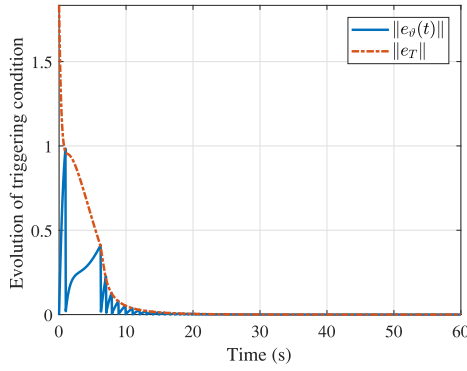

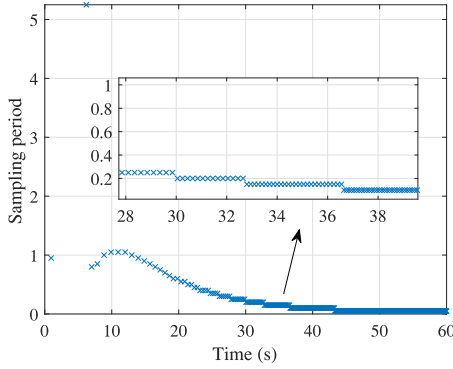Fig. 13.   Evolution of triggering condition.
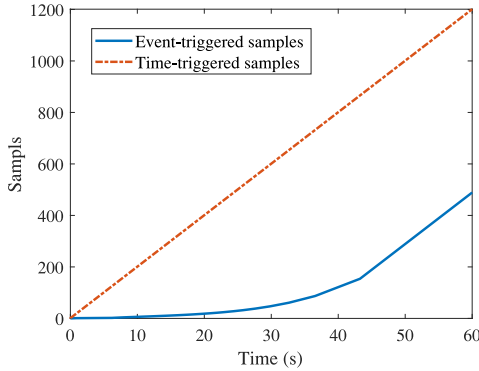

Fig. 14.   Sampling period.
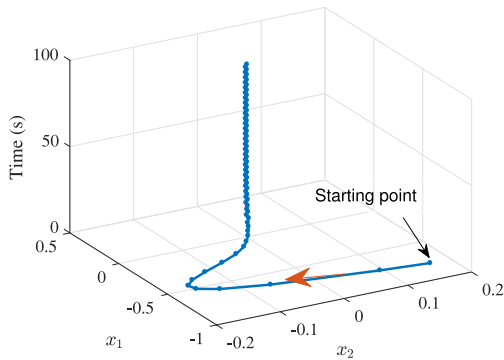

Fig. 15.   Samples.

Fig. 16.   State trajectory of multiplayer system.

14, where the minimum sampling period is 0.05 s. The samples number illustrated in Fig. 15 indicates that the ADP-based RETC method greatly reduces the updating times in contrast to
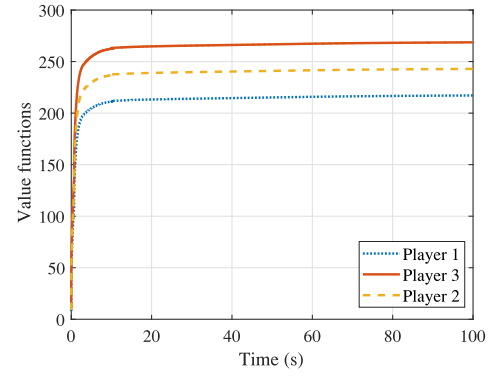

Fig. 17.   Evolution of value functions.

the time-triggered method. Therefore, the computational and communication burden are all reduced. Fig. 16 demonstrates that the developed controller can guarantee the system (65) to be stable. The convergence of three player's value functions is shown in Fig. 17, where all of them converge to different values. From the simulation results, we can conclude that the developed ADP-based RETC method not only achieves the individual objective of each player but also guarantee the stability of the torsional pendulum system with input constraints and dynamical uncertainties.
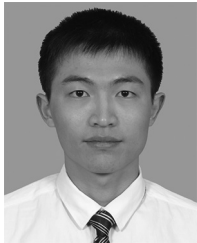
V.   CONCLUSION

In this article, the MNSG problem for unknown nonlinear CT systems with uncertainties and constrained inputs is addressed by using the ADP-based RETC method. An NN-based identifier is established to rebuild the system dynamics by utilizing the measured system data. Then, the approximated event-triggered optimal control for each player is obtained by the solution of HJ equation with the critic NN. By introducing the ER technique, the PE condition is relaxed. In order to reduce computational and communication burden, a new triggering condition for the MNSG problem is presented without control information. Moreover, the UUB stability of the critic NN weight estimate error dynamics and the multiplayer system are demonstrated by the Lyapunov stability theorem. Finally, the effectiveness of the ADP-based RETC approach is validated by two simulation examples. The main innovation of this article is on developing a robust stabilization scheme with a new nonquadratic value function for unknown multiplayer systems with uncertainties and constrained inputs. Furthermore, the developed event-triggered robust controllers are updated aperiodically such that the computational and communication burden is alleviated. In our future work, in order to improve the realizability of this developed approach, we will try to relax Assumptions 1 and 2 since they are strict in practice. Moreover, tracking control problems of multiplayer systems will be considered by using the ADPETC approach, since the system states have to follow the reference trajectories in an optimal manner in practice, such as hypersonic aircrafts, spacecrafts, and robots.

REFERENCES

[1] Y. Jiang and Z.-P. Jiang, "Robust adaptive dynamic programming and feedback stabilization of nonlinear systems," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 25, no. 5, pp. 882–893, May 2014.

[2] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York, NY, USA: Van Nostrand Reinhold, 1992, ch. 13.

[3] P. J. Werbos, "A menu of designs for reinforcement learning over time," in *Neural Networks for Control*, W. T. Miller, R. S. Sutton, and P. J. Werbos, Eds. Cambridge, MA, USA: MIT Press, 1991, pp. 67–95.

[4] H. Jiang and H. Zhang, "Iterative ADP learning algorithms for discrete-time multi-player games," *Artif. Intell. Rev.*, vol. 50, no. 1, pp. 75–91, Jun. 2018.

[5] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.

[6] B. Zhao, D. Wang, G. Shi, D. Liu, and Y. Li, "Decentralized control for large-scale nonlinear systems with unknown mismatched interconnections via policy iteration," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 10, pp. 1725–1735, Oct. 2018.

[7] D. Liu, Y. Xu, Q. Wei, and X. Liu, "Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming," *IEEE/CAA J. Automatica Sinica*, vol. 5, no. 1, pp. 36–46, Jan. 2018.

[8] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[9] D. Wang, D. Liu, Q. Wei, D. Zhao, and N. Jin, "Optimal control of unknown nonaffine nonlinear discrete-time systems based on adaptive dynamic programming," *Automatica,* vol. 48, no. 8, pp. 1825–1832, Aug. 2012.

[10] Q. Wei, D. Liu, Y. Liu, and R. Song, "Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming," *IEEE/CAA J. Automatica Sinica*, vol. 4, no. 2, pp. 168–176, Apr. 2017.

[11] Q. Wei, F. L. Lewis, D. Liu, R. Song, and H. Lin, "Discrete-time local value iteration adaptive dynamic programming: Convergence analysis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 6, pp. 875–891, Jun. 2018.

[12] Y. Zhu and D. Zhao, "Comprehensive comparison of online ADP algorithms for continuous-time optimal control," *Artif. Intell. Rev.*, vol. 49, no. 4, pp. 531–547, Apr. 2018.

[13] H. Lin, B. Zhao, D. Liu, and C. Alippi, "Data-based fault tolerant control for affine nonlinear systems through particle swarm optimized neural networks," *IEEE/CAA J. Automatica Sinica*, vol. 7, no. 4, pp. 954–964, Jul. 2020.

[14] B. Luo, D. Liu, T. Huang, and D. Wang, "Model-free optimal tracking control via critic-only Q-learning," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 27, no. 10, pp. 2134–2144, Oct. 2016.

[15] J. Hou, D. Wang, D. Liu, and Y. Zhang, "Model-free H_∞ optimal tracking control of constrained nonlinear systems via an iterative adaptive learning algorithm," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4097–4108, Nov. 2020.

[16] D. Wang, C. Mu, X. Yang, and D. Liu, "Event-based constrained robust control of affine systems incorporating an adaptive critic mechanism," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1602–1612, Jul. 2017.

[17] B. Zhao, D. Liu, and Y. Li, "Observer based adaptive dynamic programming for fault tolerant control of a class of nonlinear systems," *Inf. Sci.*, vol. 384, pp. 21–33, Dec. 2016.

[18] X. Yang and H. He, "Self-learning robust optimal control for continuous-time nonlinear systems with mismatched disturbances," *Neural Netw.* vol. 99, pp. 19–30, Mar. 2018.

[19] D. Liu, X. Yang, D. Wang, and Q. Wei, "Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1372–1385, Jul. 2015.

[20] D. Wang, D. Liu, C. Mu, and Y. Zhang, "Neural network learning and robust stabilization of nonlinear systems with dynamic uncertainties," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 29, no. 4, pp. 1342–1351, Apr. 2018.

[21] X. Zhong and H. He, "An event-triggered ADP control approach for continuous-time system with unknown internal states," *IEEE Trans. Cybern.*, vol. 47, no. 3, pp. 683–694, Mar. 2017.

[22] K. G. Vamvoudakis, "Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems," *IEEE/CAA J. Automatica Sinica*, vol. 1, no. 3, pp. 282–293, Jul. 2014.

[23] B. Zhao and D. Liu, "Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3054–3064, Apr. 2020.

[24] X. Yang and H. He, "Adaptive critic learning and experience replay for decentralized event-triggered control of nonlinear interconnected systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4043–4055, Nov. 2020.

[25] A. Sahoo, H. Xu, and S. Jagannathan, "Near optimal event-triggered control of nonlinear discrete-time systems using neurodynamic programming," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 27, no. 9, pp. 1801–1815, Sep. 2016.

[26] J. Yi, S. Chen, X. Zhong, W. Zhou, and H. He, "Event-triggered globalized dual heuristic programming and its application to networked control systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 3, pp. 1383–1392, Mar. 2019.

[27] D. Wang, H. He, X. Zhong, and D. Liu, "Event-driven nonlinear discounted optimal regulation involving a power system application," *IEEE Trans. Ind. Electron.*, vol. 64, no. 10, pp. 8177–8186, Oct. 2017.

[28] Q. Zhang and D. Zhao, "Data-based reinforcement learning for nonzero-sum games with unknown drift dynamics," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 2874–2885, Aug. 2019.

[29] R. Song, Q. Wei, and B. Song, "Neural-network-based synchronous iteration learning method for multi-player zero-sum games," *Neurocomputing* vol. 242, no. 73–82, Jun. 2017.

[30] H. Zhang, H. Jiang, C. Luo, and G. Xiao, "Discrete-time nonzero-sum games for multiplayer using policy-iteration-based adaptive dynamic programming algorithms," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3331–3340, Oct. 2017.

[31] H. Zhang, H. Sun, K. Zhang, and Y. Luo, "Event-triggered adaptive dynamic programming algorithm for non-zero-sum games of unknown nonlinear systems via generalized fuzzy hyperbolic models," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 11, pp. 2202–2214, Nov. 2019.

[32] B. Zhao, D. Liu, and C. Luo, "Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 31, no. 10, pp. 4330–4340, Oct. 2020.

[33] X. Yang and H. He, "Adaptive critic designs for event-triggered robust control of nonlinear systems with unknown dynamics," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2255–2267, Jun. 2019.

[34] X. Yang, D. Liu, and D. Wang, "Reinforcement learning for adaptive optimal control of unknown continuous-time nonlinear systems with input constraints," *Int. J. Control,* vol. 87, no. 3, pp. 553–566, Oct. 2013.

[35] Q. Qu, H. Zhang, C. Luo, and R. Yu, "Robust control design for multiplayer nonlinear systems with input disturbances via adaptive dynamic programming," *Neurocomputing*, vol. 334, pp. 1–10, Mar. 2019.

[36] D. Zhao, Q. Zhang, D. Wang, and Y. Zhu, "Experience replay for optimal control of nonzero-sum game systems with unknown dynamics," *IEEE Trans. Cybern.*, vol. 46, no. 3, pp. 854–865, Mar. 2016.

[37] Q. Wei, H. Li, X. Yang, and H. He, "Continuous-time distributed policy iteration for multicontroller nonlinear systems," *IEEE Trans. Cybern.*, vol. 51, no. 5, pp. 2372–2383, May 2021.

[38] X. Yang and H. He, "Event-driven H_∞-constrained control using adaptive critic learning," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 4860–4872, Oct. 2021.

[39] H. Modares, F. L. Lewis, and M. Naghibi-Sistani, "Adaptive optimal control of unknown constrained-input systems using policy iteration and neural networks," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 24, no. 10, pp. 1513–1525, Oct. 2013.

[40] K. G. Vamvoudakis and F. L. Lewis "Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton-Jacobi equations" *Automatica*, vol. 47, no. 8, pp. 1556–1569, Aug. 2011.

[41] K. G. Vamvoudakis, H. Modares, B. Kiumarsi, and F. L. Lewis, "Game theory-based control system algorithms with real-time reinforcement learning: How to solve multiplayer games online," *IEEE Control Syst. Mag.*, vol. 37, no. 1, pp. 33–52, Feb. 2017.

[42] Y. Yang, H. Zhu, Q. Zhang, B. Zhao, Z. Li, and D. Wunsch, "Sparse online kernelized actor–critic learning in reproducing kernel Hilbert space," *Artif. Intell. Rev.*, vol. 55, no. 1, pp. 23–58, 2022, [Online]. Available: https://doi.org/10.1007/s10462-021-10045-9

[43] Y. Zhang, B. Zhao, D. Liu, and S. Zhang, "Event-triggered control of discrete-time zero-sum games via deterministic policy gradient adaptive dynamic programming," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, Aug. 31, 2021, doi: 10.1109/TSMC.2021.3105663.

[44] M. Lin, B. Zhao, and D. Liu, "Policy gradient adaptive critic designs for model-free optimal tracking control with experience replay," *IEEE Trans. Syst., Man, Cybern., Syst.*, early access, Apr. 19, 2021, doi: 10.1109/TSMC.2021.3071968.

Yongwei Zhang (Member, IEEE) received the B.S. degree in automation from the School of Electronic and Information Engineering, Jiaying University, Meizhou, China, in 2016, and the Ph.D. degree in control science and engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2021.

He is currently a Postdoctoral Fellow with the Guangdong University of Technology. His current research interests include adaptive dynamic programming and optimal control.

Bo Zhao (Senior Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control science and engineering from Jilin University, Changchun, China, in 2009 and 2014, respectively.

He was a Postdoctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2014 to 2017. Then, he joined the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, from 2017 to 2018. He is currently an Associate Professor with the School of Systems Science, Beijing Normal University, Beijing. He was selected as the Beijing Normal University Tang Scholar in 2021. He has authored or coauthored over 110 journal and conference papers, and authorized three patents. His research interests include adaptive dynamic programming, robot control, fault diagnosis and tolerant control, optimal control, and artificial intelligence-based control.
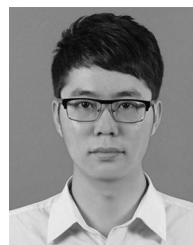
Dr. Zhao serves as an Associate Editor of IEEE TRANSACTIONS ON SYSTEMS, MAN AND CYBERNETICS: SYSTEMS and *Neurocomputing*, an Early Career Advisory Board Member of IEEE/CAA JOURNAL OF AUTOMATICA SINICA, and a Guest Editor of *Complex & Intelligent Systems*, and also served many academic conferences. He is the Secretary General of Adaptive Dynamic Programming and Reinforcement Learning Technical Committee of Chinese Automation Association, and the member of the Asian–Pacific Neural Network Society and the Chinese Association for Artificial Intelligence.

Derong Liu (Fellow, IEEE) received the B.S. degree in mechanical engineering from the East China Institute of Technology (currently Nanjing University of Science and Technology), Nanjing, China, in 1982, the M.S. degree in automatic control theory and applications from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 1987, and the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1994.

He was a Product Design Engineer with the China North Industries Corporation, Jilin, China, from 1982 to 1984. He was an Instructor with the Graduate School of Chinese Academy of Sciences, Beijing, from 1987 to 1990. He was a Staff Fellow with the General Motors Research and Development Center, Warren, MI, USA, from 1993 to 1995. He was an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, from 1995 to 1999. He joined the University of Illinois at Chicago, Chicago, IL, USA, in 1999, and became a Full Professor of Electrical and Computer Engineering and of Computer Science in 2006. He was selected for the "100 Talents Program" by the Chinese Academy of Sciences, Beijing, in 2008, and served as the Associate Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation from 2010 to 2016. He is currently a Full Professor with the School of Automation, Guangdong University of Technology, Guangzhou, China. He has published 13 books and 260 papers in international journals.

Dr. Liu received the Faculty Early Career Development Award from the National Science Foundation in 1999, the University Scholar Award from University of Illinois from 2006 to 2009, the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008, and the Outstanding Achievement Award from Asia–Pacific Neural Network Assembly in 2014. He received the International Neural Network Societys Gabor Award in 2018, the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS Outstanding Paper Award in 2018, the IEEE Systems, Man and Cybernetics Society Andrew P. Sage Best Transactions Paper Award in 2018, and the IEEE/CCA JOURNAL AUTOMATICA SINICA Hsue-Shen Tsien Paper Award in 2019. He is the recipient of the IEEE CIS Neural Network Pioneer Award in 2022. He has been named a highly cited researcher consecutively for five years from 2017 to 2021 by Clarivate. He was a plenary/keynote speaker at 32 international conferences. He was elected three times as the AdCom Member of the IEEE Computational Intelligence Society in 2006, 2015, and 2022, respectively. He was the Editor-in-Chief of IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2010 to 2015. He was elected twice as the Distinguished Lecturer of IEEE Computational Intelligence Society in 2012 and 2016, respectively. He served as a member of the Council of International Federation of Automatic Control from 2014 to 2017 and the President of the Asia–Pacific Neural Network Society in 2018. He was the General Chair of the 2014 IEEE World Congress on Computational Intelligence, the 2016 World Congress on Intelligent Control and Automation, and the 2017 International Conference on Neural Information Processing. He is currently the Editor-in-Chief of *Artificial Intelligence Review*, the Deputy Editor-in-Chief of the IEEE/CAA JOURNAL OF AUTOMATICA SINICA and *CAAI Transactions on Artificial Intelligence*, and the Chair of IEEE Guangzhou Section. He is a Fellow of the International Neural Network Society and the International Association for Pattern Recognition, and a member of Academia Europaea (The Academy of Europe).

Shunchao Zhang received the B.S. degree in measurement and control technology and instruments from the School of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan, China, in 2016, and the M.S. degree in control engineering from the Guangdong University of Technology, Guangzhou, China, in 2019, where he is currently pursuing the Ph.D. degree in control science and engineering with the School of Automation.

His current research interests include optimal control and adaptive dynamic programming.

Event-Triggered Control of Discrete-Time Zero-Sum Games via Deterministic Policy Gradient Adaptive Dynamic Programming

Yongwei Zhang, Bo Zhao, *Senior Member, IEEE*, Derong Liu, *Fellow, IEEE*, and Shunchao Zhang

Abstract—In order to address zero-sum game problems for discrete-time (DT) nonlinear systems, this article develops a novel event-triggered control (ETC) approach based on the deterministic policy gradient (PG) adaptive dynamic programming (ADP) algorithm. By adopting the input and output data, the proposed ETC method updates the control law and the disturbance law with a gradient descent algorithm. Compared with the conventional PG ADP-based control scheme, the present controller is updated aperiodically to reduce the computational and communication burden. Then, the actor-critic-disturbance framework is adopted to obtain the optimal control law and the worst disturbance law, which guarantee the input-to-state stability of the closed-loop system. Moreover, a novel neural network weight updating law which guarantees the uniform ultimate boundedness of weight estimation errors is provided based on the experience replay technique. Finally, the validity of the present method is verified by simulation of two DT nonlinear systems.

Index Terms—Adaptive dynamic programming (ADP), event-triggered control (ETC), neural networks (NNs), policy gradient (PG), zero-sum games (ZSGs).

I. INTRODUCTION

OPTIMAL feedback control, which gains extensive research in control theory and engineering, aims to design a feedback controller that minimizes the user-defined performance index function [1], [2]. In practice, the occurrence of external disturbance is ineluctable, which undermines the control performance. The H_∞ control which is taken as a two-player zero-sum game (ZSG) is widely studied to maintain a satisfactory control performance in the existence of external disturbance [3], [4]. For nonlinear systems, one needs to obtain the solution of the Hamilton–Jacobi–Isaacs (HJI) equation which is tough to solve due to its inherent nonlinearity when dealing with ZSG problems [5].

Adaptive dynamic programming (ADP), which was put forward by Werbos [6], is competitive to solve the nonlinear HJI equation and has been applied to solve optimal control problems for discrete-time (DT) systems [7]–[12] and continuous-time (CT) systems [13]–[19] with trajectory tracking [20]–[22], fault tolerance [23], and robust stabilization [24]. For ZSG problems, several ADP-based optimal control methods have been proposed [25]–[27]. In [25], the ZSG problem for DT systems was considered by developing a novel iterative ADP algorithm. The theoretical analysis illustrated that the upper and lower iterative cost functions converge to the optimal solution of the ZSG, and the existence condition of the saddle-point equilibrium was not demanded. In [26], a globalized dual heuristic programming (GDHP) was presented to solve the HJI equation for unknown DT nonlinear systems. Three neural networks (NNs) were built to obtain the approximate solution of the HJI equation. In [27], the ZSG problem was addressed through the modified policy iteration (PI) algorithm and the actor-critic-disturbance framework.

It is worth noting that the aforementioned controllers are periodically updated with a mass of transmitted data, which leads to heavy computational and communication burden. In order to break this bottleneck, the event-triggered control (ETC), which executes the control aperiodically, is investigated in the control community [28]–[34]. The ETC approach updates the control policy only when the error between the actual system state and the sampled system state violates the triggering condition. In [35], the neuro-dynamic programming-based ETC approach was developed to cope with fixed final time optimal control problems for DT nonlinear systems. The event-triggering condition was determined by the actor NN weights and the system states. Then, the control policy and all the NN weights were tuned aperiodically to reduce the computational burden. In [36], the optimal control problem for CT nonlinear systems with saturating actuators was tackled by

Yongwei Zhang and Shunchao Zhang are with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China (e-mail: yongwei_zhang@mail2.gdut.edu.cn; 1111904006@mail2.gdut.edu.cn).

Bo Zhao is with the School of Systems Science, Beijing Normal University, Beijing 100875, China (e-mail: zhaobo@bnu.edu.cn).

Derong Liu is with the School of Automation, Guangdong University of Technology, Guangzhou 510006, China, and also with the Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL 60607 USA (e-mail: derong@gdut.edu.cn).

the ADP-based ETC method. Based on the Lyapunov stability theorem, the triggering condition was derived considering the control constraints. The NN weights and the control law were tuned at the triggering moments to reduce the transmission cost. In [37], an ADP-based ETC scheme was presented to deal with the CT ZSG problem. In order to save computational resources, the controllers were renewed at triggering moments only. From the above discussions, the event-based control updates the control input at triggering instants only, which reduces the frequency of controller updating such that the computational and communication resources are economized. In practice, the external disturbance is unavoidable and the computational resource should be utilized with a higher efficiency. However, to our best knowledge, most of existing methods for DT ZSG problems are time triggered, where the controller executes periodically and consumes massive computational resources. Moreover, the training of the model NN which rebuilds the unknown system dynamics increases the computational burden. Hence, it is significant to develop a data-based ETC method to address the ZSG problem for DT systems with unknown dynamics, which motivates this work.

As it is well known, policy gradient (PG), which is a powerful approach to cope with model-free control problems, has been developed in the reinforcement learning (RL) and ADP community. In [38], Google DeepMind investigated the deterministic PG (DPG) algorithm which acts in continuous spaces. Compared with the stochastic PG algorithm (SPG), the DPG algorithm updates the control law in the direction of the value function gradient. In [39], the policy gradient ADP (PGADP) algorithm was proposed to cope with the model-free control problem for DT nonlinear systems. It is an off-policy learning method and the controller was renovated by the gradient of the Q-function with respect to the action. Inspired by existing works [38], [39], the DPG-based ETC (DPGETC) approach is developed to handle ZSG problems for DT nonlinear systems. The contributions of our work are outlined as follows.

1) A novel data-based DPGETC scheme is developed to extend the PG-based control approach to handle the ZSG problem. Different from existing methods [37], [41], the model NN is not needed to establish and the designed controller is updated aperiodically and trained by system data.

2) A triggering condition is deduced to ensure the input-to-state stability (ISS) of the closed-loop system, the control law and the disturbance law are tuned aperiodically at triggering instants only to save the computational and communication resources.

3) The actor-critic-disturbance structure is established to approximate the control law, the Q-function, and the disturbance law, respectively. Furthermore, by employing the experience replay (ER) technique, new NN weight updating laws are designed to guarantee the uniform ultimate boundedness (UUB) of the weight estimation errors.

The remainder of this article is organized as follows. In Section II, the problem statement for two-player ZSGs is given. In Section III, the actor-critic-disturbance structure is established to design the event-triggered controller. In Section IV, the ISS of the closed-loop system is analyzed. In Section V, the convergence analysis of the NN weight estimation errors is provided. In Section VI, two examples are given to verify the effectiveness of the developed method. In Section VII, concluding remark is given.

Notations: The real and non-negative real numbers are denoted as $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$, respectively. Ω is a compact set of $\mathbb{R}^n$. $\mathbb{N} = \{0, 1, 2, \ldots, \}$ denotes the set of all non-negative integers. $\mathscr{T}_1 \circ \mathscr{T}_2$ represents the composition function of $\mathscr{T}_1$ and $\mathscr{T}_2$, i.e., $\mathscr{T}_1 \circ \mathscr{T}_2(\cdot) = \mathscr{T}_1(\mathscr{T}_2(\cdot))$. $\mathcal{I}_d$ represents the identity function, i.e., $\mathcal{I}_d(s) = s$ for all $s \in \mathbb{R}_{\geq 0}$. $\mathscr{M}_p$ represents pth data in the historical data set, where $\mathscr{M}$ is a real number or a real matrix.

II. PROBLEM STATEMENT

Consider the nonaffine nonlinear systems given by

$$z_{t+1} = \mathscr{F}(z_t, u_t, d_t), \ \ t = 0, 1, 2, \ldots \tag{1}$$

where $z_t \in \mathbb{R}^n$ is the system state, $u_t \in \mathbb{R}^m$ is the control input, $d_t \in \mathbb{R}^s$ is the external disturbance, and $\mathscr{F}(\cdot)$ is the unknown nonlinear system function. In ZSG, u_t and d_t can be viewed as two players.

Assumption 1: System (1) is controllable and observable. $\mathscr{F}(z_t, u_t, d_t)$ is a Lipschitz continuous on a set Ω in $\mathbb{R}^n$ containing the origin, i.e., $\|\mathscr{F}(z_t)\| \leq b_f \|z_t\|$, where b_f is a Lipschitz constant.

Remark 1: If Assumption 1 holds, there exist feedback control laws $u_t = \varpi(z_t)$ and $d_t = \upsilon(z_t)$ with $\varpi : \mathbb{R}^n \to \mathbb{R}^m$ and $\upsilon : \mathbb{R}^n \to \mathbb{R}^s$ to stabilize the system asymptotically [29], [40].

Define a monotonically increasing subsequence of time instants $\{t_i\}_{i=0}^{\infty}$ as sampling instants. Under the event-triggering architecture, the feedback control law and the disturbance law are formulated as

$$u_t = \varpi(z_{t_i})$$
$$d_t = \upsilon(z_{t_i}).$$

The event-triggering error is given by

$$\mathcal{E}_t = z_{t_i} - z_t \tag{2}$$

for $t_i \leq t < t_{i+1}$ with $i = 0, 1, 2, \ldots$, where z_t is the current state and z_{t_i} is the sampled state. Then, system (1) can be rewritten as

$$z_{t+1} = \mathscr{F}(z_t, \varpi(\mathcal{E}_t + z_t), \upsilon(\mathcal{E}_t + z_t)). \tag{3}$$

The performance index function for system (1) is defined as

$$\mathcal{J}(z_t) = \sum_{l=t}^{\infty} \mathcal{C}(z_l, \varpi(\mathcal{E}_l + z_l), \upsilon(\mathcal{E}_l + z_l)) \tag{4}$$

where $\mathcal{C}(\cdot, \cdot, \cdot)$ is the utility function and defined as

$$\mathcal{C}(z_t, \varpi(z_{t_i}), \upsilon(z_{t_i}))$$
$$= z_t^{\mathsf{T}} \mathcal{Q} z_t + \varpi^{\mathsf{T}}(z_{t_i}) \mathcal{R} \varpi(z_{t_i}) - \beta \upsilon^{\mathsf{T}}(z_{t_i}) \upsilon(z_{t_i}) \tag{5}$$

where $\mathcal{Q} \in \mathbb{R}^{n \times n}$ and $\mathcal{R} \in \mathbb{R}^{m \times m}$ are positive definite matrices, and β is a constant describing the precribed level of disturbance attenuation.
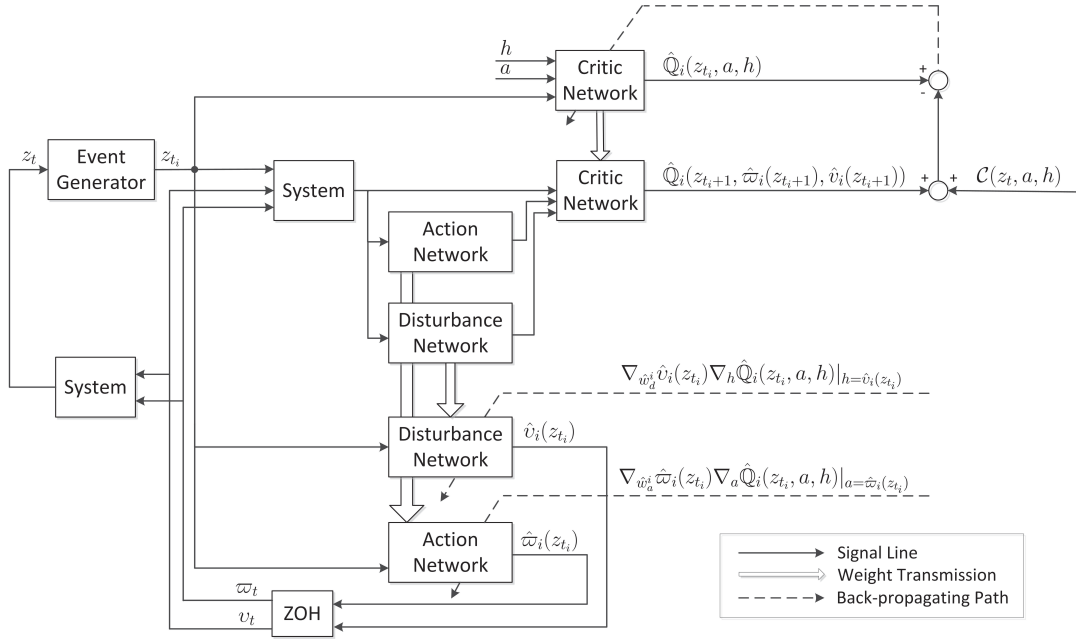
Fig. 1. Diagram of the actor-critic-disturbance framework.

According to Bellman's principle of optimality, the optimal performance index function satisfies

$$\mathcal{J}^*(z_t) = \min_{\varpi(z_{t_i})} \max_{\upsilon(z_{t_i})} \left\{ \mathcal{C}\left(z_t, \varpi\left(z_{t_i}\right), \upsilon\left(z_{t_i}\right)\right) + \mathcal{J}^*(z_{t+1}) \right\}.$$

Let $\mathcal{U}$ and $\mathcal{D}$ be policy spaces of two players, respectively. For DT ZSG problems, our goal is to find an optimal control policy $\varpi^*(z_{t_i})$ and a worst disturbance policy $\upsilon^*(z_{t_i})$ such that $\mathcal{C}(z_t, \varpi^*(z_{t_i}), \upsilon(z_{t_i})) \leq \mathcal{C}(z_t, \varpi^*(z_{t_i}), \upsilon^*(z_{t_i})) \leq \mathcal{C}(z_t, \varpi(z_{t_i}), \upsilon^*(z_{t_i}))$ [25], [26].

In order to develop the DPGETC control scheme, the action-state function, i.e., the so-called Q-function, is defined as [39]

$$\mathbb{Q}(z_t, a, h) = \mathcal{C}(z_t, a, h) + \sum_{l=t+1}^{\infty} \mathcal{C}(z_l, \varpi(\mathcal{E}_l + z_l), \upsilon(\mathcal{E}_l + z_l)) \tag{6}$$

where $\mathbb{Q}(0, 0, 0) = 0$, $a \in \mathcal{U}$, and $h \in \mathcal{D}$. Based on (4), (6) can be rewritten as

$$\begin{aligned}
\mathbb{Q}(z_t, a, h) &= \mathcal{C}(z_t, a, h) \\
&\quad + \mathbb{Q}(z_{t+1}, \varpi(\mathcal{E}_{t+1} + z_{t+1}), \upsilon(\mathcal{E}_{t+1} + z_{t+1})) \\
&= \mathcal{C}(z_t, a, h) + \mathcal{J}(z_{t+1}). \tag{7}
\end{aligned}$$

The optimal Q-function satisfies

$$\mathbb{Q}^*(z_t, a, h) = \min_a \max_h \left\{ \mathcal{C}(z_t, a, h) + \mathcal{J}^*(z_{t+1}) \right\}. \tag{8}$$

The saddle point solution $(\varpi^*(z_{t_i}), \upsilon^*(z_{t_i}))$ should satisfy the following two conditions:

$$\frac{\partial \mathbb{Q}^*\left(z_t, \varpi\left(z_{t_i}\right), \upsilon\left(z_{t_i}\right)\right)}{\partial \varpi\left(z_{t_i}\right)} = 0$$

$$\frac{\partial \mathbb{Q}^*\left(z_t, \varpi\left(z_{t_i}\right), \upsilon\left(z_{t_i}\right)\right)}{\partial \upsilon\left(z_{t_i}\right)} = 0.$$

Then, the optimal control law and the worst disturbance law are expressed as

$$\varpi^*\left(z_{t_i}\right) = \arg\min_a \mathbb{Q}^*(z_t, a, h) \tag{9}$$

$$\upsilon^*\left(z_{t_i}\right) = \arg\max_h \mathbb{Q}^*(z_t, a, h). \tag{10}$$

It indicates that the optimal control law $\varpi^*(z_{t_i})$ and the worst disturbance law $\upsilon^*(z_{t_i})$ depend on the optimal Q-function $\mathbb{Q}^*(z_t, a, h)$, but hard to obtain [26]. In the next section, the DPGETC approach is introduced to conquer this bottleneck.

III. EVENT-TRIGGERED CONTROLLER DESIGN

In this section, an event-triggered controller is developed based on PGADP. In order to approximate the control law, the Q-function and the disturbance law, the actor-critic-disturbance framework is constructed. It is noticed that the weights of the action and the disturbance networks are adjusted through the gradients of the Q-function with respect to action and disturbance, respectively. Moreover, an event generator is adopted to measure the event-triggering error. Once the event is occurred, the current state z_t is sampled as a new sampled state z_{t_i}. The diagram of the actor-critic-disturbance framework is shown in Fig. 1.

A. Critic Network

The three-layer critic network is established to approximate the Q-function $\hat{\mathbb{Q}}(z, a, h)$. Taking the sampled state z_{t_i}, the action a and the disturbance h as the input of the critic network. Then, we can obtain

$$\hat{\mathbb{Q}}_i(z_{t_i}, a, h) = \hat{w}_c^{i\mathsf{T}} \sigma\left(Y_c^\mathsf{T} Z_t\right) = \hat{w}_c^{i\mathsf{T}} \sigma\left(\Phi_c\right)$$

where $Z_t = [z_{t_i}^\mathsf{T}, a^\mathsf{T}, h^\mathsf{T}]^\mathsf{T} \in \mathbb{R}^{(n+m+s)}$, $\Phi_c = Y_c^\mathsf{T} Z_t$, $Y_c \in \mathbb{R}^{(n+m+s) \times l_c}$, $\hat{w}_c^i \in \mathbb{R}^{l_c}$ represents the weight vector from the

hidden-to-output layer, l_c denotes the number of nodes in the hidden layer, and $\sigma(\cdot) = \tanh(\cdot)$ is the activation function. The error function of the critic network is defined as

$$e_{ic} = \hat{\mathbb{Q}}_i(z_{t_i}, a, h) - \mathcal{C}(z_{t_i}, a, h)$$
$$- \gamma \hat{\mathbb{Q}}_i(z_{t_i+1}, \varpi_i(z_{t_i+1}), \upsilon_i(z_{t_i+1}))$$

where $0 < \gamma < 1$ is a discount factor. The objective function is given by

$$E_c^i = \frac{1}{2} e_{ic}^2. \tag{11}$$

By employing the ER technique, the critic NN weight vector is renovated by

$$\hat{w}_c^{i+1} = \hat{w}_c^i - \sum_{p=1}^{n_r} a_c \left[\frac{\partial E_{c,p}^i}{\partial \hat{w}_c^i} \right] \tag{12}$$

where $a_c > 0$ represents the learning rate, and n_r denotes the size of historical data.

B. Action Network

The action network is adopted to approximate the control policy $\varpi(z_{t_i})$. Consider the input z_{t_i}, the output of the action network is given by

$$\hat{\varpi}_i(z_{t_i}) = \hat{w}_a^{i\mathsf{T}} \sigma\left(Y_a^\mathsf{T} z_{t_i}\right) = \hat{w}_a^{i\mathsf{T}} \sigma(\Phi_a).$$

The involved parameters are defined similarly to the critic network. Motivated by [38] and [45], the partial gradient of the Q-function with respect to $\hat{w}_a^i$ can be calculated as

$$\nabla_{\hat{w}_a^i} \hat{\mathbb{Q}}_i = \nabla_{\hat{w}_a^i} \hat{\varpi}_i(z_{t_i}) \nabla_a \hat{\mathbb{Q}}_i(z_{t_i}, a, h)|_{a=\hat{\varpi}_i(z_{t_i})}. \tag{13}$$

Therefore, the action network weight vector is tuned by

$$\hat{w}_a^{i+1} = \hat{w}_a^i - \sum_{p=1}^{n_r} a_a \nabla_{\hat{w}_a^i} \hat{\mathbb{Q}}_{i,p} \tag{14}$$

where $a_a > 0$ is the learning rate.

C. Disturbance Network

Taking the state z_{t_i} as the input of disturbance network, the output of the disturbance network is provided as

$$\hat{\upsilon}_i(z_{t_i}) = \hat{w}_d^{i\mathsf{T}} \sigma\left(Y_d^\mathsf{T} z_{t_i}\right) = \hat{w}_d^{i\mathsf{T}} \sigma(\Phi_d)$$

where the parameters are defined analogous to the critic network. Then, the partial gradient of the Q-function with respect to $\hat{w}_d^i$ can be calculated as

$$\nabla_{\hat{w}_d^i} \hat{\mathbb{Q}}_i = \nabla_{\hat{w}_d^i} \hat{\upsilon}_i(z_{t_i}) \nabla_h \hat{\mathbb{Q}}_i(z_{t_i}, a, h)|_{h=\hat{\upsilon}_i(z_{t_i})}. \tag{15}$$

Then, the disturbance network weight vector is updated by

$$\hat{w}_d^{i+1} = \hat{w}_d^i - \sum_{p=1}^{n_r} a_d \nabla_{\hat{w}_d^i} \hat{\mathbb{Q}}_{i,p} \tag{16}$$

where $a_d > 0$ is the learning rate.

According to (13)–(16), the control law and the disturbance law are updated by using the gradients of the Q-function. Therefore, the developed control scheme is a kind of the PG approach [38], [39].

Remark 2: There are two important frameworks in PG-based control methods, i.e., DPG and SPG. It is worth mentioning that the ways of calculating the PG are different, i.e., SPG employs both state and action spaces, but DPG employs state space only. Therefore, the DPG method is more efficient due to the strategies adopted in controller update are deterministic, rather than stochastic.

Remark 3: Compared with conventional ADP-based control approaches, the advantages of the DPG-based control scheme lie in that:
1) It is a data-based approach and the controller is renewed by adopting the gradient of the Q-function; therefore, the system functions are not needed in designing the control policy;
2) The DPG-based control approach is not only suitable to affine systems, but also nonaffine systems.

Remark 4: It noticed that traditional ADP-based control methods abandon incoming data immediately after being used. It might neglect rich experience and knowledge which can be used later. Thus, to make full use of system data, the ER technique is employed to improve the data usage efficiency by reusing the sampled data and to break the temporal correlation among historical data.

Remark 5: In RL/ADP, the value function is the expected return when starting with the state z and following a policy ϖ. However, the Q-function is the expected return when starting with z and the action a, and following the policy ϖ, i.e., the information of action a is also required. In this article, the DPGETC approach is developed based on the Q-function, where both the control law and the disturbance law are updated by the corresponding gradients of the Q-function. However, the corresponding gradients cannot be obtained from the value function directly. Therefore, it is feasible to use the Q-function in this article. It is worth noticing that the Q-function and the value function are both widely used in RL/ADP. Their selection depends on the specific problem.

IV. STABILITY ANALYSIS OF THE CLOSED-LOOP SYSTEM

Before the stability analysis, the following definitions and assumptions are provided.

Definition 1: A function $\mathcal{A} : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a $\mathcal{K}$-function if it is continuous, strictly increasing and $\mathcal{A}(0) = 0$; it is a $\mathcal{K}_\infty$-function if it is a $\mathcal{K}$-function and satisfies $\mathcal{A}(s) \to \infty$ as $s \to \infty$ [43].

Definition 2: A function $\mathcal{B} : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is a $\mathcal{KL}$-function if, for each fixed $t \geq 0$, the function $\mathcal{B}(\cdot, t)$ is a $\mathcal{K}$-function, and for each fixed $s \geq 0$, the function $\mathcal{B}(s, \cdot)$ is decreasing with $\mathcal{B}(s, t) \to 0$ as $t \to \infty$ [43].

Definition 3: Let ϕ be a $\mathcal{KL}$-function, and φ_1 and φ_2 be $\mathcal{K}$-functions. For each initial state $z_0 \in \mathbb{R}^n$, all $u \in \mathbb{R}^m$ and $d \in \mathbb{R}^s$, system (3) is said to be ISS if the inequality

$$\|z_t\| \leq \phi(\|z_0\|, t) + \varphi_1(\bar{u}) + \varphi_2(\bar{d}) \tag{17}$$

holds, where $\bar{u} = \sup\{\|u_t\| : t \in \mathbb{N}\} < \infty$ and $\bar{d} = \sup\{\|d_t\| : t \in \mathbb{N}\} < \infty$ [42].

Definition 4: Let $\mathscr{C}_1$, $\mathscr{C}_2$, and $\mathscr{C}_3$ be $\mathcal{K}_\infty$-functions, and γ_e and γ_d be $\mathcal{K}$-functions. A continuous function $\mathscr{V} : \mathbb{R}^n \to \mathbb{R}_{\geq 0}$

is referred an ISS-Lyapunov function if the following inequalities:

$$\mathscr{C}_1(\|z_t\|) \leq \mathscr{V}(z_t) \leq \mathscr{C}_2(\|z_t\|) \tag{18}$$

and

$$\begin{aligned} &\mathscr{V}(\mathscr{F}(z_t, \varpi(z_t + \mathcal{E}_t), \upsilon(z_t + \mathcal{E}_t))) - \mathscr{V}(z_t) \\ &\leq -\mathscr{C}_3(\|z_t\|) + \max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\} \end{aligned} \tag{19}$$

hold.

Assumption 2: There exists positive constants l_1 and l_2, and a continuous function $\tilde{\zeta}$ such that

$$\|\mathscr{F}(z_t, \varpi(z_t + \mathcal{E}_t), \upsilon(z_t + \mathcal{E}_t))\| \leq l_1\|\mathcal{E}_t\| + l_2\|z_t\| \tag{20}$$

and

$$\|\mathcal{E}_t\| \leq \|z_t\| \leq \tilde{\zeta}(\bar{u}) \tag{21}$$

hold [41].

Lemma 1: Assume $\mathscr{C}_4(\cdot) = \mathscr{C}_3(\mathscr{C}_2^{-1}(\cdot))$, then according to (19), the inequality

$$\begin{aligned} &\mathscr{V}(\mathscr{F}(z_t, \varpi(z_t + \mathcal{E}_t), \upsilon(z_t + \mathcal{E}_t))) - \mathscr{V}(z_t) \\ &\leq -\mathscr{C}_4(\mathscr{V}(z_t)) + \max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\} \end{aligned} \tag{22}$$

holds.

Proof: According to (18), we have

$$\mathscr{C}_2^{-1}(\mathscr{V}(z_t)) \leq \|z_t\|. \tag{23}$$

Since $\mathscr{C}_3(\cdot)$ is a $\mathcal{K}_\infty$-function, we can get

$$\mathscr{C}_3\left(\mathscr{C}_2^{-1}(\mathscr{V}(z_t))\right) \leq \mathscr{C}_3(\|z_t\|). \tag{24}$$

By using (24), (19) becomes

$$\begin{aligned} &\mathscr{V}(\mathscr{F}(z_t, \varpi(z_t + \mathcal{E}_t), \upsilon(z_t + \mathcal{E}_t))) - \mathscr{V}(z_t) \\ &\leq -\mathscr{C}_3\left(\mathscr{C}_2^{-1}(\mathscr{V}(z_t))\right) + \max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\}. \end{aligned} \tag{25}$$

Letting $\mathscr{C}_4(\cdot) = \mathscr{C}_3(\mathscr{C}_2^{-1}(\cdot))$. Then, we can get (22). The proof is completed. ∎

Lemma 2: Define

$$\Sigma_z = \{z_t : \mathscr{V}(z_t) \leq \varrho\} \tag{26}$$

where $\varrho = \mathscr{C}_4^{-1} \circ \eta^{-1}(\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\})$, η is a $\mathcal{K}_\infty$-function. If there exists an integer $t_0 \in \mathbb{N}$ such that $z_{t_0} \in \Sigma_z$, then we have $z_t \in \Sigma_z, t \geq t_0$.

Proof: Suppose that $z_{t_0} \in \Sigma_z$. Then, we can get $\mathscr{V}(z_{t_0}) \leq \varrho$. From Lemma 1, we can obtain

$$\begin{aligned} &\mathscr{V}\left(\mathscr{F}\left(z_{t_0}, \varpi\left(z_{t_0} + \mathcal{E}_{t_0}\right), \upsilon\left(z_{t_0} + \mathcal{E}_{t_0}\right)\right)\right) - \mathscr{V}(z_{t_0}) \\ &\leq -\mathscr{C}_4(\mathscr{V}(z_{t_0})) + \max\{\gamma_e(\|\mathcal{E}_{t_0}\|), \gamma_d(\|d_{t_0}\|)\}. \end{aligned} \tag{27}$$

According to the proof of [43, Lemma B.1], we can conclude that $\mathcal{I}_d - \mathscr{C}_4$ and $\mathcal{I}_d - \eta$ are $\mathcal{K}$-functions. Then, (27) becomes

$$\mathscr{V}(z_{t_0+1}) \leq \mathcal{I}_d(\mathscr{V}(z_{t_0})) - \mathscr{C}_4(\mathscr{V}(z_{t_0}))$$

$$\begin{aligned} &\quad + \max\{\gamma_e(\|\mathcal{E}_{t_0}\|), \gamma_d(\|d_{t_0}\|)\} \\ &\leq (\mathcal{I}_d - \mathscr{C}_4)(\mathscr{V}(z_{t_0})) + \max\{\gamma_e(\|\mathcal{E}_{t_0}\|), \gamma_d(\|d_{t_0}\|)\} \\ &\leq (\mathcal{I}_d - \mathscr{C}_4)(\varrho) + \max\{\gamma_e(\|\mathcal{E}_{t_0}\|), \gamma_d(\|d_{t_0}\|)\}. \end{aligned} \tag{28}$$

Since $\mathcal{I}_d - \eta$ is a $\mathcal{K}$-function and using $\eta \circ \mathscr{C}_4(\varrho) = \max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\}$, we have

$$\begin{aligned} \mathscr{V}(z_{t_0+1}) &\leq (\mathcal{I}_d - \mathscr{C}_4)(\varrho) + \max\{\gamma_e(\|\mathcal{E}_{t_0}\|), \gamma_d(\|d_{t_0}\|)\} \\ &= -\mathscr{C}_4(\varrho) + \eta \circ \mathscr{C}_4(\varrho) + \varrho - \eta \circ \mathscr{C}_4(\varrho) \\ &\quad + \max\{\gamma_e(\|\mathcal{E}_{t_0}\|), \gamma_d(\|d_{t_0}\|)\} \\ &= -(\mathcal{I}_d - \eta) \circ \mathscr{C}_4(\varrho) + \varrho - \eta \circ \mathscr{C}_4(\varrho) \\ &\quad + \max\{\gamma_e(\|\mathcal{E}_{t_0}\|), \gamma_d(\|d_{t_0}\|)\} \\ &= -(\mathcal{I}_d - \eta) \circ \mathscr{C}_4(\varrho) + \varrho \\ &\leq \varrho. \end{aligned} \tag{29}$$

Hence, we know that $z_{t_0+1} \in \Sigma_z$. By using the mathematical induction, we have $\mathscr{V}(z_{t_0+j}) \leq \varrho$ for all $j \in \mathbb{N}$. This means that $z_t \in \Sigma_z$ for all $t \geq t_0$. The proof is completed. ∎

Theorem 1: Suppose that the system (3) admits an ISS-Lyapunov function $\mathscr{V}$. If there exists a continuous function ζ satisfies

$$\|\mathcal{E}_t\| \leq \zeta(\bar{u}) \tag{30}$$

then the system (3) is ISS.

Proof: Letting $t_1 = \min\{t \in \mathbb{N}: z_t \in \Sigma_z\} \leq \infty$. According to Lemma 2, we have

$$\mathscr{V}(z_t) \leq \psi(\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\}) \quad \forall t \geq t_1 \tag{31}$$

where $\psi(\cdot) = \mathscr{C}_4^{-1} \circ \eta^{-1}(\cdot)$. Then, we have $\eta \circ \mathscr{C}_4(\mathscr{V}(z_t)) \leq \max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\}$. When $t < t_1$, we can obtain

$$\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\} < \eta \circ \mathscr{C}_4(\mathscr{V}(z_t)). \tag{32}$$

Thus, considering (32) and Lemma 1, we have

$$\begin{aligned} &\mathscr{V}(\mathscr{F}(z_t, \varpi(z_t + \mathcal{E}_t), \upsilon(z_t + \mathcal{E}_t))) - \mathscr{V}(z_t) \\ &\leq -\mathscr{C}_4(\mathscr{V}(z_t)) + \max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\} \\ &= -(\mathcal{I}_d - \eta) \circ \mathscr{C}_4(\mathscr{V}(z_t)) - \eta \circ \mathscr{C}_4(\mathscr{V}(z_t)) \\ &\quad + \max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\} \\ &\leq -(\mathcal{I}_d - \eta) \circ \mathscr{C}_4(\mathscr{V}(z_t)). \end{aligned} \tag{33}$$

According to the proof of [41, Th. 1], for all $0 \leq t \leq t_1 + 1$, there exists a $\mathcal{KL}$-function $\mathscr{C}_5$ such that

$$\mathscr{V}(z_t) \leq \mathscr{C}_5(\mathscr{V}(z_0), t) \tag{34}$$

holds. According to (31) and (34), we can get (35), shown at bottom of this page. Hence, we further get

$$\mathscr{V}(z_t) \leq \max\{\mathscr{C}_5(\mathscr{V}(z_0), t), \psi(\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\})\} \tag{36}$$

$$\mathscr{V}(z_t) \leq \begin{cases} \mathscr{C}_5(\mathscr{V}(z_0), t), & 0 \leq t < t_1 \\ \max\{\mathscr{C}_5(\mathscr{V}(z_0), t), \psi(\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\})\}, & t_1 \leq t \leq t_1 + 1 \\ \psi(\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\}), & t > t_1 + 1 \end{cases} \tag{35}$$

for all $t \in \mathbb{N}$. According to (18), we have

$$
\begin{aligned}
\|z_t\| &\leq \mathscr{C}_1^{-1}(\mathscr{V}(z_t)) \\
&\leq \mathscr{C}_1^{-1}(\max\{\mathscr{C}_5(\mathscr{V}(z_0), t), \\
&\qquad\qquad \psi(\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\})\}).
\end{aligned}
$$

According to (30), we can further derive that

$$
\begin{aligned}
\|z_t\| &\leq \mathscr{C}_1^{-1}(\mathscr{C}_5(\mathscr{V}(z_0), t)) \\
&\quad + \mathscr{C}_1^{-1}(\psi(\max\{\gamma_e(\|\mathcal{E}_t\|), \gamma_d(\|d_t\|)\})) \\
&\leq \mathscr{C}_1^{-1}(\mathscr{C}_5(\mathscr{C}_2(\|z_0\|), t)) + \mathscr{C}_1^{-1} \circ \psi \circ \gamma_e \circ \zeta(\bar{u}) \\
&\quad + \mathscr{C}_1^{-1} \circ \psi \circ \gamma_d(\bar{d}). \qquad (37)
\end{aligned}
$$

According to (37) and Definition 3, we know that system (1) is ISS. The proof is completed. ∎

Lemma 3: Consider Assumption 2 and suppose the system (3) admits an ISS-Lyapunov function. The triggering condition

$$
\|\mathcal{E}_t\| \leq \mathcal{E}_T = \frac{1 - (l_1 + l_2)^{t - t_i}}{1 - (l_1 + l_2)} l_2 \|z_{t_i}\|, \ (l_1 + l_2) \neq 1 \quad (38)
$$

guarantees the system (3) to be ISS.

Proof: Based on (21) and (38), we can get

$$
\|\mathcal{E}_t\| \leq \frac{1 - (l_1 + l_2)^{t - t_i}}{1 - (l_1 + l_2)} l_2 \tilde{\zeta}(\bar{u}). \quad (39)
$$

Hence, we can find a continuous function

$$
\zeta(s) = \frac{1 - (l_1 + l_2)^{t - t_i}}{1 - (l_1 + l_2)} l_2 \tilde{\zeta}(s) \quad (40)
$$

to satisfy the inequality (30) in Theorem 1. Thus, the triggering condition (38) guarantees that the system (3) to be ISS. The proof is completed. ∎

Remark 6: This article studies the ZSG problem for DT systems, rather than CT systems. In fact, for DT systems, the intersample time Δt is greater than 0. By introducing the event-triggering mechanism, the sampling time interval in DT systems is always integral multiple of the intersample time Δt. Therefore, even if the event is triggered at every sampling time, the minimal intersample time satisfies $\tau_{\min} = \Delta t$. Hence, the Zeno behavior will not happen.

V. CONVERGENCE ANALYSIS OF NN IMPLEMENTATIONS

In this section, the convergence analysis of the NN weights is provided. Let w_c^*, w_a^*, and w_d^* be the optimal weights of the critic, the action and the disturbance networks, respectively. Define the network weight estimation errors as $\tilde{w}_c^i = \hat{w}_c^i - w_c^*$, $\tilde{w}_a^i = \hat{w}_a^i - w_a^*$, and $\tilde{w}_d^i = \hat{w}_d^i - w_d^*$, respectively.

Assumption 3: The activation function $\sigma(\cdot)$, the reconstruction error of the critic network δ_c, and the optimal weights w_c^*, w_a^* and w_d^* are norm-bounded as

$$
\|\sigma(\cdot)\| \leq \bar{\sigma}, \|\delta_c\| \leq \bar{\delta}_c, \|w_c^*\| \leq w_{cm}
$$
$$
\|w_a^*\| \leq w_{am}, \|w_d^*\| \leq w_{dm}
$$

where $\bar{\sigma}$, $\bar{\delta}_c$, w_{cm}, w_{am}, and w_{dm} are positive constants.

Theorem 2: Consider the nonlinear system (3), the critic network, the action network and the disturbance network are trained by (12), (14), and (16), respectively. Then, the network weight estimation errors $\tilde{w}_c$, $\tilde{w}_a$, and $\tilde{w}_d$ are UUB.

Proof: According to (13) and (15), we have

$$
\nabla_{\hat{w}_a^i} \hat{\mathbb{Q}}_i = \sigma(\Phi_a) \hat{w}_c^{i\mathsf{T}} \nabla \sigma(\Phi_c) Y_c^{\mathsf{T}} A_a, \quad (41)
$$

$$
\nabla_{\hat{w}_d^i} \hat{\mathbb{Q}}_i = \sigma(\Phi_d) \hat{w}_c^{i\mathsf{T}} \nabla \sigma(\Phi_c) Y_c^{\mathsf{T}} A_d, \quad (42)
$$

where $A_a = (\partial Z_t / \partial \bar{\varpi}_i) = \begin{bmatrix} 0_{n \times m} \\ \cdots \\ I_{m \times m} \\ \cdots \\ 0_{s \times m} \end{bmatrix}$, $A_d = (\partial Z_t / \partial \hat{v}_i) = \begin{bmatrix} 0_{n \times s} \\ \cdots \\ 0_{m \times s} \\ \cdots \\ I_{s \times s} \end{bmatrix}$, $\nabla \sigma(\mathcal{Y}) = ([\partial \sigma(\mathcal{Y})] / \partial \mathcal{Y}) \in \mathbb{R}^{l_c \times l_c}$ for $\mathcal{Y} \in \mathbb{R}^{l_c}$, and I is an identity matrix. Based on (41) and (42), the network weight updating laws become

$$
\tilde{w}_c^{i+1} = \tilde{w}_c^i - a_c \sum_{p=1}^{n_r} \sigma(\Phi_{c,p}) e_{c,p}^{i\mathsf{T}}, \quad (43)
$$

$$
\tilde{w}_a^{i+1} = \tilde{w}_a^i - a_a \sum_{p=1}^{n_r} \sigma(\Phi_{a,p}) \hat{w}_c^{i\mathsf{T}} \nabla \sigma(\Phi_{c,p}) Y_{c,p}^{\mathsf{T}} A_{a,p} \quad (44)
$$

$$
\tilde{w}_d^{i+1} = \tilde{w}_d^i - a_d \sum_{p=1}^{n_r} \sigma(\Phi_{d,p}) \hat{w}_c^{i\mathsf{T}} \nabla \sigma(\Phi_{c,p}) Y_{c,p}^{\mathsf{T}} A_{d,p}. \quad (45)
$$

Select a Lyapunov function candidate as

$$
\begin{aligned}
L &= L_c + L_a + L_d \\
&= \frac{1}{a_c} \mathrm{tr}\{\tilde{w}_c^{i\mathsf{T}} \tilde{w}_c^i\} + \mathrm{tr}\{\tilde{w}_a^{i\mathsf{T}} \tilde{w}_a^i\} + \mathrm{tr}\{\tilde{w}_d^{i\mathsf{T}} \tilde{w}_d^i\}. \quad (46)
\end{aligned}
$$

Consider the first term of (46), we have

$$
\begin{aligned}
\Delta L_c &= \frac{1}{a_c} \mathrm{tr}\{\tilde{w}_c^{(i+1)\mathsf{T}} \tilde{w}_c^{i+1} - \tilde{w}_c^{i\mathsf{T}} \tilde{w}_c^i\} \\
&= \frac{1}{a_c} \mathrm{tr}\left\{ \left(\tilde{w}_c^{i\mathsf{T}} - a_c \sum_{p=1}^{n_r} e_{c,p}^i \sigma^{\mathsf{T}}(\Phi_{c,p}) \right) \right. \\
&\qquad \times \left. \left(\tilde{w}_c^i - a_c \sum_{p=1}^{n_r} \sigma(\Phi_{c,p}) e_{c,p}^{i\mathsf{T}} \right) - \tilde{w}_c^{i\mathsf{T}} \tilde{w}_c^i \right\} \\
&= \mathrm{tr}\{\Delta L_{c1} + \Delta L_{c2}\} \quad (47)
\end{aligned}
$$

where

$$
\Delta L_{c1} = \mathrm{tr}\left\{ -2\tilde{w}_c^{i\mathsf{T}} \sum_{p=1}^{n_r} \sigma(\Phi_{c,p}) e_{c,p}^{i\mathsf{T}} \right\}
$$

$$
\Delta L_{c2} = \mathrm{tr}\left\{ a_c \sum_{p=1}^{n_r} e_{c,p}^i \sigma^{\mathsf{T}}(\Phi_{c,p}) \sum_{p=1}^{n_r} \sigma(\Phi_{c,p}) e_{c,p}^{i\mathsf{T}} \right\}.
$$

Define $\Psi_{ci} = \tilde{w}_c^{i\mathsf{T}} \sigma(\Phi_c)$ and $\delta_c = \Psi_{ci} - e_c^i$. Considering ΔL_{c1} and applying the Cauchy–Schwarz inequality, we can obtain

$$\Delta L_{c1} = \mathrm{tr}\left\{-2\sum_{p=1}^{n_r} \tilde{w}_c^{i\mathsf{T}}\sigma(\Phi_{c,p})e_{c,p}^{i\mathsf{T}}\right\}$$

$$= \mathrm{tr}\left\{-2\sum_{p=1}^{n_r}\Psi_{ci,p}\Psi_{ci,p}^{\mathsf{T}} + 2\sum_{p=1}^{n_r}\Psi_{ci,p}\delta_{c,p}^{\mathsf{T}}\right\}$$

$$\leq -2\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + \sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + \sum_{p=1}^{n_r}\left\|\delta_{c,p}^{\mathsf{T}}\right\|^2$$

$$= -\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + \sum_{p=1}^{n_r}\left\|\delta_{c,p}^{\mathsf{T}}\right\|^2$$

$$\leq -\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + n_r\bar{\delta}_c^2. \tag{48}$$

Considering the second term ΔL_{c2}, we can derive

$$\Delta L_{c2} = \mathrm{tr}\left\{a_c\sum_{p=1}^{n_r} e_{c,p}^{i}\sigma^{\mathsf{T}}(\Phi_{c,p})\sum_{p=1}^{n_r}\sigma(\Phi_{c,p})e_{c,p}^{i\mathsf{T}}\right\}$$

$$= a_c\left\|\sum_{p=1}^{n_r} e_{c,p}^{i}\sigma^{\mathsf{T}}(\Phi_{c,p})\right\|^2$$

$$= a_c\left\|\sum_{p=1}^{n_r}\Psi_{ci,p}\sigma^{\mathsf{T}}(\Phi_{c,p}) - \sum_{p=1}^{n_r}\delta_{c,p}\sigma^{\mathsf{T}}(\Phi_{c,p})\right\|^2$$

$$\leq 2a_c\left\|\sum_{p=1}^{n_r}\Psi_{ci,p}\sigma^{\mathsf{T}}(\Phi_{c,p})\right\|^2 + 2a_c\left\|\sum_{p=1}^{n_r}\delta_{c,p}\sigma^{\mathsf{T}}(\Phi_{c,p})\right\|^2. \tag{49}$$

Let $\Psi_M = [\Psi_{ci,1},\ldots,\Psi_{ci,n_r}]$, $\sigma_{cM} = [\sigma^{\mathsf{T}}(\Phi_{c,1}),\ldots,\sigma^{\mathsf{T}}(\Phi_{c,n_r})]$, and $\delta_M = [\delta_{c,1},\ldots,\delta_{c,n_r}]$, and assume that $\bar{\sigma}_{cM}$ and $\bar{\delta}_M$ are the norm bounds of σ_{cM} and δ_M, respectively. Then, we have

$$\Delta L_{c2} \leq 2a_c\left\|\Psi_M\sigma_{cM}^{\mathsf{T}}\right\|^2 + 2a_c\left\|\delta_M\sigma_{cM}^{\mathsf{T}}\right\|^2$$

$$\leq 2a_c\bar{\sigma}_{cM}^2\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + 2a_c\bar{\sigma}_{cM}^2\sum_{p=1}^{n_r}\left\|\delta_{c,p}\right\|^2$$

$$\leq 2a_c\bar{\sigma}_{cM}^2\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + 2n_r a_c\bar{\sigma}_{cM}^2\bar{\delta}_c^2. \tag{50}$$

According to (48) and (50), we can get

$$\Delta L_c \leq -\left(1 - 2a_c\bar{\sigma}_{cM}^2\right)\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + 2n_r a_c\bar{\sigma}_{cM}^2\bar{\delta}_c^2$$

$$+ n_r\bar{\delta}_c^2. \tag{51}$$

Consider the second term of (46), we have

$$\Delta L_a = \mathrm{tr}\{\Delta L_{a1} + \Delta L_{a2}\} \tag{52}$$

where

$$\Delta L_{a1} = \mathrm{tr}\left\{-2a_a\tilde{w}_a^{i\mathsf{T}}\sum_{p=1}^{n_r}\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right\}$$

$$\Delta L_{a2} = \mathrm{tr}\left\{a_a^2\sum_{p=1}^{n_r} A_{a,p}^{\mathsf{T}}Y_{c,p}\nabla\sigma^{\mathsf{T}}(\Phi_{c,p})\hat{w}_c^{i}\sigma^{\mathsf{T}}(\Phi_{a,p})\right.$$

$$\left. \times \sum_{p=1}^{n_r}\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right\}.$$

For ΔL_{a1}, we can derive that

$$\Delta L_{a1} = a_a\mathrm{tr}\left\{\sum_{p=1}^{n_r}\left(\tilde{w}_a^i - \sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right)^{\mathsf{T}}\right.$$

$$\left. \times \left(\tilde{w}_a^i - \sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right)\right\}$$

$$- a_a\mathrm{tr}\left\{\sum_{p=1}^{n_r}\left(\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right)^{\mathsf{T}}\right.$$

$$\left. \times \left(\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right)\right\}$$

$$- a_a n_r\mathrm{tr}\left\{\tilde{w}_a^{i\mathsf{T}}\tilde{w}_a^i\right\}$$

$$= a_a\sum_{p=1}^{n_r}\left\|\tilde{w}_a^i - \sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right\|^2$$

$$- a_a\sum_{p=1}^{n_r}\left\|\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right\|^2$$

$$- a_a n_r\left\|\tilde{w}_a^i\right\|^2$$

$$\leq a_a\sum_{p=1}^{n_r}\left\|\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right\|^2$$

$$+ a_a n_r\left\|\tilde{w}_a^i\right\|^2. \tag{53}$$

For ΔL_{a2}, we can get

$$\Delta L_{a2} = \mathrm{tr}\left\{a_a^2\sum_{p=1}^{n_r} A_{a,p}^{\mathsf{T}}Y_{c,p}\nabla\sigma^{\mathsf{T}}(\Phi_{c,p})\hat{w}_c^{i}\sigma^{\mathsf{T}}(\Phi_{a,p})\right.$$

$$\left. \times \sum_{p=1}^{n_r}\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right\}$$

$$= a_a^2\left\|\sum_{p=1}^{n_r}\sigma(\Phi_{a,p})\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,p})Y_{c,p}^{\mathsf{T}}A_{a,p}\right\|^2. \tag{54}$$

Let $\sigma_{aM} = [\sigma(\Phi_{a,1}),\ldots,\sigma(\Phi_{a,n_r})]$, $\Xi_a = [\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,1})Y_{c,1}^{\mathsf{T}}A_{a,1},\ldots,\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,n_r})Y_{c,n_r}^{\mathsf{T}}A_{a,n_r}]$, $\bar{\sigma}_{aM}$ be the norm bound of σ_{aM}, and λ_{Ξ_a} be the maximum eigenvalue of $\Xi_a\Xi_a^{\mathsf{T}}$. Then, ΔL_{a2} becomes

$$\Delta L_{a2} = a_a^2\left\|\sigma_{aM}\Xi_a^{\mathsf{T}}\right\|^2 \leq a_a^2\lambda_{\Xi_a}\bar{\sigma}_{aM}^2. \tag{55}$$

According the proof of [44, Th. 2], let $\bar{w}_c$, $\bar{w}_a$, $\bar{w}_d$, $\bar{\sigma}_1$, $\bar{y}_c$, $\bar{A}_a$, and $\bar{A}_d$ be the upper-bounds of $\tilde{w}_c$, $\tilde{w}_a$, $\tilde{w}_d$, $\nabla\sigma(\cdot)$, Y_c, A_a, and A_d, respectively. Then, we have

$$\left\|\sigma\left(\Phi_{a,p}\right)\tilde{w}_c^{i\mathsf{T}}\nabla\sigma\left(\Phi_{c,p}\right)Y_{c,p}^{\mathsf{T}}A_{a,p}\right\|^2 \le \bar{\sigma}^2\bar{w}_c^2\bar{\sigma}_1^2\bar{y}_{c,p}^2\bar{A}_{a,p}^2 = \Theta_{1,p}$$

$$\left\|\sigma\left(\Phi_{a,p}\right)w_c^{i\mathsf{T}}\nabla\sigma\left(\Phi_{c,p}\right)Y_{c,p}^{\mathsf{T}}A_{a,p}\right\|^2 \le \bar{\sigma}^2 w_{cm}^2\bar{\sigma}_1^2\bar{y}_{c,p}^2\bar{A}_{a,p}^2 = \Theta_{2,p}$$

$$\left\|\sigma\left(\Phi_{d,p}\right)\tilde{w}_c^{i\mathsf{T}}\nabla\sigma\left(\Phi_{c,p}\right)Y_{c,p}^{\mathsf{T}}A_{d,p}\right\|^2 \le \bar{\sigma}^2\bar{w}_c^2\bar{\sigma}_1^2\bar{y}_{c,p}^2\bar{A}_{d,p}^2 = \Theta_{3,p}$$

$$\left\|\sigma\left(\Phi_{d,p}\right)w_c^{i\mathsf{T}}\nabla\sigma\left(\Phi_{c,p}\right)Y_{c,p}^{\mathsf{T}}A_{d,p}\right\|^2 \le \bar{\sigma}^2 w_{cm}^2\bar{\sigma}_1^2\bar{y}_{c,p}^2\bar{A}_{d,p}^2 = \Theta_{4,p}.$$

Then, we can get

$$\Delta L_a \le a_a n_r \bar{w}_a^2 + a_a \sum_{p=1}^{n_r}\Theta_{1,p} + a_a\sum_{p=1}^{n_r}\Theta_{2,p} + a_a^2\lambda_{\Xi_a}\bar{\sigma}_{aM}^2. \tag{56}$$

Let $\sigma_{dM} = [\sigma(\Phi_{d,1}),\ldots,\sigma(\Phi_{d,n_r})]$, $\Xi_d = [\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,1})Y_{c,1}^{\mathsf{T}}A_{d,1},\ldots,\hat{w}_c^{i\mathsf{T}}\nabla\sigma(\Phi_{c,n_r})Y_{c,n_r}^{\mathsf{T}}A_{d,n_r}]$, $\bar{\sigma}_{dM}$ be the norm bound of σ_{dM}, and λ_{Ξ_d} be the maximum eigenvalue of $\Xi_d\Xi_d^{\mathsf{T}}$. Then, by using the similar process, we have

$$\Delta L_d \le a_d n_r \bar{w}_d^2 + a_d\sum_{p=1}^{n_r}\Theta_{3,p} + a_d\sum_{p=1}^{n_r}\Theta_{4,p} + a_d^2\lambda_{\Xi_d}\bar{\sigma}_{dM}^2. \tag{57}$$

Denote

$$\mathcal{P}_1 = a_a n_r\bar{w}_a^2 + a_a\sum_{p=1}^{n_r}\Theta_{1,p} + a_a\sum_{p=1}^{n_r}\Theta_{2,p} + a_a^2\lambda_{\Xi_a}\bar{\sigma}_{aM}^2$$

$$\mathcal{P}_2 = a_d n_r\bar{w}_d^2 + a_d\sum_{p=1}^{n_r}\Theta_{3,p} + a_d\sum_{p=1}^{n_r}\Theta_{4,p} + a_d^2\lambda_{\Xi_d}\bar{\sigma}_{dM}^2.$$

Then, according to (51), (56), and (57), we can get

$$\begin{aligned}\Delta L &= \Delta L_c + \Delta L_a + \Delta L_d\\ &\le -\left(1 - 2a_c\bar{\sigma}_{cM}^2\right)\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2\\ &\quad + 2n_r a_c\bar{\sigma}_{cM}^2\bar{\delta}_c^2 + n_r\bar{\delta}_c^2 + \mathcal{P}_1 + \mathcal{P}_2\\ &\le -\left(1 - 2a_c\bar{\sigma}_{cM}^2\right)\sum_{p=1}^{n_r}\left\|\Psi_{ci,p}\right\|^2 + \mathcal{P}_3\end{aligned} \tag{58}$$

where $\mathcal{P}_3 = 2n_r a_c\bar{\sigma}_{cM}^2\bar{\delta}_c^2 + n_r\bar{\delta}_c^2 + \mathcal{P}_1 + \mathcal{P}_2$. Therefore, if $a_c < (1/[2\bar{\sigma}_{cM}^2])$ holds, and $\Psi_{ci,p}$ lies outside the compact set

$$\Omega_{\Psi_{ci,p}} = \left\{\Psi_{ci,p} : \left\|\Psi_{ci,p}\right\| \le \sqrt{\frac{\mathcal{P}_3}{1 - 2a_c\bar{\sigma}_{cM}^2}}\right\}$$

we have $\Delta L < 0$. The proof is completed. ∎

Remark 7: In fact, Assumption 3 is widely used in previous papers [41], [44], [45]. The rationalities of Assumption 3 are explained as follows.

1) In the NN implementation, the activation function $\sigma(\cdot)$ is usually selected as $\tanh(\cdot)$. Therefore, it is reasonable to assume that it is norm bounded.
2) It is noticed that the NN weights ω_c^*, ω_a^*, and ω_d^* are optimal weights. In addition, δ_c is the reconstruction error of the critic NN. Since they cannot be infinite in practice, it is reasonable to assume that they are norm-bounded.

TABLE I
PARAMETERS OF TORSIONAL PENDULUM SYSTEM

Symbols	Description	Value
M	Mass of the pendulum bar	$\frac{1}{3}$kg
l_t	Length of the pendulum bar	$\frac{2}{3}$m
θ	Angle	$\theta(0) = 0.2$
$\mathscr{W}$	Angular velocity	$\mathscr{W}(0) = -0.2$
J_t	Rotary inertia	$\frac{4}{3}Ml_t^2$
f_d	Frictional factor	2
g	Gravity	9.8m/s^2

Remark 8: The difference between the developed DPGETC method and the existing methods [26], [37], [39], [41], and [45] are outlined as follows.

1) On the one hand, unlike [37] which solved the ZSG problem for CT systems, the developed DPGETC method can be applied to DT systems. On the other hand, different from [41] which tackled the optimal control problem with the dual heuristic dynamic programming (DHP)-based ETC approach, this article considered the ZSG problem. The developed data-based DPGETC approach updates the control law and the disturbance law by the corresponding gradients of the Q-function. Therefore, the model NN is not required anymore. Furthermore, the ER technique is adopted to design novel NN weight updating laws to improve the data usage efficiency.
2) This article extends the DPGADP method to solve the ZSG problem under the event-triggering mechanism. Compared with [26], [39], and [45], the control law and the disturbance law are renovated aperiodically at triggering instants only to reduce the computational and communication burden.

Remark 9: This article investigates the ZSG problem rather than the non-ZSG problems. In ZSG, two players compete with each other since the goals of the two players are completely opposite, i.e., one tries to minimize the performance index function and another one tries to maximize it. However, for nonzero-sum games, all players have their individual control objectives and an overall goal, i.e., they are not only competitive, but also cooperative [46]–[49].

VI. SIMULATION STUDIES

In this section, two simulation examples are employed to verify the effectiveness of the DPGETC scheme.

A. Example 1

Consider the torsional pendulum system given by

$$\frac{d\theta}{dt} = d\mathscr{W}$$

$$J_t\frac{d\mathscr{W}}{dt} = u_t - Mgl_t\sin\theta - f_d\frac{d\theta}{dt} + d_t.$$

The parameters and corresponding initial values of the torsional pendulum system are given in Table I. Inspired by [45],

TABLE II
CONTROL PARAMETERS

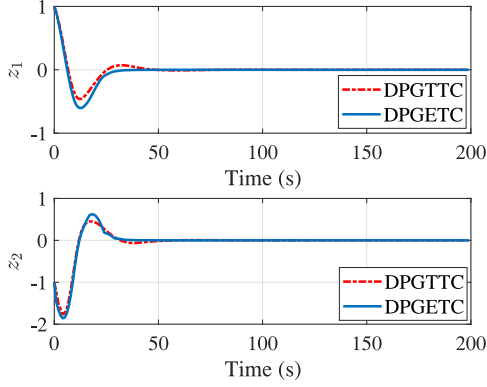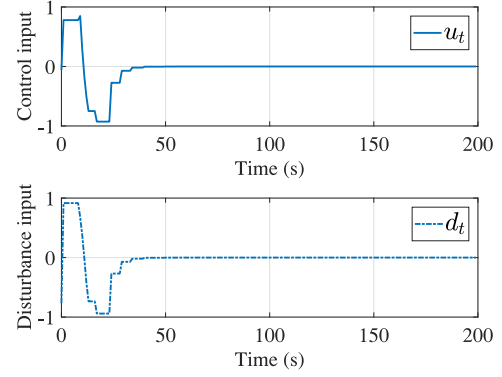|	Initial state	$\mathcal{Q}$	$\mathcal{R}$	β	n_r	l_1	l_2	a_c	a_a	a_d
Example 1	$[1, -1]$	$0.01I_2$	$0.01I$	1	20	0.3	0.3	0.2	0.1	0.15
Example 2	$[1, -1]$	$0.01I_2$	$0.01I$	1	20	0.2	0.2	0.3	0.3	0.3

Fig. 2.   Trajectories of system states z_1 and z_2 under DPGETC and DPGTTC methods of Example 1.

Fig. 3.   Norm value of weight vectors of Example 1.

Fig. 4.   Trajectories of the control input and the disturbance input of Example 1.

Fig. 5.   Evolution of the triggering condition of Example 1.
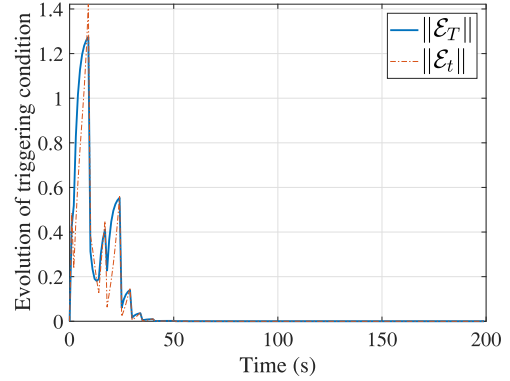

Fig. 6.   Comparison of sample numbers in Example 1.

the torsional pendulum system is discretized as

$$\begin{bmatrix} z_{1(t+1)} \\ z_{2(t+1)} \end{bmatrix} = \begin{bmatrix} 0.1z_{2t} + z_{1t} \\ -0.49\sin(z_{1t}) - 0.1f_d z_{2t} + z_{2t} \end{bmatrix} + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} u_t + \begin{bmatrix} 0 \\ 0.1 \end{bmatrix} d_t$$

where $z_t = [z_{1t}, z_{2t}]^{\mathsf{T}} = [\theta_t, \mathscr{W}_t]^{\mathsf{T}}$, u_t is the control input, and d_t is the disturbance. The initial state of the torsional pendulum system and the control parameters are provided in Table II. The structure of the critic network, the action network, and the disturbance network are selected as 4–10–1, 2–10–1, and 2–10–1, respectively. The activation functions are selected as tanh. The initial weights of all NNs are randomly chosen within $[-1, 1]$.

The simulation results are exhibited in Figs. 2–6. The evolution of system states under the DPGETC method and DPG-based time triggered control (DPGTTC) approach is revealed in Fig. 2. We can find that system states approach to a small region of zero after 50 s. The norms of the weight vectors are provided in Fig. 3, which displays that the weights
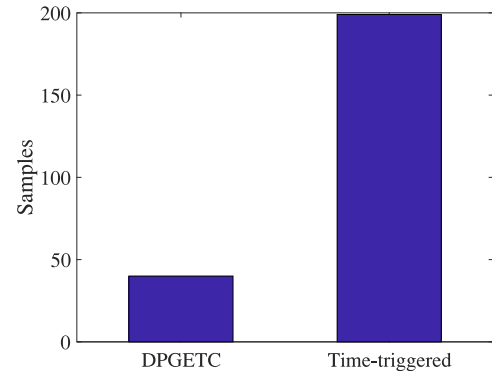
of NNs remain unchanged after 100 s. Fig. 4 indicates that the control law and the disturbance law are aperiodic updated and reach zero after 50 s. The event-triggering error $\|\mathcal{E}_t\|$ and threshold $\mathcal{E}_T$ are depicted in Fig. 5, which explicitly displays that the event-triggering error will converge to zero. A comparison of the sample number between the DPGETC method
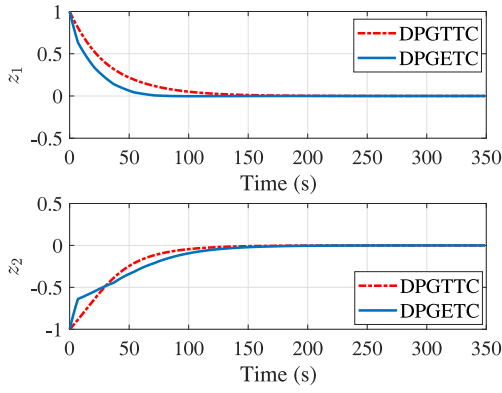
Fig. 7.   Trajectories of system states z_1 and z_2 under DPGETC and DPGTTC methods of Example 2.

Fig. 8.   Norm value of weight vectors of Example 2.

and the DPGTTC method is revealed in Fig. 6. Compared with 200 samples in the time-triggered method, the DPGETC method needs 40 samples only, which means that it saves 80% computation. Hence, the validity of the DPGETC approach is verified.

Remark 10: The parameters l_1 and l_2 in the triggering condition (38) play an important role on the control performance since they determine the triggering threshold $\mathcal{E}_T$ directly. If the triggering threshold $\mathcal{E}_T$ is too small, the control inputs are updated with high frequency, which brings a large amount of computation. However, if the triggering threshold is too large, the control inputs are updated with a low frequency, but the system may be unstable. Therefore, we need to select these two parameters appropriately with repeatitive simulations to tradeoff the control performance and the computational burden.

Remark 11: It is noticed that the selection of the NN structure and the historical data size are challenging since they affect the control performance directly. In this article, we select the NN structure and the historical data size by "trial and error" with repetitive simulations.

B. Example 2

Consider the nonaffine nonlinear system

$$z_{1(t+1)} = 0.97z_{1t} + 0.97z_{2t}u_t + 0.97d_t$$
$$z_{2(t+1)} = 0.97z_{2t} + 0.97\left(1 + z_{1t}^2\right)u_t + 0.97u_t^3 + 0.97d_t^2.$$

The control parameters are displayed in Table II. In this example, the structure of NNs and the activation functions are the same as those of Example 1. We randomly initialize the NN weights within $[-0.5, 0.5]$. Simulation results are provided in Figs. 7–11. Fig. 7 presents that the state trajectories under the DPGETC method and DPGTTC approach stationary points after 250 s. We can discover that the norms of weight vectors are convergent in Fig. 8. The curves of the ETC law and the disturbance law are given in Fig. 9. It is distinct that both of them are segmented signals and arrive to a small region of zero after 200 s. The event-triggering error $\|\mathcal{E}_t\|$ and threshold $\mathcal{E}_T$ are depicted in Fig. 10, which shows that the event-triggering error will converge to zero. The sample numbers of the DPGETC approach and the DPGTTC approach are comparatively displayed in Fig. 11. We can conclude
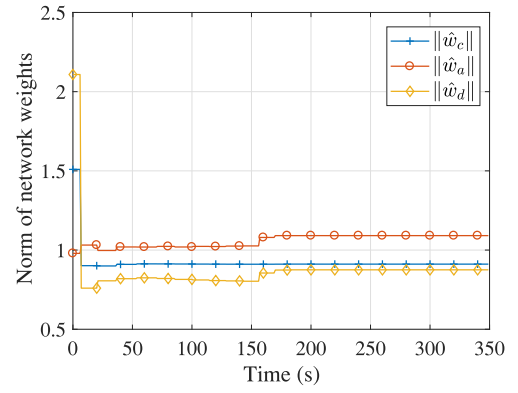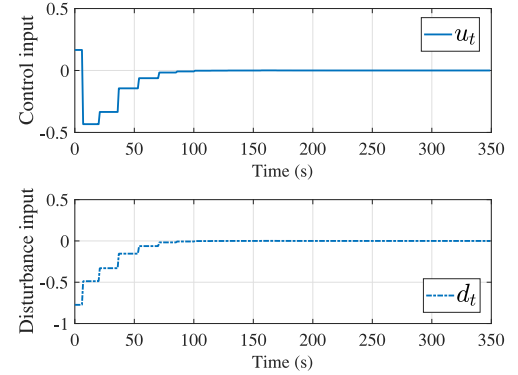

Fig. 9.   Trajectories of the control input and the disturbance input of Example 2.
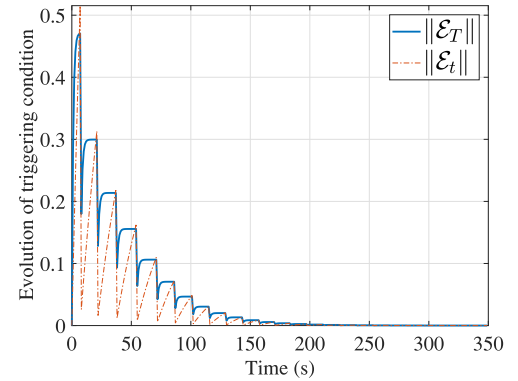
Fig. 10.   Evolution of the triggering condition of Example 2.

that the time triggered-based controller updates 350 times, and the DPGETC-based controller updates 24 times only. From the above analysis, the developed DPGETC method greatly reduces the computational and communication burden.

Remark 12: The convergence rate and the algorithm complexity depend on the parameters of the NN, i.e., the NN structure and the learning rate. On the one hand, a simple NN structure can reduce the complexity but may result in poor approximation performance. Otherwise, if the NN structure is complex, the approximation performance will be improved, but the complexity will be increased. On the other hand, a large learning rate will accelerate the convergence, but the NN weights may not converge, and vice versa. Therefore, we need to select these parameters based on repeatitive simulations to

Fig. 11.   Comparison of sample numbers in Example 2.

get a tradeoff. In fact, the convergence rate and the algorithm complexity are mainly reflected in the training stage. Once the optimal weights are obtained, we can implement the optimal controller directly.
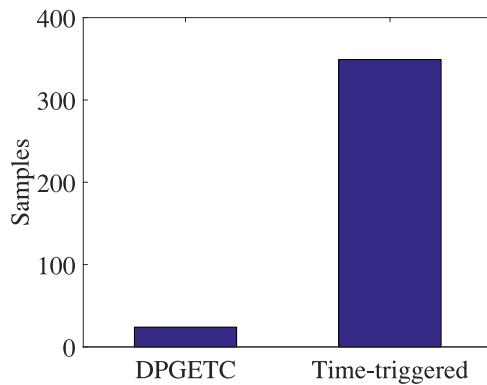
Remark 13: In the ETC approach, an event generator is employed to monitor the event-triggering error. Once the error exceeds the triggering threshold, an event is triggered and the current state is sampled as a new sampled state. Moreover, the control input is updated and remains unchanged until the next triggering instant.

Remark 14: The main difference between the ETC approach and the time-triggered approach is that the ETC is updated aperiodically, rather than periodically as the time-triggered one. In the ETC approach, only when the magnitude of the event-triggering error reaches the prescribed threshold, an event is triggered and the control input is updated. Therefore, compared with the time-triggered method, the number of control input updating are reduced, and the computation and communication resources are saved.

VII. Conclusion

In this article, the ZSG problem for DT nonlinear systems is investigated by using the DPGETC approach. A proper event-triggering condition is deduced to guarantee the ISS of the closed-loop system. The control law and the disturbance law are approximated by the action network and the disturbance network, and are updated by the corresponding gradients of the Q-function. Therefore, it is a data-based method and the system dynamics are not required. Moreover, based on the ER technique, new NN weight tuning laws are designed to ensure the NN weight estimation errors to be UUB. Simulation studies show that the developed DPGETC method alleviates the computational and communication burden. The main novelties of the developed DPGETC control scheme lie in that: 1) the ZSG problem of unknown nonaffine nonlinear systems is addressed without requiring the model NN and 2) the designed controllers are updated aperiodically to reduce the computational and communication burden. The inadequacies of this control approach and the related future work are provided as follows.

1) It is noticed that Assumptions 2 and 3 are necessary to ensure the stability of the closed-loop system. In our future work, we will try to relax these assumptions.

2) This article adopts the ETC approach to reduce the computational and communication burden. However, it requires hardware equipment to monitor whether the triggering condition is satisfied, and then triggering the next sampling in real time, which is not available in many practical systems. In our future work, we will try to develop the ADP-based self-triggered control approach since it calculates the next sampling time based on the latest triggering instant and system dynamics without additional monitoring equipments.

References

[1] Y. Li, X. Min, and S. Tong, "Observer-based fuzzy adaptive inverse optimal output feedback control for uncertain nonlinear systems," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 6, pp. 1484–1495, Jun. 2021.

[2] Y. Li, T. Yang, and S. Tong, "Adaptive neural networks finite-time optimal control for a class of nonlinear systems," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 31, no. 11, pp. 4451–4460, Nov. 2020.

[3] L. Liu, T. Gao, Y.-J. Liu, S. Tong, C. L. P. Chen, and L. Ma, "Time-varying IBLFs-based adaptive control of uncertain nonlinear systems with full state constraints," *Automatica*, vol. 129, Jul. 2021, Art. no. 109595.

[4] L. Liu, Y.-J. Liu, S. C. Tong, and C. L. P. Chen, "Integral barrier Lyapunov function-based adaptive control for switched nonlinear systems," *Sci. China Inf. Sci.*, vol. 63, no. 3, pp. 1–14, Mar. 2020.

[5] D. Liu, Q. Wei, D. Wang, X. Yang, and H. Li, *Adaptive Dynamic Programming With Applications in Optimal Control*. Cham, Switzerland: Springer, 2017.

[6] P. J. Werbos, "Approximate dynamic programming for real-time control and neural modeling," in *Handbook of Intelligent Control: Neural, Fuzzy, and Adaptive Approaches*, D. A. White and D. A. Sofge, Eds. New York, NY, USA: Van Nostrand Reinhold, 1992, ch. 13.

[7] H. Jiang and H. Zhang, "Iterative ADP learning algorithms for discrete-time multi-player games," *Artif. Intell. Rev.*, vol. 50, no. 1, pp. 75–91, 2018.

[8] Q. Wei, D. Liu, Y. Liu, and R. Song, "Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 4, no. 2, pp. 168–176, Apr. 2017.

[9] D. Liu, Y. Xu, Q. Wei, and X. Liu, "Residential energy scheduling for variable weather solar energy based on adaptive dynamic programming," *IEEE/CAA J. Autom. Sinica*, vol. 5, no. 1, pp. 36–46, Jan. 2018.

[10] D. Liu, S. Xue, B. Zhao, B. Luo, and Q. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[11] H. Lin, B. Zhao, D. Liu, and C. Alippi, "Data-based fault tolerant control for affine nonlinear systems through particle swarm optimized neural networks," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 4, pp. 954–964, Jul. 2020.

[12] D. Liu and Q. Wei, "Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 25, no. 3, pp. 621–634, Mar. 2014.

[13] D. Liu, D. Wang, and H. Li, "Decentralized stabilization for a class of continuous-time nonlinear interconnected systems using online learning optimal control approach," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 25, no. 2, pp. 418–428, Feb. 2014.

[14] D. Liu, X. Yang, D. Wang, and Q. Wei, "Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints," *IEEE Trans. Cybern.*, vol. 45, no. 7, pp. 1372–1385, Jul. 2015.

[15] B. Zhao, D. Liu, and C. Alippi, "Sliding-mode surface-based approximate optimal control for uncertain nonlinear systems with asymptotically stable critic structure," *IEEE Trans. Cybern.*, vol. 51, no. 6, pp. 2858–2869, Jun. 2021.

[16] B. Zhao, D. Liu, and C. Luo, "Reinforcement learning-based optimal stabilization for unknown nonlinear systems subject to inputs with uncertain constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4330–4340, Oct. 2020.

[17] B. Niu, D. Wang, M. Liu, X. Song, H. Wang, and P. Duan, "Adaptive neural output-feedback controller design of switched nonlower triangular nonlinear systems with time delays," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 10, pp. 4084–4093, Oct. 2020.

[18] B. Zhao, D. Wang, G. Shi, D. Liu, and Y. Li, "Decentralized control for large-scale nonlinear systems with unknown mismatched interconnections via policy iteration," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 48, no. 10, pp. 1725–1735, Oct. 2018.

[19] Y. Zhu and D. Zhao, "Comprehensive comparison of online ADP algorithms for continuous-time optimal control," *Artif. Intell. Rev.*, vol. 49, no. 4, pp. 531–547, 2018.

[20] B. Zhao and D. Liu, "Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3054–3064, Apr. 2020.

[21] B. Niu, Y. Liu, W. Zhou, H. Li, P. Duan, and J. Li, "Multiple Lyapunov functions for adaptive neural tracking control of switched nonlinear nonlower-triangular systems," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 1877–1886, May 2020.

[22] B. Niu, D. Wang, N. D. Alotaibi, and F. E. Alsaadi, "Adaptive neural state-feedback tracking control of stochastic nonlinear switched systems: An average dwell-time method," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 4, pp. 1076–1087, Apr. 2019.

[23] B. Zhao, D. Liu, and Y. Li, "Online fault compensation control based on policy iteration algorithm for a class of affine non-linear systems with actuator failures," *IET Control Theory Appl.*, vol. 10, no. 15, pp. 1816–1823, 2016.

[24] D. Wang, H. He, and D. Liu, "Adaptive critic nonlinear robust control: A survey," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3429–3451, Oct. 2017.

[25] Q. Wei, D. Liu, Q. Lin, and R. Song, "Adaptive dynamic programming for discrete-time zero-sum games," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 957–969, Apr. 2018.

[26] X. Zhong, H. He, D. Wang, and Z. Ni, "Model-free adaptive control for unknown nonlinear zero-sum differential game," *IEEE Trans. Cybern.*, vol. 48, no. 5, pp. 1633–1646, May 2018.

[27] Y. Zhu, D. Zhao, and X. Li, "Iterative adaptive dynamic programming for solving unknown nonlinear zero-sum game based on online data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 3, pp. 714–725, Mar. 2017.

[28] X. Yang and H. He, "Adaptive critic designs for event-triggered robust control of nonlinear systems with unknown dynamics," *IEEE Trans. Cybern.*, vol. 49, no. 6, pp. 2255–2267, Jun. 2019.

[29] M. Ha, D. Wang, and D. Liu, "Event-triggered adaptive critic control design for discrete-time constrained nonlinear systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 9, pp. 3158–3168, Sep. 2020.

[30] D. Wang, C. Mu, D. Liu, and H. Ma, "On mixed data and event driven design for adaptive-critic-based nonlinear H_∞ control," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 4, pp. 993–1005, Apr. 2018.

[31] H. Ma, H. Li, R. Lu, and T. Huang, "Adaptive event-triggered control for a class of nonlinear systems with periodic disturbances," *Sci. China Inf. Sci.*, vol. 63, no. 5, pp. 161–175, Mar. 2020.

[32] L. Liu, X. Li, Y.-J. Liu, and S. Tong, "Neural network based adaptive event trigger control for a class of electromagnetic suspension systems," *Control Eng. Pract.*, vol. 106, Jan. 2021, Art. no. 104675.

[33] H. Wang, K. Xu, and J. Qiu, "Event-triggered adaptive fuzzy fixed-time tracking control for a class of nonstrict-feedback nonlinear systems," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 7, pp. 3058–3068, Jul. 2021.

[34] H. Wang, S. Ling, P. Liu, and Y.-X. Li, "Control of high-order nonlinear systems under error-to-actuator based event-triggered framework," *Int. J. Control*, to be published, doi: 10.1080/00207179.2021.1934734.

[35] A. Sahoo, H. Xu, and S. Jagannathan, "Near optimal event-triggered control of nonlinear discrete-time systems using neurodynamic programming," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 9, pp. 1801–1815, Sep. 2016.

[36] L. Dong, X. Zhong, C. Sun, and H. He, "Event-triggered adaptive dynamic programming for continuous-time systems with control constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 8, pp. 1941–1952, Aug. 2017.

[37] S. Xue, B. Luo, and D. Liu, "Event-triggered adaptive dynamic programming for zero-sum game of partially unknown continuous-time nonlinear systems," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 9, pp. 3189–3199, Sep. 2020.

[38] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proc. 31st Int. Conf. Mach. Learn.*, vol. 32. Beijing, China, 2014, pp. 387–395.

[39] B. Luo, D. Liu, H.-N. Wu, D. Wang, and F. L. Lewis, "Policy gradient adaptive dynamic programming for data-based optimal control," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3341–3354, Oct. 2017.

[40] L. Dong, X. Zhong, C. Sun, and H. He, "Adaptive event-triggered control based on heuristic dynamic programming for nonlinear discrete-time systems," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 7, pp. 1594–1605, Jul. 2017.

[41] D. Wang, M. Ha, and J. Qiao, "Self-learning optimal regulation for discrete-time nonlinear systems under event-driven formulation," *IEEE Trans. Autom. Control*, vol. 65, no. 3, pp. 1272–1279, Mar. 2020.

[42] M. Lazar, W. P. M. H. Heemels, and A. R. Teel, "Further input-to-state stability subtleties for discrete-time systems," *IEEE Trans. Autom. Control*, vol. 58, no. 6, pp. 1609–1613, Jun. 2013.

[43] Z.-P. Jiang and Y. Wang, "Input-to-state stability for discrete-time nonlinear systems," *Automatica*, vol. 37, no. 6, pp. 857–869, 2001.

[44] Y. Sokolov, R. Kozma, L. D. Werbos, and P. J. Werbos, "Complete stability analysis of a heuristic approximate dynamic programming control design," *Automatica*, vol. 59, pp. 9–18, Sep. 2015.

[45] Y. Zhang, B. Zhao, and D. Liu, "Deterministic policy gradient adaptive dynamic programming for model-free optimal control," *Neurocomputing*, vol. 387, pp. 40–50, Apr. 2020.

[46] Y. Fu and T. Chai, "Online solution of two-player zero-sum games for continuous-time nonlinear systems with completely unknown dynamics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2577–2587, Dec. 2016.

[47] A. Al-Tamimi, M. Abu-Khalaf, and F. L. Lewis, "Adaptive critic designs for discrete-time zero-sum games with application to H_∞ control," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 37, no. 1, pp. 240–247, Feb. 2007.

[48] K. G. Vamvoudakis and F. L. Lewis "Multi-player non-zero-sum games: Online adaptive learning solution of coupled Hamilton-Jacobi equations," *Automatica*, vol. 24, no. 10, pp. 1556–1569, Aug. 2011.

[49] K. G. Vamvoudakis, H. Modares, B. Kiumarsi, and F. L. Lewis, "Game theory-based control system algorithms with real-time reinforcement learning: How to solve multiplayer games online," *IEEE Control Syst. Mag.*, vol. 37, no. 1, pp. 33–52, Feb. 2017.
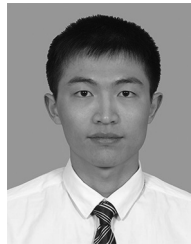
Yongwei Zhang received the B.S. degree in automation from the School of Electronic and Information Engineering, Jiaying University, Meizhou, China, in 2016, and the M.S. and Ph.D. degrees in control science and engineering from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2018 and 2021, respectively.

He is a Postdoctoral Fellow with the Guangdong University of Technology. His current research interests include adaptive dynamic programming and optimal control.

Bo Zhao (Senior Member, IEEE) received the B.S. degree in automation and the Ph.D. degree in control science and engineering from Jilin University, Changchun, China, in 2009 and 2014, respectively.

He was a Postdoctoral Fellow with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China, from 2014 to 2017, where he joined the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation from 2017 to 2018. He is currently an Associate Professor with the School of Systems Science, Beijing Normal University, Beijing. He has authored or coauthored over 100 journal and conference articles. His research interests include adaptive dynamic programming, robot control, fault diagnosis and tolerant control, optimal control, and neural-network-based control.

Dr. Zhao serves as an Associate Editor for *Neurocomputing* and a Guest Editor for *Complex and Intelligent Systems*. He is the Secretary of the Adaptive Dynamic Programming and Reinforcement Learning Technical Committee of Chinese Association of Automation (CAA) and was the Secretary of 2017 the 24th International Conference on Neural Information Processing. He is an Asia–Pacific Neural Network Society Member and a member of CAA.

Derong Liu (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 1994.

He was a Staff Fellow of General Motors Research and Development Center, Warren, MI, USA, from 1993 to 1995. He was an Assistant Professor with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA, from 1995 to 1999. He joined the University of Illinois at Chicago, Chicago, IL, USA, in 1999, and became a Full Professor of Electrical and Computer Engineering, and Computer Science in 2006. He served as an Associate Director of the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Beijing, China, from 2010 to 2015. He was a Full Professor with the School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing, China, from 2015 to 2017. He is currently a Full Professor with the School of Automation, Guangdong University of Technology, Guangzhou, China. He has authored or coauthored 19 books.

Prof. Liu was a recipient of the Michael J. Birck Fellowship from the University of Notre Dame, in 1990, the Harvey N. Davis Distinguished Teaching Award from the Stevens Institute of Technology, in 1997, the Faculty Early Career Development Award from the National Science Foundation in 1999, the University Scholar Award from the University of Illinois, from 2006 to 2009, and the Overseas Outstanding Young Scholar Award from the National Natural Science Foundation of China in 2008. He was selected for the "100 Talents Program" by the Chinese Academy of Sciences in 2008. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS from 2010 to 2015. He is the Editor-in-Chief of *Artificial Intelligence Review* (Springer). He is a Fellow of the International Neural Network Society and the International Association of Pattern Recognition.

Shunchao Zhang received the B.S. degree in measurement and control technology and instrument from the School of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan, China, in 2016, and the M.S. degree in control engineering from the Guangdong University of Technology, Guangzhou, China, in 2019, where he is currently pursuing the Ph.D. degree in control science and engineering with the School of Automation.

His current research interests include optimal control and adaptive dynamic programming.

Riemannian Mean Shift-Based Data Fusion Scheme for Multi-Antenna Cooperative Spectrum Sensing

Yongwei Zhang, Shunchao Zhang, Yonghua Wang, *Member, IEEE*, Jiawei Zhuang, and Pin Wan

Abstract—In this article, the multi-antenna cooperative spectrum sensing problem in cognitive radio networks is investigated over Riemannian manifold. At the beginning, a signal matrix is constructed by using the sensing signals from secondary users (SUs) and the corresponding covariance matrix is calculated. Subsequently, the covariance matrices are transmitted to the fusion center and mapped to points on the manifold. In order to reduce the impact of aberrant SUs, a data fusion scheme based on Riemannian mean shift algorithm is developed. After data fusion, the representative points are obtained to train a classifier. In order to realize clustering directly over Riemannian manifold, a novel Riemannian distance based particle swarm optimization (RDPSO) algorithm, is proposed to train a classifier, which is employed to determine the state of primary user (PU). Finally, in simulation part, the validity of the proposed scheme is verified under different scenarios.

Index Terms—Cognitive radio networks, multi-antenna cooperative spectrum sensing, Riemannian mean, Riemannian distance.

I. INTRODUCTION

WITH the continuous development of wireless communication network, spectrum resource is becoming increasingly scarcity [1]–[3]. However, most of the spectrum resources are not using sufficient [4]. In order to alleviate the shortage of spectrum resources, cognitive radio networks (CRN) have emerged [5], [6]. Different from traditional methods, CRN introduces cognitive radio (CR) technology to enable secondary users (SUs) to perceive the status of authorized channel. The SU can opportunistically access the available channel for communication when the primary user (PU) is not on the channel [7]. There are several spectrum sensing algorithms, such as energy detection (ED) [8], [9], matched filtering detection (MFD) [10], [11] and cyclostationary feature detection (CFD) [12], [13] have been proposed. However, these methods are restricted in practical application due to some shortcomings, such as require the prior

information of the PU, susceptible to noise uncertainty and require a large amount of computation.

In the modern CRN, more and more communication devices adopt multiple antennas to improve link reliability and communication quality [14]–[16]. Multi-antenna spectrum sensing has become a research hotspot in the field of spectrum sensing over the past decade. In [17], a improved energy detector which uses an arbitrary positive power of amplitudes of the PU's signal samples was developed for multi-antenna cooperative spectrum sensing. The mathematical expressions of the probabilities of the false alarm and the missed detection for the developed approach were deduced. In [18], a detection scheme based on eigenvalue weight was proposed for multi-antenna CRN. Based on Neyman-Pearson criterion, the eigenvalue based-likelihood ratio test was analyzed and a simple closed-form expression was deduced. In [19], the optimal Wald test based sequential Bartlett spectral detector was designed for multi-antenna cooperative spectrum sensing (CSS). It is suitable for single/multiple PU scenario or multiple-input multiple-output wireless systems. In [20], a multi-antenna CSS approach based on the expectation maximization (EM) algorithm was developed to detect the PU signal. This method can be regarded as joint detection and estimation, which used EM algorithm to detect PU signal and estimated the unknown channel frequency response and the noise variance of multiple subbands iteratively. In [21], a multi-user multiple-input and multiple-output based CR approach was developed for Internet of Things, which used weighted-eigenvalue detection technique to analyze sensing, system throughput, energy efficiency and expected lifetime. It is noticed that the above CSS methods need to derive an accurate threshold to judge the state of the channel, which is difficult to achieve in a complex environment.

In recent years, there are some machine leaning based spectrum sensing approaches have been present to solve this problem [22]–[24]. In [22], several familiar machine learning algorithms, such as K-means clustering, Gaussian mixture model and K-nearest-neighbor were adopted to develop a CSS scheme. This method used signal energy as a feature and trained a classifier to judge whether the PU exists or not. In [23], a support vector machine-based CSS model was presented. In order to reduce the cooperation cost and improve the sensing performance, CR users were appropriately grouped using energy data samples and support vector machine models. In [24], a multiple-antenna CSS was proposed by using the wavelet transform algorithm and the Gaussian mixture model. The noise of the signal was removed by wavelet transform

and the Gaussian mixture model was adopted to obtain a classifier to detect the spectrum hole. In general, compared with traditional sensing methods, the machine learning based spectrum sensing method has the advantages of high sensing accuracy and strong adaptability to environment. By using the feature vector and the clustering algorithm, a classifier is trained to judge the channel state. Therefore, choosing suitable feature vector and clustering algorithm are essential to acquire expected detection performance.

It is well known that spectrum sensing can be regarded as a signal detection problem. By analyzing the probability distribution of the detection data, the existence of the PU can be determined. Under the background of information geometry (IG), the spectrum sensing problem can be converted to a geometric problem in manifold, and the geometric method can be adopted to analyze the properties of probability distribution function clusters. In recent years, several scholars study the spectrum sensing problem based on IG theory. In [26], a Riemannian distance (RD) detector which does not require the noise statistical characteristics and the priori knowledge of PU, was designed to estimate the channel state by using the RD and the Riemannian mean. In [27], a CSS scheme based on empirical mode decomposition (EMD) and IG was proposed to enhance the detection performance under complex electromagnetic environments. The EMD algorithm was adopted to denoise the signals collected by the SUs and the covariance matrices were calculated and mapped onto the manifold. Subsequently, the geodesic distance between two points on the manifold was used as a signal feature, and the K-medoids clustering algorithm was adopted to train a classifier to determine whether the PU exists or not. However, these methods use the geometric distance value on the manifold as a signal feature only and cannot be used directly in the manifold space.

In this paper, unlike previous CSS approaches which are suitable for vector space only, we propose a novel spectrum sensing method that can be employed on manifold space directly. In actual scenarios, SU may be interfered by the environment and receives aberrant signal data. If these data are used directly, the sensing performance of the whole CRN will be affected. Therefore, a novel data fusion method based on Riemannian mean shift (RMS) algorithm is developed to eliminate the aberrant data. Finally, a RD-based particle swarm optimization (RDPSO) algorithm which works on manifold space, is designed to acquire a classifier to perceive the state of PU. The innovations of this paper are given as follows.

1) This article develops a novel data fusion approach to eliminate the aberrant data by using the RMS algorithm. Traditional methods [24] and [27] use denoising algorithms to reduce environmental interference, but it may remove some useful information in the signal. The proposed data fusion scheme can exclude the aberrant data directly and is suitable on manifold space.

2) A novel RDPSO clustering algorithm is developed, which uses RD to estimate the two points on the manifold.

3) By using the sample points on the manifold and the RDPSO algorithm, an RMS based RDPSO (RMS-RDPSO) approach is proposed to train a classifier for
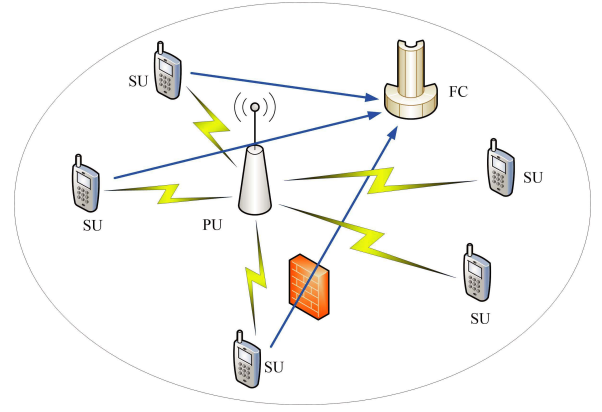

Fig. 1. The scenario of CRN.

determining the state of the channel. Different from existing methods [25] and [26], the developed CSS scheme is adaptive, it does not need to derive a precise threshold and can be used on manifold space.

The structure of this article is given follows. The scenario of the multi-antenna CSS is introduced in Section II. Section III develops an RMS-RDPSO approach for CSS. The validity of the RMS-RDPSO approach is tested under different scenarios in Section IV. Finally, a conclusion is provided in Section V.

II. Multi-Antenna Cooperative Spectrum Sensing in Cognitive Radio Network

In this section, we study a CRN system which contains one PU with single antenna and $\mathcal{L}$ SUs with $\mathcal{P}$ antennas [28]. The structure of the CRN system is provided in Fig. 1. In CSS scenario, in order to guarantee the communication of the PU, the CRN needs to judge the channel state before allowing the SUs to access the spectrum. Therefore, SUs should transmit their sensing data to fusion center (FC) for acquiring a global decision to identify whether the PU is exist. In summary, the spectrum sensing can be regarded as a binary hypotheses testing problem, which is formulated as

$$
\begin{aligned}
&\mathcal{H}_0 : y_l^p(n) = z_l^p(n), \ n = 1, \ldots, N, \\
&\mathcal{H}_1 : y_l^p(n) = h_l^p(n)x(n) + z_l^p(n), \ n = 1, \ldots, N, \quad (1)
\end{aligned}
$$

where $y_l^p(n)$ represents the signal received from the pth antenna of lth SU, $x(n)$ represents the signal transmitted by the PU and obeys the Gaussian distribution with mean zero and variance σ_x^2, i.e., $x(n) \sim \mathcal{N}(0, \sigma_x^2)$, and $z_l^p(n)$ represents the Gaussian white noise and satisfies $z_l^p(n) \sim \mathcal{N}(0, \sigma_{lp}^2)$, $l = 1, 2, \ldots, \mathcal{L}, \ p = 1, 2, \ldots, \mathcal{P}$. N represents the number of sample points, $h_l^p(n) = 1$ represents the channel gain [29], $\mathcal{H}_0$ and $\mathcal{H}_1$ stand for the absence and presence of PU, respectively.

In this paper, we use the probabilities of detection P_d and the false alarm P_f to reflect the sensing performance, which are defined as

$$
P_d = P[\mathcal{H}_1 | \mathcal{H}_1], \quad (2)
$$

$$
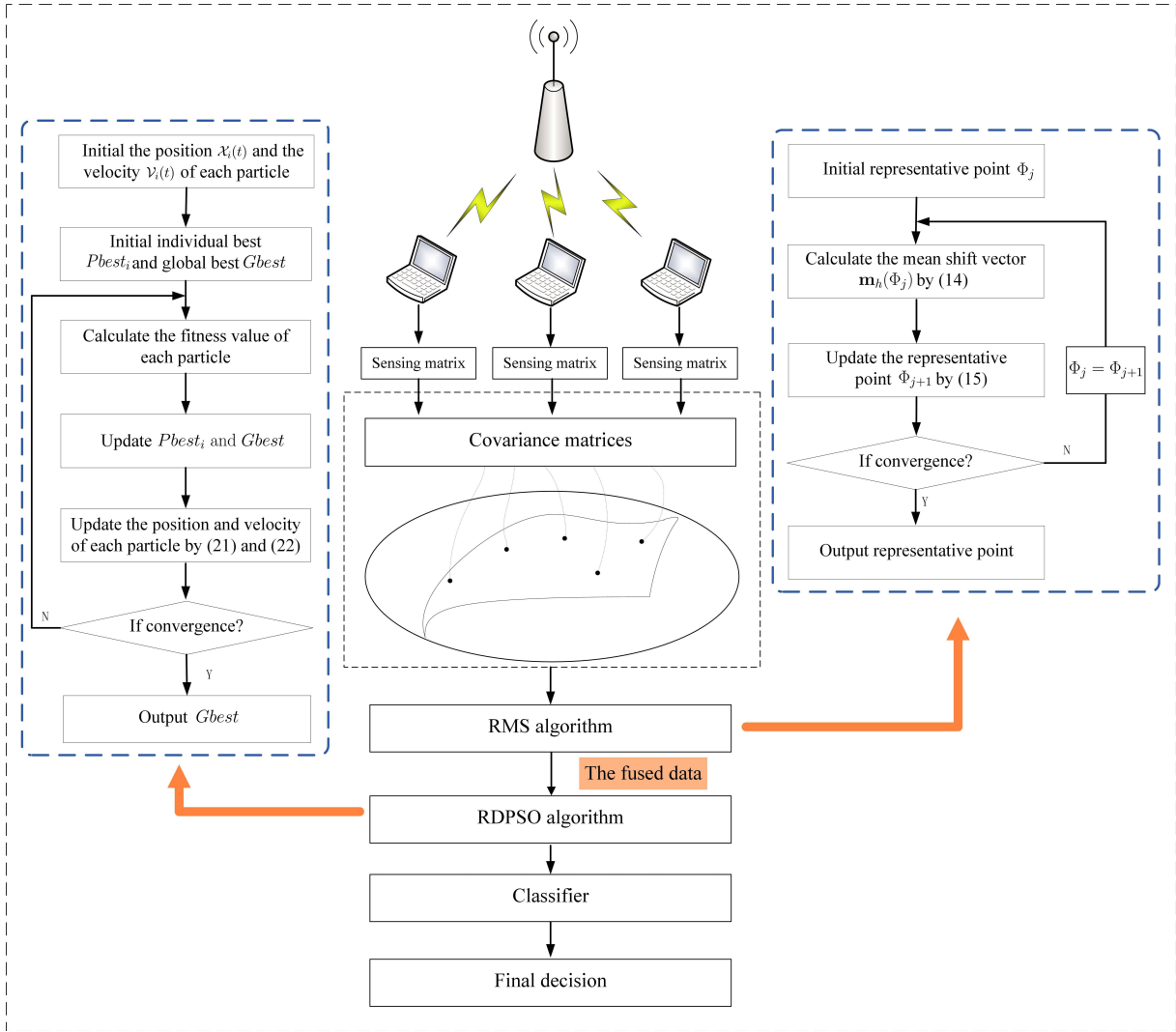P_f = P[\mathcal{H}_1 | \mathcal{H}_0]. \quad (3)
$$

Fig. 2.   The structure of the developed RMS-RDPSO scheme.

III. MULTI-ANTENNA COOPERATIVE SPECTRUM SENSING BASED ON RMS-RDPSO SCHEME

In this section, the structure of the developed RMS-RDPSO scheme is provided in Fig. 2. To begin with, the FC collects the sensing data from each SU and maps the corresponding covariance matrices to coordinate points on the manifold. In order to reduce the interference of aberrant SUs, the RMS algorithm is developed to fuse the coordinate points to a reference point. Moreover, the RDPSO algorithm is adopted to cluster directly on manifold. Finally, a data set which contains the reference points is collected to obtain a classifier to determine whether the PU exists.

Remark 1: The received signal from each SU with $\mathcal{P}$ antennas can form a signal matrix. Then, we can calculate the covariance matrices of all the signal matrices. It is noticed that under the background of IG theory, the covariance matrix can be regarded as a point on the Riemannian manifold. Therefore, we can use the RD to measure the distance between two points on the manifold. Since the SUs may be interfered by the environment and receive aberrant signal data. We develop an RMS algorithm to remove the abnormal data. It is well known that

the spectrum sensing is a binary hypothesis problem, i.e., the PU is present or PU is absent. Hence, we propose the RDPSO algorithm to train a classifier to judge whether the PU exists. It is worth mentioning that the developed RDPSO algorithm is worked on manifold space directly. Therefore, the spectrum sensing problem is solved on the Riemannian manifold.

A. Multi-Antenna Cooperative Spectrum Sensing Over Riemannian Manifold

In this section, the multi-antenna spectrum sensing problem is analyzed in manifold space. The received signal of the *p*th antenna of the *l*th SU is expressed as

$$y_l^p = [y_l^p(1), \ldots, y_l^p(N)]^{\mathrm{T}}. \qquad (4)$$

Therefore, the signal matrix of *l*th SU is given as

$$\mathbf{Y}_l = \begin{bmatrix} y_l^1(1) & y_l^1(2) & \cdots & y_l^1(N) \\ y_l^2(1) & y_l^2(2) & \cdots & y_l^2(N) \\ \vdots & \vdots & \ddots & \vdots \\ y_l^{\mathcal{P}}(1) & y_l^{\mathcal{P}}(2) & \cdots & y_l^{\mathcal{P}}(N) \end{bmatrix}. \qquad (5)$$

The corresponding covariance matrix is calculated by

$$\hat{\mathbf{C}}_l = \frac{1}{N}\mathbf{Y}_l\mathbf{Y}_l^T, \tag{6}$$

where $\hat{\mathbf{C}}_l \in \mathbb{R}^{\mathcal{P}\times\mathcal{P}}$. Then, the binary hypotheses can be reformulated as

$$\mathcal{H}_0 : \hat{\mathbf{C}}_l = \sigma_R^2\mathbf{I},$$
$$\mathcal{H}_1 : \hat{\mathbf{C}}_l = \hat{\mathbf{C}}_\beta + \sigma_R^2\mathbf{I}, \tag{7}$$

where $\hat{\mathbf{C}}_\beta \in \mathbb{R}^{\mathcal{P}\times\mathcal{P}}$ represents the covariance matrix of PU signal, $\mathbf{I} \in \mathbb{R}^{\mathcal{P}\times\mathcal{P}}$ represents the identity matrix, $\sigma_R^2 = \mathrm{diag}[\sigma_{l1}^2, \ldots, \sigma_{l\mathcal{P}}^2]$. According [26], we find that $\hat{\mathbf{C}}_l$ obeys the Wichter distributions $\mathbb{W}(\mathcal{P}, \sigma_R^2\mathbf{I})$ and $\mathbb{W}(\mathcal{P}, \hat{\mathbf{C}}_\beta + \sigma_R^2\mathbf{I})$ in the case of $\mathcal{H}_0$ and $\mathcal{H}_1$, respectively. Under the background of IG theory, the Wichter distributions $\mathbb{W}(\mathcal{P}, \sigma_R^2\mathbf{I})$ and $\mathbb{W}(\mathcal{P}, \hat{\mathbf{C}}_\beta + \sigma_R^2\mathbf{I})$ can be mapped into the statistical manifold and the covariance matrices $\sigma_R^2\mathbf{I}$ and $\hat{\mathbf{C}}_\beta + \sigma_R^2\mathbf{I}$ can be considered as the corresponding coordinates.

In this paper, the RD is employed to estimate the distance between two points on the manifold. Assume that Ξ_1 and Ξ_2 are two points on the manifold. Then, the RD between Ξ_1 and Ξ_2 is calculated as

$$
\begin{aligned}
\mathcal{R}_D^2(\Xi_1, \Xi_2) &= \left\| \log\left(\Xi_1^{-\frac{1}{2}}\Xi_2\Xi_1^{-\frac{1}{2}} \right) \right\|^2 \\
&= \left\| \log\left(\Xi_1^{-1}\Xi_2 \right) \right\|^2 \\
&= \mathrm{tr}\left(\log^2\left(\Xi_1^{-1}\Xi_2 \right) \right) \\
&= \sum_{i=1}^{N_e} \log^2(\lambda_i),
\end{aligned} \tag{8}
$$

where $\|\cdot\|$ is the Frobenius norm and λ_i is ith eigenvalue of the matrix $\Xi_1^{-1}\Xi_2$.

B. RMS-Based Data Fusion

In actual environment, SUs may be affected by the environment noise and send aberrant data to the FC. To ensure the performance of spectrum sensing, an RMS-based data fusion approach is adopted to reduce the interference of aberrant points. It is well known that mean shift is a crucial clustering algorithm for obtaining representative centers of the disturbed data and widely used in many fields since it easy to implement while owning good convergence property. However, traditional mean shift algorithm works on vector space only. To achieve fusing directly on manifold, inspired by [30], the RMS algorithm is adopted.

Assume that the matrix set in the FC is expressed as

$$\Psi = \left\{ \hat{\mathbf{C}}_1, \ldots, \hat{\mathbf{C}}_M \right\}, \tag{9}$$

where M represents the size of data set. The kernel density estimator is defined as

$$\hat{f}(\Phi) = \frac{z_{k,h}}{M}\sum_{i=1}^{M} k\left(\frac{\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right)}{h^2} \right), \tag{10}$$

Algorithm 1: RMS Algorithm

Input: Random initial center point Φ_j. Let $j = 0$ and $\varepsilon > 0$. Initial the maximum number of iteration T_1.

Output: Φ_j

1   **while** $j < T_1$ **do**
2   Calculate the mean shift vector $\mathbf{m}_h(\Phi_j)$ by

$$\mathbf{m}_h(\Phi_j) = \frac{\sum_{i=1}^M g(\frac{\mathcal{R}_D^2(\Phi_j, \hat{\mathbf{C}}_i)}{h^2})\log_{\Phi_j}(\hat{\mathbf{C}}_i)}{\sum_{i=1}^M g(\frac{\mathcal{R}_D^2(\Phi_j, \hat{\mathbf{C}}_i)}{h^2})}. \tag{16}$$

     Update the center point as

$$\Phi_{j+1} = \exp(\log(\Phi_j) + \mathbf{m}_h(\Phi_j)). \tag{17}$$

     if $\|\mathbf{m}_h(\Phi_j)\| < \varepsilon$ **then**
3   break
4   **end**
5 **end**

---

where Φ is the center of the data, $k(\cdot)$ is a profile function satisfying $k(t) \geq 0$ for $t \geq 0$, h is bandwidth parameter, and $z_{k,h}$ is a constant which ensures that $\hat{f}(\Phi)$ integrates to one. The gradient of the $\hat{f}(\Phi)$ with respect to Φ is calculated as

$$
\begin{aligned}
\nabla\hat{f}(\Phi) &= \frac{1}{M}\sum_{i=1}^{M} \nabla k\left(\frac{\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right)}{h^2} \right) \\
&= -\frac{1}{M}\sum_{i=1}^{M} g\left(\frac{\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right)}{h^2} \right)\frac{\nabla\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right)}{h^2},
\end{aligned} \tag{11}
$$

where $g(\cdot) = -k'(\cdot)$ and $k'(\cdot)$ is the tangent at $k(\cdot)$. Inspired by [30], we know that the gradient of the RD satisfies

$$\nabla\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right) = -2\log_\Phi\left(\hat{\mathbf{C}}_i\right). \tag{12}$$

Therefore, according to (11) and (12), we can obtain

$$\nabla\hat{f}(\Phi) = \frac{2}{M}\sum_{i=1}^{M} g\left(\frac{\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right)}{h^2} \right)\frac{\log_\Phi\left(\hat{\mathbf{C}}_i\right)}{h^2}. \tag{13}$$

Therefore, the mean shift vector can be defined as

$$\mathbf{m}_h(\Phi) = \frac{\sum_{i=1}^M g\left(\frac{\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right)}{h^2} \right)\log_\Phi\left(\hat{\mathbf{C}}_i\right)}{\sum_{i=1}^M g\left(\frac{\mathcal{R}_D^2\left(\Phi, \hat{\mathbf{C}}_i\right)}{h^2} \right)}. \tag{14}$$

It is noticed that the mean shift vector always point toward the direction of maximum increase in the density. Thus, the center Φ_j is updated by

$$\Phi_{j+1} = \exp\left(\log\left(\Phi_j\right) + \mathbf{m}_h\left(\Phi_j\right)\right). \tag{15}$$

The complete algorithm of the RMS is given in Algorithm 1.

The developed RMS algorithm can directly fuse data points on manifold. After data fusion, the interference of the aberrant

points can be avoided. The convergence of the RMS algorithm is analyzed in Lemma 1.

Lemma 1: Suppose that the kernel K has a convex and monotonically decreasing profile, the sequence $\{\hat{f}(\Phi_j)\}_{j=1,2,\ldots}$, is convergent and monotonically non-decreasing if the bandwidth h is less than the injectivity radius.

Proof: Since the size of data set M is finite, the value of $\hat{f}(\Phi_j)$ is bounded. Hence, if $\hat{f}(\Phi_{j+1}) \geq \hat{f}(\Phi_j)$ is satisfies, then the sequence $\{\hat{f}(\Phi_j)\}_{j=1,2,\ldots}$, is convergent and monotonically non-decreasing. According to (10), we can get

$$\hat{f}(\Phi_{j+1}) - \hat{f}(\Phi_j)$$
$$= \frac{z_{k,h}}{M} \sum_{i=1}^{M} \left[k\left(\frac{\mathcal{R}_D^2(\Phi_{j+1}, \hat{\mathbf{C}}_i)}{h^2} \right) - k\left(\frac{\mathcal{R}_D^2(\Phi_j, \hat{\mathbf{C}}_i)}{h^2} \right) \right].$$

The convexity of the profile k implies that

$$k(x_2) \geq k(x_1) + k'(x_1)(x_2 - x_1) \tag{18}$$

for any $x_1, x_2 \in \mathbb{R}$. By using $g(\cdot) = -k'(\cdot)$, (18) becomes

$$k(x_2) - k(x_1) \geq g(x_1)(x_1 - x_2). \tag{19}$$

Based on (19), we have

$$\hat{f}(\Phi_{j+1}) - \hat{f}(\Phi_j)$$
$$= \frac{z_{k,h}}{Mh^2} \sum_{i=1}^{M} \left[g\left(\frac{\mathcal{R}_D^2(\Phi_j, \hat{\mathbf{C}}_i)}{h^2} \right) \left(\mathcal{R}_D^2(\Phi_j, \hat{\mathbf{C}}_i) \right. \right.$$
$$\left. \left. - \mathcal{R}_D^2(\Phi_{j+1}, \hat{\mathbf{C}}_i) \right) \right].$$

From [30], we know that if the bandwidth h is less than the injectivity radius, then we have

$$\frac{z_{k,h}}{Mh^2} \sum_{i=1}^{M} \left[g\left(\frac{\mathcal{R}_D^2(\Phi_j, \hat{\mathbf{C}}_i)}{h^2} \right) \left(\mathcal{R}_D^2(\Phi_j, \hat{\mathbf{C}}_i) \right. \right.$$
$$\left. \left. - \mathcal{R}_D^2(\Phi_{j+1}, \hat{\mathbf{C}}_i) \right) \right] \geq 0.$$

Therefore $\hat{f}(\Phi_{j+1}) \geq \hat{f}(\Phi_j)$ is satisfied. The proof is completed. ∎

C. Cooperative Spectrum Sensing Based on RDPSO Algorithm

In this part, a classifier is trained to judge whether SUs can access the channel. It is well known that particle swarm optimization (PSO) is a popular optimization algorithm, which finds the optimal solution through cooperation and information sharing among individuals in the group. Moreover, PSO algorithm can find the class center of the sample by minimizing the sum of the distances from each sample to the particle. Traditional PSO algorithm is used in vector space only. In order to achieve clustering on manifold, the RDPSO algorithm
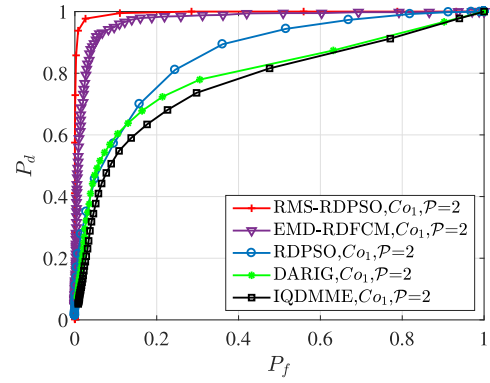

Fig. 3. The ROC curves of several approaches in Co_1.

is developed. Before moving on, a data set which includes the fused samples is built as

$$\bar{\Phi} = \{ \Phi_1, \ldots, \Phi_Q \}, \tag{20}$$

where $\Phi_q \in \mathbb{R}^{\mathcal{P} \times \mathcal{P}}$ denotes the qth sample in $\bar{\Phi}$ and Q denotes the size of fused data set. For the spectrum sensing problem, the RDPSO requires to find two centers on the manifold which represent PU is absent and PU is active, respectively.

Suppose that there $\mathcal{N}_p$ particles in Riemannian manifold. $\mathcal{X}_i$ and $\mathcal{V}_i$ represent the position and velocity of ith particle, respectively. In each iteration, the position and velocity of each particle are renovated by

$$\mathcal{V}_i(t+1) = \omega \mathcal{V}_i(t) + r_1 c_1 (Pbest_i(t) - \mathcal{X}_i(t))$$
$$+ r_2 c_2 (Gbest - \mathcal{X}_i(t)), \tag{21}$$
$$\mathcal{X}_i(t+1) = \mathcal{X}_i(t) + \mathcal{V}_i(t+1), \tag{22}$$

where t represents the iteration index, ω represents the inertia weight, $Pbest_i$ represents the individual best of ith particle, and $Gbest$ represents the global best. c_1 and c_2 are non-negative acceleration coefficients, r_1 and r_2 are positive constants and randomly selected within [0, 1].

In traditional PSO algorithm, it is inevitable to calculate the fitness value of the particle to evaluate the performance of the particle. In fact, the particles represent the class centers for spectrum sensing problem. Therefore, we use the sum of the geodesic distances between the particles and each sample as the fitness value. The details of the RDPSO algorithm is described in Algorithm 2.

After the training is completed, we can obtain a classifier as

$$\mathcal{F}(\hat{T}) = \frac{\mathcal{R}_D^2(\Upsilon_1, \hat{T})}{\mathcal{R}_D^2(\Upsilon_2, \hat{T})}, \tag{23}$$

where $\hat{T}$ represents the data on the manifold which required to be classified, Υ_1 and Υ_2 are the cluster centers. If $\mathcal{F}(\hat{T}) > \xi$, then the PU is presence and the SUs cannot use the channel. Otherwise, the SUs can access the spectrum. ξ is a positive constant, which used to determine P_d and P_f.

D. Complexity Analysis

The complexity of the RMS algorithm is given by $O(M \times T_1)$, where M is the number of data set, T_1 is the number of

---

Algorithm 2: RDPSO Algorithm

Input: Initial the individual best $Pbest_i$, the global best $Gbest$, the position and velocity of each particle $\mathcal{X}_i(t)$ and $\mathcal{V}_i(t)$, and the maximum number of iteration T_2.

Output: $Gbest$

1 Let $t = 0$;
2 **while** $t < T_2$ **do**
3   Obtain the fitness value of the $Gbest$ by
$$\mathcal{F}_G = \sum_{q=1}^{Q} \mathcal{R}_D^2(\Phi_q, Gbest).$$
4   **for** $i = 0 \rightarrow \mathcal{N}_p$ **do**
5   Obtain the fitness values of the ith particle and the $Pbest_i$ by
6 $$\mathcal{F}_i = \sum_{q=1}^{Q} \mathcal{R}_D^2(\Phi_q, \mathcal{X}_i(t)),$$
7 $$\mathcal{F}_{ip} = \sum_{q=1}^{Q} \mathcal{R}_D^2(\Phi_q, Pbest_i).$$
8   **if** $\mathcal{F}_i < \mathcal{F}_{ip}$ **then**
9            $Pbest_i \leftarrow \mathcal{X}_i(t)$
10   **if** $\mathcal{F}_{ip} < \mathcal{F}_G$ **then**
11              $Gbest \leftarrow Pbest_i$
12   **end**
13   **end**
14   **end**
15   **for** $i = 0 \rightarrow \mathcal{N}_p$ **do**
16   Update the position and velocity of each particle by (21) and (22)
17   **end**
18   **if** *Gbest is convergent* **then**
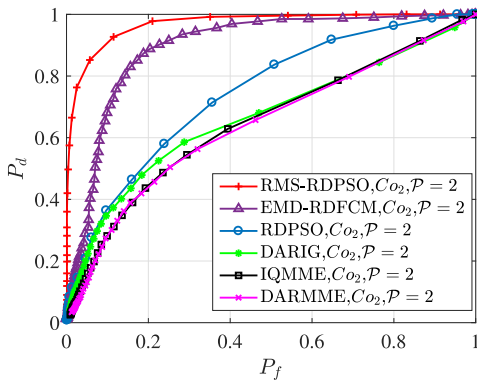19   break
20   **end**
21 **end**

Fig. 4. The ROC curves of several approaches in Co_2.

iteration. It is noticed that the complexity of the RMS algorithm can be rewritten as $O(M)$ as the number of data set increase.

The complexity of the RDPSO algorithm is calculated as $O(T_2 \times (Q + \mathcal{N}_p \times 2Q + \mathcal{N}_p))$, where T_2 is the number of
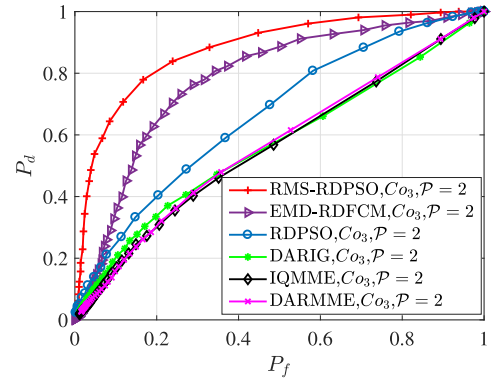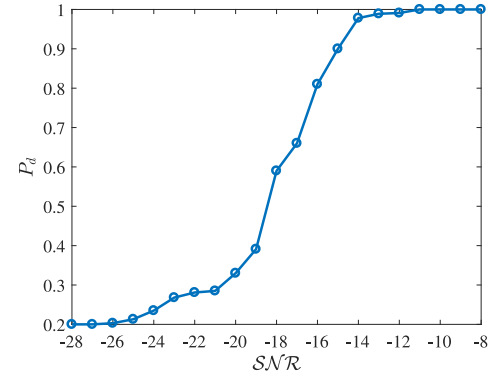

Fig. 5. The ROC curves of several approaches in Co_3.

Fig. 6. The ROC curve of the RMS-RDPSO approach with different $\mathcal{SNR}$.

iteration, $\mathcal{N}_p$ is the number of particle, and Q is the number of fused date set. Therefore, with the Q increase, the complexity of the RDPSO algorithm becomes $O(Q)$.

Remark 2: Different from the existing approaches [28] and [31] which used unsupervised clustering algorithms, such as K-mean and fuzzy-c mean algorithms for CSS, this paper develops an RDPSO algorithm. It is noticed that the unsupervised clustering algorithms are sensitive to the choice of initial point, which require several attempts to achieve good performance. However, the RDPSO algorithm can obtain excellent performance by randomly initializing the velocity and the position of the particles. Moreover, this paper develops an RMS algorithm to eliminate the aberrant data, which further improves the sensing performance.

Remark 3: It is noticed that the distribution of the test-statistic and the threshold are not required in this approach. In this paper, we adopted sensing data and the developed RDPSO algorithm to obtain a classifier to determine whether the PU exists. Compared with traditional approaches which require to derive a precise threshold, this approach is more adaptive. This is because a precise threshold is hard to obtain on complex environment.

Remark 4: Compared with the single-input and single-output based scheme, the advantages of the multiple-input and multiple-output based scheme lies in that: 1) multiple-input and multiple-output based scheme can obtain higher data transmission rate and more reliable link transmission through multiplexing gain and diversity gain without increasing the
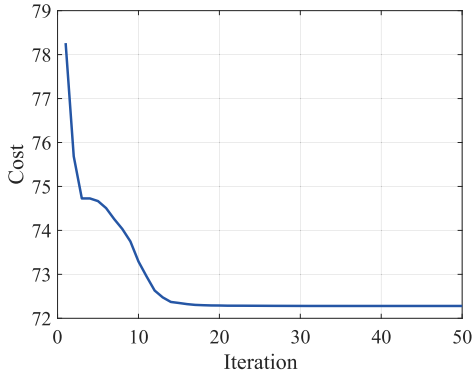
Fig. 7. The cost value of the RMS-RDPSO approaches in Co_1.
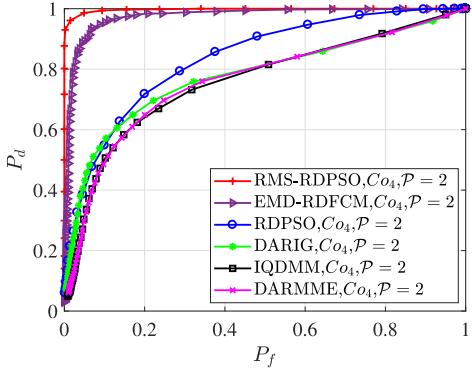
Fig. 8. The ROC curves of several approaches in Co_4.

Fig. 9. The ROC curves of several approaches in Co_5.

Fig. 10. The ROC curves of several approaches in Co_6.
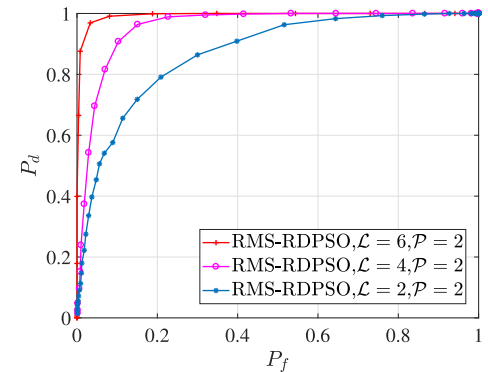
Fig. 11. The ROC curves of the RMS-RDPSO scheme under different number of SUs.

Fig. 12. The ROC curves of the RMS-RDPSO scheme under different number of antennas.

system bandwidth and transmission power. 2) multiple-input and multiple-output technology allows SUs to obtain useful data and reduces the effects of multipath fading, shading and receiver uncertainty.

IV. SIMULATION ANALYSIS

A. SUs in Same Condition

In this section, the validity of the RMS-RDPSO based CSS approach is analyzed on the conditions of Co_1, Co_2 and Co_3, where Co_1, Co_2 and Co_3 represent $\mathcal{SNR} = [-12\text{dB}, -12\text{dB}]$, $\mathcal{SNR} = [-14\text{dB}, -14\text{dB}]$ and $\mathcal{SNR} = [-16\text{dB}, -16\text{dB}]$, respectively. The $\mathcal{SNR}$ is the signal to noise ratio of the antennas of the SU. Let the number of the SUs that participate the data fusion be $M = 10$, the number of antennas

be $\mathcal{P} = 2$, and the number of sample points be $N = 1000$, respectively. By applying the RMS algorithm, the data set $\bar{\Phi} = \{\Phi_1, \ldots, \Phi_{1000}\}$ which contains 1000 fused samples is prepared. After RDPSO optimization, the class centers are given as

$$\Upsilon_1 = \begin{bmatrix} 11.2338, 0.7418 \\ 0.7418, 11.3071 \end{bmatrix}, \quad \Upsilon_2 = \begin{bmatrix} 10.3258, -0.2695 \\ -0.2695, 10.3258 \end{bmatrix}.$$

Under the conditions of Co_1, Co_2 and Co_3, the experiment results of the RMS-RDPSO based CSS scheme and other previous methods, i.e., EMD-RDFCM, DARDMM, IQDMM, and IQRMET [24], [27], [28], [31], are shown in Fig. 3–5, respectively. The simulation results expound that the RDPSO

TABLE I
THE P_d OF DIFFERENT APPROACHES WHEN SUs IN SAME CONDITION

	RMS-RDPSO	EMD-RDFCM	RDPSO	DARDMM	DARIG	IQDMM
$Co_1, P_f = 0.1$	0.995	0.952	0.638	0.625	0.590	0.585
$Co_1, P_f = 0.2$	0.998	0.981	0.781	0.720	0.706	0.695
$Co_2, P_f = 0.1$	0.852	0.706	0.374	0.346	0.281	0.280
$Co_2, P_f = 0.2$	0.978	0.888	0.512	0.478	0.436	0.435
$Co_3, P_f = 0.1$	0.707	0.363	0.313	0.210	0.170	0.168
$Co_3, P_f = 0.2$	0.801	0.650	0.404	0.334	0.308	0.305

TABLE II
THE P_d OF DIFFERENT APPROACHES WHEN SUs IN DIFFERENT CONDITION

	RMS-RDPSO	EMD-RDFCM	RDPSO	DARDMM	DARIG	IQDMM
$Co_4, P_f = 0.1$	0.994	0.955	0.539	0.475	0.539	0.472
$Co_4, P_f = 0.2$	0.997	0.984	0.718	0.648	0.695	0.645
$Co_5, P_f = 0.1$	0.886	0.671	0.407	0.231	0.269	0.268
$Co_5, P_f = 0.2$	0.951	0.830	0.567	0.403	0.401	0.425
$Co_6, P_f = 0.1$	0.576	0.328	0.215	0.139	0.138	0.136
$Co_6, P_f = 0.2$	0.781	0.612	0.398	0.305	0.309	0.301

method obtain superior performance in contrast to conventional approaches. It is worth noticed that the RMS-RDPSO achieve better result than RDPSO, which means that the RMS-based data fusion method is valid. The detailed data are given in Table I. In the case of $Co_1, P_f = 0.1$, compared with previous approaches, the P_d of the RMS-RDPSO scheme is improved by 4.51%, 55.95%, 59.20%, 68.64%, and 70.08%, respectively. Fig. 6 shows the receiver operating characteristic (ROC) curve of the RMS-RDPSO approach with different $\mathcal{SNR}$. We can observe that when $P_f = 0.2$, the P_d will rise with the $\mathcal{SNR}$ increase. The convergence of the cost value for the RDPSO algorithm under Co_1 is displayed in Fig. 7. We can know that the developed RDPSO algorithm is converged.

Remark 5: It is worth mentioning that the existing algorithms DARDMM, DARIG, and IQDDM are worked on Euclidean space, which use Euclidean distance to measure the distance between two points. However, the developed RMS-RDPSO approach can be applied on manifold space, which is an extension of the previous methods. Furthermore, the RMS algorithm is developed to eliminate the aberrant data, which further improves the sensing performance.

Remark 6: In this paper, the cost value of the RDPSO algorithm is the sum of the distances between the particle and all the sample points. Therefore, as the number of iteration increase, the particle is closer to the class center, so the cost value will decrease.

B. SUs in Different Conditions

In this part, we will verify the validity of the RMS-RDPSO scheme on the conditions of Co_4, Co_5 and Co_6, where Co_4, Co_5 and Co_6 represent $\mathcal{SNR} = [-12\text{dB}, -12.5\text{dB}]$, $\mathcal{SNR} = [-14\text{dB}, -14.5\text{dB}]$ and $\mathcal{SNR} = [-16\text{dB}, -16.5\text{dB}]$, respectively. The experimental results are given in Fig. 8–10, where we can conclude that compared with previous approaches, the

TABLE III
THE P_d OF THE RMS-RDPSO SCHEME IN DIFFERENT NUMBER OF SU

	$\mathcal{L} = 2$	$\mathcal{L} = 4$	$\mathcal{L} = 6$
$Co_6, P_f = 0.1$	0.991	0.908	0.656
$Co_6, P_f = 0.2$	0.999	0.964	0.791

TABLE IV
THE P_d OF THE RMS-RDPSO SCHEME IN DIFFERENT NUMBER OF ANTENNAS

	$\mathcal{P} = 2$	$\mathcal{P} = 4$	$\mathcal{P} = 6$
$Co_6, P_f = 0.1$	0.639	0.931	0.998
$Co_6, P_f = 0.2$	0.801	0.975	0.999

RMS-RDPSO scheme can acquire the best sensing results. The P_d of serval approaches are displayed in Table II. In the case of $Co_4, P_f = 0.1$, we can calculate that the P_d of the RMS-RDPSO scheme is enhanced by 4.08%, 84.41%, 109.26%, 84.41%, and 110.59%, respectively. According to the results of Sections IV-A and IV-B, we can find that the RMS-RDPSO approach acquires the best detection performance than conventional schemes under two different scenario.

C. Different Number of SUs

In this section, we verify the performance of the developed RMS-RDPSO scheme when the number of the SU is different. The ROC curves and P_d are shown in Fig. 11 and Table III, respectively. According to the experimental data, we can find that the sensing performance is better with the increase of SU.

D. Different Number of Antennas

In this section, we analyze the influence of the number of antennas on the performance of the RMS-RDPSO scheme. The simulation results are shown in Fig. 12 and Table IV. We can

find that with the number of antennas increase, the P_d of the RMS-RDPSO scheme is raised.

V. Conclusion

In this article, an RMS-RDPSO scheme is proposed to deal with the spectrum sensing problem in CRN. To improve the sensing performance, a data fusion scheme based on RMS algorithm is adopted to eliminate the aberrant data in FC. Then, an RDPSO clustering algorithm is developed to cluster samples on manifold space. After the classifier is trained, we adopt it to judge whether the PU exists or not. Finally, the detection performance of the RMS-RDPSO scheme is verified under two different conditions. The main contributions of this paper are on adopting an RMS-based data fusion scheme to eliminate the aberrant data and developing an RDPSO clustering algorithm to train a classifier on manifold. It is worth mentioning that the RMS-RDPSO scheme approach can use on manifold directly and acquires the better sensing performance than previous approaches. In our future work, since the scenario of this paper is ideal, we will consider more complex scenarios, such as the primary network traffic model, the energy consumption, the spatial diversity of the SU is not enough, and the SUs are mobile.

References

[1] J. Ren, Y. Zhang, Q. Ye, K. Yang, K. Zhang, and X. S. Shen, "Exploiting secure and energy-efficient collaborative spectrum sensing for cognitive radio sensor networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6813–6827, Oct. 2016.

[2] N. Tadayon and S. Aissa, "A multichannel spectrum sensing fusion mechanism for cognitive radio networks: Design and application to IEEE 802.22 WRANs," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 4, pp. 359–371, Dec. 2015.

[3] A. Zaeemzadeh, M. Joneidi, N. Rahnavard, and G.-J. Qi, "Co-SpOT: Cooperative spectrum opportunity detection using Bayesian clustering in spectrum-heterogeneous cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 2, pp. 206–219, Jun. 2018.

[4] J. Mitola and G. Q. Maguire, "Cognitive radio: Making software radios more personal," *IEEE Pers. Commun.*, vol. 6, no. 4, pp. 13–18, Aug. 1999.

[5] P. Qi, Y. Du, D. Wang, and Z. Li, "Wideband spectrum sensing based on bidirectional decision of normalized spectrum for cognitive radio networks," *IEEE Access*, vol. 7, pp. 140833–140845, 2019.

[6] G. Ganesan and Y. Li, "Cooperative spectrum sensing in cognitive radio, part I: Two user networks," *IEEE Trans. Wireless Commun.*, vol. 6, no. 6, pp. 2204–2213, Jun. 2007.

[7] A. Mariani, S. Kandeepan, and A. Giorgetti, "Periodic spectrum sensing with non-continuous primary user transmissions," *IEEE Trans. Wireless Commun.*, vol. 14, no. 3, pp. 1636–1649, Mar. 2015.

[8] S. Atapattu, C. Tellambura, and H. Jiang, "Energy detection based cooperative spectrum sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 4, pp. 1232–1241, Apr. 2011.

[9] J.-Y. Wu, C.-H. Wang, and T.-Y. Wang, "Performance analysis of energy detection based spectrum sensing with unknown primary signal arrival time," *IEEE Trans. Commun.*, vol. 59, no. 7, pp. 1779–1784, Jul. 2011.

[10] F. Salahdine, H. E. Ghazi, N. Kaabouch, and W. F. Fihri, "Matched filter detection with dynamic threshold for cognitive radio networks," in *Proc. Int. Conf. Wireless Netw. Mobile Commun. (WINCOM)*, 2015, pp. 1–6.

[11] X. Zhang, R. Chai, and F. Gao, "Matched filter based spectrum sensing and power level detection for cognitive radio network," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Atlanta, GA, USA, 2014, pp. 1267–1270.

[12] P. Sepidband and K. Entesari, "A CMOS spectrum sensor based on quasi-cyclostationary feature detection for cognitive radios," *IEEE Trans. Microw. Theory Techn.*, vol. 63, no. 12, pp. 4098–4109, Dec. 2015.

[13] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Sequential detection of cyclostationary signal for cognitive radio systems," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4480–4485, Sep. 2009.

[14] D. Morales-Jimenez, R. H. Y. Louie, M. R. McKay, and Y. Chen, "Analysis and design of multiple-antenna cognitive radios with multiple primary user signals," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4925–4939, Sep. 2015.

[15] L. Shen, H. Wang, W. Zhang, and Z. Zhao, "Multiple antennas assisted blind spectrum sensing in cognitive radio channels," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 92–94, Jan. 2012.

[16] A. Taherpour, M. Nasiri-Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814–823, Feb. 2010.

[17] A. Singh, M. R. Bhatnagar, and R. K. Mallik, "Cooperative spectrum sensing in multiple antenna based cognitive radio network using an improved energy detector," *IEEE Commun. Lett.*, vol. 16, no. 1, pp. 64–67, Jan. 2012.

[18] C. Liu, H. Li, J. Wang, and M. Jin, "Optimal eigenvalue weighting detection for multi-antenna cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2083–2096, Apr. 2017.

[19] S. Dwivedi, A. Kota, and A. K. Jagannatham, "Optimal bartlett detector based SPRT for spectrum sensing in multi-antenna cognitive radio systems," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1409–1413, Sep. 2015.

[20] A. Assra, J. Yang, and B. Champagne, "An EM approach for cooperative spectrum sensing in multiantenna CR networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 3, pp. 1229–1243, Mar. 2016.

[21] M. Miah, M.Schukat, and E. Barrett, "Sensing and throughput analysis in MU-MIMO based cognitive radio Internet of Things with energy constraints," *Comput. Commun.*, vol. 154, pp. 442–454, Mar. 2020.

[22] K. M. Thilina, K. W. Choi, N. Saquib, and E. Hossain, "Machine learning techniques for cooperative spectrum sensing in cognitive radio networks," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2209–2221, Nov. 2013.

[23] Z. Li, W. Wu, X. Liu, and P. Qi, "Improved cooperative spectrum sensing model based on machine learning for cognitive radio networks," *IET Commun.*, vol. 12, no. 19, pp. 2485–2492, 2018.

[24] S. Zhang, Y. Wang, H. Yuan, P. Wan, and Y. Zhang, "Multiple-antenna cooperative spectrum sensing based on the wavelet transform and Gaussian mixture model," *Sensors*, vol. 19, no. 18, p. 3863, 2019.

[25] Q. Chen, P. Wan, and Y. Wang, "Research on cognitive radio spectrum sensing method based on information geometry," in *Proc. Int. Conf. Cloud Comput. Security*, 2017, pp. 554–564.

[26] Q. Lu, S. Yang, and F. Liu, "Wideband spectrum sensing based on riemannian distance for cognitive radio networks," *Sensors*, vol. 17, no. 4, pp. 661–679, Mar. 2017.

[27] Y. Wang, S. Zhang, Y. Zhang, P. Wan, J. Li, and N. Li, "A cooperative spectrum sensing method based on empirical mode decomposition and information geometry in complex electromagnetic environment," *Complexity*, vol. 2019, pp. 1–13, Feb. 2019, doi: 10.1155/2019/5470974.

[28] S. Zhang, Y. Wang, Y. Zhang, P. Wan, and J. Zhuang, "A novel clustering algorithm based on information geometry for cooperative spectrum sensing," *IEEE Syst. J.*, vol. 15, no. 2, pp. 3121–3130, Jun. 2021.

[29] V. Kumar, D. C. Kandpal, M. Jain, R. Gangopadhyay, and S. Debnath, "K-mean clustering based cooperative spectrum sensing in generalized $\kappa - \mu$ fading channels," in *Proc. 22th Nat. Conf. Commun.*, 2016, pp. 1–5.

[30] R. Subbarao and P. Meer, "Nonlinear mean shift over Riemannian manifolds," *Int. J. Comput. Vis.*, vol. 84, p. 1, Mar. 2009.

[31] Y. Zhang, C. Ma, Y. Wang, S. Zhang, and P. Wan, "Information geometry-based fuzzy-c means algorithm for cooperative spectrum sensing," *IEEE Access*, vol. 8, pp. 155742–155752, 2020.

Yongwei Zhang received the B.S. degree from the School of Electronic and Information Engineering, Jiaying University, Meizhou, China, in 2016, and the Ph.D. degree from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2021. His current research interests include cognitive radio, spectrum sensing, adaptive dynamic programming, and optimal control.

Shunchao Zhang received the B.S. degree from the School of Electrical and Information Engineering, Hunan Institute of Engineering, Xiangtan, China, in 2016, the M.S. degree from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2019, where he is currently pursuing the Ph.D. degree. His current research interests include cooperative spectrum sensing, clustering algorithms, and adaptive dynamic programming.

Jiawei Zhuang received the B.S. degree from the School of Electronic and Information Engineering, Jiaying University, Meizhou, China, in 2017, the M.S. degree from the School of Automation, Guangdong University of Technology, Guangzhou, China, in 2020, where he is currently pursuing the Ph.D. degree. His current research interests include cognitive radio and spectrum sensing.

Yonghua Wang (Member, IEEE) received the B.S. degree in electrical engineering and automation from the Hebei University of Technology, Tianjin, China, in 2001, the M.S. degree in control theory and control engineering from the Guangdong University of Technology, Guangzhou, China, in 2006, and the Ph.D. degree in communication and information system from Sun Yat-sen University, Guangzhou, China, in 2009. He is currently with the School of Automation, Guangdong University of Technology.

Pin Wan received the B.S. degree in electronic engineering and the M.S. degree in circuit and system from Southeast University in 1984 and 1990, respectively, and the Ph.D. degree in control theory and control engineering from the Guangdong University of Technology in 2011, where he is currently a Professor with the School of Automation.
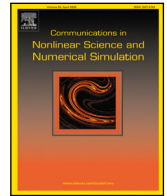
Research paper

Dynamic event-triggered neuro-optimal control for uncertain nonlinear systems with unknown dead-zone constraint ☆

Shunchao Zhang [a], Jiawei Zhuang [a], Yongwei Zhang [b],*

[a] *School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou 510521, China*
[b] *College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China*

ARTICLE INFO

ABSTRACT

In this article, we propose a dynamic event-triggered neuro-optimal control scheme (DETNOC) for uncertain nonlinear systems subject to unknown dead-zone and disturbances through the design of a composite control law. An integral sliding mode-based discontinuous control law is utilized to compensate for the effects of unknown dead-zone, disturbance, and a component of uncertainties. As a result, a system dynamics that evolves free of these effects during the sliding mode is obtained. Then, an adaptive dynamic programming-based dynamic event-triggered optimal control law is designed to stabilize the sliding mode dynamics with the help of critic-only neural network architecture. Finally, stability analysis of the closed-loop system is provided and the effectiveness of the developed DETNOC scheme is verified.

1. Introduction

Due to the inherent properties of actuators, dead-zone is one of the prevalent constraints in practical systems. It widely exists in electrical servomotors, hydraulic actuator, power generators, etc. The existence of dead-zone can severely degrade the control performance, cause the system instability, and even lead to safety accidents if not appropriately compensated or tackled. Owing to its distinguished features, such as a simple structure, ease of implementation, inherent robustness, and fast dynamic response, sliding mode control (SMC) has been extensively studied and applied to address dead-zone constraints [1,2]. However, it is worth mentioning that the stability of the closed-loop system is a fundamental requirement in controller design. Additionally, the designed controller is expected to achieve a significant level of optimality.

Optimal control is an effective methodology to optimize a pre-defined performance index function [3,4]. To achieve the above control objective, many methods design a composite control law consists of a discontinuous control law and an optimal control law by integrating SMC and optimal control approaches [5]. For uncertain linear systems, Das and Mahanta [6] proposed an optimal second-order SMC method. In this method, the discontinuous control law was designed by combining the terminal sliding mode and integral sliding mode surfaces (SMSs). Additionally, the optimal control law was obtained by solving the algebraic Riccati equation (ARE). In [7], an optimal integral SMC (ISMC) method was proposed, in which the discontinuous and optimal control laws were designed by using ISMC and solving the ARE, respectively. In the aforementioned methods, the discontinuous control law is employed to mitigate the impact of uncertainties and disturbances, resulting in the attainment of linear sliding mode dynamics. On the other hand, the optimal control law is utilized to stabilize the linear sliding mode dynamics (SMD) while achieving a significant

level of optimality. In the case of nonlinear systems, the resulting sliding mode dynamics are nonlinear. To obtain the optimal control law, the Hamilton–Jacobi-Bellman (HJB) equation needs to be solved. However, the HJB equation is a partial differential equation that is often impossible to solve analytically due to its complexity [8–10]. To address this challenge, the adaptive dynamic programming (ADP) technique, inspired by reinforcement learning (RL), has been developed. ADP leverages approximators, such as neural networks (NNs) [11] or fuzzy logic systems [12] to compute forward-in-time and finds an approximate solution to the HJB equation.

Recently, numerous RL or ADP-based methods have emerged to solve the HJB equation and derive the optimal control law [13–19]. It should be mentioned that these approaches typically employ time-triggered control (TTC) strategies, which can result in unnecessary computational burdens and inefficient use of communication resources and bandwidth [20–22]. Fortunately, there has been a growing interest in the development of event-triggered control (ETC) methods, as they offer the capability to address these challenges. Numerous promising results have been reported in this area [23–25]. Nevertheless, the event-triggering condition will become conservative as the sampling error decreases, leading to several unnecessary triggering [26–30]. To overcome this deficiency, Mu et al. [26] developed an ADP-based dynamic event-triggered optimal control (DETOC) method for nonlinear systems, where an internal dynamic variable was introduced to establish a dynamic event-triggering condition for determining the occurrence of events. Zou et al. [27] developed a dynamic event-triggered optimal tracking control (DETOTC) method by using ADP. This method utilized an auxiliary dynamic variable to construct a dynamic triggering rule. Yang et al. [28] developed an RL-based decentralized DETOC method for interconnected nonlinear systems, where the triggering rule relied on the system states and the variables generated by time-based differential equations. Tan et al. [29] proposed an ADP-based DETOTC method for uncertain nonlinear systems. This method designed a composite control law, in which a discontinuous law was used to compensate uncertain terms and acquire the nonlinear SMD, and the ADP-based DETOC law was designed to stabilize the nonlinear SMD.

Although there have been several reports on ADP-based DETOC methods for addressing optimal control problems, it is important to note that research in this area is still in its early stages. The practical and theoretical significance motivates us to develop the ADP-based dynamic event-triggered neuro-optimal control (DETNOC) scheme for uncertain nonlinear systems with unknown dead-zone and disturbances. The novelties and contributions of this work are outlined below.

1. Unlike the existing methods [26–29,31] that addressed optimal control problem without input dead-zone, the ISMC and ADP are combined to investigate optimal control problems for uncertain nonlinear systems with unknown dead-zone and disturbances.
2. Different from existing method [32,33] which constructed auxiliary systems to address the control problems for nonlinear systems with mismatched uncertainties, this paper designs a modified value function to address the effect of the mismatched component of uncertainties.
3. Compared to existing event-triggered mechanisms [20,23,24], this scheme adopts a dynamic event-triggered mechanism by introducing a dynamic variable, which effectively reduces the computational burden and saves communication resources.

The remaining sections of this paper are structured as follows. Section 2 presents the problem statement. In Section 3, we develop an ADP-based DETNOC method and provide a stability analysis. Section 4 includes two simulation examples to illustrate the effectiveness of the proposed ADP-based DETNOC method. Finally, Section 5 provides a brief conclusion.

2. Problem statement

Consider a class of nonlinear systems

$$\dot{\mathcal{Z}} = \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})\big(\mathfrak{D}(u) + \varrho\big) + \mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z}), \tag{1}$$

where $\mathcal{Z} \in \mathbb{R}^n$ is the system state with $\mathcal{Z}_0 = \mathcal{Z}(0)$, $u = [u_1, \dots, u_i, \dots, u_m]^\mathsf{T} \in \mathbb{R}^m$ is the control input, $\mathcal{I}_a(\mathcal{Z}) \in \mathbb{R}^n$, $\mathcal{I}_b(\mathcal{Z}) \in \mathbb{R}^{n \times m}$ and $\mathcal{K}(\mathcal{Z}) \in \mathbb{R}^{n \times q}$ are continuously differentiable functions, $\varrho \in \mathbb{R}^m$ is the disturbance, $\mathcal{F}(\mathcal{Z}) \in \mathbb{R}^q$ is the unknown perturbation, and $\mathfrak{D}(u) = [\mathfrak{D}_1(u_1), \dots, \mathfrak{D}_i(u_i), \dots, \mathfrak{D}_m(u_m)]^\mathsf{T} \in \mathbb{R}^m$ is the output of the dead-zone with

$$\mathfrak{D}_i(u_i) = \begin{cases} m_r(u_i - o_r) & \text{if} \quad u_i \geq o_r, \\ 0 & \text{if} \quad o_l < u_i < o_r, \\ m_l(u_i - o_l) & \text{if} \quad u_i \leq o_l, \end{cases} \tag{2}$$

where $m_r, m_l \in \mathbb{R}$ are the slops of dead-zone and $m_r = m_l = m$, and $o_r, o_l \in \mathbb{R}$ denote the bounded unknown parameters of dead-zone. Furthermore, the model (2) can be described as

$$\mathfrak{D}_i(u_i) = mu_i + d_i(u_i) \tag{3}$$

where $d_i(u_i)$ is written as

$$d_i\big(u_i(t)\big) = \begin{cases} -mo_r & \text{if} \quad u_i \geq o_r, \\ -mu_i & \text{if} \quad o_l < u_i < o_r, \\ -mo_l & \text{if} \quad u_i \leq o_l. \end{cases} \tag{4}$$

Let $d(u) = [d_1(u_1), \dots, d_i(u_i), \dots, d_m(u_m)]^\mathsf{T}$. Then, the system (1) can be formulated as

$$\dot{\mathcal{Z}} = \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})\big(mu + d(u) + \varrho\big) + \mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z}). \tag{5}$$

For the uncertain term $\mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z})$, it can be composed of two components

$$\mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z}) = \overbrace{\mathcal{I}_b(\mathcal{Z})\mathcal{I}_b^+(\mathcal{Z})\mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z})}^{\text{matched component}} + \overbrace{(I_n - \mathcal{I}_b(\mathcal{Z})\mathcal{I}_b^+(\mathcal{Z}))\mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z})}^{\text{mismatched component}},\tag{6}$$

Hence, the system (1) can further be rewritten as

$$\dot{\mathcal{Z}} = \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})(mu + \rho) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z}),\tag{7}$$

where $\rho = d(u) + \varrho + \mathcal{I}_b^+(\mathcal{Z})\mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z})$ and $h(\mathcal{Z}) = (I_n - \mathcal{I}_b(\mathcal{Z})\mathcal{I}_b^+(\mathcal{Z}))\mathcal{K}(\mathcal{Z})$.

3. Controller design based on ISMC and ADP

In this paper, to guarantee the stability of system (1) with considerable optimal performance, a composite control input is designed as

$$u = u_d + \hat{\theta}u_c,\tag{8}$$

where u_d is the discontinuous control input employed to eliminate the influences of dead-zone, disturbance, and the matched component of uncertainties via ISMC, u_c is the DETOC input designed to stabilize the SMD by using ADP-based dynamic ETC (DETC) method, and $\hat{\theta}$ is the adaptive term to approximate the reciprocal of the unknown slope parameter m.

Remark 1. Indeed, it is well known that ISMC is an effective technique to deal with the dead-zone, disturbance, and uncertainties effects [1,2,34,35]. Furthermore, the stability of the closed-loop system is only the foundation, and the control performance should also be considered in the process of controller design. Recently, ADP has emerged as a powerful method widely employed to solve optimal control problems. Hence, this paper combines ISMC and ADP to develop a DETNOC method and design a composite control law for nonlinear systems with the unknown dead-zone, disturbance, and uncertainties.

Remark 2. Recently, many ADP-based ETC methods have been investigated to solve optimal control, robust control, and fault-tolerant control problems [3]. It is worth noting that these control schemes are implemented under the static event-triggering strategy relating on the current values of system state and sampling error [26]. Different from the ADP-based static event-triggering strategy, the ADP-based dynamic event-triggered strategy designs an internal signal in the basis of static event-triggering rule. As a result, larger sampling intervals can be generated by the ADP-based dynamic event-triggering rule compared to the static one, which can further reduce the wastage of computation and communication resources.

3.1. Discontinuous control design

On the basis of the SMC technique, an integral type sliding mode function is formulated as

$$\mathfrak{S}(\mathcal{Z},t) = \mathcal{G}(\mathcal{Z}) - \mathcal{G}(\mathcal{Z}_0) - \int_0^t \mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c(\mathcal{Z})\big)d\tau,\tag{9}$$

where $\mathcal{G}(\mathcal{Z}) \in \mathbb{R}^m$ is a design function and $\mathfrak{M}(\mathcal{Z}) = \partial\mathcal{G}(\mathcal{Z})/\partial\mathcal{Z} \in \mathbb{R}^{m \times n}$. The time derivative of $\mathfrak{S}(\mathcal{Z},t)$ is derived by

$$\begin{aligned}\dot{\mathfrak{S}}(\mathcal{Z},t) &= \mathfrak{M}(\mathcal{Z})\dot{\mathcal{Z}} - \mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c\big)\\ &= \mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})(mu + \rho) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\big) - \mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c\big)\\ &= \mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_b(\mathcal{Z})(mu + \rho - u_c) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\big)\\ &= \mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_b(\mathcal{Z})(mu_d + m\hat{\theta}u_c + \rho - u_c) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\big).\end{aligned}\tag{10}$$

To maintain the system trajectory on the SMS, the discontinuous control law is designed as

$$u_d = -K\text{sgn}\big(\varXi(\mathcal{Z})\big),\tag{11}$$

where $\varXi(\mathcal{Z}) = \mathcal{I}_b^{\mathsf{T}}(\mathcal{Z})\mathfrak{M}^{\mathsf{T}}(\mathcal{Z})\mathfrak{S}$, K denotes the sliding mode gain, and

$$\text{sgn}\big(\varXi(\mathcal{Z})\big) = \big[\text{sgn}_1(\varXi_1(\mathcal{Z})), \ldots, \text{sgn}_m(\varXi_m(\mathcal{Z}))\big]^{\mathsf{T}},$$

where $\text{sgn}_i(\cdot)$ is the sign function.

Remark 3. The control law $u_d = -K\text{sgn}(\varXi(\mathcal{Z}))$ aims to maintain the system trajectory on the SMS and achieve the SMD without the effects from dead-zone, disturbance, and the matched component of uncertainties, thereby enhancing the robustness of the nonlinear systems. Specifically, the function $\text{sgn}(\cdot)$ is employed to resist parameter changes and perturbations, the sliding mode gain K is selected in accordance with the magnitude of these changes and perturbations. Moreover, the form of (11) is straightforward to implement and is widely used in the ISMC control field [29,34,35].

Theorem 1. *Considering the nonlinear system* (1), *the sliding mode function* (9), *Assumptions* 1 *and* 2, *the adaptive term is turned by* $\dot{\hat{\theta}} = -\frac{1}{\gamma}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c$, *the discontinuous control law can maintain the system trajectory on the SMS, that is,* $\mathfrak{S} = 0$.

Proof. Select a Lyapunov function candidate as

$$L_1 = \frac{1}{2m}\mathfrak{S}^{\mathsf{T}}\mathfrak{S} + \gamma\tilde{\theta}^2. \tag{12}$$

where $\tilde{\theta} = \hat{\theta} - \theta$ is the approximation error. Taking the time derivation of L_1, we have

$$
\begin{aligned}
\dot{L}_1 &= \frac{1}{m}\mathfrak{S}^{\mathsf{T}}\left\{\mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})(mu+\rho) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\big) - \mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c\big)\right\} + \gamma\tilde{\theta}\dot{\hat{\theta}} \\
&= \frac{1}{m}\mathfrak{S}^{\mathsf{T}}\left\{\mathfrak{M}(\mathcal{Z})\big(\mathcal{I}_b(\mathcal{Z})(mu+\rho-u_c) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\big)\right\} + \gamma\tilde{\theta}\dot{\hat{\theta}} \\
&= \mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u + \frac{1}{m}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\rho + \frac{1}{m}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) - \frac{1}{m}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c + \gamma\tilde{\theta}\dot{\hat{\theta}},
\end{aligned} \tag{13}
$$

Let $\theta = \frac{1}{m}$, we have

$$
\begin{aligned}
\dot{L}_1 &= \mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\rho + \gamma\tilde{\theta}\dot{\hat{\theta}} + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) - \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c \\
&= -K\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\mathrm{sgn}\big(\mathcal{I}_b^{\mathsf{T}}(\mathcal{Z})\mathfrak{M}^{\mathsf{T}}(\mathcal{Z})\mathfrak{S}\big) + \hat{\theta}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\rho + \gamma\tilde{\theta}\dot{\hat{\theta}} \\
&\quad + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) - \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c \\
&= -K\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\mathrm{sgn}\big(\mathcal{I}_b^{\mathsf{T}}(\mathcal{Z})\mathfrak{M}^{\mathsf{T}}(\mathcal{Z})\mathfrak{S}\big) + \tilde{\theta}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\rho + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) + \gamma\tilde{\theta}\dot{\hat{\theta}},
\end{aligned} \tag{14}
$$

Then, recalling the turned rule $\dot{\hat{\theta}} = -\frac{1}{\gamma}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c$, $\dot{L}_1$ becomes

$$
\begin{aligned}
\dot{L}_1 &= -K\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\mathrm{sgn}\big(\mathcal{I}_b^{\mathsf{T}}(\mathcal{Z})\mathfrak{M}^{\mathsf{T}}(\mathcal{Z})\mathfrak{S}\big) + \tilde{\theta}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\rho + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) - \tilde{\theta}\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c \\
&= -K\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\mathrm{sgn}\big(\mathcal{I}_b^{\mathsf{T}}(\mathcal{Z})\mathfrak{M}^{\mathsf{T}}(\mathcal{Z})\mathfrak{S}\big) + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\rho + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) \\
&\leq -K\|\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\|_1 + \theta\|\rho\|\|\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\| + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) \\
&\leq -(K - \theta\|\rho\|)\|\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\| + \theta\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}).
\end{aligned} \tag{15}
$$

Inspired by [31,34], $\mathfrak{M}(\mathcal{Z})$ is chosen as $\mathcal{I}_b^+(\mathcal{Z})$, (15) becomes

$$
\begin{aligned}
\dot{L}_1 &\leq -\|\mathfrak{S}\|(K - \theta\|\rho\|) + \theta\mathfrak{S}^{\mathsf{T}}\mathcal{I}_b^+(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) \\
&\leq -\|\mathfrak{S}\|(K - \bar{\theta}\bar{\rho} - \bar{\theta}\bar{h}),
\end{aligned} \tag{16}
$$

where $\theta < \bar{\theta}$, $\bar{\rho}$ and $\bar{h}$ are the norm-bound of ρ and $\mathcal{I}_b^+(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z})$, respectively. Thus, if $K > \bar{\theta}(\bar{\rho} + \bar{h})$ is satisfied, the system trajectory is maintained on SMS. □

According to the SMC theory, from $\mathfrak{S}(\mathcal{Z},t) = \dot{\mathfrak{S}}(\mathcal{Z},t) = 0$, we can obtain the formulation of equivalent control law as

$$u_{deq} = -\frac{1}{m}\rho - (\hat{\theta}-1)u_c - \frac{1}{m}\big(\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\big)^{-1}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}). \tag{17}$$

Substituting (17) and (8) into (1), the SMD is obtained as

$$
\begin{aligned}
\dot{\mathcal{Z}} &= \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})\big(m(u_d + \hat{\theta}u_c) - \rho\big) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) \\
&= \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c - \mathcal{I}_b(\mathcal{Z})\big(\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\big)^{-1}\mathfrak{M}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) \\
&= \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c + \Big(I_n - \mathcal{I}_b(\mathcal{Z})\big(\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\big)^{-1}\mathfrak{M}(\mathcal{Z})\Big)h(\mathcal{Z})\mathcal{F}(\mathcal{Z}).
\end{aligned} \tag{18}
$$

Considering $\mathfrak{M}(\mathcal{Z}) = \mathcal{I}_b^+(\mathcal{Z})$, (18) becomes

$$
\begin{aligned}
\dot{\mathcal{Z}} &= \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c + \big(I_n - \mathcal{I}_b(\mathcal{Z})\mathcal{I}_b^+(\mathcal{Z})\big)h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) \\
&= \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c + \big(I_n - \mathcal{I}_b(\mathcal{Z})\mathcal{I}_b^+(\mathcal{Z})\big)\big(I_n - \mathcal{I}_b(\mathcal{Z})\mathcal{I}_b^+(\mathcal{Z})\big)\mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z}) \\
&= \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c + \big(I_n - \mathcal{I}_b(\mathcal{Z})\mathcal{I}_b^+(\mathcal{Z})\big)\mathcal{K}(\mathcal{Z})\mathcal{F}(\mathcal{Z}).
\end{aligned} \tag{19}
$$

Then, the SMD (19) is revised as

$$\dot{\mathcal{Z}} = \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c + h(\mathcal{Z})\mathcal{F}(\mathcal{Z}). \tag{20}$$

Assumption 1. The mismatched component $h(\mathcal{Z})\mathcal{F}(\mathcal{Z})$ satisfies $\|h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\| \leq \Gamma(\mathcal{Z})$ and the input matrix function $\mathcal{I}_b(\mathcal{Z})$ is norm-bounded by $\|\mathcal{I}_b(\mathcal{Z})\| \leq \bar{g}$, where $\Gamma(\mathcal{Z})$ is a known function and $\bar{g}$ is a positive constant.

3.2. ADP-based DETOC design

In the following, a DETOC law u_c is designed to stabilize the SMD (20). Through the design of a modified value function, the control problem of the SMD is transformed into a DETOC problem of its nominal version, which is given by

$$\dot{\mathcal{Z}} = \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c. \tag{21}$$

For the system (21), the value function is derived by

$$\mathcal{V}(\mathcal{Z}, u_c) = \int_t^\infty \left(\Gamma(\mathcal{Z}) + \nabla\mathcal{V}^\mathsf{T}(\mathcal{Z})\nabla\mathcal{V}(\mathcal{Z}) + \mathcal{Z}^\mathsf{T}Q\mathcal{Z} + u_c^\mathsf{T}Ru_c \right)d\tau, \tag{22}$$

where $Q \in \mathbb{R}^{n \times n}$ and $R \in \mathbb{R}^{m \times m}$ are positive definite symmetric matrices. Denote the optimal value function as $\mathcal{V}^*(\mathcal{Z})$, which satisfies

$$\mathcal{V}^*(\mathcal{Z}) = \min_{u_c \in R(\Omega)} \mathcal{V}(\mathcal{Z}, u_c), \tag{23}$$

where $R(\Omega)$ is the admissible control set over Ω. The $\mathcal{V}^*(\mathcal{Z})$ is the solution of the following HJB equation

$$\min_{u_c \in R(\Omega)} \mathcal{H}\left(\mathcal{Z}, \nabla\mathcal{V}^*(\mathcal{Z}), u_c \right) = 0, \tag{24}$$

where $\mathcal{H}\left(\mathcal{Z}, \nabla\mathcal{V}^*(\mathcal{Z}), u_c \right)$ is the Hamiltonian for $\nabla\mathcal{V}^*(\mathcal{Z})$ and u_c, written as

$$\mathcal{H}\left(\mathcal{Z}, \nabla\mathcal{V}^*(\mathcal{Z}), u_c \right) = \mathcal{V}^{*\mathsf{T}}(\mathcal{Z})(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c) + \mathcal{Z}^\mathsf{T}Q\mathcal{Z} + u_c^\mathsf{T}Ru_c + \Gamma(\mathcal{Z}) + \nabla\mathcal{V}^\mathsf{T}(\mathcal{Z})\nabla\mathcal{V}(\mathcal{Z}), \tag{25}$$

where $\nabla\mathcal{V}^*(\mathcal{Z}) \triangleq \partial\mathcal{V}^*(\mathcal{Z})/\partial\mathcal{Z}$. From (24). we have

$$\left. \frac{\partial\mathcal{H}\left(\mathcal{Z}, \nabla\mathcal{V}^*(\mathcal{Z}), u_c \right)}{\partial u_c} \right|_{u_c = u_c^*} = 0, \tag{26}$$

Then, according to (26) and (25), it yields

$$u_c^*(\mathcal{Z}) = -\frac{1}{2}R^{-1}\mathcal{I}_b^\mathsf{T}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}), \tag{27}$$

Substituting (27) into (24), we further obtain

$$\mathcal{H}\left(\mathcal{Z}, \nabla\mathcal{V}^*(\mathcal{Z}), u_c^* \right) = \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c^*) + \mathcal{Z}^\mathsf{T}Q\mathcal{Z} + u_c^{*\mathsf{T}}Ru_c^* + \Gamma(\mathcal{Z}) + \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}) = 0. \tag{28}$$

It is noted that the HJB equation (28) is a nonlinear partial differential equation, the analytical solution is extremely difficult or even impossible to be obtained. Many methods have been developed to solve the HJB equation using ADP-based time-triggered mechanism, and these methods design the TTC law which often involves heavy computational burden and wastes communication resources. To overcome this shortcoming, we develop a DETNOC method in the following.

Denote $\{t_s\}_{s=0}^\infty$ as a monotonically increasing sequence of triggering instants, where s is the sth triggering instant. Let the event-triggered state as $\hat{\mathcal{Z}}_s = \mathcal{Z}(t_s)$, the event-triggering error can be defined as

$$e_s(t) = \hat{\mathcal{Z}}_s - \mathcal{Z}(t). \tag{29}$$

Then, the DETOC law is derived by

$$u_c^*(\hat{\mathcal{Z}}_s) = -\frac{1}{2}R^{-1}\mathcal{I}_b^\mathsf{T}(\hat{\mathcal{Z}}_s)\nabla\mathcal{V}^*(\hat{\mathcal{Z}}_s). \tag{30}$$

Inspired by [26,28], a dynamic event-triggering condition is designed as

$$t_{s+1} = \inf\left\{ t > t_s \middle| \mathcal{P}(t) + \beta\left((1 - \varpi^2)\mathcal{Z}^\mathsf{T}Q\mathcal{Z} - L_u^2\bar{g}^2\|e_s(t)\|^2\right) \le 0 \right\} \tag{31}$$

with $t_0 = 0$, where β is a design parameter and $\mathcal{P}(t)$ is updated by

$$\dot{\mathcal{P}}(t) = -\alpha\mathcal{P}(t) + (1 - \varpi^2)\mathcal{Z}^\mathsf{T}Q\mathcal{Z} - L_u^2\bar{g}^2\|e_s(t)\|^2 \tag{32}$$

with $\mathcal{P}(0) \ge 0$, where $\alpha > 0$ and $0 < \varpi < 1$ are two design parameters, $L_u > 0$ is the Lipschitz constant from the following Assumption 2.

Assumption 2. For the optimal control law u_c^*, there exists a Lipschitz constant L_u such that

$$\|u_c^*(\mathcal{Z}) - u_c^*(\hat{\mathcal{Z}}_s)\| \le L_u\|\mathcal{Z} - \hat{\mathcal{Z}}_s\| = L_u\|e_s(t)\|. \tag{33}$$

Lemma 1. Let $\mathcal{P}(t)$ be turned by (32), and the events be generated using the condition (31), then $\mathcal{P}(t) \ge 0$ for any $t \in [0, \infty)$.

Proof. Based on the triggering condition (31), for any $t \in [0, \infty)$, one has

$$\mathcal{P}(t) + \beta\left((1 - \varpi^2)\mathcal{Z}^\mathsf{T}Q\mathcal{Z} - L_u^2\bar{g}^2\|e_s(t)\|^2\right) \ge 0. \tag{34}$$

If $\beta = 0$, then $\mathcal{P}(t) \geq 0$ is true. If $\beta \neq 0$, by combining (31) and (34), we can derive

$$\dot{\mathcal{P}}(t) + \alpha\mathcal{P}(t) = (1 - \varpi^2)\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - L_u^2\bar{g}^2\|e_s(t)\|^2 \geq -\frac{\mathcal{P}(t)}{\beta} \tag{35}$$

with $\mathcal{P}(0) \geq 0$. Then, by using the comparison lemma, we have

$$\mathcal{P}(t) \geq \mathcal{P}(0)e^{-(\alpha+\frac{1}{\beta})t}, \quad t \in [0, \infty). \tag{36}$$

It implies that $\mathcal{P}(t) \geq 0$ for any $t \in [0, \infty)$. $\quad\square$

Theorem 2. *Considering the system* (1)*, the SMD* (20)*, Assumptions* 1 *and* 2*, the composite control law* (8) *with* (11) *and* (30)*. By designing the dynamic event-triggering condition as* (31)*, the closed-loop system* (1) *can be ensured to be asymptotically stable.*

Proof. Select a Lyapunov function candidate as

$$L_2(t) = \mathcal{V}^*(\mathcal{Z}) + \mathcal{P}(t). \tag{37}$$

According to Theorem 1, the trajectory of system (1) is maintained on the ISM surface, and then, the sliding mode dynamic system (19) can be obtained. Taking the time derivative of the $L_2(t)$, it yields

$$\dot{L}_2(t) = \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c^*(\hat{\mathcal{Z}}_s) + h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\big) + \dot{\mathcal{P}}(t). \tag{38}$$

According to (27), it reveals

$$\nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_a(\mathcal{Z}) = -\nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c^* - \mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}Ru_c^* - \Gamma(\mathcal{Z}) - \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}), \tag{39}$$

Together with (39), we have (40).

$$\begin{aligned}\dot{L}_2(t) = {} & \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c^*(\hat{\mathcal{Z}}_s) + \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) - \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c^* - \mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) - \Gamma(\mathcal{Z}) \\ & - \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}) + \dot{\mathcal{P}}(t) \\ \leq {} & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) + \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\big(u_c^*(\hat{\mathcal{Z}}_s) - u_c^*\big) + \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})h(\mathcal{Z})\mathcal{F}(\mathcal{Z}) - \Gamma(\mathcal{Z}) - \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}) + \dot{\mathcal{P}}(t) \\ \leq {} & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) + \frac{1}{2}\|\nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\|^2 + \frac{1}{2}\|\mathcal{I}_b(\mathcal{Z})\big(u_c^*(\hat{\mathcal{Z}}_s) - u_c^*(\mathcal{Z})\big)\|^2 + \frac{1}{2}\|\nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\|^2 + \frac{1}{2}\|h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\|^2 \\ & - \Gamma(\mathcal{Z}) - \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}) + \dot{\mathcal{P}}(t) \\ \leq {} & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) + \frac{1}{2}\bar{g}^2\|\big(u_c^*(\hat{\mathcal{Z}}_s) - u_c^*(\mathcal{Z})\big)\|^2 + \frac{1}{2}\|h(\mathcal{Z})\mathcal{F}(\mathcal{Z})\|^2 - \Gamma(\mathcal{Z}) + \dot{\mathcal{P}}(t). \end{aligned} \tag{40}$$

Let $R = r^{\mathsf{T}}r$, and recalling Assumption 2, we get

$$\begin{aligned}\dot{L}_2(t) \leq {} & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} + L_u^2\bar{g}^2\|e_s(t)\|^2 - \|r\|^2\|u_c^*(\mathcal{Z})\|^2 + \dot{\mathcal{P}}(t) \\ \leq {} & -\varpi^2\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - (1 - \varpi^2)\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} + L_u^2\bar{g}^2\|e_s(t)\|^2 - \|r\|^2\|u_c^*(\mathcal{Z})\|^2 + \dot{\mathcal{P}}(t) \\ \leq {} & -\varpi^2\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - (1 - \varpi^2)\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} + L_u^2\bar{g}^2\|e_s(t)\|^2 - \|r\|^2\|u_c^*(\mathcal{Z})\|^2 - \alpha\mathcal{P}(t) + (1 - \varpi^2)\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - L_u^2\bar{g}^2\|e_s(t)\|^2 \\ \leq {} & -\varpi^2\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - \alpha\mathcal{P}(t) - \|r\|^2\|u_c^*(\mathcal{Z})\|^2, \end{aligned} \tag{41}$$

Since $\mathcal{P}(t) \geq 0$ based on Lemma 1, we further derive from (41) as

$$\dot{L}_2(t) \leq -\varpi^2\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - \|r\|^2\|u_c^*(\mathcal{Z})\|^2 \leq 0. \tag{42}$$

Therefore, the closed-loop system (1) can be guaranteed to be asymptotically stable. $\quad\square$

In the sequence, a three feedforward NN is used as a critic NN to approximate the optimal value function $\mathcal{V}^*(\mathcal{Z})$ on the compact Ω by

$$\mathcal{V}^*(\mathcal{Z}) = \vartheta_v^{\mathsf{T}}\phi_v(\mathcal{Z}) + \zeta_v(\mathcal{Z}). \tag{43}$$

where $\vartheta_v \in \mathbb{R}^{l_c}$ is the ideal weight vector, l_v is the number of neurons, $\phi_v(\mathcal{Z}) = [\phi_{v1}(\mathcal{Z}), \phi_{v2}(\mathcal{Z}), \dots, \phi_{vl_c}(\mathcal{Z})]^{\mathsf{T}}$ is the activation function, and its element $\phi_{vi}(\mathcal{Z})$ satisfy $\phi_{vi}(\mathcal{Z}) \in C^1(\Omega)$ with $\phi_{vi}(0) = 0$ and $\nabla\phi_{vi}(0) = 0$, $i = 1, 2, \dots, l_c$, and $\zeta_v(\mathcal{Z}) \in \mathbb{R}$ is the approximation error. Differentiating $\mathcal{V}^*(\mathcal{Z})$ with respect to $\mathcal{Z}$, it yields

$$\nabla\mathcal{V}^*(\mathcal{Z}) = \nabla\phi_v^{\mathsf{T}}(\mathcal{Z})\vartheta_v + \nabla\zeta_v(\mathcal{Z}). \tag{44}$$

Inserting (44) into (30), we have

$$u_c(\hat{\mathcal{Z}}_s) = -\frac{1}{2}R^{-1}\mathcal{I}_b^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\big(\nabla\phi_v^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\vartheta_v + \nabla\zeta_v(\hat{\mathcal{Z}}_s)\big). \tag{45}$$

Note that the ideal weight vector ϑ is unavailable, which causes the inability of implementing $u_c(\hat{\mathcal{Z}}_s)$. To solve the issue, an estimated weight vector $\hat{\vartheta}$ is employed to replace ϑ, and then the approximation $\mathcal{V}^*(\mathcal{Z})$ is expressed as

$$\hat{\mathcal{V}}(\mathcal{Z}) = \hat{\vartheta}_v^{\mathsf{T}}\phi_v(\mathcal{Z}). \tag{46}$$
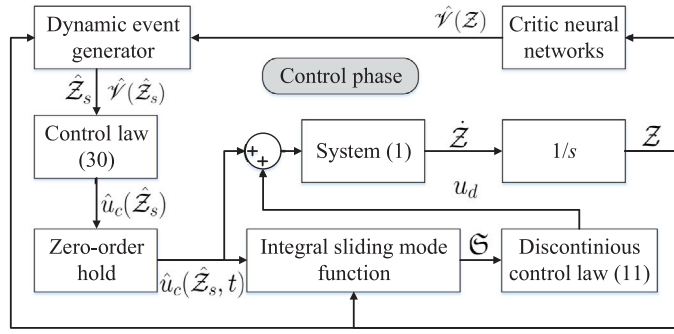
Fig. 1. The control phase of the developed ADP-based DETNOC method.

In similar to (44), we deduce

$$\nabla \hat{\mathscr{V}}(\mathcal{Z}) = \nabla \phi_v^{\mathsf{T}}(\mathcal{Z}) \hat{\vartheta}_v. \tag{47}$$

Then, the approximation DETOC law is given by

$$\hat{u}_c(\hat{\mathcal{Z}}_s) = -\frac{1}{2} R^{-1} \mathcal{I}_b^{\mathsf{T}}(\hat{\mathcal{Z}}_s) \nabla \hat{\phi}_v^{\mathsf{T}}(\hat{\mathcal{Z}}_s) \vartheta_v. \tag{48}$$

Replacing simultaneously $\mathscr{V}^*(\mathcal{Z})$ and $u_c^*(\mathcal{Z})$ in Hamiltonian (27) with $\hat{\mathscr{V}}^*(\mathcal{Z})$ and $u_c(\hat{\mathcal{Z}}_s)$, the approximate Hamiltonian is expressed as

$$\mathscr{H}\big(\mathcal{Z}, \nabla \hat{\mathscr{V}}(\mathcal{Z}), \hat{u}_c^*(\hat{\mathcal{Z}}_s)\big) = \hat{\vartheta}_v^{\mathsf{T}} \nabla \phi_v(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})\hat{u}_c(\hat{\mathcal{Z}}_s)\big) + \mathcal{Z}^{\mathsf{T}} Q \mathcal{Z} + \hat{u}_c^{\mathsf{T}}(\hat{\mathcal{Z}}_s) R \hat{u}_c(\hat{\mathcal{Z}}_s) + \Gamma(\mathcal{Z}) + \hat{\vartheta}_v^{\mathsf{T}} \nabla \phi_v(\mathcal{Z}) \nabla \phi_v^{\mathsf{T}}(\mathcal{Z}) \hat{\vartheta}_v. \tag{49}$$

According to (28), $\mathscr{H}\big(\mathcal{Z}, \nabla \mathscr{V}^*(\mathcal{Z}), u_c^*\big) = 0$. Thus, there exists an error between (28) and (49). Let e_v be the error, we have

$$e_v = \hat{\mathscr{H}}\big(\mathcal{Z}, \nabla \hat{\mathscr{V}}(\mathcal{Z}), \hat{u}_c(\hat{\mathcal{Z}}_s)\big) - \mathscr{H}\big(\mathcal{Z}, \nabla \mathscr{V}^*(\mathcal{Z}), u_c^*\big) = \hat{\vartheta}_v^{\mathsf{T}} \varphi_v + \mathcal{Z}^{\mathsf{T}} Q \mathcal{Z} + \hat{u}_c^{\mathsf{T}}(\hat{\mathcal{Z}}_s) R \hat{u}_c(\hat{\mathcal{Z}}_s) + \Gamma(\mathcal{Z}) + \hat{\vartheta}_v^{\mathsf{T}} \nabla \phi_v(\mathcal{Z}) \nabla \phi_v^{\mathsf{T}}(\mathcal{Z}) \hat{\vartheta}_v, \tag{50}$$

where $\varphi_v = \nabla \phi_v(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})\hat{u}_c\big)$. The training objective of the critic NN is to make $e_v \to 0$ through tuning $\hat{\vartheta}_v \to \vartheta_v$. To achieve this objective, we should minimize the square function defined as $E_v = 0.5 e_v^{\mathsf{T}} e_v$ via updating $\hat{\vartheta}_v$. Based on the gradient descent algorithm, $\hat{\vartheta}_v$ is updated by

$$\dot{\hat{\vartheta}}_v = -\alpha_c \frac{1}{(1 + \varphi_v^{\mathsf{T}} \varphi_v)^2}\left(\frac{\partial E_v}{\partial \hat{\vartheta}_v}\right) = -\alpha_c \frac{\varphi_v + 2\nabla \phi_v(\mathcal{Z}) \nabla \phi_v^{\mathsf{T}}(\mathcal{Z}) \hat{\vartheta}_v}{(1 + \varphi_v^{\mathsf{T}} \varphi_v)^2} e_v. \tag{51}$$

Let $\tilde{\vartheta}_v = \vartheta_v - \hat{\vartheta}_v$ be the weight error vector, which can be guaranteed to be UUB with the updating law (51). The proof has been provided in [11,20,24], so the detail is omitted here.

To illustrate the developed ADP-based DETNOC method, the control architecture is shown in Fig. 1, and the process of controller design is provided in Algorithm 1.

3.3. Stability analysis

Assumption 3. $\nabla \phi_v(\mathcal{Z})$, $\nabla \zeta_v(\mathcal{Z})$ and $\dot{\vartheta}$ are norm-bounded, i.e., $\|\nabla \phi_v(\mathcal{Z})\| \le \bar{\phi}_v$, $\|\nabla \zeta_v(\mathcal{Z})\| \le \bar{\zeta}_v$ and $\|\dot{\vartheta}\| \le \bar{\vartheta}$, where $\bar{\phi}_v$, $\bar{\zeta}_v$ and $\bar{\vartheta}_v$ are positive constants.

Theorem 3. *Considering the system* (21), *Assumptions* 1–3, *the approximate DETOC law* (48), *if the dynamic event-triggering condition* (31) *is satisfied, the closed-loop system* (21) *is guaranteed to be UUB and* $\mathcal{P}(t)$ *is asymptotically stable.*

Proof. Choose the Lyapunov function candidate as

$$L_3(t) = \underbrace{\mathscr{V}^*(\hat{\mathcal{Z}}_s)}_{L_{31}} + \underbrace{\mathscr{V}^*(\mathcal{Z})}_{L_{32}} + \underbrace{\mathcal{P}(t)}_{L_{33}} \tag{52}$$

Case 1: There is no event triggered, i.e., $\forall t \in [t_s, t_{s+1})$. Differentiating (52), yields

$$\dot{L}_{31} = 0, \tag{53a}$$

$$\dot{L}_{32} = \nabla \mathscr{V}^{*\mathsf{T}}(\mathcal{Z})\big(\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})\hat{u}_c(\hat{\mathcal{Z}}_s)\big), \tag{53b}$$

$$\dot{L}_{33} = \dot{\mathcal{P}}(t), \tag{53c}$$

Algorithm 1 ADP-based DETNOC method

Step 1: Switch control law design

1-1: Construct the sliding mode function as

$$\mathfrak{S}(\mathcal{Z},t) = \mathcal{G}(\mathcal{Z}) - \mathcal{G}(\mathcal{Z}_0) - \int_0^t \mathfrak{M}(\mathcal{Z})\big(\mathcal{Z}\big)\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c(\mathcal{Z})\big)d\tau.$$

1-2: Design the adaptive term θ, which is updated by $\dot{\theta} = -(1/\gamma)\mathfrak{S}^{\mathsf{T}}\mathfrak{M}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c$. Moreover, design a discontinuous control law $u_d = -K\mathrm{sgn}\big(\Xi(\mathcal{Z})\big)$ to maintain the system trajectory on SMS for obtaining the SMD

$$\dot{\mathcal{Z}} = \mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})u_c + h(\mathcal{Z})\mathcal{F}(\mathcal{Z}).$$

Step 2: Dynamic event-triggered control law design

2-1: According to the nominal version of the SMD, define the value function as

$$\mathcal{V}(\mathcal{Z},u_c) = \int_t^{\infty} \big(\Gamma(\mathcal{Z}) + \nabla\mathcal{V}^{\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}(\mathcal{Z}) + \mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} + u_c^{\mathsf{T}}Ru_c\big)d\tau.$$

2-1: Design a dynamic event-triggering strategy as

$$t_{s+1} = \inf\left\{t > t_s \,\middle|\, \mathcal{P}(t) + \beta\big((1-\varpi^2)\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - L_u^2\bar{g}^2\|e_s(t)\|^2\big) \le 0\right\}.$$

2-2: Introduce a critic NN to approximate the optimal value function, the weight vector of the critic NN is updated by

$$\dot{\vartheta}_c = -\alpha_c \frac{1}{(1+\varphi_v^{\mathsf{T}}\varphi_v)^2}\left(\frac{\partial E_v}{\partial\hat{\vartheta}_v}\right).$$

2-3: Obtain the dynamic event-triggered control law as

$$\hat{u}_c(\hat{\mathcal{Z}}_s) = -\frac{1}{2}R^{-1}\mathcal{I}_b^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\nabla\hat{\phi}_v^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\vartheta_v.$$

Step 3: Integrate the discontinuous control law u_d and the ADP-based DETOC law $\hat{u}_c(\hat{\mathcal{Z}}_s)$

3-1: From composite control law (8), integrate the control laws u_d and $\hat{u}_c(\hat{\mathcal{Z}}_s)$ as

$$u = -K\mathrm{sgn}\big(\Xi(\mathcal{Z})\big) - \frac{1}{2}\hat{\theta}R^{-1}\mathcal{I}_b^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\nabla\hat{\phi}_v^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\vartheta_v.$$

3-2: Apply the composite control law to nonlinear system (1) with uncertainties, disturbances, and dead-zone.

From (28) and (60), we have

$$
\begin{aligned}
\dot{L}_3 = & -\nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})u_c^* - \mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) + \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})\hat{u}_c(\hat{\mathcal{Z}}_s) + \dot{\mathcal{P}}(t) - \Gamma(\mathcal{Z}) - \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}) \\
\le & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) + \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\mathcal{I}_b(\mathcal{Z})(\hat{u}_c(\hat{\mathcal{Z}}_s) - u_c^*(\mathcal{Z})) + \dot{\mathcal{P}}(t) - \nabla\mathcal{V}^{*\mathsf{T}}(\mathcal{Z})\nabla\mathcal{V}^*(\mathcal{Z}) \\
\le & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) + \frac{1}{2}\|\mathcal{I}_b(\mathcal{Z})(\hat{u}_c(\hat{\mathcal{Z}}_s) - u_c^*(\mathcal{Z}))\|^2 + \dot{\mathcal{P}}(t) \\
\le & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - u_c^{*\mathsf{T}}(\mathcal{Z})Ru_c^*(\mathcal{Z}) + \frac{1}{2}\bar{g}^2\underbrace{\|u_c^*(\mathcal{Z}) - \hat{u}_c(\hat{\mathcal{Z}}_s)\|^2}_{\Theta} + \dot{\mathcal{P}}(t).
\end{aligned}
\tag{54}
$$

Considering $\vartheta_v = \hat{\vartheta}_v + \tilde{\vartheta}_v$, we get

$$
\begin{aligned}
\Theta = & \left\|\big(u_c^*(\mathcal{Z}) - u_c^*(\hat{\mathcal{Z}}_s)\big) + \big(u_c^*(\hat{\mathcal{Z}}_s) - \hat{u}_c(\hat{\mathcal{Z}}_s)\big)\right\|^2 \\
\le & 2\left\|u_c^*(\mathcal{Z}) - u_c^*(\hat{\mathcal{Z}}_s)\right\|^2 + 2\left\|u_c^*(\hat{\mathcal{Z}}_s) - \hat{u}_c(\hat{\mathcal{Z}}_s)\right\|^2 \\
\le & \frac{1}{2}\|r^{-1}\|^2\bar{g}\left\|\nabla\phi_v^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\hat{\vartheta}_v - \nabla\phi_v^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\vartheta_v - \nabla\zeta_v(\hat{\mathcal{Z}}_s)\right\|^2 + 2L_u^2\|e_s(t)\|^2 \\
\le & \frac{1}{2}\|r^{-1}\|^2\bar{g}^2\|-\nabla\zeta_v(\hat{\mathcal{Z}}_s) - \nabla\phi_c^{\mathsf{T}}(\hat{\mathcal{Z}}_s)\tilde{\vartheta}_v\|^2 + 2L_u^2\|e_s(t)\|^2 \\
\le & 2L_u^2\|e_s(t)\|^2 + \underbrace{\|r^{-1}\|^2\bar{g}^2\big(\bar{\phi}_v^2\tilde{\vartheta}_v^2 + \bar{\zeta}_v^2\big)}_{\bar{\Pi}}.
\end{aligned}
\tag{55}
$$

According to (32) and (55), (54) becomes

$$
\begin{aligned}
\dot{L}_3 = & -\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - r^2\|u_c^*(\mathcal{Z})\|^2 + L_u^2\bar{g}^2\|e_s(t)\|^2 + \bar{\Pi} + \dot{\mathcal{P}}(t) \\
= & -\varpi^2\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - (1-\varpi^2)\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - \|r\|^2\|u_c^*(\mathcal{Z})\|^2 + L_u^2\bar{g}^2\|e_s(t)\|^2 + \bar{\Pi} - \alpha\mathcal{P}(t) + (1-\varpi^2)\mathcal{Z}^{\mathsf{T}}Q\mathcal{Z} - L_u^2\bar{g}^2\|e_s(t)\|^2.
\end{aligned}
\tag{56}
$$

8

Therefore, it yields

$$\dot{L}_3 \leq -\varpi^2 \lambda_{\min}(Q)\|\mathcal{Z}\|^2 - \|r\|^2\|u_c^*(\mathcal{Z})\|^2 - \alpha\mathcal{P}(t) + \bar{\Pi}. \tag{57}$$

Then, if $\mathcal{Z}$ lies outside the compact set $\Omega_{\mathcal{Z}} = \left\{ \mathcal{Z} : \|\mathcal{Z}\| \leq \sqrt{\bar{\Pi}/(\varpi^2 \lambda_{\min}(Q))} \right\}$, we have

$$\dot{L}_3 \leq -\|r\|^2\|u_c^*(\mathcal{Z})\|^2 - \alpha\mathcal{P}(t) \leq 0, \tag{58}$$

Furthermore, it shows $\dot{L}_3 \leq -\|r\|^2\|u_c^*(\mathcal{Z})\|^2 < 0, \forall \mathcal{Z} \neq 0$ and $\dot{L}_3 \leq -\alpha\mathcal{P}(t) < 0, \forall \mathcal{P}(t) \neq 0$. Hence, it reveals that the closed-loop system (21) is guaranteed to be UUB and $\mathcal{P}(t)$ is asymptotically stable using the Lyapunov theorem [26,28].

Case 2: Events are triggered, i.e., $\forall t = t_{s+1}$. Calculating the difference of the Lyapunov function candidate, we have

$$\Delta L_3 = \Delta L_{31} + \Delta L_{32} + \Delta L_{33}, \tag{59}$$

with

$$\Delta L_{31} = \mathscr{V}^*(\hat{\mathcal{Z}}_{s+1}) - \mathscr{V}^*(\hat{\mathcal{Z}}_s), \tag{60a}$$

$$\Delta L_{32} = \mathscr{V}^*\left(\mathcal{Z}(t_{s+1})\right) - \mathscr{V}^*\left(\mathcal{Z}(t_{s+1}^-)\right), \tag{60b}$$

$$\Delta L_{33} = \mathcal{P}(t_{s+1}) - \mathcal{P}(t_{s+1}^-), \tag{60c}$$

where $\hbar(t_{s+1}^-) = \lim_{\Delta t \to 0} \hbar(t_{s+1} - \Delta t)$, and $\hbar(\cdot)$ denotes $\mathcal{Z}(\cdot)$ and $\mathcal{P}(\cdot)$, respectively. According to the proof presented in Case 1, $\dot{L}_3 < 0$ for $t \in [t_s, t_{s+1})$ if $\mathcal{Z}$ lies outside the compact set (58). This gives that L_3 is strictly decreasing over $t \in [t_s, t_{s+1})$. With all systems signals are continuous, it means that $\Delta L_{31} \leq 0$, $\Delta L_{32} \leq 0$ and $\Delta L_{33} \leq 0$. Then, we have

$$\Delta L_3 \leq -\chi(\|e_{s+1}(t_s)\|) \leq 0, \tag{61}$$

where $e_{s+1}(t_s) = \hat{\mathcal{Z}}_{s+1} - \hat{\mathcal{Z}}_s$ and $\chi(\cdot)$ is a class-$\mathcal{K}$ function. Thus, it can be concluded that the Lyapunov candidate is decreasing $\forall t = t_{s+1}$.

Combining Cases 1 and 2, one can obtain the conclusion that the UUB stability of $\mathcal{Z}$ and the asymptotic stability of $\mathcal{P}(t)$ are guaranteed under the dynamic event-triggering condition (31) and the approximate DETOC law (48). $\quad\square$

Remark 4. Recall the property that multilayer NNs are universal approximators, capable of approximating smooth functions on a compact set [26,36]. However, there is always a deviation between the approximate function and the target function in practice. Observing Theorem 3, this deviation results in an error term given by $0.5\|r^{-1}\|^2\bar{g}^2\| - \nabla\zeta_v(\hat{\mathcal{Z}}_s) - \nabla\phi_c^\top(\hat{\mathcal{Z}}_s)\hat{\vartheta}_v\|^2$ occurrence in (55), which is upper-bounded by a positive constant $\bar{\Pi}$. In Theorem 3, according to Lyapunov theorem, only UUB stability can be guaranteed for the closed-loop system (21), rather than asymptotically stable, due to the existence of $\bar{\Pi}$.

Assumption 4. The nonlinear function $\mathcal{I}_a(\mathcal{Z})$ is uniformly bounded on Ω, i.e., $\sup_{\mathcal{Z} \in \Omega} \|\mathcal{I}_a(\mathcal{Z})\| \leq K_f\|\mathcal{Z}\|$, where K_f is a positive constant [28,37].

In the following sequence, we will provide a proof that the Zeno behavior is excluded.

Theorem 4. *Let Assumption 4 holds, considering the system (21), the dynamic event-triggering condition (31), the minimal intersampling time $\Delta t_{\min}$ has lower bound by*

$$\Delta t_{\min} \geq \frac{1}{K_f}\ln\left(1 + \frac{\sqrt{\mathcal{P}(t)/\beta + (1-\varpi^2)\mathcal{Z}^\top Q\mathcal{Z}}}{\mathcal{B}_z}\right) > 0, \tag{62}$$

where K_f and $\mathcal{B}_z$ are positive constants.

Proof. Differentiating the event-triggered error $e_p(t)$, yields

$$\frac{d\left(e_s(t)\right)}{dt} = \dot{e}_s(t) = \dot{\hat{\mathcal{Z}}}_s - \dot{\mathcal{Z}} \equiv -\dot{\mathcal{Z}}, \forall t \in [t_s, t_{s+1}).$$

By using the (29) and (48), we have

$$\begin{aligned}
\|\dot{e}_s\| &= \|\dot{\mathcal{Z}}\| \\
&= \left\|\mathcal{I}_a(\mathcal{Z}) + \mathcal{I}_b(\mathcal{Z})\hat{u}_c\right\| \\
&\leq \left\|\mathcal{I}_a(\mathcal{Z})\right\| + \left\|\mathcal{I}_b(\mathcal{Z})\hat{u}_c\right\| \\
&\leq K_f\|x\| + K_g \\
&\leq K_f\|\hat{\mathcal{Z}}_s - e_s\| + K_g \\
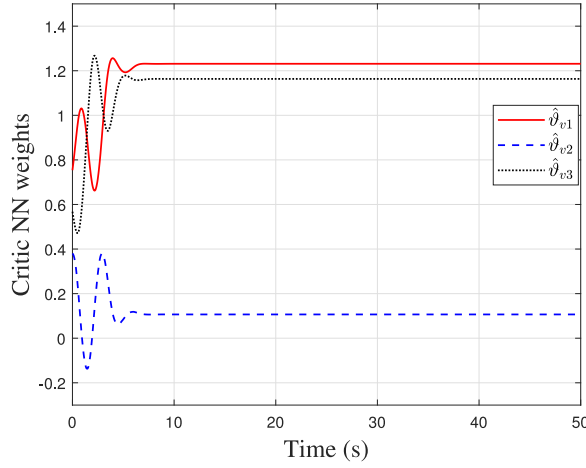&\leq K_f\|e_s\| + K_f\|\hat{\mathcal{Z}}_s\| + K_g,
\end{aligned} \tag{63}$$

Fig. 2. Critic NN weights.

where $\|\mathcal{I}_b(\mathcal{Z})\hat{u}_c\| \le K_g$, $\mathcal{I}_a(\mathcal{Z})$ is a Lipschitz function which satisfies $\|\mathcal{I}_a(\mathcal{Z})\| \le K_f\|\mathcal{Z}\|$, where K_f is a positive constant. According to [28,37], we can obtain from (63)

$$\|e_s\| \le \frac{K_f\|\hat{\mathcal{Z}}_s\| + K_g}{K_f}\left(e^{K_f(t-t_s)} - 1\right) \tag{64}$$

for all $t \in [t_s, t_{s+1})$. Recalling the dynamic event-triggering condition (31), we have

$$\|e_s\|^2 \ge \frac{\mathcal{P}(t)/\beta + (1 - \varpi^2)\mathcal{Z}^\mathsf{T} Q \mathcal{Z}}{L_u^2 \bar{g}^2} \tag{65}$$

for $t = t_{s+1}$. According to (64) and (65), it indicates that the sth intersampling time satisfies

$$t_{s+1} - t_s \ge \frac{1}{K_f}\ln\left(1 + \frac{\sqrt{\mathcal{P}/\beta + (1 - \varpi^2)\mathcal{Z}^\mathsf{T} Q \mathcal{Z}}}{B_z}\right) > 0,$$

where $B_z = \frac{K_f\|\hat{\mathcal{Z}}_s\| + K_g}{K_f}$. That is to say, $\Delta t_{\min} = \min\{t_{s+1} - t_s\} > 0$ in (62). $\quad\square$

4. Simulation studies

4.1. Example 1

Consider the following nonlinear system

$$\dot{\mathcal{Z}} = \begin{bmatrix} \mathcal{Z}_2 \\ -\mathcal{Z}_1 + 0.5(1 - \mathcal{Z}_2^2)\mathcal{Z}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}\left(\mathfrak{D}(u) + \overbrace{\sin(2t)\cos(t)}^{\varrho(t)}\right) + \overbrace{\begin{bmatrix} 0.15 \\ 0 \end{bmatrix}}^{\mathcal{K}(\mathcal{Z})}\overbrace{\left(0.25\mathcal{Z}_2\sin(\mathcal{Z}_1\mathcal{Z}_2)\cos(\mathcal{Z}_2)\right)}^{\mathcal{F}(\mathcal{Z})}, \tag{66}$$

where $\mathcal{Z} = [\mathcal{Z}_1, \mathcal{Z}_2]^\mathsf{T} \in \mathbb{R}^2$ is the system state, u is the control input, the input dead-zone $\mathfrak{D}(u)$ is given as $m_l = m_r = 0.5$, $o_r = 0.25$ and $o_l = -0.25$, and the nominal system can be given by

$$\dot{\mathcal{Z}} = \begin{bmatrix} \mathcal{Z}_2 \\ -\mathcal{Z}_1 + 0.5(1 - \mathcal{Z}_2^2)\mathcal{Z}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u. \tag{67}$$

First, the ADP-based DETOC law is designed for nominal system (67). Define the value function as (22), and its parameters are set as $Q = I$, $R = 1$ and $\Gamma(\mathcal{Z}) = \|\mathcal{Z}\|^2$. For the critic NN, $\sigma_v(\mathcal{Z}) = [\mathcal{Z}_1^2, \mathcal{Z}_1\mathcal{Z}_2, \mathcal{Z}_2^2]^\mathsf{T}$, $\hat{\vartheta}_v = [\hat{\vartheta}_{v1}, \hat{\vartheta}_{v2}, \hat{\vartheta}_{v3}]^\mathsf{T}$, $\alpha_c = 1.8$. The parameters of the triggering condition are selected as $L_u = 15$, $\varpi = 06$, $\bar{g} = 1.2$, $\beta = 0.6$, $\mathcal{P}(0) = 4$, and $\alpha = 0.3$. Simulation results are illustrated in Figs. 2–10.

From Fig. 2, it can be observed that the weight vector of the critic neural network $\hat{\vartheta}_c$ converges to the values $[1.2312, 0.1066, 1.1633]^\mathsf{T}$ finally. Fig. 3 provides the state trajectories of the nominal system. As shown in Fig. 4, the DETOC input $\hat{u}_c$ is a piecewise signal which is updated at t_s only and keeps unchanged during $[t_s, t_{s+1})$ under the DETC mechanism. The curves of dynamic event-triggered threshold and error are shown in Fig. 5. The sampling period of TTC mechanism is set as 0.05 s, the updating times of the DETC, ETC and TTC are depicted in Fig. 6, it is clear from this figure that the less updating frequency of the control signal in DETC than ETC and TTC mechanism, and easy to conclude that the computational and communication resources can be saved. Fig. 7 presents the curves of dynamic variable and its low bound function, one can find that the dynamic variable
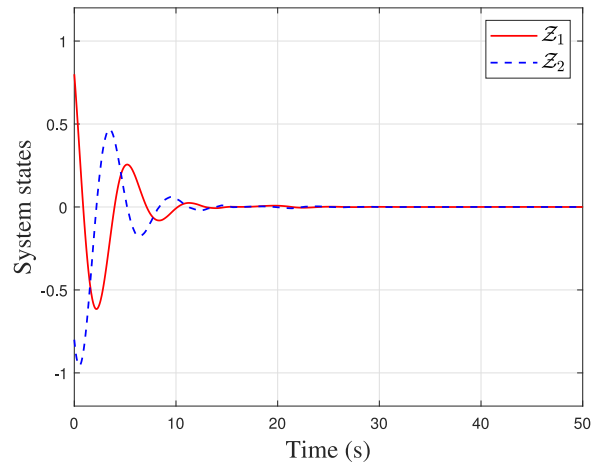
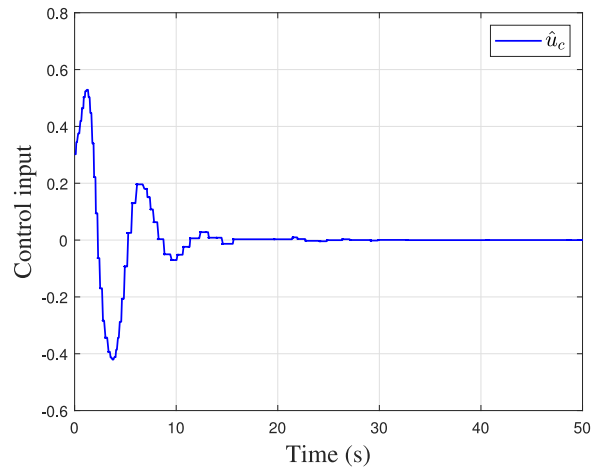**Fig. 3.** System states of system (67).

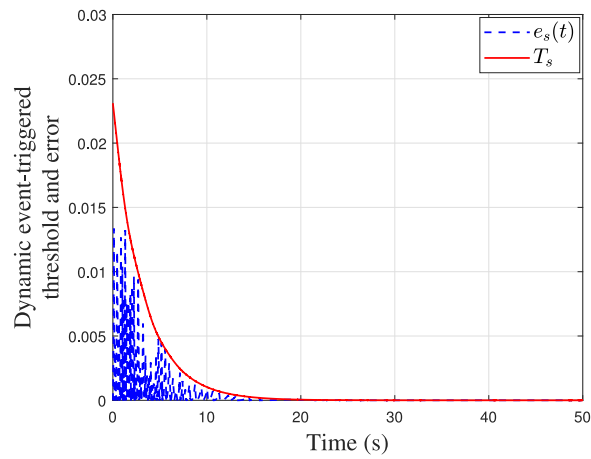**Fig. 4.** Approximation DETOC input $\hat{u}_c$ of system (67).

Fig. 5. Dynamic event-triggered threshold and error.

is always greater than or equal to the bound function with the time increasing. The triggering time instants are shown in Fig. 8, which demonstrates that Zeno behavior is avoided.
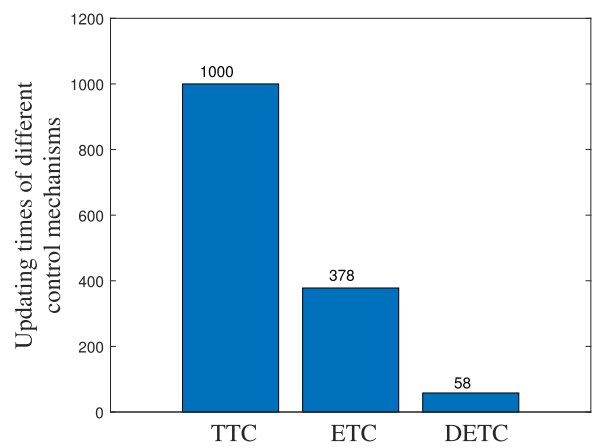
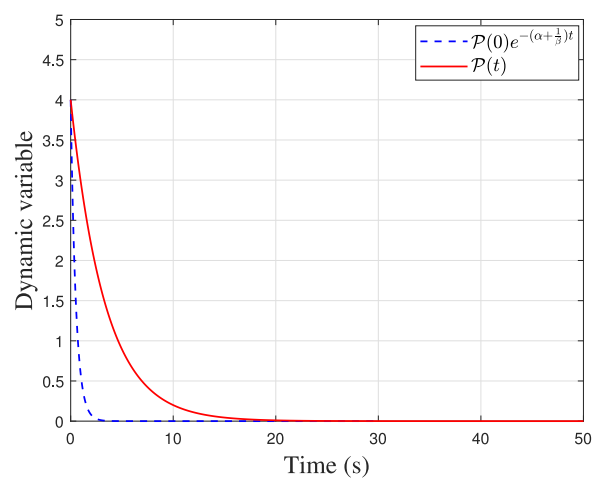**Fig. 6.** Updating times of different control mechanisms.
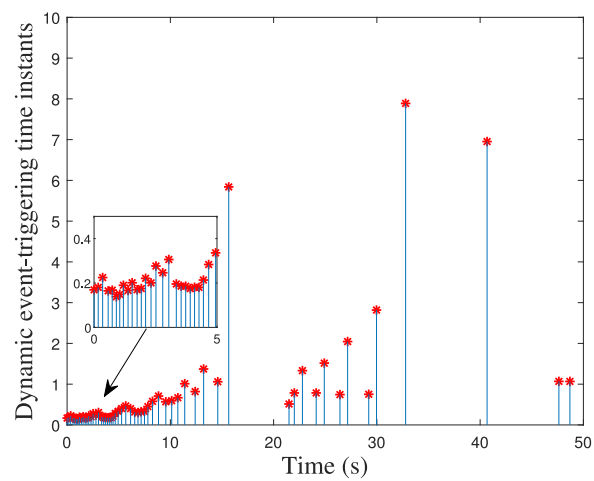

Fig. 7. Dynamic variable and its low bound function.

Fig. 8. Dynamic event-triggering time instants.

12
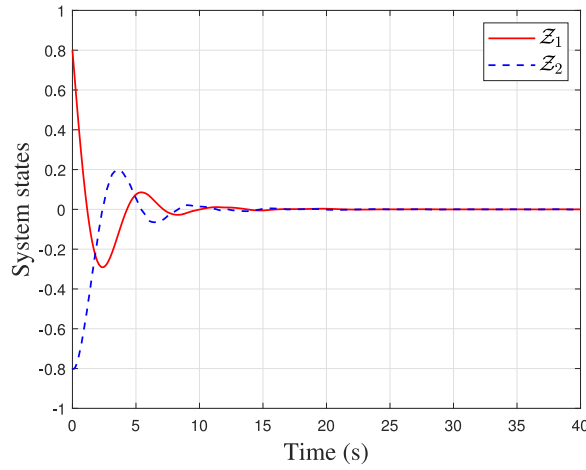
Fig. 9. System states of closed-loop system (66).
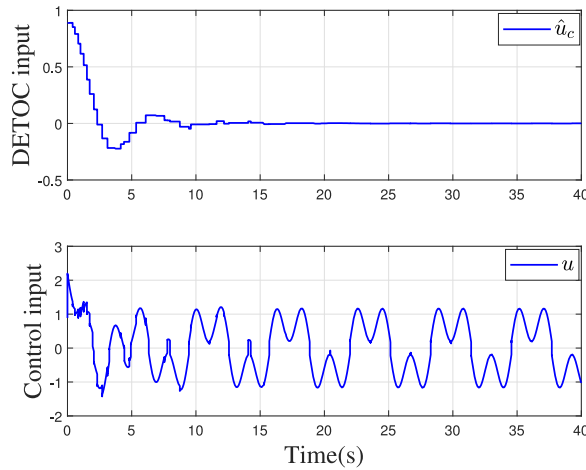
Fig. 10. Approximation DETOC and composite control inputs.

Table 1
Parameters of the pendulum system.

Parameter	J	L	M	g	f_d
Value	4 kg m^2	1.5 m	4/3 kg	9.8 m/s^2	0.8 N m s/rad

Next, the sliding mode function is designed as (9) with $\mathcal{G}(\mathcal{Z}) = \mathcal{Z}_2$ and $\mathfrak{M}(\mathcal{Z}) = [0, 1]$. In order to weaken the chattering, the discontinuous control law (11) is designed as $u_d = -K \tanh(\mathcal{I}_b^\mathsf{T}(\mathcal{Z})\mathfrak{M}^\mathsf{T}(\mathcal{Z})\mathfrak{S}(\mathcal{Z})/\xi)$ to replace $-K \operatorname{sgn}(\mathcal{I}_b^\mathsf{T}(\mathcal{Z})\mathfrak{M}^\mathsf{T}(\mathcal{Z})\mathfrak{S}(\mathcal{Z}))$ with $\xi = 0.0001$ and $K = 3$. The composite control input (8) with (11) and (48) is employed to drive the system (66) for simulation. As depicted in Fig. 9, the states of the closed-loop system converge to a small region of zero (SRZ) within 20 s. Fig. 10 presents the curves of the DETOC and the composite control input. Based on the aforementioned discussions, it is evident that the proposed ADP-based DETNOC method effectively addresses the challenges posed by dead-zone, disturbance, and uncertainties.

4.2. Example 2

Consider the following pendulum system [38]

$$\ddot{\theta} = -\frac{f_d}{J}\dot{\theta} - \frac{MgL}{J}\sin(\theta) + \frac{1}{J}u,$$

where $\theta \in \mathbb{R}$ is the pendulum's angle position, u is the control input, and the system parameters are shown in Table 1. Define $\mathcal{Z}_1 = \theta$, $\mathcal{Z}_2 = \dot{\theta}$, and considering dead-zone, disturbance, and uncertainties, we have
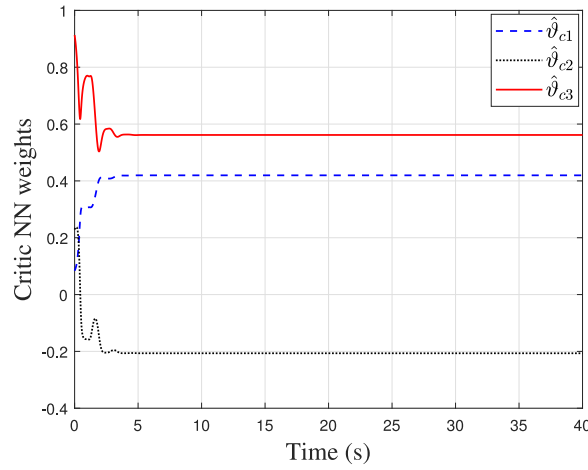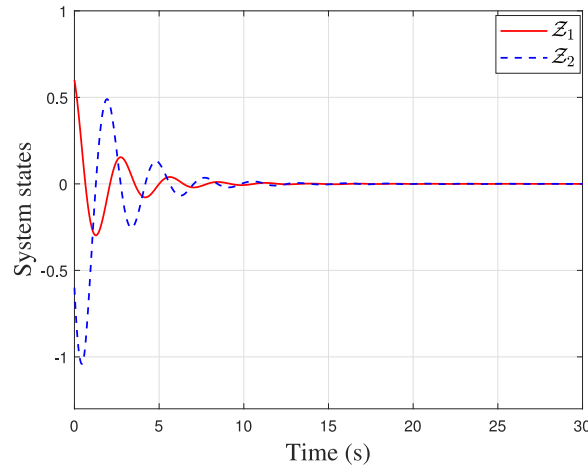
Fig. 11. The critic NN weights.

Fig. 12. System states of system (69).

$$\dot{\mathcal{Z}} = \begin{bmatrix} \mathcal{Z}_2 \\ -4.9\sin(\mathcal{Z}_1) - 0.2\mathcal{Z}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.25 \end{bmatrix} \overbrace{\left(\mathfrak{D}(u) + \overbrace{\sin(3t)}^{\varrho(t)}\right)}^{} + \overbrace{\begin{bmatrix} 0.1 \\ 0 \end{bmatrix}}^{\mathcal{K}(\mathcal{Z})} \overbrace{\left(0.25\mathcal{Z}_2\sin(\mathcal{Z}_1\mathcal{Z}_2)\right)}^{\mathcal{F}(\mathcal{Z})}, \tag{68}$$

where $\mathcal{Z} = [\mathcal{Z}_1, \mathcal{Z}_2]^{\mathsf{T}} \in \mathbb{R}^2$ is the system state, and the parameters of the dead-zone are set as $m_l = m_r = 0.5$, $o_r = 0.25$ and $o_l = -0.25$.

First, an ADP-based DETOC law is designed to the nominal system

$$\dot{\mathcal{Z}} = \begin{bmatrix} \mathcal{Z}_2 \\ -4.9\sin(\mathcal{Z}_1) - 0.2\mathcal{Z}_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.25 \end{bmatrix} u. \tag{69}$$

The value function parameters are chosen as $Q = 2I$, $R = 0.05$, and $\Gamma(\mathcal{Z}) = \|\mathcal{Z}\|^2$. The structure of critic NN is chosen as the same to Example 1, and the learning rate of critic NN is set as $\alpha_c = 2$.

The learning process of critic NN is depicted in Fig. 11, and the vector of the critic NN weight finally converges to $[0.4196, -0.2065, 0.5616]^{\mathsf{T}}$. Fig. 12 shows the convergence of states of system (68), it is evident that the states tend to an SRZ after 15s. From Fig. 13, we can observe that the DETOC input is updated in an aperiodic manner. Fig. 14 presents the curves of dynamic event-triggered threshold and error, which indicates the effectiveness of the designed triggering condition. In comparison with the TTC method, a equidistant period of 0.05 s is employed. From Fig. 15, we observe that DETC requires a lower updating frequency of the control signal compared to ETC and TTC mechanisms, which means that DETC can help save computational and communication resources. In Fig. 16, it is emphasized that the dynamic signal is indeed restricted by an exponential signal. From Fig. 17, it can be seen that the minimal intersampling time $\Delta t_{\min} = 0.047$ is greater than zero, indicating the avoidance of Zeno phenomenon.
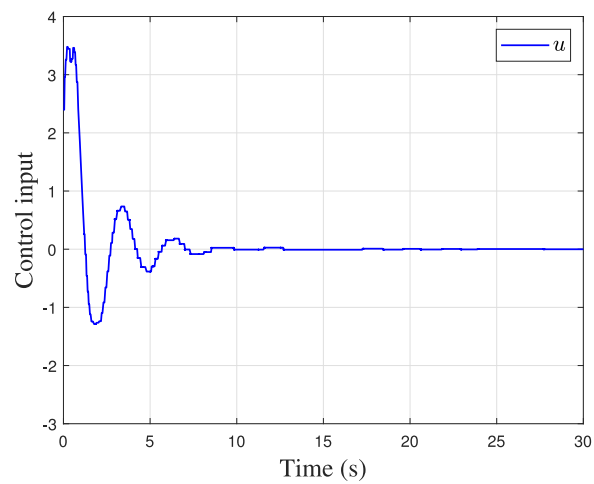
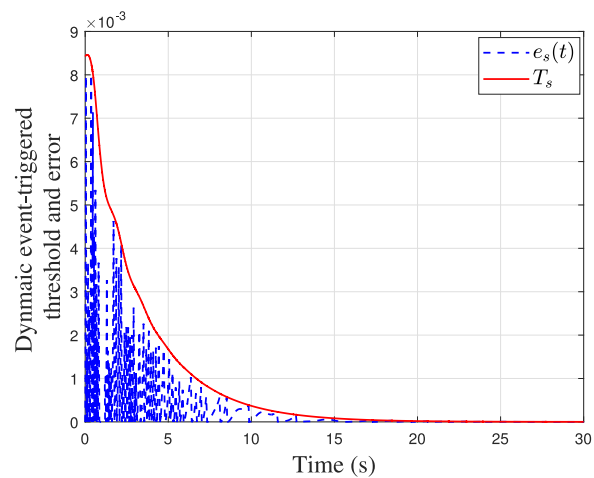**Fig. 13.** Approximation DETOC input $\hat{u}_c$ of system (69).
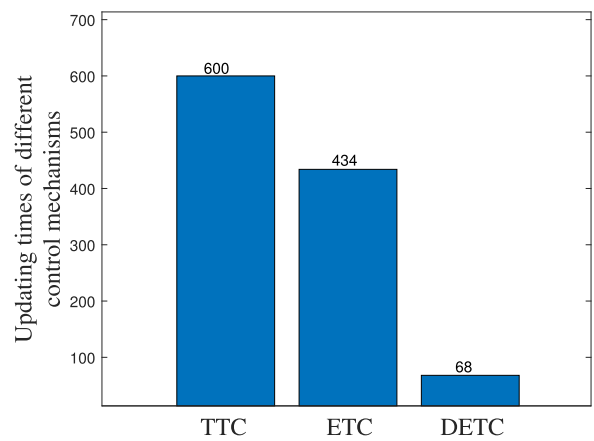

Fig. 14. Dynamic event-triggered threshold and error.

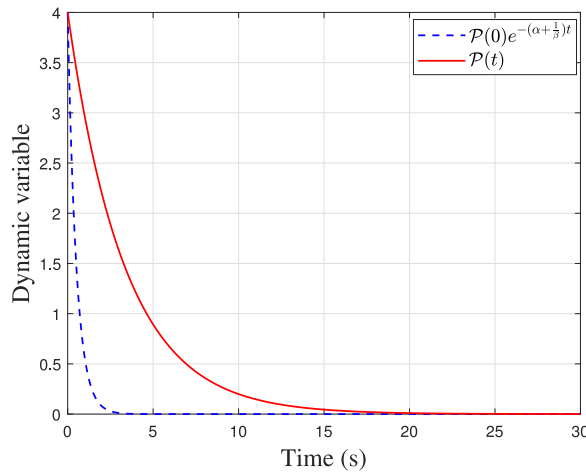**Fig. 15.** Updating times of different control mechanisms.

15

162

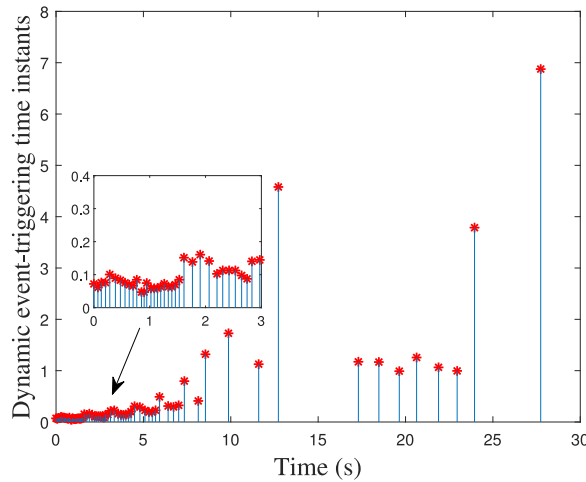**Fig. 16.** Dynamic variable and its low bound function.

Fig. 17. Dynamic event-triggering time instants.

Then, in the sliding mode function, $\mathcal{G}(\mathcal{Z}) = 4\mathcal{Z}_2$, $\mathfrak{M}(\mathcal{Z}) = [0, 4]$. Similar to Example 1, in the discontinuous control law (11), $-K\mathrm{sgn}(\mathcal{I}_b^\mathsf{T}(\mathcal{Z})\mathfrak{M}^\mathsf{T}(\mathcal{Z})\mathfrak{S})$ is replaced by $-K\tanh(\mathcal{I}_b^\mathsf{T}(\mathcal{Z})\mathfrak{M}^\mathsf{T}(\mathcal{Z})\mathfrak{S}/\xi)$ with $\xi = 0.0001$ and $K = 4$ to reduce the chattering phenomenon. The composite control input (8), composed of (11) and (48), is employed to control the uncertain system (69) for simulation. As depicted in Fig. 18, one can observe that the states of closed-loop system converge to an SRZ as time increases. The curves of the DETOC and the composite control inputs are presented in Fig. 19. The simulation results illustrate the effectiveness of the developed ADP-based DETNOC method.

5. Conclusion

The paper presents a DETNOC method for addressing the optimal control problem for uncertain nonlinear systems with unknown dead-zone and disturbance. By combining ADP and ISMC techniques, the proposed method involves the design of a composite control law comprising discontinuous and DETOC laws. The discontinuous control law effectively eliminates the effects of dead-zone, disturbance, and the matched component uncertainties, while obtaining the SMD. Subsequently, the ADP-based approximate DETOC law guarantees the stability of the SMD. The Lyapunov stability theorem is employed to prove the UUB of the closed-loop system. Simulation results are presented to demonstrate the effectiveness of the proposed DETNOC method.

Fixed-time and predefined-time control methods have attracted widespread attention in the control field, as they can guarantee the closed-loop systems stability within fixed-time and predefined-time, respectively. In future work, fixed-time optimal and predefined-time optimal control will be further investigated for nonlinear systems with input constraints.
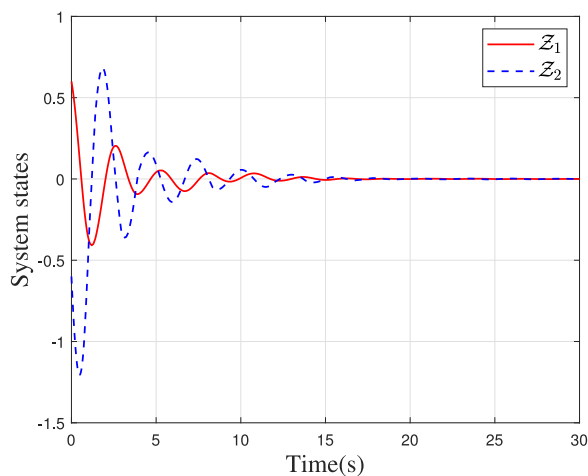
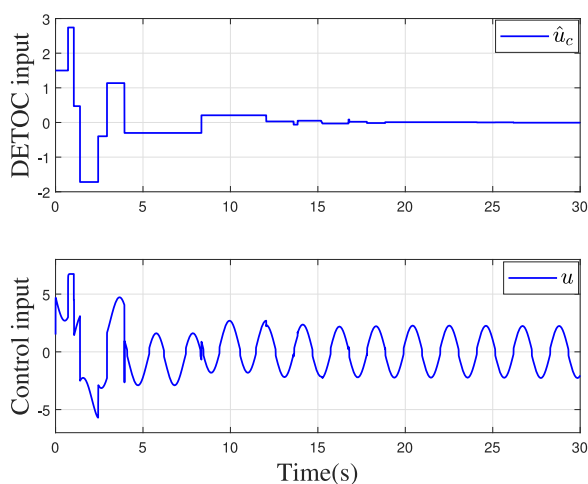**Fig. 18.** System states of closed-loop system (68).

Fig. 19. Approximation DETOC and the composite control inputs.

CRediT authorship contribution statement

Shunchao Zhang: Writing – review & editing, Writing – original draft. **Jiawei Zhuang:** Methodology, Investigation. **Yongwei Zhang:** Writing – review & editing, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

References

[1] Wang J, Yan Y, Liu J, Chen CP, Liu Z, Zhang C. NN event-triggered finite-time consensus control for uncertain nonlinear multi-agent systems with dead-zone input and actuator failures. ISA Trans 2023;137:59–73.

[2] Huang Q, Zhang E, Dai X, Wu Q, Su S. Sliding-mode fault-tolerant control of the six-rotor UAV with dead-zone-input under event-triggered mechanism. ISA Trans 2024;145:19–31.

[3] Liu D, Wei Q, Wang D, Yang X, Li H. Adaptive dynamic programming with applications in optimal control. Springer; 2017.

[4] Liu D, Xue S, Zhao B, Luo B, Wei Q. Adaptive dynamic programming for control: A survey and recent advances. IEEE Trans Syst Man Cybern: Syst 2020;51(1):142–60.

[5] Abdulraheem KK, Korolev SA. Robust optimal-integral sliding mode control for a pressurized water nuclear reactor in load following mode of operation. Ann Nucl Energy 2021;158:108288.

[6] Das M, Mahanta C. Optimal second order sliding mode control for linear uncertain systems. ISA Trans 2014;53(6):1807–15.

[7] Surjagade PV, Tiwari AP, Shimjith SR. Robust optimal integral sliding mode controller for total power control of large PHWRs. IEEE Trans Nucl Sci 2018;65(7):1331–44.

[8] Lin D, Xue S, Liu D, Liang M, Wang Y. Adaptive dynamic programming-based hierarchical decision-making of non-affine systems. Neural Netw 2023;167:331–41.

[9] Xia H, Zhao B, Guo P. Synergetic learning structure-based neuro-optimal fault tolerant control for unknown nonlinear systems. Neural Netw 2022;155:204–14.

[10] Zhang S, Zhao B, Liu D, Zhang Y. Observer-based event-triggered control for zero-sum games of input constrained multi-player nonlinear systems. Neural Netw 2021;144:101–12.

[11] Wang D, Liu D, Li H, Ma H. Neural-network-based robust optimal control design for a class of uncertain nonlinear systems via adaptive dynamic programming. Inform Sci 2014;282:167–79.

[12] Zhang H, Su H, Zhang K, Luo Y. Event-triggered adaptive dynamic programming for non-zero-sum games of unknown nonlinear systems via generalized fuzzy hyperbolic models. IEEE Trans Fuzzy Syst 2019;27(11):2202–14.

[13] Zhao H, Wang H, Niu B, Zhao X, Alharbi KH. Event-triggered fault-tolerant control for input-constrained nonlinear systems with mismatched disturbances via adaptive dynamic programming. Neural Netw 2023;164:508–20.

[14] Chen Z, Chen S-Z, Chen K, Zhang Y. Constrained decoupling adaptive dynamic programming for a partially uncontrollable time-delayed model of energy systems. Inform Sci 2022;608:1352–74.

[15] Chen Z, Chen S-Z, Zhang Y, Deng Q, Zeng X. Online and hard constrained adaptive dynamic programming algorithm for energy storage control in smart buildings. Optim Control Appl Methods 2023;44(3):1074–91.

[16] Zhao B, Wang D, Shi G, Liu D, Li Y. Decentralized control for large-scale nonlinear systems with unknown mismatched interconnections via policy iteration. IEEE Trans Syst Man Cybern: Syst 2017;48(10):1725–35.

[17] Lin M, Zhao B, Liu D. Policy gradient adaptive critic designs for model-free optimal tracking control with experience replay. IEEE Trans Syst Man Cybern: Syst 2021;52(6):3692–703.

[18] Dong B, An T, Zhu X, Li Y, Liu K. Zero-sum game-based neuro-optimal control of modular robot manipulators with uncertain disturbance using critic only policy iteration. Neurocomputing 2021;450:183–96.

[19] Zhao B, Zhang Y, Liu D. Adaptive dynamic programming-based cooperative motion/force control for modular reconfigurable manipulators: A joint task assignment approach. IEEE Trans Neural Netw Learn Syst 2022. http://dx.doi.org/10.1109/TNNLS.2022.3171828.

[20] Xue S, Luo B, Liu D. Event-triggered adaptive dynamic programming for zero-sum game of partially unknown continuous-time nonlinear systems. IEEE Trans Syst Man Cybern: Syst 2018;50(9):3189–99.

[21] Yao D, Li H, Shi Y. Event-based average consensus of disturbed MASs via fully distributed sliding mode control. IEEE Trans Autom Control 2024;69(3):2015–22.

[22] Yao D, Li H, Shi Y. SMO-based distributed tracking control for linear MASs with event-triggering communication. IEEE Trans Control Netw Syst 2024;11(1):18–29.

[23] Vamvoudakis KG. Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems. IEEE/CAA J Autom Sin 2014;1(3):282–93.

[24] Wang D, Mu C, Yang X, Liu D. Event-based constrained robust control of affine systems incorporating an adaptive critic mechanism. IEEE Trans Syst Man Cybern: Syst 2017;47(7):1602–12.

[25] Xue S, Luo B, Liu D, Gao Y. Event-triggered integral reinforcement learning for nonzero-sum games with asymmetric input saturation. Neural Netw 2022;152:212–23.

[26] Mu C, Wang K, Qiu T. Dynamic event-triggering neural learning control for partially unknown nonlinear systems. IEEE Trans Cybern 2020;52(4):2200–13.

[27] Zou H, Zhang G. Dynamic event-triggered-based single-network ADP optimal tracking control for the unknown nonlinear system with constrained input. Neurocomputing 2023;518:294–307.

[28] Yang X, Xu M, Wei Q. Dynamic event-sampled control of interconnected nonlinear systems using reinforcement learning. IEEE Trans Neural Netw Learn Syst 2022. http://dx.doi.org/10.1109/TNNLS.2022.3178017.

[29] Tan W, Yu W, Wang H. Dynamic event-triggered integral sliding mode adaptive optimal tracking control for uncertain nonlinear systems. Symmetry 2022;14(6):1264.

[30] Zhang Y, Zhao B, Liu D, Zhang S. Distributed fault tolerant consensus control of nonlinear multiagent systems via adaptive dynamic programming. IEEE Trans Neural Netw Learn Syst 2024;35(7):9041–53.

[31] Zhang H, Qu Q, Xiao G, Cui Y. Optimal guaranteed cost sliding mode control for constrained-input nonlinear systems with matched and unmatched disturbances. IEEE Trans Neural Netw Learn Syst 2018;29(6):2112–26.

[32] Zhao J, Na J, Gao G. Robust tracking control of uncertain nonlinear systems with adaptive dynamic programming. Neurocomputing 2022;471:21–30.

[33] Zhang Q, Zhao D, Wang D. Event-based robust control for uncertain nonlinear systems using adaptive dynamic programming. IEEE Trans Neural Netw Learn Syst 2016;29(1):37–50.

[34] Zhang S, Zhao B, Liu D, Zhang Y. Event-triggered decentralized integral sliding mode control for input-constrained nonlinear large-scale systems with actuator failures. IEEE Trans Syst Man Cybern: Syst 2024;54(3):1914–25.

[35] Yao D, Wu Y, Ren H, Li H, Shi Y. Event-based adaptive sliding-mode containment control for multiple networked mechanical systems with parameter uncertainties. IEEE Trans Autom Sci Eng 2024. http://dx.doi.org/10.1109/TASE.2024.3349634.

[36] Hornik K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Netw 1990;3(5):551–60.

[37] Cui L, Xie X, Guo H, Luo Y. Dynamic event-triggered distributed guaranteed cost FTC scheme for nonlinear interconnected systems via ADP approach. Appl Math Comput 2022;425:127082.

[38] Yang X, Wei Q. Adaptive critic learning for constrained optimal event-triggered control with discounted cost. IEEE Trans Neural Netw Learn Syst 2020;32(1):91–104.

165

RESEARCH ARTICLE

WILEY

Integral sliding mode-based event-triggered nearly optimal tracking control for uncertain nonlinear systems

Shunchao Zhang[1] | **Yonghua Wang[2]** | **Dacai Liu[1]** | **Jiawei Zhuang[1]** | **Yongwei Zhang[3]**

[1]School of Internet Finance and Information Engineering, Guangdong University of Finance, Guangzhou, China

[2]School of Automation, Guangdong University of Technology, Guangzhou, China

[3]College of Mathematics and Informatics, South China Agricultural University, Guangzhou, China

Correspondence
Yongwei Zhang. College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510642, China.
Email:
Yongwei_Zhang@mail2.gdut.edu.cn

Summary
In this paper, an event-triggered nearly optimal tracking control method is investigated for a class of uncertain nonlinear systems by integrating adaptive dynamic programming (ADP) and integral sliding mode (ISM) control techniques. An ISM-based discontinuous control law with a neural network (NN) adaptive term is designed to eliminate the influence of the uncertainties and obtain the sliding mode dynamics which is equivalent to the tracking error dynamics without uncertainties, and relax the known upper-bounded condition of uncertainties. In order to guarantee the stability of tracking error system and the considerable optimality, under the ADP technique, a critic NN is applied to approximate the optimal value function for solving the event-triggered Hamilton-Jacobi-Bellman equation and the event-triggered nearly optimal feedback control is obtained. The feedback control law is updated and transmitted to plant only when events occur, thus both the communication and the computational resources can be saved. Furthermore, the stability of tracking error is proven thanks to Lyapunov's direct method. Finally, we provide two simulation examples to validate the developed control scheme.

KEYWORDS
adaptive dynamic programming, event-triggered mechanism, integral sliding mode control, neural networks, uncertain systems

1 | INTRODUCTION

With the existing of model uncertainties and disturbance, there will always be a deviation between practical control systems and their nominal systems employed for controllers design.[1,2] It is necessary to investigate a robust control method for guaranteeing the stability and desired performance of systems in the presence of deviation. During the past few years, many advanced control methods, such as adaptive control,[3,4] robust control,[5] H_∞ control,[6] and sliding mode control (SMC),[7,8] have been used to design robust controller. Among these methods, as an effective technique, SMC has attracted much attention due to the insensitive of parameter changes and the ability of fast respond.[9-12] Liu et al.[11] proposed an adaptive SMC method for nonlinear systems with parametric uncertainties and external disturbances by combining immersion and invariance adaptive scheme. Ding et al.[12] developed a discontinuous and a quasi-continuous second-SMC methods for uncertain nonlinear systems, and the chattering phenomenon was reduced in the last method to some extent.

The traditional SMC composes two parts, that is, the initial reaching and the sliding motion phases, and the robustness is only occurred during the sliding motion. In order to avoid the reaching phase and improve the robustness, many integral SMC (ISMC) methods[13-15] have been developed in recent years. Cao et al.[13] developed an ISMC method for nonlinear systems with uncertainties by designing a nonlinear integral-type sliding mode surface (SMS). In these methods, the system trajectory starts on the sliding manifold for any initial system state by designing an integral sliding mode function.

Although the aforementioned methods have been widely employed to design robust controllers, they are required to not only stabilize the systems with uncertainties, but also satisfy the considerable optimality in practical applications.[16] By integrating ISMC technique and optimal control (OC) approaches, many approaches designed a composite control law to achieve the objective for linear systems.[17-22] Surjagade et al.[21] developed an optimal ISMC method for a pressurized heavy water reactor system, this method combined the optimal control law with ISMC law to guarantee the stability of the closed-loop system when the existing of uncertainties and external disturbances. Das and Mahanta[22] proposed an optimal second-order SMC method for uncertain linear systems by combining the terminal SMS and the integral SMS. On the whole, a discontinuous control law is employed to eliminate the effect of uncertainties or disturbances and obtain sliding mode dynamics, and the OC law from solving algebraic Riccati equation is obtained to stabilize the linear sliding mode dynamics. However, for the nonlinear systems, these methods are not easy to implement since they are difficult to obtain the OC law for nonlinear sliding mode dynamics by solving the Hamilton-Jacobi-Bellman (HJB) equation, which is difficult or even impossible to obtain the analysis solution.

Fortunately, as two effective techniques, adaptive dynamic programming (ADP) and reinforcement learning (RL) are viewed as synonyms which overcome this difficulty by computing forward-in-time.[23-25] Many significant ADP-based control methods have been reported to solve the OC problem for nonlinear systems.[26,27] Vamvoudakis and Lewis[28] developed an actor-critic (AC) strategy to solve the OC problem for nonlinear systems. Vrabie and Lewis[29] developed an integral RL-based method to obtain the solution of HJB equation and solve the OC problem of partially known nonlinear systems. It is easy to find that the aforementioned results are achieved for optimal regulation problems. However, in many practical systems, the objective of controller design is to guarantee the system state tracking an user-defined reference trajectory rather than regulate the system state approaching the origin.[30,31] Hence, it is significant to track the user-defined reference trajectory with optimal performance and is also one of the common problem in ADP- or RL-based control community. For discrete-time (DT) nonlinear systems, the optimal tracking control problem was converted into an OC problem for tracking error dynamics and a neuro-optimal tracking control scheme was developed for nonlinear systems via the ADP technique.[32] Wei et al.[33] developed a data-based optimal tracking control method for DT nonlinear systems and to apply the coal gasification system. For continuous-time (CT) nonlinear systems, Modares and Lewis[34] developed an integral RL-based tracking control method for CT nonlinear systems. Zhao et al.[35] developed an ADP-based robust tracking control method for CT nonlinear systems with uncertainties, where the tracking control problem was transformed into an OC problem for the augmented system. Wang et al.[36] developed an adaptive-critic-based robust tracking control method for uncertain nonlinear systems, and this method was applying to a spring-mass-damper system.

However, these methods adopted time-triggered mechanism, the updating of the control law with a fixed period may increase the energy consumption, and waste computational and communication resources. In order to save the computational and communication resources on the basis of satisfying some control performance, many researchers have introduced the event-triggered mechanism to ADP, and developed many ADP-based ETC methods,[37-40] where the event was defined as the event-triggering error exceeded the designed event-triggering condition and the control law was updated only when the occurrence of the events. For example, Vamvoudakis[39] developed an event-triggered OC (ETOC) method for CT nonlinear systems, this method was implemented based on AC structure, a critic and an actor neural networks (NNs) were employed to approximate the cost function and the ETOC law, respectively. Wang et al.[40] developed an event-triggered robust control method for uncertain CT nonlinear systems, where the robust control problem was transformed into an ETOC problem by designed a modified value function. For the tracking control problem, Zhang et al.[41] developed an event-triggered tracking control (ETTC) scheme for CT nonlinear systems, the designed control law composited with a feedforward and a feedback control laws which were employed to track the reference trajectory and stabilize tracking error dynamics, respectively.

Based on the above-mentioned literature, these methods involved precise system dynamics only, research in ADP-based ETTC has not been fully taken into account. However, the uncertainties is widely existed between actual plant and its nominal system. On the other hand, among existing methods[35,40-42] required the upper-bounded function of the uncertainties which is difficult to be obtained. Inspired by the aforementioned literature, this paper focus on developing

an event-triggered nearly optimal tracking control (ETNOTC) method for uncertain nonlinear systems by integrating ADP and ISMC. The main contributions of this scheme is summarized in the following three aspects.

1. In contrast to existing methods[17-22] combined the ISMC and OC approaches to develop robust control methods for uncertain linear systems, this paper develops an ISMC and ADP-based ETNOTC method for uncertain nonlinear systems with the considerable optimality.
2. Unlike works[35,36,40-42] developing robust control methods for nonlinear systems required the known upper-bound of uncertainties, the developed method adopts the ISMC technique with a neural network-based adaptive term to eliminate the effect of uncertainties with unknown upper bound.
3. Different from works[17-22,43,44] which adopted time-triggered mechanism to design the nearly optimal continuous control law, this paper develops an ETC method to save the computational and communication resources.

The reminder of this paper is organized as follows. Section 2 presents the problem statement. Section 3 introduces the composite control law design in detail. In Section 4, a numerical and a practical examples are employed to verify the effectiveness of ETNOTC method. In Section 5, conclusion is given.

2 | PROBLEM STATEMENT

Consider the uncertain nonlinear system described by

$$\dot{s}(t) = \mathcal{F}(s(t)) + \mathcal{G}(s(t))u(t) - \Delta\mathcal{F}(s(t)), \tag{1}$$

where $s \in \mathbb{R}^n$ is the system state, $u \in \mathbb{R}^m$ is the control input, $\Delta\mathcal{F}(s) = \mathcal{G}(s)d(s) \in \mathbb{R}^n$ is the uncertainties, $\mathcal{F}(s) \in \mathbb{R}^n$ and $\mathcal{G}(s) \in \mathbb{R}^{n\times m}$ are continuously differentiable matrix functions, and $\mathcal{G}(s)$ is invertible.

Assumption 1. The system (1) is controllable, and the system dynamic $\mathcal{F}(s) + \mathcal{G}(s)u$ is Lipschitz continuous on a compact set Ω and $\mathcal{F}(0) = 0$.[1-7]

For the tracking control, the system state is expected to track an user-defined reference trajectory which is give by

$$\dot{x}_d(t) = \phi(x_d), \tag{2}$$

where $x_d \in \mathbb{R}^n$ is the reference state, and $\phi(x_d) \in \mathbb{R}^n$ is an Lipschitz continuous function. According to (1) and (2), the tracking error is defined as $\delta(t) = s(t) - x_d(t)$, and the tracking error system is given by

$$\begin{aligned} \dot{\delta}(t) &= \dot{s}(t) - \dot{x}_d(t) \\ &= \mathcal{F}(s) + \mathcal{G}(s)u(t) - \Delta\mathcal{F}(s) - \phi(x_d). \end{aligned} \tag{3}$$

In the following, a composite control law u is designed to guarantee the system state tracking the reference trajectory and minimize a given value function as far as possible.

3 | COMPOSITE CONTROL LAW DESIGN VIA ISMC AND ADP

For the tracking error system (3), an ETNOTC method which integrates ADP and ISMC techniques is developed to design a composite control law as

$$u = u_c + u_d + w, \tag{4}$$

where $u_c \in \mathbb{R}^m$ is the discontinuous component to eliminate the influence of the uncertainties, $u_d \in \mathbb{R}^m$ is the continuous feedforward control component to track the trajectory, $w \in \mathbb{R}^m$ is the continuous ADP-based feedback control component to guarantee the tracking error stabilization.

3.1 | Discontinuous control law design via ISMC

The integral sliding mode (ISM) function is designed as

$$S(\delta(t), t) = A\delta - A\delta_0 - \int_0^t A(\mathcal{F}(s(\tau)) + \mathcal{G}(s(\tau))W(\tau) - \phi(x_d))d\tau, \tag{5}$$

where $\delta_0 = s(0)$, $W = u_d + w$, $A \in \mathbb{R}^{m \times n}$ is a design matrix. It is worth pointing out that the ISM function satisfies $S(\delta_0, 0) = 0$ for any initial state δ_0, the state of tracking error system (3) starts on the ISM surface, thus the reaching phase can be removed.

Differentiating $S(\delta(t), t)$ with respect to t, it yields

$$\begin{aligned}
\dot{S}(\delta, t) &= A\dot{\delta} - A(\mathcal{F}(s) + \mathcal{G}W - \phi(x_d)) \\
&= A(\mathcal{F}(s) + \mathcal{G}(s)(u - d(s)) - \phi(x_d)) - A(\mathcal{F}(s) + \mathcal{G}(s)W - \phi(x_d)) \\
&= A\mathcal{G}(s)(u_c - d(s)).
\end{aligned}$$

According to SMC theory, let $\dot{S}(s, t) = 0$, the equivalent control law u_{ceq} is derived as

$$u_{ceq} = d(s). \tag{6}$$

Substituting (6) into (3), we get the sliding mode dynamics as

$$\dot{\delta}(t) = \mathcal{F}(s) + \mathcal{G}(s)W - \phi(x_d). \tag{7}$$

However, the u_{ceq} cannot be obtained since the unknown $d(s)$. To keep the integral sliding mode function as zero, that is, $S(\delta, t) = 0$, the discontinuous control law u_c is designed as

$$u_c = -\mathcal{K}\text{sgn}(\Xi), \tag{8}$$

where $\Xi = \mathcal{G}^\mathsf{T}(s)A^\mathsf{T}S$, $\text{sgn}(\cdot)$ is the sign function, $\mathcal{K} > \overline{d}$ is a sliding mode gain, $\overline{d}$ is the norm-bound of $d(s)$. In order to relax the requirement of the known $\overline{d}$, a radial basis function (RBF) NN-based adaptive team is designed to estimate the uncertainties as

$$d(s) = \theta^{*\mathsf{T}}h(s) + \epsilon,$$

where $\theta^* \in \mathbb{R}^{l_d \times m}$ is the ideal weight, l_d is the number of neurons, $h(s) \in \mathbb{R}^{l_d}$ is a RBF, and ϵ is the approximation error. Denote $\hat{\theta} \in \mathbb{R}^{l_d \times m}$ be the estimation of θ^*, we have

$$\hat{d}(s) = \hat{\theta}^\mathsf{T}h(s).$$

Furthermore, the discontinuous control law u_c in (8) is changed as

$$u_c = -K\text{sgn}(\Xi) + \hat{d}(s), \tag{9}$$

where K is the improved sliding mode gain satisfying $K > \epsilon_b$ and ϵ_b is the norm-bound of ϵ.

Theorem 1. *For the nonlinear system (1), the designed integral sliding mode function (5), and Assumption 1, the discontinuous control law u_c (9) can maintain the system state trajectory on the ISM surface $S = 0$ with the adaptive law*

$$\dot{\hat{\theta}} = -\frac{1}{\gamma}h(s)S^\mathsf{T}A\mathcal{G}(s), \tag{10}$$

where $\gamma > 0$ is the updating rate.

169

Proof. Consider the Lyapunov function candidate given as

$$\Sigma_1(t) = \frac{1}{2}S^\mathsf{T}S + \frac{\gamma}{2}\mathrm{tr}\{\tilde{\theta}^\mathsf{T}\tilde{\theta}\}, \tag{11}$$

where $\tilde{\theta} = \hat{\theta} - \theta^*$. The time derivative of the Σ_1 is deduced as

$$\begin{aligned}
\dot{\Sigma}_1(t) &= S^\mathsf{T}A(\mathcal{F}(s) + \mathcal{G}(s)(u - d(s)) - \phi(x_d)) - S^\mathsf{T}A(\mathcal{F}(s) + \mathcal{G}(s)W - \phi(x_d)) + \gamma\mathrm{tr}\{\tilde{\theta}^\mathsf{T}\dot{\tilde{\theta}}\}\\
&= S^\mathsf{T}A(\mathcal{F}(s) + \mathcal{G}(s)(u_c + W - d(s))) - S^\mathsf{T}A(s)(\mathcal{F}(s) + \mathcal{G}(s)W) + \gamma\mathrm{tr}\{\tilde{\theta}^\mathsf{T}\dot{\tilde{\theta}}\}\\
&= S^\mathsf{T}A\Big(\mathcal{G}(s)\Big(-K\mathrm{sgn}(\Xi) + \hat{d}(s) - d(s)\Big)\Big) + \gamma\mathrm{tr}\{\tilde{\theta}^\mathsf{T}\dot{\tilde{\theta}}\}\\
&= -KS^\mathsf{T}A\mathcal{G}(s)\mathrm{sgn}(\Xi) + \gamma\mathrm{tr}\{\tilde{\theta}^\mathsf{T}\dot{\tilde{\theta}}\} + S^\mathsf{T}A(s)\mathcal{G}(s)\Big(\hat{d}(s) - d(s)\Big)\\
&= -KS^\mathsf{T}A\mathcal{G}(s)\mathrm{sgn}(\Xi) + \gamma\mathrm{tr}\{\tilde{\theta}^\mathsf{T}\dot{\tilde{\theta}}\} + S^\mathsf{T}A\mathcal{G}(s)\Big(\hat{\theta}^\mathsf{T}h(s) - \theta^{*\mathsf{T}}h(s) - \epsilon\Big). \tag{12}
\end{aligned}$$

Considering the adaptive law (10) and $\dot{\tilde{\theta}} = \dot{\hat{\theta}}$, (12) becomes

$$\begin{aligned}
\dot{\Sigma}_1(t) &= -KS^\mathsf{T}A\mathcal{G}(s)\mathrm{sgn}(\Xi) + S^\mathsf{T}A\mathcal{G}(s)\Big(\tilde{\theta}^\mathsf{T}h(s) - \epsilon\Big) - \mathrm{tr}\{\tilde{\theta}^\mathsf{T}h(s)S^\mathsf{T}A\mathcal{G}(s)\}\\
&= -KS^\mathsf{T}A\mathcal{G}(s)\mathrm{sgn}(\Xi) - S^\mathsf{T}A\mathcal{G}(s)\epsilon\\
&\leq -K\|S^\mathsf{T}A\mathcal{G}(s)\|_1 - \|S^\mathsf{T}A\mathcal{G}(s)\epsilon\|\\
&\leq -(K - \epsilon_b)\|S^\mathsf{T}A\mathcal{G}(s)\|_1. \tag{13}
\end{aligned}$$

Therefore, if $K > \epsilon_b$ holds, the system state trajectory is maintained on sliding mode surface. ∎

Remark 1. It is noticed that the improved sliding mode gain K is different from the gain $\mathcal{K}$, where K depends on the norm-bound of approximation error ϵ instead of the norm-bound of the uncertain term $d(s)$. In practical applications, it is difficult to obtain the norm-bound of of the uncertain term. The approximation error ϵ can be guaranteed to be arbitrary small by selecting sufficient number of neurons.[45,46] Although the selection of gain K is challenging, there is no guiding method to select an optimal sliding mode gain, and it can be selected based on repeated "trial and error".

Remark 2. From (13) if the sliding mode gain K is chosen as $K > \epsilon_b$, we have $\dot{\Sigma}_1(t) < 0$, the Lyapunov candidate function (11) will decrease gradually and the sliding mode surface S will converge to zero.

Remark 3. In this paper, a tracking control problem is investigated for nonlinear uncertain systems. Indeed, it is well known that ISMC is an effective technique to deal with the uncertainties of nonlinear systems. However, the stability of the closed-loop system is only the basis, and the control performance and cost should be further considered in the process of controller design. Recently, as a powerful method, ADP has been widely employed to solve optimal problems. As a result, this paper combines ISMC and ADP to develop an ETNOTC method.

Remark 4. For the ETNOTC problem, the main technical difficulty lies in that (1) a neural network adaptive term is designed to relax the unknown bound of uncertainties; (2) a discontinuous control law is developed to eliminate the effect of the uncertainties; (3) an ADP-based event-triggered feedback control law is designed to satisfy the considerable optimality.

3.2 | Continuous control law design via ADP

Assume that the desired trajectory satisfies

$$\dot{x}_d(t) = \mathcal{F}(x_d) + \mathcal{G}(x_d)u_d, \tag{14}$$

where u_d is the feedforward control law. Combining (2) and (14), we have

$$u_d = \mathcal{G}^+(x_d)(\phi(x_d) - \mathcal{F}(x_d)), \tag{15}$$

where $\mathcal{G}^+(x_d)$ denotes the generalized inverse of $\mathcal{G}(x_d)$. Substituting (15) into (7), the tracking error dynamics is given by

$$\begin{aligned}
\dot{\delta}(t) &= \dot{x}(t) - \dot{x}_d(t) \\
&= \mathcal{F}(s) + \mathcal{G}(s)(u_d + w) - \phi(x_d) \\
&= \mathcal{F}(s) + \mathcal{G}(s)\mathcal{G}^+(x_d)(\phi(x_d) - \mathcal{F}(x_d)) + \mathcal{G}(s)w - \phi(x_d).
\end{aligned}$$

Letting $\mathcal{F}_\delta = \mathcal{F}(s) + \mathcal{G}(s)\mathcal{G}^+(x_d)(\phi(x_d) - \mathcal{F}(x_d)) - \phi(x_d)$, we have

$$\dot{\delta}(t) = \mathcal{F}_\delta + \mathcal{G}(s)w. \tag{16}$$

Then, under the event-triggered mechanism, an ADP-based control method is developed to design the feedback control law w. The value function of (16) is defined as

$$V(\delta) = \int_t^\infty \left(\delta^\mathsf{T}(\tau)Q_\delta\delta(\tau) + w^\mathsf{T}(\tau)Rw(\tau) \right)d\tau, \tag{17}$$

where $Q_\delta \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ are symmetric positive definite matrices. Based on (17), we have

$$0 = \delta^\mathsf{T}Q_\delta\delta + w^\mathsf{T}Rw + \nabla V^\mathsf{T}(\delta)(\mathcal{F}_\delta + \mathcal{G}(s)w)$$

with $V(0) = 0$, where $\nabla V(\delta) \triangleq \partial V(\delta)/\partial\delta$. The Hamiltonian of system (16) is given by

$$H(\nabla V(\delta), \delta, w) = \delta^\mathsf{T}Q_\delta\delta + w^\mathsf{T}Rw + \nabla V^\mathsf{T}(\delta)(\mathcal{F}_\delta + \mathcal{G}(s)w).$$

The optimal value function $V^*(\delta)$ satisfy the following HJB equation

$$0 = \min_w H(\nabla V^*(\delta), \delta, w), \tag{18}$$

where $\nabla V^*(\delta) \triangleq \partial V^*(\delta)/\partial\delta$. We drive from (18) that

$$\left.\frac{\partial H(\delta, \nabla V^*(\delta), w)}{\partial w}\right|_{w=w^*} = 0,$$

where w^* is the optimal tracking control law and given by

$$w^*(\delta) = -\frac{1}{2}R^{-1}\mathcal{G}^\mathsf{T}(\delta)\nabla V^*(\delta). \tag{19}$$

Substituting (19) into (18), we further obtain

$$\begin{aligned}
H(\nabla V_i^*(\delta), \delta, w^*) &= \delta^\mathsf{T}Q_\delta\delta + w^{*\mathsf{T}}Rw^* + \nabla V^{*\mathsf{T}}(\delta)(\mathcal{F}_\delta + \mathcal{G}(s)w^*) \\
&= 0.
\end{aligned} \tag{20}$$

From (20), it is a time-triggered HJB equation whose solution often involves heavy computational burden and the waste of communication resource by using ADP-based time-triggered mechanism. Hence, we developed an ADP-based ETC method to obviate this shortcoming. Under the ETC framework, the sampled state is denoted as

$$\hat{s}_\kappa = s(t_s), \quad \forall t \in [t_\kappa, t_{\kappa+1}),$$

where t_κ represents the κth sampling instant, $\kappa \in \mathbb{N}$. The corresponding tracking error is given by

$$\hat{\delta}_\kappa = \hat{s}_\kappa - x_d(t_\kappa), \quad \forall t \in [t_\kappa, t_{\kappa+1}).$$

Then, introduce an triggering error function as

$$\mathbf{E}_\kappa(t) = \hat{\delta}_\kappa - \delta(t), \quad \forall t \in [t_\kappa, t_{\kappa+1}). \tag{21}$$

According to (21), the ETC law is expressed by

$$w(\hat{\delta}_\kappa) = w(\mathbf{E}_\kappa(t) + \delta(t)), \tag{22}$$

Based on (22), the system (16) becomes

$$\dot{\delta}(t) = \mathcal{F}_\delta + \mathcal{G}(s)w(\hat{\delta}_\kappa). \tag{23}$$

Furthermore, the event-triggered optimal tracking control (ETOTC) can be obtained from (23) as

$$w^*(\hat{\delta}_s) = -\frac{1}{2}R^{-1}\mathcal{G}^\mathsf{T}(\hat{\delta}_\kappa)\nabla V^*(\hat{\delta}_\kappa) \tag{24}$$

for all $t \in [t_\kappa, t_{\kappa+1})$, where $\nabla V^*(\hat{\delta}_\kappa) \triangleq \partial V^*(\hat{\delta}_\kappa)/\partial \hat{\delta}_\kappa$. By replacing w in (18) with $w^*(\hat{\delta}_s)$, the event-triggered version of HJB equation at $t = t_\kappa$ is written as

$$H\big(\nabla V^*(\delta), \delta, w^*(\hat{\delta}_\kappa)\big) = \delta^\mathsf{T}Q_\delta\delta + w^{*\mathsf{T}}(\hat{\delta}_\kappa)Rw^*(\hat{\delta}_\kappa) + \nabla V^{*\mathsf{T}}(\delta)\big(\mathcal{F}_\delta + \mathcal{G}(s)w^*(\hat{\delta}_\kappa)\big).$$

Assumption 2. $w^*(\delta)$ is Lipschitz continuous, that is, $\|w^*(\delta(t)) - w^*(\hat{\delta}_\kappa)\| \leq \mathcal{L}_w\|\mathbf{E}_\kappa(t)\|$, where $\mathcal{L}_{wi} > 0$ is a constant.[38-42]

Theorem 2. *For the tracking error system given by* (3), *the sliding mode dynamics* (16), *Assumptions* 1 *and* 2, *the composite control law* (4) *with* (9), (15) *and* (24), *if the triggering condition is designed as*

$$\|\mathbf{E}_\kappa\|^2 \leq \frac{(1-\beta^2)\delta^\mathsf{T}Q_\delta\delta + \|r\|^2\|w^*(\hat{\delta}_\kappa)\|^2}{\mathcal{L}_w^2} = T_\kappa^2, \tag{25}$$

where $\mathcal{L}_w$ is a positive constant, T_κ is the event-triggering threshold, the closed-loop tracking error system (3) *is guaranteed to be asymptotically stable.*

Proof. Choose a Lyapnuov function candidate as

$$\Sigma_2(t) = V^*(\delta).$$

Based on Theorem 1, by using the discontinuous control law u_c, the system state trajectory can be forced on integral sliding mode surface $S = 0$ and maintained on it. And then, applying the feedforward control law, the tracking error system is obtained as (16). Using the trajectories of system (16), we find

$$\dot{\Sigma}_2(t) = \nabla V^{*\mathsf{T}}(\delta)\big(\mathcal{F}_\delta + \mathcal{G}(s)w^*(\hat{\delta}_\kappa)\big). \tag{26}$$

Based on (19), we have

$$\nabla V^{*\mathsf{T}}(\delta)\mathcal{G}(s) = -2w^{*\mathsf{T}}(\delta)R. \tag{27}$$

From (20), it reveals that

$$\nabla V^{*\mathsf{T}}(\delta)\mathcal{F}_\delta = -\delta^\mathsf{T}Q_\delta\delta - w^{*\mathsf{T}}(\delta)Rw^*(\delta) - \nabla V^{*\mathsf{T}}(\delta)\mathcal{G}(s)w^*(\delta). \tag{28}$$

Substituting (27) and (28) into (26), we obtain

$$
\begin{aligned}
\dot{\Sigma}_2(t) &= -\delta^\mathsf{T} Q_\delta \delta - w^{*\mathsf{T}}(\delta) R w^*(\delta) - \nabla V^{*\mathsf{T}}(\delta) \mathcal{G}(s) w^*(\delta) + \nabla V^{*\mathsf{T}}(\delta) \mathcal{G}(s) w^*(\hat{\delta}_\kappa) \\
&= -\delta^\mathsf{T} Q_\delta \delta + w^{*\mathsf{T}}(\delta) R w^*(\delta) - 2 w^{*\mathsf{T}}(\delta) R w^*(\hat{\delta}_\kappa) \\
&= -\delta^\mathsf{T} Q_\delta \delta + \left(w^*(\delta) - w^*(\hat{\delta}_\kappa) \right)^\mathsf{T} R \left(w^*(\delta) - w^*(\hat{\delta}_\kappa) \right) - w^{*\mathsf{T}}(\hat{\delta}_\kappa) R w^*(\hat{\delta}_\kappa).
\end{aligned}
$$

According to Assumption 2, we have

$$
\begin{aligned}
\dot{\Sigma}_2(t) &\le -\delta^\mathsf{T} Q_\delta \delta + \mathcal{L}_w^2 \|r\|^2 \|\mathbf{E}_\kappa(t)\|^2 - \|r\|^2 \|w^*(\hat{\delta}_\kappa)\|^2 \\
&\le -\beta^2 \lambda_{\min}(Q_\delta) \|\delta\|^2 + (\beta^2 - 1) \lambda_{\min}(Q_\delta) \|\delta\|^2 + \mathcal{L}_w^2 \|r\|^2 \|\mathbf{E}_\kappa(t)\|^2 - \|r\|^2 \|w^*(\hat{\delta}_\kappa)\|^2,
\end{aligned}
$$

where $R = r^\mathsf{T} r$, $r \in R^{m \times m}$ is a square matrix. Then, if condition (25) holds, we have

$$
\dot{\Sigma}_2(t) \le -\beta^2 \lambda_{\min}(Q_\delta) \|\delta\|^2 < 0
$$

for any $\delta \ne 0$, it means the closed-loop tracking error system (3) is asymptotically stable. ∎

3.3 | Critic-only structure implementation

The optimal value function $V^*(\delta)$ can be represented via a critic NN with l_c hidden neurons as

$$
V^*(\delta) = \varphi_c^\mathsf{T} \sigma_c(\delta) + \xi_c(\delta), \tag{29}
$$

where $\varphi_c \in \mathbb{R}^{l_c}$ is the ideal weight vector, $\sigma_c(\delta) \in \mathbb{R}^{l_c}$ is the activation function, and $\xi_c(\delta)$ is the reconstruction error. Differentiating $V^*(\delta)$ in (29) with respect to δ, it yields

$$
\nabla V^*(\delta) = \nabla \sigma_c^\mathsf{T}(\delta) \varphi_e + \nabla \xi_c(\delta). \tag{30}
$$

According to (19) and (30), we have

$$
w^*(\hat{\delta}_\kappa) = -\frac{1}{2} R^{-1} \mathcal{G}^\mathsf{T}(s) \left(\nabla \sigma_c^\mathsf{T}(\hat{\delta}_\kappa) \varphi_c + \nabla \xi_c(\hat{\delta}_\kappa) \right). \tag{31}
$$

Letting $\hat{\varphi}_c \in \mathbb{R}^{l_c}$ be the estimate of φ_c, the approximate $V^*(\delta)$ is given by

$$
\hat{V}(\delta) = \hat{\varphi}_c^\mathsf{T} \sigma_c(\delta),
$$

and its partial derivative is given by

$$
\nabla \hat{V}(\delta) = \nabla \sigma_c^\mathsf{T}(\delta) \hat{\varphi}_c. \tag{32}
$$

Based on (31) and (32), the approximate ETOTC law is obtained as

$$
\hat{w}(\hat{\delta}_\kappa) = -\frac{1}{2} R^{-1} \mathcal{G}^\mathsf{T}(s) \nabla \sigma_c^\mathsf{T}(\hat{\delta}_\kappa) \hat{\varphi}_c. \tag{33}
$$

Noticing (32), the approximate Hamiltonian is defined as

$$
\begin{aligned}
H_i(\hat{\varphi}_c, \delta, \hat{w}(\hat{\delta}_\kappa)) &= \delta^\mathsf{T} Q_\delta \delta + \hat{w}^{*\mathsf{T}}(\hat{\delta}_\kappa) R \hat{w}^*(\hat{\delta}_\kappa) + \nabla V^{*\mathsf{T}}(\delta) \left(\mathcal{F}_\delta + \mathcal{G}(s) \hat{w}^*(\hat{\delta}_\kappa) \right) \\
&= \mathcal{E}_c.
\end{aligned}
$$

Obviously, we can obtain

$$\frac{\partial \mathcal{E}_c}{\partial \hat{\varphi}_c} = \nabla \sigma_c(\delta)\big(\mathcal{F}_\delta + \mathcal{G}(s)\hat{w}(\hat{\delta}_\kappa)\big) \triangleq \pi,$$

where π is a l_c-dimension column vector. To minimize the objective function $\mathcal{O}_c = (1/2)\mathcal{E}_c^\mathsf{T}\mathcal{E}_c$, $\hat{\varphi}_c$ is updated by

$$\dot{\hat{\varphi}}_c = -\alpha_c \frac{1}{(1 + \pi^\mathsf{T}\pi)^2}\left(\frac{\partial \mathcal{O}_c}{\partial \hat{\varphi}_c}\right) = -\alpha_c \frac{\pi}{(1 + \pi^\mathsf{T}\pi)^2}\mathcal{E}_c, \tag{34}$$

where $\alpha_c > 0$ is the learning rate.

Lemma 1. *Let $\tilde{\varphi} = \varphi - \hat{\varphi}$ be the weight error vector, then the weight error dynamics is derived as $\dot{\tilde{\varphi}} = -\dot{\hat{\varphi}}$. The weight error dynamics is guaranteed to be UUB with the updating law* (34).

Proof. The related proof of Lemma 1 is similar to that in References 47,48, so it is omitted here. ∎

3.4 | Stability analysis

Assumption 3. $\nabla \sigma_c(\delta), \nabla \xi_c(\delta), \mathcal{G}(s)$ *and $\tilde{\varphi}$ are norm-bounded, that is,* $\|\nabla \sigma_c(\delta)\| \leq \overline{\sigma}_c, \|\nabla \xi_c(\delta)\| \leq \overline{\xi}_c, \|\mathcal{G}(s)\| \leq \overline{g}$ *and* $\|\tilde{\varphi}\| \leq \overline{\varphi}$, *where $\overline{\sigma}_c, \overline{\xi}_c, \overline{g}$ and $\overline{\varphi}$ are positive constants.*[9,48,49]

Theorem 3. *Take the system* (16) *into account, if Assumptions* 1, 2, *and* 3 *hold and the event-triggering condition is designed as*

$$\|\mathcal{E}_\kappa(t)\|^2 \leq \frac{(1 - \rho^2)\lambda_{\min}(Q_\delta)\|\delta\|^2}{2\mathcal{L}_w^2} = \hat{T}_\kappa^2, \tag{35}$$

where $0 < \rho < 1$, and $\hat{T}_\kappa$ is the event-triggering threshold. Then, the approximate ETOTC law (33) *can guarantee the closed-loop system* (16) *to be UUB.*

Proof. Choose a Lyapunov function candidate as

$$\Sigma_3(t) = \Sigma_{31}(t) + \Sigma_{32}(t),$$

where $\Sigma_{31}(t) = V^*(\delta)$ and $\Sigma_{32}(t) = V^*(\hat{\delta}_\kappa)$. The stability analysis is presented as the following two cases.

Case 1: $\forall t \in [t_\kappa, t_{\kappa+1})$, we have

$$\dot{\Sigma}_{32}(t) = 0, \tag{36}$$

According to (20), we can derive

$$\begin{aligned}
\dot{\Sigma}_{31}(t) &= -\delta^\mathsf{T} Q_\delta \delta - w^{*\mathsf{T}}(\delta)Rw^*(\delta) - \nabla V^{*\mathsf{T}}(\delta)\mathcal{G}(s)w^*(\delta) + \nabla V^{*\mathsf{T}}(\delta)\mathcal{G}(s)\hat{w}(\hat{\delta}_\kappa) \\
&= -\delta^\mathsf{T} Q_\delta \delta - w^{*\mathsf{T}}(\delta)Rw^*(\delta) + \nabla V^{*\mathsf{T}}(\delta)\mathcal{G}(s)(\hat{w}(\hat{\delta}_\kappa) - w^*(\delta)).
\end{aligned} \tag{37}$$

Based on (19), (37) becomes

$$\begin{aligned}
\dot{\Sigma}_{31}(t) &= -\delta^\mathsf{T} Q_\delta \delta - w^{*\mathsf{T}}(\delta)Rw^*(\delta) + 2w^{*\mathsf{T}}(\delta)R(w^*(\delta) - \hat{w}(\hat{\delta}_\kappa)) \\
&= -\delta^\mathsf{T} Q_\delta \delta + w^{*\mathsf{T}}(\delta)Rw^*(\delta) - 2w^{*\mathsf{T}}(\delta)R\hat{w}(\hat{\delta}_\kappa) \\
&= -\delta^\mathsf{T} Q_\delta \delta + \big(w^*(\delta) - \hat{w}(\hat{\delta}_\kappa)\big)^\mathsf{T} R\big(w^*(\delta) - \hat{w}(\hat{\delta}_\kappa)\big) - \hat{w}^\mathsf{T}(\hat{\delta}_\kappa)R\hat{w}(\hat{\delta}_\kappa) \\
&= -\delta^\mathsf{T} Q_\delta \delta + \|r\|^2 \big\|w^*(\delta) - \hat{w}(\hat{\delta}_\kappa)\big\|^2 - \|r\|^2 \|\hat{w}^*(\hat{\delta}_\kappa)\|^2.
\end{aligned} \tag{38}$$

Considering $\varphi_{ic} = \hat{\varphi}_{ic} + \tilde{\varphi}_{ic}$, we get

$$
\begin{aligned}
\|w^*(\delta) - w(\hat{\delta}_\kappa)\|^2 &= \left\| \left(w^*(\delta) - w^*(\hat{\delta}_\kappa) \right) + \left(w^*(\hat{\delta}_\kappa) - \hat{w}(\hat{\delta}_\kappa) \right) \right\|^2 \\
&\leq 2 \left\| w^*(\delta) - w^*(\hat{\delta}_\kappa) \right\|^2 + 2 \left\| w^*(\hat{\delta}_\kappa) - \hat{w}(\hat{\delta}_\kappa) \right\|^2 \\
&\leq \frac{1}{2} \|r^{-1}\|^2 \overline{g} \left\| \nabla \sigma_c^\mathsf{T}(\hat{\delta}_\kappa) \hat{\varphi}_c - \nabla \sigma_c^\mathsf{T}(\hat{\delta}_\kappa) \varphi_c - \nabla \xi_c(\hat{\delta}_\kappa) \right\|^2 + 2\mathcal{L}_w^2 \|\mathbf{E}_\kappa(t)\|^2 \\
&\leq \frac{1}{2} \|r^{-1}\|^2 \overline{g}^2 \| - \nabla \xi_c(\hat{\delta}_\kappa) - \nabla \sigma_c^\mathsf{T}(\hat{\delta}_\kappa) \tilde{\varphi}_c \|^2 + 2\mathcal{L}_w^2 \|\mathbf{E}_\kappa(t)\|^2 \\
&\leq 2\mathcal{L}_w^2 \|\mathbf{E}_\kappa(t)\|^2 + \|r^{-1}\|^2 \overline{g}^2 \left(\overline{\sigma}_c^2 \overline{\varphi}_c^2 + \overline{\xi}_c^2 \right).
\end{aligned}
\tag{39}
$$

According to (39), we further derive from (38) as

$$
\dot{\Sigma}_{31}(t) \leq -\delta^\mathsf{T} Q_\delta \delta + 2\mathcal{L}_w^2 \|r\|^2 \|\mathbf{E}_\kappa(t)\|^2 - \|r\|^2 \|\hat{w}(\hat{\delta}_\kappa)\|^2 + \underbrace{\|r^{-1}\|^2 \overline{g}^2 \left(\overline{\sigma}_{ic}^2 \overline{\varphi}_{ic}^2 + \overline{\xi}_{ic}^2 \right)}_{\Theta}.
\tag{40}
$$

By combining (36) and (40), we obtain

$$
\dot{\Sigma}_3(t) \leq -\rho^2 \lambda_{\min}(Q_\delta) \|\delta\|^2 + (\rho^2 - 1)\lambda_{\min}(Q_\delta)\|\delta\|^2 + 2\|r\|^2 \mathcal{L}_w^2 \|\mathbf{E}_\kappa(t)\|^2 + \Theta.
$$

Therefore, if the condition (35) holds and δ lies outside the compact set

$$
\Omega_\delta = \left\{ \delta : \|\delta\| \leq \sqrt{\frac{\Theta}{\rho^2 \lambda_{\min}(Q_\delta)}} \right\},
$$

we can find that $\dot{\Sigma}_3(t) \leq -\rho^2 \lambda_{\min}(Q_\delta)\|\delta\|^2 < 0$ for any $\delta \neq 0$.

Case 2: $\forall t = t_{s+1}$, we have

$$
\begin{aligned}
\Delta\Sigma_3(t) &= \Sigma_3(\hat{\delta}_{\kappa+1}) - \Sigma_3(\delta(t_{\kappa+1}^-)) \\
&= \Delta\Sigma_{31}(t) + \Delta\Sigma_{32}(t).
\end{aligned}
$$

Noting the fact that δ and $V^*(\cdot)$ are both continuous, we derive

$$
\Delta\Sigma_{31}(t) = V^*(\hat{\delta}_{\kappa+1}) - V^*(\delta(t_{\kappa+1}^-)) \leq 0,
\tag{41a}
$$

$$
\Delta\Sigma_{32}(t) = V^*(\hat{\delta}_{\kappa+1}) - V^*(\hat{\delta}_\kappa) \leq -\vartheta(\|\mathbf{E}_{\kappa+1}(t_\kappa)\|),
\tag{41b}
$$

where $\delta(t_{\kappa+1}^-) = \lim_{\Delta t \to 0} \delta(t_{\kappa+1} - \Delta t)$, $\vartheta(\cdot)$ is a class-$\mathcal{K}$ function and $\mathbf{E}_{\kappa+1}(t_\kappa) = \hat{\delta}_{\kappa+1} - \hat{\delta}_\kappa$. Based on (41), we derive $\Delta\Sigma_3(t) \leq 0$. In the end, from the two aspects, if (35) holds, the closed-loop system is UUB. ∎

Remark 5. Assumption 1 which is also provided in References 1-7 is a basic assumption for the nonlinear systems. For Assumption 2, $w^*(\delta)$ represent the optimal tracking control law. Based on Assumption 1, one can find that $w^*(\delta)$ is Lipschitz continuous. Thus, there exists a Lipschitz constant such that $\|w^*(\delta(t)) - w^*(\hat{\delta}_k)\| \leq \mathcal{L}_w \|\mathbf{E}_k(t)\|$ (References 38-42). For Assumption 3, the optimal value function $V^*(\delta)$ is bounded, which implies φ_c is norm-bounded. Then, according to Lemma 1, the critic NN weight error dynamics is guaranteed to be UUB. Thus, $\tilde{\varphi}_c$ can be further assumed norm-bounded by $\|\tilde{\varphi}_c\| \leq \overline{\varphi}_c$, where $\overline{\varphi}_c$ is a constant. The term $\nabla \sigma_c^\mathsf{T}(\delta)$ is the partial derivation of activation function respect δ by selecting a suitable activation function $\sigma_c(\delta)$ and the term $\nabla \xi_c^\mathsf{T}(\delta)$ is the partial derivation of reconstruction error, they are reasonable to assume norm-bounded by $\|\nabla \sigma_c(\delta)\| \leq \overline{\sigma}_c$ and $\|\nabla \xi_c(\delta)\| \leq \overline{\xi}_c$, where $\overline{\sigma}_c$ and $\overline{\xi}_c$ are constants. $\mathcal{G}(s)$ is the input gain function, it reasonable to assume that it is norm-bounded by a constant.[45-48] In fact, Assumption 3 is usually employed in the ADP-based control field.[1-11,23-29,35-48,49]

4 | SIMULATION RESULTS

The effectiveness of the proposed ETNOTC method is demonstrated by employing a numerical and a realistic nonlinear systems.

4.1 | Example 1

Consider the nonlinear system with uncertainties as

$$\dot{s} = \begin{bmatrix} s_2 \\ -0.5s_1^3 - 0.5s_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}(u + d), \tag{42}$$

where $s \in [s_1, s_2]^\mathsf{T}$ is the system state, $d = \sin(0.6s_1)\cos(s_2)\cos(0.6s_1)$ is the uncertainties. The reference trajectory is chosen as

$$\dot{x}_d = \begin{bmatrix} -0.5\sin(t) + 0.6\cos(3t) \\ -0.5\cos(t) - 1.8\sin(3t) \end{bmatrix}. \tag{43}$$

The feedback control input in continuous control component is designed by using ADP-based ETC control method for system (16). The parameters of the value function are set as $Q_\delta = 4I$ and $R = 0.1$. In the critic NN, the learning rate $\alpha_c = 2$, the activation function is chosen as $\sigma_c(\delta) = \begin{bmatrix} \delta_1^2, \delta_1\delta_2, \delta_2^2 \end{bmatrix}^\mathsf{T}$, the weight vector is defined as $\hat{\varphi}_c = \begin{bmatrix} \hat{\varphi}_{c1}, \hat{\varphi}_{c2}, \hat{\varphi}_{c3} \end{bmatrix}^\mathsf{T}$. Figure 1 displays that the weight vector of the critic NN $\hat{\varphi}_i$ finally converge to $[0.1361, 0.0691, 0.1613]^\mathsf{T}$. Figure 2 shows the feedback control and the continuous control inputs, the feedback control input w is updated at t_κ only, and keeps unchanged during $[t_\kappa, t_{\kappa+1})$. Figure 3 describes that the tracking errors converge to a small region of zero (SRZ) after 7 s. From Figure 4, it is found that the less updating frequency of the feedback control signal is required by using ETC than TTC mechanism, which implies that the computational and communication resources can be saved.

Then, in the discontinuous control component, the initial weight vector $\hat{\theta}$ is randomly selected within $[-1, 1]$, the RBF $h(s) = [h_1(s), h_2(s), \ldots, h_{ld}(s)]^\mathsf{T}$ is chosen as

$$h_l(s) = \exp\left(\frac{-\|s - c_l\|^2}{b^2}\right), \tag{44}$$
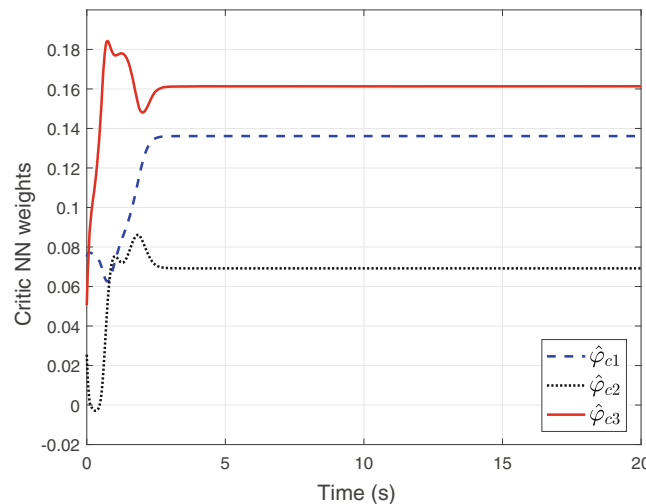
FIGURE 1 The learning process of critic NN weights.
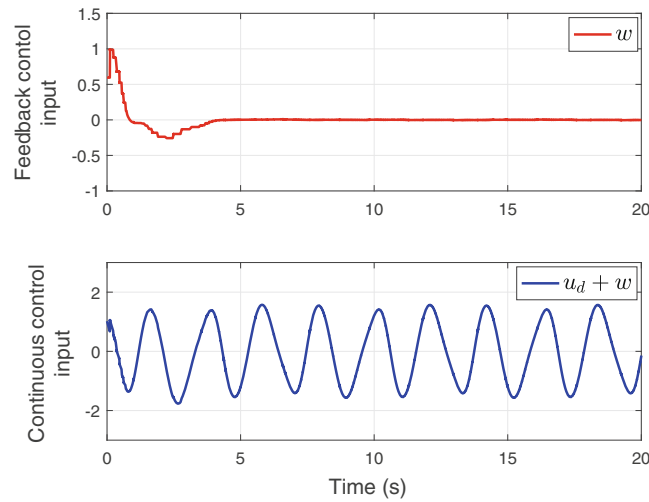
FIGURE 2 Feedback and continuous control inputs.
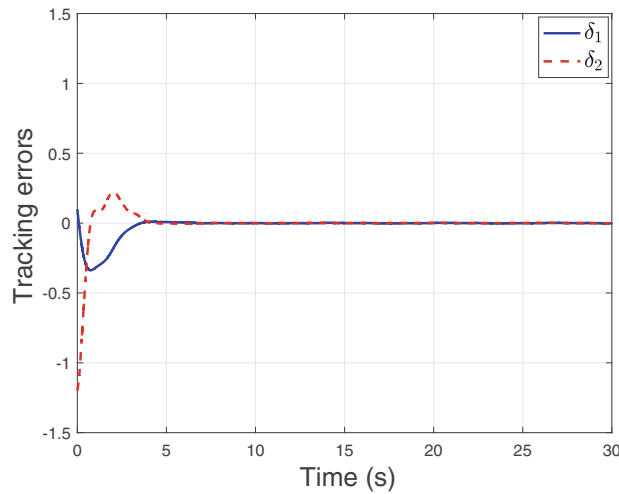
FIGURE 3 Tracking errors of system (16).

where $l = 1, 2, \ldots, l_d$, and c_l is the lth colum vector of the matrix

$$C_d = \begin{bmatrix} -3 & -2 & -1 & 0 & 1 & 2 & 3 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{bmatrix}.$$

The sliding mode gain is chosen as $K = 0.02$, $A = [0, 1]$. The discontinuous control law (9) is given as $u_c = -K\mathrm{sign}(A^\mathsf{T}(s)$ $g^\mathsf{T}(s)S) + \hat{d}(s)$. The composite control input (4) is employed to drive the tracking error dynamics (3) for simulation. Figure 5 displays the tracking performance. As shown in Figure 6, the the tracking errors converge to a SRZ after 7 s. The curves of composite control input and sliding mode function are presented in Figure 7. Figure 8 shows the curves of $d(s)$ and $\hat{d}(s)$ and their difference, we can conclude that the adaptive term is effective to approximate the $d(s)$.

4.2 | Example 2

The pendulum system[50] is formulated as

$$\ddot{\theta} = -\frac{f_d}{J}\dot{\theta} - \frac{MgL}{J}\sin(\theta) + \frac{1}{J}(u + d),$$

177
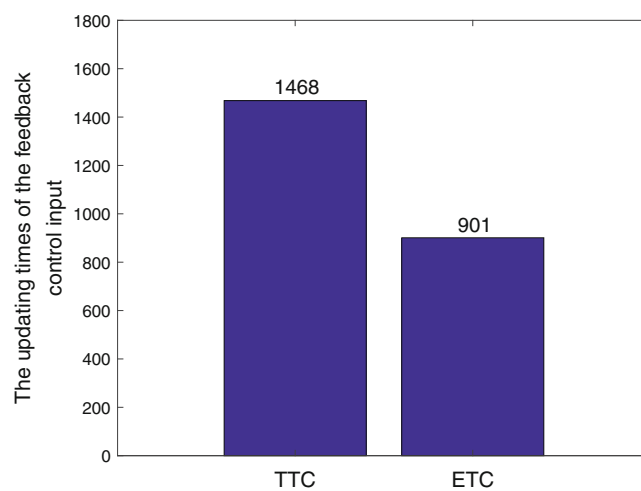
FIGURE 4 The updating times of the feedback control input.
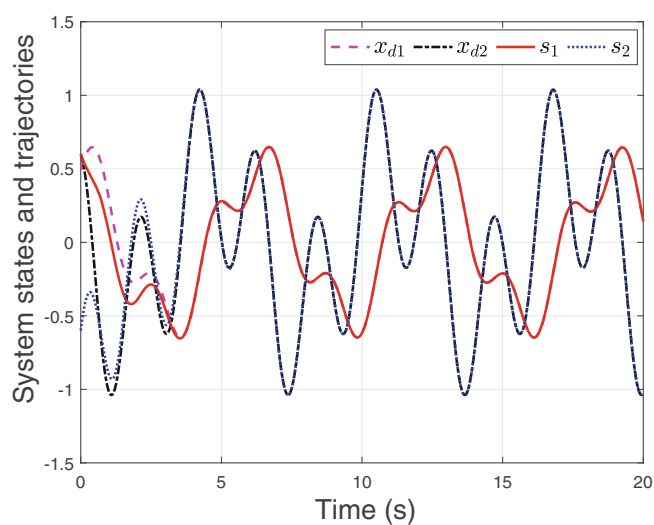
FIGURE 5 Tracking control performance.
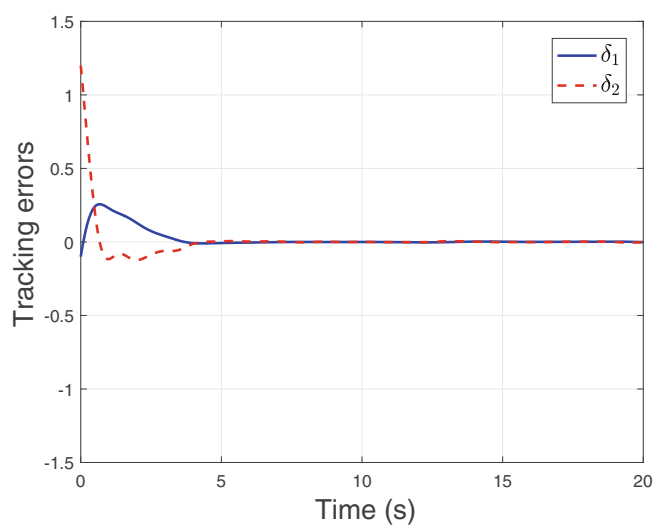
FIGURE 6 The tracking errors of system (3).

178
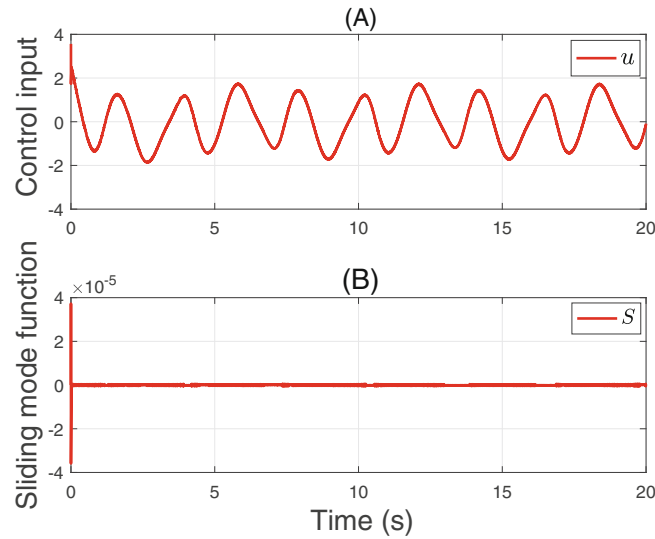
FIGURE 7 The curves of composite control input and sliding mode function.
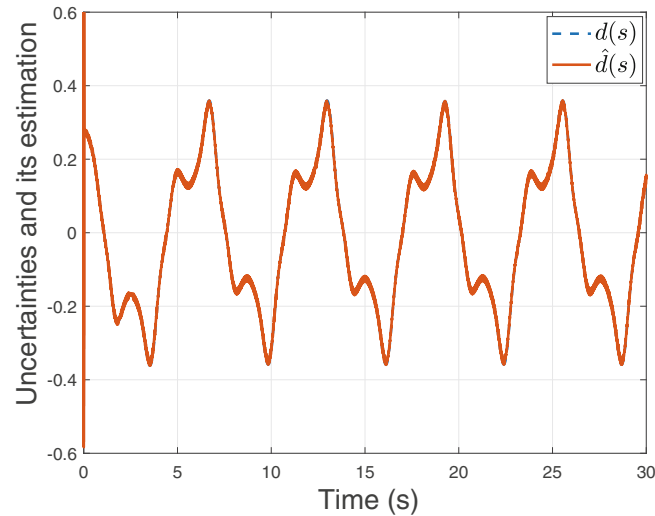
FIGURE 8 The uncertainties and its estimation.

where $\theta \in \mathbb{R}$ denotes the angle position of the pendulum, and the parameters are given in Table 1. Let $s_1 = \theta$, $s_2 = \dot{\theta}$, we have

$$\dot{s} = \begin{bmatrix} s_2 \\ -4.9\sin(s_1) - 0.2s_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 0.25 \end{bmatrix}(u + d), \tag{45}$$

where $s = [s_1, s_2]^{\mathsf{T}} \in \mathbb{R}^2$ is the system state, $d = \sin(s_1)\cos(s_2)\sin(s_2)$. The reference system is chosen as

$$\dot{x}_d = \begin{bmatrix} -0.6\sin(t) + 0.4\cos(2t) \\ -0.6\cos(t) - 0.8\sin(2t) \end{bmatrix}. \tag{46}$$

First, ADP-based ETC method is developed to design the feedback control input. The value function is given as (16) with $Q_\delta = 5I$ and $R = 0.05$. The structure of the critic NN is the same in Example 1. As shown in Figure 9, the weight vector $\hat{\varphi}$ finally converges to $[0.4230, 0.1627, 0.5893]^{\mathsf{T}}$. Figure 10 shows the feedback control and the continuous control

179

T A B L E 1   Parameters of the pendulum system.

Parameter	J	L	M	g	f_d
Value	$4 \text{ kg} \cdot \text{m}^2$	1.5 m	$4/3 \text{ kg}$	9.8 m/s^2	$0.8 \text{ N} \cdot \text{m} \cdot \text{s/rad}$
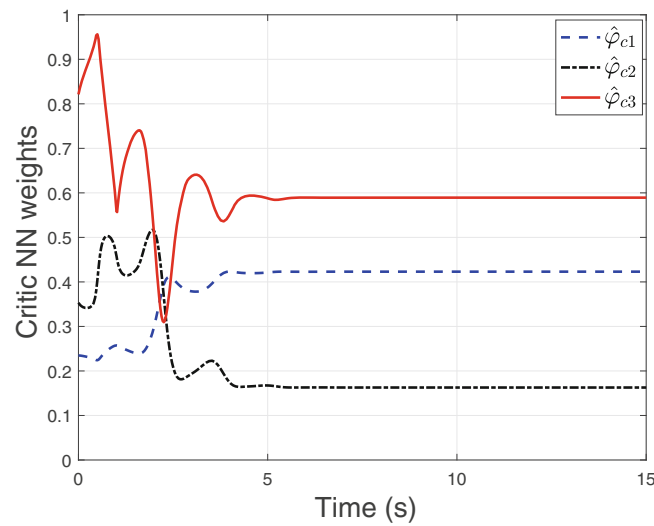
F I G U R E 9   The learning process of critic NN weights.
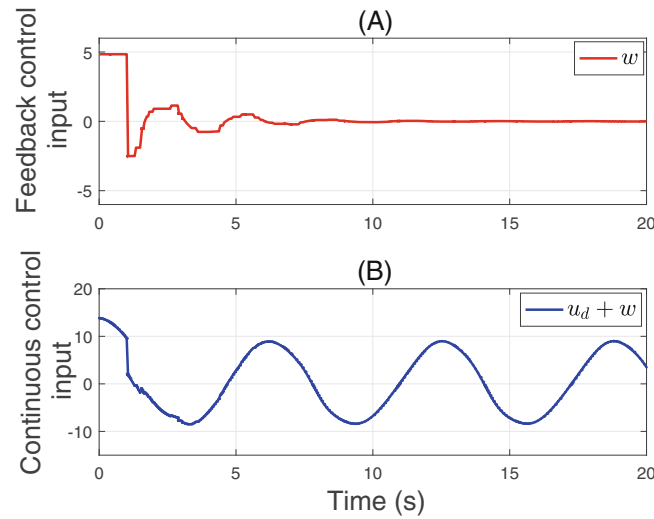
F I G U R E 10   Feedback and continuous control inputs.

inputs, the feedback control input w is a piecewise signal, which implies it only updated when events occur. Figure 11 displays that the tracking errors converge to a SRZ after 15 s. Figure 12 shows the updating times of the feedback control input, the TTC and ETC methods are required 1200 and 399 times, respectively. Thus, the computation and communication resources are saved.

Then, in the discontinuous control component, $K = 0.2$, $A = [0, 1]$, and the parameters and structure of the adaptive term are selected as the same in Example 1. Furthermore, The composite control input is used to drive the tracking error dynamics. Figure 13 shows the tracking performance. From Figure 14, we can find that the tracking errors converge to a SRZ after 15 s. Figure 15 displays the curves of the composite control input and the sliding mode function. According to Figure 16, we known the adaptive term $\hat{d}(s)$ estimate $d(s)$ successfully.
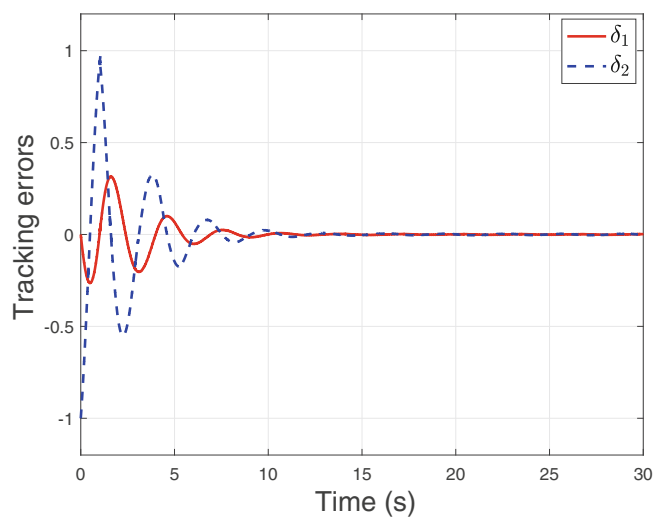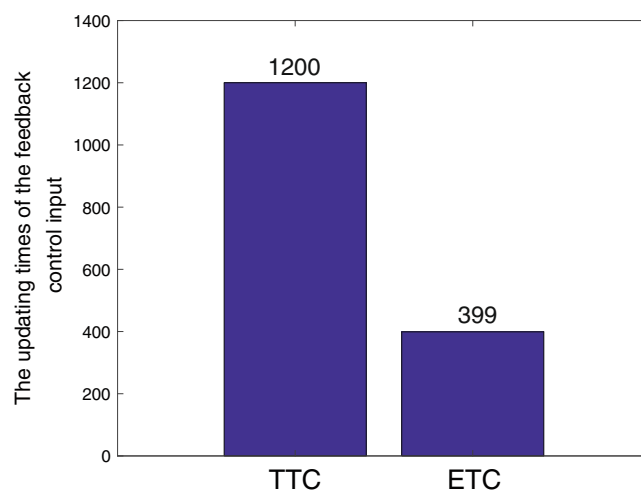
FIGURE 11 Tracking errors of system (16).

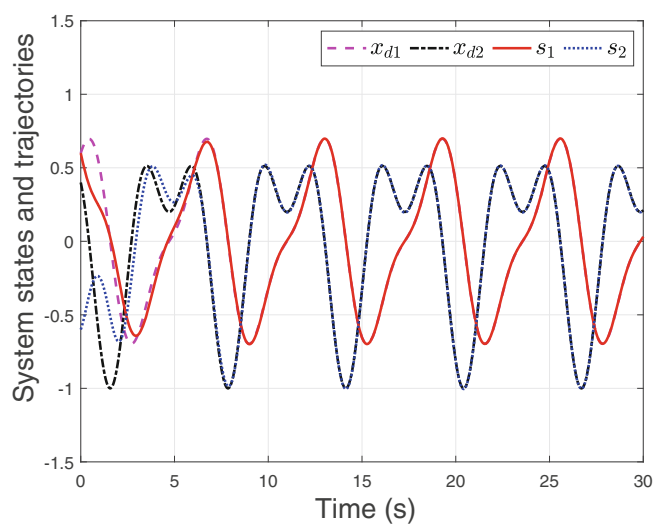**FIGURE 12** The updating times of the feedback control input.

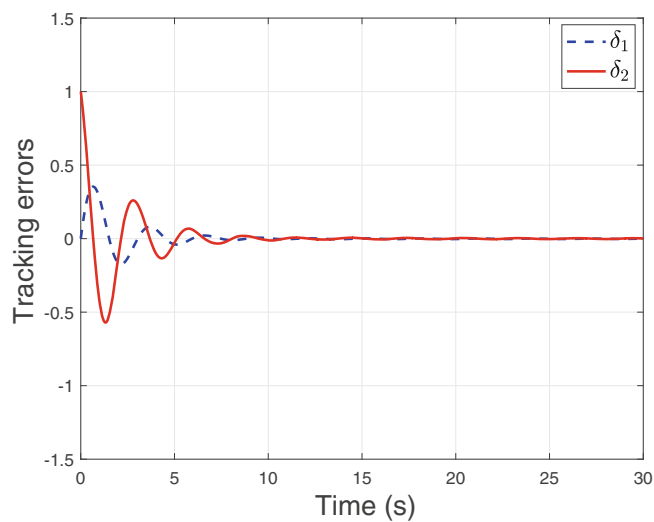**FIGURE 13** Tracking control performance.

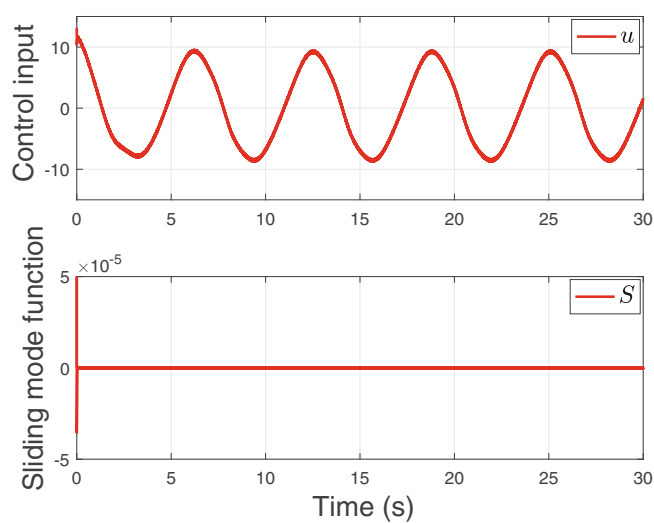**FIGURE 14** Tracking errors of system (3).

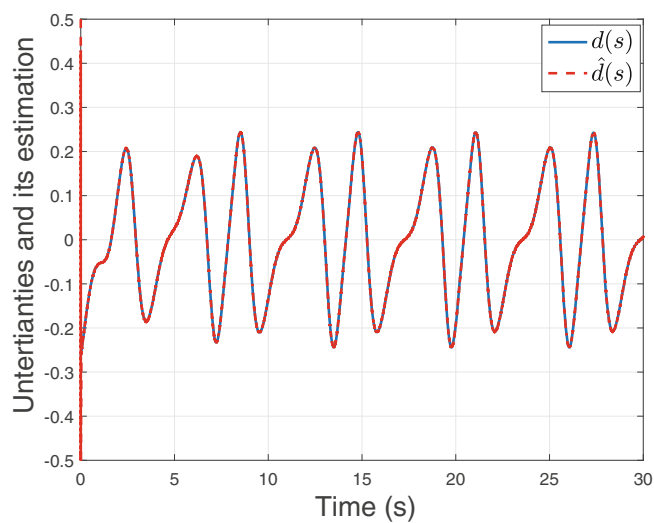**FIGURE 15** The curves of composite control input and sliding mode function.

FIGURE 16 The uncertainty and its estimation.

5 | CONCLUSIONS

In this paper, we develop an ETNOTC method for nonlinear uncertain systems by integrating ADP and ISMC techniques. The discontinuous control input with an adaptive term is developed to eliminate the influence of uncertainties and obtain the sliding mode dynamics, this method can relax the assumption of known upper-bounded function of uncertainties, and the designed continuous control input composed of feedforward and feedback control inputs is employed to achieve the tracking task. The ADP-based feedback control input is updated only when events occur, thus the updating frequency is reduced, and the computational and communication burdens are reduced. According to Lyapunov stability theorem, we prove that the closed-loop tracking error system is asymptotically stable. Finally, the simulation results declare that the developed ETNOTC method is effective.

It is noticed that dead-zone is one of the most commonly encountered non-smooth non-linearities and widely exists in many mechanical and electrical systems, such as mechanical transmissions, hydraulic actuators, and power generators. Furthermore, finite-time and predefined-time control methods have attracted widespread attention in the control field, which can guarantee the close-loop systems stability under finite-time and predefined-time, respectively. In the future work, finite-time optimal and predefined-time optimal tracking control will be further investigated for nonlinear systems with dead-zone.

CONFLICT OF INTEREST STATEMENT
The authors declare that there is no conflict of interest regarding the publication of this article.

DATA AVAILABILITY STATEMENT
Research data are not shared.

ORCID
Shunchao Zhang https://orcid.org/0000-0003-1303-1781
Yongwei Zhang https://orcid.org/0000-0003-3381-6340

REFERENCES
1. Liu D, Wei Q, Wang D, Yang X, Li H. *Adaptive Dynamic Programming with Applications in Optimal Control.* Springer; 2017.
2. Wang D, He H, Liu D. Adaptive critic nonlinear robust control: a survey. *IEEE Trans. Cybern.* 2017;47(10):3429-3451.
3. Lin M, Zhao B, Liu D. Policy gradient adaptive critic designs for model-free optimal tracking control with experience replay. *IEEE Trans. Syst. Man Cybern. Syst.* 2022;52(6):3692-3703.
4. Lu K, Liu Z, Yu H, Chen CLP, Zhang Y. Adaptive fuzzy inverse optimal fixed-time control of uncertain nonlinear systems. *IEEE Trans Fuzzy Syst.* 2022;30(9):1-12.
5. Wang D, Liu D, Li H, Luo B, Ma H. An approximate optimal control approach for robust stabilization of a class of discrete-time nonlinear systems with uncertainties. *IEEE Trans. Syst. Man Cybern. Syst.* 2016;46(5):713-717.
6. Xue S, Luo B, Liu D, Yang Y. Constrained event-triggered H_∞ control based on adaptive dynamic programming with concurrent learning. *IEEE Trans Syst Man Cybern Syst.* 2022;52(1):357-369.
7. Xia H, Guo P. Sliding mode-based online fault compensation control for modular reconfigurable robots through adaptive dynamic programming. *Complex Intell Syst.* 2021;8:1963-1973.
8. Li H, Yu J, Hilton C, Liu H. Adaptive sliding-mode control for nonlinear active suspension vehicle systems using T-S fuzzy approach. *IEEE Trans Ind Electron.* 2013;60(8):3328-3338.
9. Zhao B, Liu D, Alippi C. Sliding-mode surface-based approximate optimal control for uncertain nonlinear systems with asymptotically stable critic structure. *IEEE Trans. Cybern.* 2021;51(6):2858-2869.
10. Zhou Q, Yao D, Wang J, Wu C. Robust control of uncertain semi-Markovian jump systems using sliding mode control method. *Appl Math Comput.* 2016;286:72-87.
11. Liu Z, Su H, Pan S. A new adaptive sliding mode control of uncertain nonlinear systems. *Asian J Control.* 2014;16(1):198-208.

12. Ding S, Wang J, Zheng WX. Second-order sliding mode control for nonlinear uncertain systems bounded by positive functions. *IEEE Trans. Ind. Electron.* 2015;62(9):5899-5909.

13. Cao WJ, Xu JX. Nonlinear integral-type sliding surface for both matched and unmatched uncertain systems. *IEEE Trans. Autom. Control.* 2011;49(8):1355-1360.

14. Rubagotti M, Estrada A, Castanos F, Ferrara A, Fridman L. Integral sliding mode control for nonlinear systems with matched and unmatched perturbations. *IEEE Trans. Autom. Control.* 2011;56(11):2699-2704.

15. Castanos F, Fridman L. Analysis and design of integral sliding manifolds for systems with unmatched perturbations. *IEEE Trans. Autom. Control.* 2006;51(5):853-858.

16. Zhang H, Qu Q, Xiao G, Cui Y. Optimal guaranteed cost sliding mode control for constrained-input nonlinear systems with matched and unmatched disturbances. *IEEE Trans. Neural Netw. Learn. Syst.* 2018;29(6):2112-2126.

17. Janardhanan S, Kariwala V. Multirate-output-feedback-based LQ-optimal discrete-time sliding mode control. *IEEE Trans. Autom. Control.* 2008;53(1):367-373.

18. Sun N, Niu Y, Chen B. Optimal integral sliding mode control for a class of uncertain discrete-time systems. *Optim Control Appl Methods.* 2014;35(4):468-478.

19. Basin M, Rodriguez-Ramirez P, Ferrara A, Calderon-Alvarez D. Sliding mode optimal control for linear systems. *J Franklin Inst.* 2012;349(4):1350-1363.

20. Poznyak A, Fridman L, Bejarano FJ. Mini-max integral sliding-mode control for multimodel linear uncertain systems. *IEEE Trans Autom Control.* 2004;49(1):97-102.

21. Surjagade PV, Tiwari AP, Shimjith SR. Robust optimal integral sliding mode controller for total power control of large PHWRs. *IEEE Trans Nucl Sci.* 2018;65(7):1331-1344.

22. Das M, Mahanta C. Optimal second order sliding mode control for linear uncertain systems. *ISA Trans.* 2014;53(6):1807-1815.

23. Zhao B, Wang D, Shi G, Liu D, Li Y. Decentralized control for large-scale nonlinear systems with unknown mismatched interconnections via policy iteration. *IEEE Trans Syst Man Cybern Syst.* 2018;48(10):1725-1735.

24. Liu D, Xue S, Zhao B, Luo B, Wei Q. Adaptive dynamic programming for control: a survey and recent advances. *IEEE Trans. Syst. Man Cybern. Syst.* 2021;51(1):142-160.

25. Chen Z, Chen SZ, Chen K, Zhang Y. Constrained decoupling adaptive dynamic programming for a partially uncontrollable time-delayed model of energy systems. *Inform Sci.* 2022;608:1352-1374.

26. Wei Q, Liu D, Liu Y, Song R. Optimal constrained self-learning battery sequential management in microgrid via adaptive dynamic programming. *IEEE/CAA J Automa Sinica.* 2017;4(2):168-176.

27. Liu D, Wei Q. Policy iteration adaptive dynamic programming algorithm for discrete-time nonlinear systems. *IEEE Trans. Neural Netw. Learn. Syst.* 2014;25(3):621-634.

28. Vamvoudakis KG, Lewis FL. Online actor-critic algorithm to solve the continuous-time infinite horizon optimal control problem. *Automatica.* 2010;46(5):878-888.

29. Vrabie D, Lewis FL. Neural network approach to continuous-time direct adaptive optimal control for partially unknown nonlinear systems. *Neural Netw.* 2009;22(3):237-246.

30. Wang H, Xu K, Zhang H. Adaptive finite-time tracking control of nonlinear systems with dynamics uncertainties. *IEEE Trans. Autom. Control.* 2022;68(9):5737-5744.

31. Wang H, Wang M, Zhao X, Niu B, Yang M. Predefined-time adaptive neural tracking control of switched nonlinear systems. *IEEE Trans. Cybern.* 2022;53(10):6538-6548.

32. Wang D, Liu D, Wei Q. Finite-horizon neuro-optimal tracking control for a class of discrete-time nonlinear systems using adaptive dynamic programming approach. *Neurocomputing.* 2012;78:14-22.

33. Wei Q, Liu D. Adaptive dynamic programming for optimal tracking control of unknown nonlinear systems with application to coal gasification. *IEEE Trans Autom Sci Eng.* 2014;11(4):1020-1036.

34. Modares H, Lewis FL. Optimal tracking control of nonlinear partially-unknown constrained-input systems using integral reinforcement learning. *Automatica.* 2014;50(7):1780-1792.

35. Zhao J, Na J, Gao G. Robust tracking control of uncertain nonlinear systems with adaptive dynamic programming. *Neurocomputing.* 2021;471:21-30.

36. Wang D, Mu C. Adaptive-critic-based robust trajectory tracking of uncertain dynamics and its application to a spring-mass-damper system. *IEEE Trans. Ind. Electron.* 2018;65(1):654-663.

37. Wang Z, Wei Q, Liu D. Event-triggered adaptive dynamic programming for discrete-time multi-player games. *Inform Sci.* 2020;506:457-470.

38. Ha M, Wang D, Liu D. Event-triggered adaptive critic control design for discrete-time constrained nonlinear systems. *IEEE Trans. Syst. Man Cybern. Syst.* 2020;50(9):3158-3168.

39. Vamvoudakis KG. Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems. *IEEE/CAA J Automa Sinica.* 2014;1(3):282-293.

40. Wang D, Mu C, Yang X, Liu D. Event-based constrained robust control of affine systems incorporating an adaptive critic mechanism. *IEEE Trans Syst Man Cybern Syst.* 2017;47(7):1602-1612.

41. Zhang K, Zhang H, Jiang H, Wang Y. Near-optimal output tracking controller design for nonlinear systems using an event-driven ADP approach. *Neurocomputing.* 2018;309:168-178.

42. Xue S, Luo B, Liu D, Gao Y. Event-triggered ADP for tracking control of partially unknown constrained uncertain systems. *IEEE Trans. Cybern.* 2022;52(9):9001-9012.

43. Li Y, Yang G. Adaptive integral sliding mode control fault tolerant control for a class of uncertain nonlinear systems. *IET Control Theory Appl.* 2018;12(13):1864-1872.

44. Fan QY, Yang GH. Adaptive actor-critic design-based integral sliding-mode control for partially unknown nonlinear systems with input disturbances. *IEEE Trans. Neural Netw. Learn. Syst.* 2016;27(1):165-177.

45. Lewis FL, Yesildirek A, Liu K. Multilayer neural-net robot controller with guaranteed tracking performance. *IEEE Trans Neural Netw.* 1996;7(2):388-399.

46. Hornik K, Stinchcombe M, White H. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. *Neural Netw.* 1990;3(5):551-560.

47. Yang X, He H. Decentralized event-triggered control for a class of nonlinear-interconnected systems using reinforcement learning. *IEEE Trans. Cybern.* 2021;51(2):635-648.

48. Zhang Q, Zhao D, Zhu Y. Event-triggered H_∞ control for continuous-time nonlinear system via concurrent learning. *IEEE Trans. Syst. Man Cybern. Syst.* 2017;47(7):1071-1081.

49. Zhao B, Liu D. Event-triggered decentralized tracking control of modular reconfigurable robots through adaptive dynamic programming. *IEEE Trans. Ind. Electron.* 2020;67(4):3054-3064.

50. Yang X, Wei Q. Adaptive critic learning for constrained optimal event-triggered control with discounted cost. *IEEE Trans. Neural Netw. Learn. Syst.* 2021;32(1):91-104.

获奖证书

certificate of award

华南农业大学的参赛作品《农业生产中的可持续性评估和优化》，在2024年（第十届）全国大学生统计建模大赛 广东赛区（广东、海南、香港、澳门）赛区选拔赛中，荣获本科生组三等奖。

参赛队员：邵欣媛、钟倩、韦淇凯

指导老师：张勇威

证书编号：20242901C0102

中国统计教育学会

二〇二四年十月

全国大学生大数据分析技术技能大赛

获奖证书
AWARD CERTIFICATE

证书编号：BBDA2024JSJ1639

张勇威 老师：

指导学生参加第三届全国大学生大数据分析技术技能大赛索

获 数据分析赛项 本科组，广东省 赛

Excel

区指导教师一等奖。

参赛院校：华南农业大学

参赛选手：陈洪琳

指导单位：

教育部高等学校统计学类专业教学指导委员会

全国高校大数学位研究生数据教育指导委员会

全国大学生大数据分析技术技能大赛

主办单位：

中国统计教育学会

全国应用统计专业学位研究生教育指导委员会

承办单位：

北京天拓数据新闻学会数据专委会

全国大学生大数据分析技术技能大赛组织委员会

二〇二四年十月十日

奖 证 书

AWARD CERTIFICATE

证书编号：BBDA2024JSJ1638

张勇威 老师：

指导学生参加第三届全国大学生大数据分析技术技能大赛荣

获 数据分析赛项 本科组 ， 广东省 赛

Excel

区指导教师二等奖。

参赛院校：华南农业大学

参赛选手：叶榆彤

指导单位：
教育部高等学校统计学类专业教学指导委员会
全国应用统计专业学位研究生教育指导委员会

主办单位：
中国高校科技新闻学会数据专委会
北京天数据协会（华章）

全国大学生大数据分析技术技能大赛
组织委员会

二〇二四年四月三十日

188

证书编号：BBDA2024SJ1637

获奖证书
AWARD CERTIFICATE

张勇威 老师：

指导学生参加第三届全国大学生大数据分析技术技能大赛
Excel 数据分析赛项 本科组 ，广东省 赛

区指导教师三等奖。

参赛院校：华南农业大学

参赛选手：许斯淇

指导单位：
教育部高等学校统计学类专业教学指导委员会
全国应用统计专业学位研究生教育指导委员会

主办单位：
中国新闻学会体育数据专委会
北京天泽数据协会(华章)

全国大学生大数据分析技术技能大赛
组织委员会

二〇二四年四月三十日

全国大学生大数据分析技术技能大赛

奖证书
AWARD CERTIFICATE

证书编号：BBDA2024JSJ1640

张勇威 老师：

指导学生参加第三届全国大学生大数据分析技术技能大赛

Excel 数据分析赛项 本科组，广东省

区指导教师三等奖。

参赛院校：华南农业大学

参赛选手：王培豪

190

证书编号：BBDA2024JSJ1641

奖 证 书
AWARD CERTIFICATE

张勇威　老师：

指导学生参加第三届全国大学生大数据分析技术技能大赛

获 Python 数据分析赛项 本科组 ， 广东省 赛

区指导教师三等奖。

参赛院校：华南农业大学

参赛选手：黄梓帆

指导单位：
教育部高等学校统计学类专业教学指导委员会
全国应用统计专业学位研究生教育指导委员会
主办单位：
中国科技新闻学会大数据专委会
北京天数据协会（华章）

全国大学生大数据分析技术技能大赛
组织委员会
二○二四年七月十日