華東理工大學学报 自然科学版 Natural Science Edition
JOURNAL OF EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY

预定性能约束下非线性博弈系统的最优事件触发控制

张志祥 王梦灵

Optimal Event-Triggered Control for Nonlinear Game Systems with Prescribed Performance Constraints

ZHANG Zhixiang, WANG Mengling

在线阅读 View online: http://doi.org/10.14135/j.cnki.1006-3080.20260126001

您可能感兴趣的其他文章

Articles you may be interested in

网络攻击下非线性CPSs的事件触发自适应控制

Event-Triggered Adaptive Control of Nonlinear Cyber-Physical Systems Under Network Attacks

华东理工大学学报（自然科学版）. 2024, 50(3): 411-417 http://doi.org/10.14135/j.cnki.1006-3080.20230518001

基于自适应事件触发的交通网络预测控制

Traffic Network Predictive Control Based on An Adaptive Event-Triggered Scheme

华东理工大学学报（自然科学版）. 2021, 47(3): 316-322 http://doi.org/10.14135/j.cnki.1006-3080.20191230002

基于周期事件触发的永磁同步电机滑模控制

Periodic Event-Triggered Sliding Mode Control of Permanent Magnet Synchronous Motor

华东理工大学学报（自然科学版）. 2022, 48(6): 832-840 http://doi.org/10.14135/j.cnki.1006-3080.20210702002

一种基于事件触发的异构WSN分布式定位方法

An Event-Triggering Distributed Positioning Method of A Heterogeneous WSN

华东理工大学学报（自然科学版）. 2023, 49(4): 554-561 http://doi.org/10.14135/j.cnki.1006-3080.20220211001

基于组稀疏优化的强化学习稀疏表征

Reinforcement Learning with Sparse Representation via Sparse Overlapping Group Lasso

华东理工大学学报（自然科学版）. 2024, 50(6): 913-919 http://doi.org/10.14135/j.cnki.1006-3080.20231223001

基于强化学习的正弦优化算法求解能耗分布式流水车间节能调度问题

Sine Optimization Algorithm Based on Reinforcement Learning for Energy-Efficient Distributed Flow Shop Scheduling Problem

华东理工大学学报（自然科学版）. 2025, 51(5): 623-632 http://doi.org/10.14135/j.cnki.1006-3080.20250117002

扫码关注公众号，获取更多信息！

文章编号：1006-3080(2026)05-0686-09 DOI: 10.14135/j.cnki.1006-3080.20260126001

预定性能约束下非线性博弈系统的最优事件触发控制

张志祥， 王梦灵

（华东理工大学能源化工过程智能制造教育部重点实验室，上海 200237）

摘要：针对动态未知的非线性零和博弈系统，提出一种基于预定性能约束的最优事件触发控制方法，在降低计算资源的条件下实现系统动态未知的预定性能控制。首先，引入预定性能函数，将受约束的系统控制问题转化为等效的无约束最优控制问题；其次，基于离策略(off policy)数据迭代机制，结合多项式拟合技术与积分强化学习(Integral Reinforcement Learning, IRL)方法，在纯数据驱动的控制律求解框架下推导出无模型迭代控制策略，解决系统动态未知带来的求解困难问题。同时，本文设计了一种基于博弈理论的事件触发机制，首先，通过将事件采样误差与控制策略建模为零和博弈，有效降低了采样数量，从而减少计算负担；然后，采用单评价神经网络逼近最优控制解，利用设计的策略更新网络权值；最后，通过理论分析和仿真验证说明了方法的收敛性和有效性。

关键词：零和博弈；自适应动态规划；事件触发；预定性能控制；积分强化学习

中图分类号：TP13 **文献标志码：**A

微分博弈是处理具有复杂交互动态系统最优控制问题的有力工具，在多智能体系统、对抗性控制等场景中具有广泛应用。当参与者目标完全对立时，系统常被建模为零和博弈[1]。然而，实际工业系统中的动态往往难以精确建模，且需同时兼顾计算效率、通信资源与闭环性能(如收敛速度、稳态误差)等多重约束[2]，这为零和博弈的实时、安全与高效求解带来了算法稳定性、参数辨识及计算复杂度等多方面挑战。为应对性能约束问题，预定性能控制(Prescribed Performance Control, PPC)通过预设性能函数对系统误差边界及收敛过程进行严格规划，能有效提升系统的鲁棒性与瞬态性能[3-4]。例如，文献 [3] 通过预定性能强化学习控制将约束最优控制问题转化为无约束优化问题；文献 [4] 则结合博弈论思想设计了事件触发 PPC 策略，保证了系统的暂态性能。然而，现有研究在处理 PPC 约束时多依赖于精确的系统先验知识，例如文献 [3] 需要已知系统信息，文献 [4] 的 PPC 与事件触发阈值设定仍然建立在系统模型部分已知的假设上。在系统动态完全未知且存在输入耦合的情况下，利用纯数据流同时保证 PPC 性能，在现有文献中鲜见提供有效的解决方案。

强化学习，亦称自适应动态规划(Adaptive Dynamic Programming, ADP)，已成为近似求解非线性系统最优控制问题的主流方法[5-9]。在模型已知的情况下，ADP 方法结合神经网络与策略迭代技术，能够有效逼近系统最优价值函数与控制策略。例如，文献 [5] 针对模型已知的输入约束非线性零和博弈系统，结合事件驱动框架与自适应评判技术实现了权重的在线学习。针对模型未知的情形，无模型 ADP 方法主要分为模型辨识与数据迭代学习两种类型。模型辨识通过预先辨识系统参数重构动态模型，但往往带来较高的计算负担且存在误差传递问题[6]。而数据迭代学习方法，如文献 [7-9] 所示，基于离策略积分强化学习(off-policy IRL)框架，可直接从数据中学习控制策略，无需显式辨识系统模型。文献 [7] 利用 off-policy IRL 框架，将求解 Hamilton-

收稿日期：2026-01-26

基金项目：上海市科技创新行动计划项目 (24511104302)

作者简介：张志祥(2000—)，男，河南人，硕士生，主要研究方向为近似动态规划和强化学习，微分博弈系统。E-mail: Y30230925@mail.ecust.edu.cn

通信联系人：王梦灵，E-mail: wml_ling@ecust.edu.cn

引用本文：张志祥, 王梦灵. 预定性能约束下非线性博弈系统的最优事件触发控制 [J]. 华东理工大学学报(自然科学版), 2026, 52(5): 686-694.

Citation: ZHANG Zhixiang, WANG Mengling. Optimal Event-Triggered Control for Nonlinear Game Systems with Prescribed Performance Constraints[J]. Journal of East China University of Science and Technology, 2026, 52(5): 686-694.

Jacobi-Isaacs(HJI)方程的问题转化为一类无模型方法, 并应用于特定结构的非零和博弈系统。文献 [8] 面向具有状态与输入约束的未知非仿射系统, 结合 IRL 与单评价网络实现了安全驱动控制。文献 [9] 则针对部分未知非线性系统, 融合 IRL 与动态事件触发自适应策略进行控制设计。然而, 现有 ADP 方法多针对特定非线性结构[7] 或部分模型已知的系统[8-9], 对于输入动态耦合完全未知的一般情形, 其逼近能力与泛化性能仍有待提升。

此外, ADP 方法在迭代学习中通常涉及密集的计算过程。为降低此类资源消耗, 事件触发控制通过必要时执行更新来提升效率。根据触发机制设计方式的不同, 传统事件触发方法可分为静态触发[3-4]、动态触发[9] 以及自触发等类型。例如文献 [10] 提出的网络系统自适应静态事件触发模型预测控制, 以及文献 [11] 研究的网络攻击下事件触发输出反馈控制。近年来, 将事件触发机制与最优控制理论相结合, 构建最大化采样间隔的最优事件触发策略成为研究热点[12-13]。其中, 文献 [12] 针对线性系统基于博弈黎卡提方程进行采样求解镇定问题, 文献 [13] 针对多输入非线性系统提出的双重博弈框架, 均针对模型已知系统研究了最优触发控制方法。进一步审视上述两类事件触发机制, 传统事件触发设计独立于系统性能指标, 可能导致触发次数次优, 而现有的最优事件触发研究核心中均预设了系统动态模型精确已知的前提条件。在系统动态完全未知的非线性零和博弈背景下, 如何设计无模型的最优事件触发策略, 相关研究仍较为有限。

综上, 针对现有研究对精确模型信息[3,13]、部分模型[4,8-9] 或特定系统结构[7] 的依赖局限, 本文旨在填补动态完全未知非线性零和博弈系统在预定性能约束下的最优事件触发控制这一空白。具体而言, 本文针对动态未知的非线性零和博弈系统, 研究其事件触发最优控制问题。通过融合预定性能函数、离策略积分强化学习与博弈论方法, 提出一种基于多项式拟合的单评价网络逼近方案。该方法将带约束的控制问题转化为无约束优化问题, 利用纯数据驱动机制克服系统模型未知的困难, 并设计基于博弈理论的最优事件触发条件, 在保证闭环系统稳定性和预定性能的同时, 显著降低计算资源负担。

在本文中, R、$\mathbb{R}^n$、$\mathbb{R}^{n\times m_u}$、$\mathbb{R}^{n\times m_d}$ 分别表示实数空间、n 维实向量组成的欧氏空间、$n\times m_u$ 实矩阵空间和 $n\times m_d$ 实矩阵空间。对于数学运算符号, T 表示转置; $|\cdot|$ 和 $\|\cdot\|$ 表示范数; $\otimes$ 表示克罗内克积; $\lambda_{\min}(\cdot)$ 和 $\lambda_{\max}(\cdot)$ 分别表示最小和最大的特征值; $(\cdot)^{[*]}$ 表示优化最优值; $\inf\{\cdot\}$ 表示集合最大下界。

1　问题描述

1.1　零和博弈系统问题表示

考虑一个具有两人参与的非线性零和博弈系统, 其动态描述为

$$\dot{\boldsymbol{x}}(t)=\boldsymbol{f}(\boldsymbol{x}(t))+\boldsymbol{g}(\boldsymbol{x}(t))\boldsymbol{u}(t)+\boldsymbol{k}(\boldsymbol{x}(t))\boldsymbol{d}(t) \tag{1}$$

其中, $\boldsymbol{x}(t)\in\mathbb{R}^n$ 表示系统状态, $\boldsymbol{f}(\boldsymbol{x})\in\mathbb{R}^n$ 表示满足 $\boldsymbol{f}(0)=0$ 的未知漂移动态, $\boldsymbol{g}(\boldsymbol{x})\in\mathbb{R}^{n\times m_u}$ 和 $\boldsymbol{k}(\boldsymbol{x})\in\mathbb{R}^{n\times m_d}$ 分别表示与控制器 $\boldsymbol{u}(t)$ 和干扰 $\boldsymbol{d}(t)$ 相关的输入矩阵, $\boldsymbol{u}(t)\in\mathbb{R}^{m_u}$ 为控制输入, $\boldsymbol{d}(t)\in\mathbb{R}^{m_d}$ 为外部干扰输入。

假设 1[4]　所有系统动态均为局部 Lipschitz 连续, 且闭环状态被约束在紧集 $\Omega\subset\mathbb{R}^n$ 内演化。

系统性能函数定义为[1,6]

$$J(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d})=\int_0^{\infty}\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}+U(\boldsymbol{u})-\gamma^2\|\boldsymbol{d}\|^2\right)\mathrm{d}\nu \tag{2}$$

其中, $\boldsymbol{Q}\in\mathbb{R}^{n\times n}$ 为对称正定矩阵, 控制惩罚项为 $U(\boldsymbol{u})=\boldsymbol{u}^{\mathrm{T}}R_u\boldsymbol{u}$, $R_u>0$, $\gamma>0$, ν 为积分变量。参数 γ 的作用类似于 H_∞ 控制中的 $\mathcal{L}_2$ 增益[1]。当 γ 超过某一临界阈值时, 与零和博弈对应的 HJI 方程存在唯一镇定解, 从而保证最优对抗策略的存在性。

该零和博弈问题的目标是在系统动力学约束(式 (1)) 下, 求解如下极小极大优化问题:

$$\min_{\boldsymbol{u}}\max_{\boldsymbol{d}}J(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d})$$

定义 1[5]　若存在一组策略对 $(\boldsymbol{u}^{[*]},\boldsymbol{d}^{[*]})$, 使得对任意可行控制策略 $\boldsymbol{u}$ 与扰动策略 $\boldsymbol{d}$, 均满足

$$J(\boldsymbol{x},\boldsymbol{u},\boldsymbol{d}^{[*]})\geqslant J(\boldsymbol{x},\boldsymbol{u}^{[*]},\boldsymbol{d}^{[*]})\geqslant J(\boldsymbol{x},\boldsymbol{u}^{[*]},\boldsymbol{d}) \tag{3}$$

则称 $(\boldsymbol{u}^{[*]},\boldsymbol{d}^{[*]})$ 为式 (1)~(2) 的 Nash 均衡解, 亦即零和博弈意义下的鞍点解。

1.2　预定性能函数

本文采用预定性能控制方案在控制设计中同时约束系统的暂态与稳态响应特性。定义跟踪误差为 $\boldsymbol{e}(t)=\boldsymbol{x}(t)$, 并要求其始终满足预先给定的时变性能边界。

性能包络函数的选取为一严格递减且光滑的函数[3]

$$\mu(t)=(\mu_0-\mu_\infty)\mathrm{e}^{-\beta t}+\mu_\infty \tag{4}$$

其中, $\mu_0>\mu_\infty>0$, $\beta>0$。误差约束表示为 $-\underline{\varsigma}\mu(t)<\boldsymbol{e}(t)<\overline{\varsigma}\mu(t)$, $t\geqslant 0$, 其中, $\underline{\varsigma},\overline{\varsigma}>0$ 均为不对称性能系数。为了转换性能约束为在无约束空间下求解, 引入归一化误差 $\zeta(t)=\boldsymbol{e}(t)/\mu(t)$ 及变换

$$P(z)=\frac{\overline{\varsigma}\mathrm{e}^{z(t)}-\underline{\varsigma}\mathrm{e}^{-z(t)}}{\mathrm{e}^{z(t)}+\mathrm{e}^{-z(t)}} \tag{5}$$

其逆变换可表示为

$$z(t)=\frac{1}{2}\ln\left(\frac{\zeta+\underline{\varsigma}}{\overline{\varsigma}-\zeta}\right) \tag{6}$$

$z(t)$ 为无约束状态, 由此可将误差 $\boldsymbol{e}(t)$ 表示为

$$\boldsymbol{e}(t)=\frac{\overline{\varsigma}\mathrm{e}^{2z(t)}-\underline{\varsigma}}{1+\mathrm{e}^{2z(t)}}\mu(t) \tag{7}$$

对 $z(t)$ 求导, 省略时间 t, 结合式 (1) 得到转换后的无约束动力学方程

$$\dot{z}=\boldsymbol{F}(z)+\boldsymbol{G}(z)\boldsymbol{u}+\boldsymbol{K}(z)\boldsymbol{d} \tag{8}$$

其中,

$$\begin{aligned}&\boldsymbol{F}(z)=\frac{T}{2\mu}\left(\boldsymbol{f}(\boldsymbol{x})-\frac{\boldsymbol{e}\dot{\mu}}{\mu}\right),\\&\boldsymbol{G}(z)=\frac{T}{2\mu}\boldsymbol{g}(\boldsymbol{x}),\\&\boldsymbol{K}(z)=\frac{T}{2\mu}\boldsymbol{k}(\boldsymbol{x}),\\&T=1/(\zeta+\underline{\varsigma})+1/(\overline{\varsigma}-\zeta)。\end{aligned}$$

引理 1[4] 考虑状态向量 $\boldsymbol{x}(t)$ 和转换向量 $z(t)$, 当 $z(t)$ 有界时, 对任意 $t>0$, $\boldsymbol{x}(t)$ 的预定性能被保证。

1.3 最优事件触发机制

本文将事件触发引入的输入误差与系统扰动统一建模为零和博弈中的对抗参与方, 从而在控制性能与触发频率之间建立权衡关系。为减少不必要的策略更新与计算负担, 引入事件触发机制。设 $\tau_{jj=0}^{\infty}$ 为触发时刻序列, 其中 $\tau_0=0$。对于 $t\in[\tau_j,\tau_{j+1})$, 定义状态测量误差为 $\boldsymbol{e}^z(t)=\hat{\boldsymbol{z}}-\boldsymbol{z}(t)$, 其中 $\hat{\boldsymbol{z}}=\boldsymbol{z}(\tau_j)$ 为最近触发时刻采样的状态。

相应的控制与干扰输入误差定义为

$$\boldsymbol{e}^u(t)=\hat{\boldsymbol{u}}-\boldsymbol{u}(t),\ \boldsymbol{e}^d(t)=\hat{\boldsymbol{d}}-\boldsymbol{d}(t) \tag{9}$$

其中, $\hat{\boldsymbol{u}}=\boldsymbol{u}(\tau_j)$, $\hat{\boldsymbol{d}}=\boldsymbol{d}(\tau_j)$ 为采样控制输入, $\boldsymbol{u}(t)$, $\boldsymbol{d}(t)$ 为连续控制输入, 下面 $\boldsymbol{e}^u(t)$、$\boldsymbol{e}^d(t)$ 分别简写为 $\boldsymbol{e}^u$, $\boldsymbol{e}^d$。

在采样数据控制框架下, 输入误差 $\boldsymbol{e}^u$ 与 $\boldsymbol{e}^d$ 可等效视为附加外部输入, 式 (8) 可表示为

$$\dot{\boldsymbol{z}}=\boldsymbol{F}(z)+\boldsymbol{G}(z)\boldsymbol{u}+\boldsymbol{K}(z)\boldsymbol{d}+\boldsymbol{G}(z)\boldsymbol{e}^u+\boldsymbol{K}(z)\boldsymbol{e}^d \tag{10}$$

显然, 输入误差越大, 采样间隔区间越长。为刻画该权衡关系, 重新定义式 (2) 为

$$J(\boldsymbol{z},\boldsymbol{u},\boldsymbol{d},\boldsymbol{e}^u,\boldsymbol{e}^d)=\int_0^{\infty}r\mathrm{d}\nu \tag{11}$$

其中, $r=\boldsymbol{z}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{z}+U(\boldsymbol{u})-\gamma^2|\boldsymbol{d}|^2-\gamma_u^2|\boldsymbol{e}^u|^2-\gamma_d^2|\boldsymbol{e}^d|^2$, $\gamma_u,\gamma_d>0$ 为设计参数, 用以保证性能指标的有界性[12]。不同于已有工作 [12-13], 式 (11) 同时考虑扰动策略与事件输入误差在最坏情形的联合影响。

基于上述性能指标, 定义值函数 $V=\int_t^{\infty}r\mathrm{d}\nu$, 则构造 Hamilton 函数为

$$H=\nabla\boldsymbol{V}^{\mathrm{T}}\left(\boldsymbol{F}(z)+\boldsymbol{G}(z)\boldsymbol{u}+\boldsymbol{K}(z)\boldsymbol{d}+\boldsymbol{G}(z)\boldsymbol{e}^u+\boldsymbol{K}(z)\boldsymbol{e}^d\right)+r \tag{12}$$

其中 $\nabla\boldsymbol{V}=\partial\boldsymbol{V}/\partial\boldsymbol{z}$。依据最优性条件 ($\partial H/\partial\boldsymbol{u}=0$, $\partial H/\partial\boldsymbol{d}=0$), 可得最优控制策略为:

$$\boldsymbol{u}^{[*]}=-\frac{1}{2}R_u^{-1}\boldsymbol{G}(\hat{\boldsymbol{z}})^{\mathrm{T}}\nabla\boldsymbol{V}^{[*]}(\hat{\boldsymbol{z}}),\boldsymbol{d}^{[*]}=\frac{1}{2\gamma^2}\boldsymbol{K}(\hat{\boldsymbol{z}})^{\mathrm{T}}\nabla\boldsymbol{V}^{[*]}(\hat{\boldsymbol{z}}) \tag{13}$$

进一步将最坏情况下事件触发控制输入误差视为对抗策略, 同理可得到:

$$(\boldsymbol{e}^u)^{[*]}=\frac{1}{2\gamma_u^2}\boldsymbol{G}(\hat{\boldsymbol{z}})^{\mathrm{T}}\nabla\boldsymbol{V}^{[*]}(\hat{\boldsymbol{z}}),(\boldsymbol{e}^d)^{[*]}=\frac{1}{2\gamma_d^2}\boldsymbol{K}(\hat{\boldsymbol{z}})^{\mathrm{T}}\nabla\boldsymbol{V}^{[*]}(\hat{\boldsymbol{z}}) \tag{14}$$

将式 (13~14) 代入式 (12) 中, 可得到一类 HJI 偏微分方程。然而, 由于系统动态未知且方程高度非线性耦合, 解析求解通常不可行。为此, 本文提出一种基于积分强化学习的事件触发数据驱动学习方法。

2 最优事件触发预定性能控制器设计

2.1 最优事件触发控制器框架

求解与零和博弈对应的 HJI 方程, 传统上采用策略迭代方法[7]。但是该方法依赖系统动态精确数学模型, 这在实际系统中常难以满足。

为此, 本文采用 off-policy IRL 框架, 直接基于系统运行数据对最优策略进行学习。针对无约束系统 (式 (10)), 首先引入 off-policy 技巧, 将系统动力学重写为

$$\begin{aligned}\dot{\boldsymbol{z}}=&\boldsymbol{F}(z)+\boldsymbol{G}(z)(\boldsymbol{u}-\boldsymbol{u}^k)+\boldsymbol{G}(z)\boldsymbol{u}^k+\boldsymbol{G}(z)\boldsymbol{e}^u+\\&\boldsymbol{K}(z)(\boldsymbol{d}-\boldsymbol{d}^k)+\boldsymbol{K}(z)\boldsymbol{d}^k+\boldsymbol{K}(z)\boldsymbol{e}^d\end{aligned} \tag{15}$$

其中, $\boldsymbol{u}^k$, $\boldsymbol{d}^k$ 为具有任意初始值的目标策略, 引入行为策略 $\boldsymbol{u}$, $\boldsymbol{d}$ 作为可接受初始控制策略[14]。这允许在满足持续激励(Persistence of Excitation, PE)下自由设计探索信号。

然后, 将连续时间的最优控制问题转化为有限时间区间 $[p,p+\delta p]$ 的积分形式。对值函数 $\boldsymbol{V}^{k+1}$ 在该区间内积分, 有:

$$\boldsymbol{V}^{k+1}(\boldsymbol{z}(p+\delta p))-\boldsymbol{V}^{k+1}(\boldsymbol{z}(p))=\int_p^{p+\delta p}\nabla\boldsymbol{V}^{k+1}(z)^{\mathrm{T}}\dot{\boldsymbol{z}}\mathrm{d}\nu \tag{16}$$

结合式 (10)、式 (12) 及最优性原理, 整理得到

$$\begin{aligned}&\boldsymbol{V}^{k+1}(p)-\boldsymbol{V}^{k+1}(p+\delta p)+\int_p^{p+\delta p}(\nabla\boldsymbol{V}^{k+1})[\boldsymbol{G}(z)(\boldsymbol{u}-\boldsymbol{u}^k+\boldsymbol{e}^u)+\\&\boldsymbol{K}(z)(\boldsymbol{d}-\boldsymbol{d}^k+\boldsymbol{e}^d)]\mathrm{d}\nu=\int_p^{p+\delta p}r\mathrm{d}\nu\end{aligned} \tag{17}$$

为处理未知输入动态, 引入辅助多项式函数 $l_g=\boldsymbol{G}(z)^{\mathrm{T}}\nabla\boldsymbol{V}^{k+1}$, $l_k=\boldsymbol{K}(z)^{\mathrm{T}}\nabla\boldsymbol{V}^{k+1}$, 将式 (17) 改写为:

$$\begin{aligned}&\boldsymbol{V}^{k+1}(p)-\boldsymbol{V}^{k+1}(p+\delta p)+\int_p^{p+\delta p}l_g(z)^{\mathrm{T}}(\boldsymbol{u}-\boldsymbol{u}^k+\boldsymbol{e}^u)\mathrm{d}\nu+\\&\int_p^{p+\delta p}l_k(z)^{\mathrm{T}}(\boldsymbol{d}-\boldsymbol{d}^k+\boldsymbol{e}^d)\mathrm{d}\nu=\int_p^{p+\delta p}r\mathrm{d}\nu\end{aligned} \tag{18}$$

通过方程 (18), 将策略评估与策略改进同步进行, 仅需系统运行状态信息, 无需预先进行模型参数辨识, 即可通过学习获得近似最优控制策略。式 (18) 结合了多项式函数, 避免了文献 [8-9] 中对部分系统模型的需求。根据文献 [14] 中定理 1, 可证明式 (12) 与式 (18) 在理论上是等价的。为表达简便, 将 $\boldsymbol{V}(z)$ 与输入动态 $\boldsymbol{G}(z)$ 分别简写为 $\boldsymbol{V}$ 和 $\boldsymbol{G}$。

根据最坏情形下事件驱动误差 (式 (14)), 定义式 (8) 的最优事件触发条件为:

$$\|\boldsymbol{e}^u(t)\|^2 \leqslant T_s^k, T_s^k = \frac{1}{4{\gamma_u}^4}\nabla \boldsymbol{V}^{[*]\mathrm{T}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}\nabla \boldsymbol{V}^{[*]} \tag{19}$$

其中, T_s^k 为当前迭代步下的最优事件触发阈值, 控制输入与干扰输入均可构造对应的阈值条件, 为简化表述, 此处将仅给出控制输入误差 $\boldsymbol{e}^u$ 的触发条件与后续分析, 干扰输入的触发条件与后续分析可同理构造。

为进一步避免静态阈值可能导致的过于保守触发, 引入动态事件触发机制。借鉴动态事件触发思想[15], 定义内部动态变量 $\eta_d(p)$, 其演化规律为

$$\dot{\eta}_d(p) = -\lambda\eta_d(p) + 1/4{\gamma_u}^4\nabla \boldsymbol{V}^{[*]\mathrm{T}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}\nabla \boldsymbol{V}^{[*]} - {\gamma_u}^2\boldsymbol{e}^{u\mathrm{T}}\boldsymbol{e}^u \tag{20}$$

其中, $\eta_d(0) = \eta_d^0 \geqslant 0$, $\lambda > 0$ 为滤波系数, 用于调节历史误差信息对当前触发条件的影响[9]。基于 $\eta_d(p)$, 定义事件触发时刻为式 (21), 其中 $\beta > 0$ 设计为连接静态与动态条件的参数[9]。

$$\begin{aligned}\tau_{j+1} = \inf\Big\{p > \tau_j:\ &\eta_d(p) + \beta(1/4{\gamma_u}^4\nabla \boldsymbol{V}^{[*]\mathrm{T}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}\nabla \boldsymbol{V}^{[*]} - \\ &{\gamma_u}^2\boldsymbol{e}^{u\mathrm{T}}\boldsymbol{e}^u) \leqslant 0\Big\}\end{aligned} \tag{21}$$

引理 2[9]　若动态变量 $\eta_d(p)$ 由式 (20) 定义, 且事件触发规则由式 (21) 给出, 则对任意 $p \in [0, \tau_\infty)$, $\eta_d(p)$ 恒非负。

证明　由触发规则式 (21), 对任意 p 有

$$\eta_d(p) + \beta\left(\frac{1}{4{\gamma_u}^4}\nabla \boldsymbol{V}^{[*]\mathrm{T}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}\nabla \boldsymbol{V}^{[*]} - {\gamma_u}^2\boldsymbol{e}^{u\mathrm{T}}\boldsymbol{e}^u\right) \geqslant 0$$

当 $\beta = 0$ 时, 显然有 $\eta_d(p) \geqslant 0$。当 $\beta \neq 0$ 时, 由式 (20) 可得

$$\dot{\eta}_d(p) \geqslant -\left(\lambda + \frac{1}{\beta}\right)\eta_d(p), \eta_d(0) \geqslant 0$$

上式根据比较引理(见文献 [9] 内 Lemma 1), 可得

$$\eta_d(p) \geqslant \eta_d(0)\exp\left(-\left(\lambda + \frac{1}{\beta}\right)p\right) \geqslant 0$$

从而完成证明。

2.2　神经网络近似实现

在算法实现阶段, 采用神经网络结构作为函数逼近器, 对值函数及其辅助函数进行近似。根据通用逼近定理[16], 具有足够隐藏层与神经元数量的神经网络能够以任意精度逼近连续非线性函数。本文采用单评价网络结构进行函数拟合, 通过迭代学习最小化逼近误差, 分别获得最优值函数 $\boldsymbol{V}$、其梯度 $\nabla\boldsymbol{V}$ 以及辅助函数 l_g 和 l_k。

$$\boldsymbol{V}^k = (\boldsymbol{W}_c^k)^{\mathrm{T}}\boldsymbol{\varphi}_c(z) + \epsilon_c^k(z), \nabla\boldsymbol{V}^k = (\nabla\boldsymbol{\varphi}_c(z))^{\mathrm{T}}\boldsymbol{W}_c^k + \nabla\epsilon_c(z) \tag{22}$$

$$l_g^k = (\boldsymbol{W}_{lg}^k)^{\mathrm{T}}\boldsymbol{\varphi}_{lg}(z) + \epsilon_{lg}^k(z), l_k^k = (\boldsymbol{W}_{lk}^k)^{\mathrm{T}}\boldsymbol{\varphi}_{lk}(z) + \epsilon_{lk}^k(z) \tag{23}$$

其中, $\boldsymbol{\varphi}_c, \boldsymbol{\varphi}_{lg}, \boldsymbol{\varphi}_{lk} \in \mathbb{R}^S$ 为激活函数, S 为神经元个数, $\boldsymbol{W}_c, \boldsymbol{W}_{lg}, \boldsymbol{W}_{lk}$ 为理想权重, $\epsilon_c, \epsilon_{lg}, \epsilon_{lk}$ 为逼近残差。定义近似值函数 $\hat{\boldsymbol{V}}^k$ 及辅助函数 $\hat{l}_g^k$、$\hat{l}_k^k$:

$$\hat{\boldsymbol{V}}^k = (\hat{\boldsymbol{W}}_c^k)^{\mathrm{T}}\boldsymbol{\varphi}_c(z), \nabla\hat{\boldsymbol{V}}^k = (\nabla\boldsymbol{\varphi}_c(z))^{\mathrm{T}}\hat{\boldsymbol{W}}_c^{\,k} \tag{24}$$

$$\hat{l}_g^k = (\hat{\boldsymbol{W}}_{lg}^k)^{\mathrm{T}}\boldsymbol{\varphi}_{lg}(z), \hat{l}_k^k = (\hat{\boldsymbol{W}}_{lk}^k)^{\mathrm{T}}\boldsymbol{\varphi}_{lk}(z) \tag{25}$$

将上述近似关系代入积分等式 (18), 可定义 Bellman 近似误差为:

$$\begin{aligned}\epsilon_B^{k+1}(z(\nu), \boldsymbol{u}(\nu), \boldsymbol{d}(\nu)) = {}&(\boldsymbol{\varphi}_c(z(p+\delta p)) - \boldsymbol{\varphi}_c(z(p)))^{\mathrm{T}}\hat{\boldsymbol{W}}_c^k - \\ &\int_p^{p+\delta p}\boldsymbol{\varphi}_{lg}^{\mathrm{T}}(z)\hat{\boldsymbol{W}}_{lg}^k(\boldsymbol{u}(\nu) - \boldsymbol{u}^k(\nu) + \boldsymbol{e}^u)\mathrm{d}\nu - \\ &\int_p^{p+\delta p}\boldsymbol{\varphi}_{lk}^{\mathrm{T}}(z)\hat{\boldsymbol{W}}_{lk}^k(\boldsymbol{d}(\nu) - \boldsymbol{d}^k(\nu) + \boldsymbol{e}^d)\mathrm{d}\nu + \\ &\int_p^{p+\delta p} r\mathrm{d}\nu \triangleq \epsilon_B^{k+1}(p)\end{aligned} \tag{26}$$

定义回归向量与权重向量如下:

$$\boldsymbol{\Pi}(p) \triangleq \begin{bmatrix} \boldsymbol{\varphi}_c(z(p)) - \boldsymbol{\varphi}_c(z(p+\delta p)) \\ \int_p^{p+\delta p}(\boldsymbol{u}(\nu) - \boldsymbol{u}^k(\nu) + \boldsymbol{e}^u)\otimes\boldsymbol{\varphi}_{lg}(z)\mathrm{d}\nu \\ \int_p^{p+\delta p}(\boldsymbol{d}(\nu) - \boldsymbol{d}^k(\nu) + \boldsymbol{e}^d)\otimes\boldsymbol{\varphi}_{lk}(z)\mathrm{d}\nu \end{bmatrix}$$

$$\hat{\boldsymbol{W}} \triangleq \left[(\hat{\boldsymbol{W}}_c^k)^{\mathrm{T}}, \mathrm{vec}(\hat{\boldsymbol{W}}_{lg}^k)^{\mathrm{T}}, \mathrm{vec}(\hat{\boldsymbol{W}}_{lk}^k)^{\mathrm{T}}\right]^{\mathrm{T}} \tag{27}$$

则近似误差可表示为 $\epsilon_B^{k+1}(p) = \boldsymbol{\Pi}(p)^{\mathrm{T}}\hat{\boldsymbol{W}} - \Lambda(p)$, 其中 $\Lambda(p) = \int_p^{p+\delta p} r\mathrm{d}\nu$, 为了简化, 后文省略 p。通过最小化 ϵ_B^{k+1} 可更新网络权重, 进而实现值函数与策略的迭代学习。本文采用最小二乘法(LS)对评价网络权重进行估计, 即通过最小化平方误差 E 求解 $\hat{\boldsymbol{W}}$:

$$E = \frac{1}{2}(\epsilon_B^{k+1})^{\mathrm{T}}\epsilon_B^{k+1} \tag{28}$$

本文所用 LS 方法需利用离策略控制在线采集足够数据[7]。定义数据量为 m, 在单调递增的时间序列 $\{p_n\}_{n=0}^m$ 上可获取采样集 $\{z(p_n), \boldsymbol{u}(p_n), \boldsymbol{d}(p_n)\}_{n=0}^m$。基于采样数据批量计算权重, 式 ϵ_B^{k+1} 的权值更新为:

$$\hat{\boldsymbol{W}} = \left(\boldsymbol{\Pi}_m^{\mathrm{T}}\boldsymbol{\Pi}_m\right)^{-1}\boldsymbol{\Pi}_m^{\mathrm{T}}\boldsymbol{\Lambda}_m \tag{29}$$

其中,

$$\boldsymbol{\Pi}_m=\left[\boldsymbol{\Pi}(p_1)^{\mathrm{T}},\cdots,\boldsymbol{\Pi}(p_m)^{\mathrm{T}}\right]^{\mathrm{T}},\boldsymbol{\Lambda}_m=[\Lambda(p_1),\cdots,\Lambda(p_m)]^{\mathrm{T}} \tag{30}$$

所提方法使用单评价网络架构,无需显示求解策略函数。需要指出的是,为保证采用LS方法更新相应权重时式(29)中矩阵($\boldsymbol{\Pi}_m^{\mathrm{T}}\boldsymbol{\Pi}_m$)的可逆性,在采样期间收集数据需要满足的PE条件[7]。在在线数据采集过程中引入小幅高斯激励可有效保证该条件成立,实际应用通常需要满足采样点 $m>\mathrm{rank}(\boldsymbol{\Pi}_m)$。

2.3 稳定性分析

在动态最优事件驱动机制(式(21))的背景下严格检验系统式(15)的闭环稳定性。

假设 2[9] 系统动态函数 $\boldsymbol{F}(z),\boldsymbol{G}(z),\boldsymbol{K}(z)$ 满足一致有界,即存在正常数 L_f,g_M,k_M,使得对任意 $z\in\Omega$,有 $\boldsymbol{F}(z)\leqslant L_f,\boldsymbol{G}(z)\leqslant g_M,\boldsymbol{K}(z)\leqslant k_M$。

假设 3[8] 评价网络权重误差 $\tilde{\boldsymbol{W}}$、理想权重 $\boldsymbol{W}^{[*]}$、激活函数梯度 $\nabla\varphi$ 及近似误差梯度 $\nabla\epsilon$ 均一致有界,即存在正常数 $\tilde{\boldsymbol{W}}_M$、$\boldsymbol{W}_M^{[*]}$、$\nabla\varphi_M$、$\nabla\epsilon_M$,使得 $\|\tilde{\boldsymbol{W}}\|\leqslant\tilde{W}_M,\|\boldsymbol{W}^{[*]}\|\leqslant W_M^{[*]},\|\nabla\varphi\|\leqslant\nabla\varphi_M,\|\nabla\epsilon\|\leqslant\nabla\epsilon_M$。

定理 1 考虑到式(8),当评价网络结构如式(27)所示,权重更新律由式(29)给出,事件触发条件由式(21)设计时,若存在正常数 $\beta_1\in(0,1)$,使得状态满足式(31),无约束事件触发系统将渐进稳定。其中 Υ_1 在后续(式(44))中定义。

$$\Omega_z=\left\{z:\ \|z\|\leqslant\sqrt{\frac{\Upsilon_1}{(1-\beta_1^2)\lambda_{\min}(\boldsymbol{Q})}}\right\} \tag{31}$$

证明 给定下列李雅普诺夫函数[8-9]:

$$L(t)=L_{11}(t)+L_{12}(t)+L_{13}(t)+\overline{L}(t) \tag{32}$$

其中,

$$\begin{aligned}L_{11}(t)&=\int_p^{p+\delta p}\boldsymbol{V}^{[*]}(z(\tau))\mathrm{d}\tau\\ L_{12}(t)&=\int_p^{p+\delta p}\boldsymbol{V}^{[*]}(\hat{z}(\tau))\mathrm{d}\tau\\ L_{13}(t)&=1/2\tilde{\boldsymbol{W}}^{\mathrm{T}}\tilde{\boldsymbol{W}}\\ \overline{L}(t)&=\int_p^{p+\delta p}\eta_d(\tau)\mathrm{d}\tau\end{aligned}$$

对各项求导可得:

$$\begin{aligned}\dot{L}_{11}(t)&=\int_p^{p+\delta p}\dot{\boldsymbol{V}}^{[*]}(z(\tau))\mathrm{d}\tau\\ \dot{L}_{12}(t)&=\int_p^{p+\delta p}\dot{\boldsymbol{V}}^{[*]}(\hat{z}(\tau))\mathrm{d}\tau\\ \dot{L}_{13}(t)&=\tilde{\boldsymbol{W}}^{\mathrm{T}}\dot{\tilde{\boldsymbol{W}}},\ \dot{\overline{L}}(t)=-\lambda\int_p^{p+\delta p}\eta_{\mathrm{d}}(\tau)\mathrm{d}\tau\\ \tilde{\boldsymbol{W}}&=\boldsymbol{W}-\hat{\boldsymbol{W}}\end{aligned} \tag{33}$$

情况 1 在无事件触发区间 $t\in[\tau_j,\tau_{j+1})$ 内,有 $\hat{z}=z(\tau_j)$,因此 $\dot{L}_{12}=0$。由系统动力学及最优策略定义:

$$\begin{aligned}&\dot{\boldsymbol{V}}^{[*]}(z)=(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}(\boldsymbol{F}(z)+\boldsymbol{G}(z)\hat{\boldsymbol{u}}(\hat{z})+\boldsymbol{K}(z)\hat{\boldsymbol{d}}(\hat{z})+\boldsymbol{G}(z)\hat{\boldsymbol{e}}^u+\\ &\boldsymbol{K}(z)\hat{\boldsymbol{e}}^d)=(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}(\boldsymbol{F}(z)+\boldsymbol{G}(z)\boldsymbol{u}^{[*]}(z)+\boldsymbol{K}(z)\boldsymbol{d}^{[*]}(z))-\\ &(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}\boldsymbol{G}(z)(\boldsymbol{u}^{[*]}(z)-\hat{\boldsymbol{u}}(\hat{z}))-(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}\boldsymbol{K}(z)(\boldsymbol{d}^{[*]}(z)-\hat{\boldsymbol{d}}(\hat{z}))+\\ &(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}\boldsymbol{G}(z)\boldsymbol{e}^{u[*]}+(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}\boldsymbol{K}(z)\boldsymbol{e}^{d[*]}-\\ &(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}\boldsymbol{G}(z)(\boldsymbol{e}^{u[*]}-\hat{\boldsymbol{e}}^u)-(\nabla\boldsymbol{V}^{[*]})^{\mathrm{T}}\boldsymbol{K}(z)(\boldsymbol{e}^{d[*]}-\hat{\boldsymbol{e}}^d)=\\ &-z^{\mathrm{T}}\boldsymbol{Q}z-\boldsymbol{u}^{[*]\mathrm{T}}R_u\boldsymbol{u}^{[*]}+\gamma^2\boldsymbol{d}^{[*]\mathrm{T}}\boldsymbol{d}^{[*]}+\gamma_u^2\boldsymbol{e}^{u[*]\mathrm{T}}\boldsymbol{e}^{u[*]}-\\ &\gamma_d^2\boldsymbol{e}^{d[*]\mathrm{T}}\boldsymbol{e}^{d[*]}+2R_u\boldsymbol{u}^{[*]\mathrm{T}}(\boldsymbol{u}^{[*]}-\hat{\boldsymbol{u}}(\hat{z}))-2\gamma^2\boldsymbol{d}^{[*]\mathrm{T}}(\boldsymbol{d}^{[*]}-\hat{\boldsymbol{d}}(\hat{z}))-\\ &2\gamma_u^2\boldsymbol{e}^{u[*]\mathrm{T}}(\boldsymbol{e}^{u[*]}-\hat{\boldsymbol{e}}^u)+2\gamma_d^2\boldsymbol{e}^{d[*]\mathrm{T}}(\boldsymbol{e}^{d[*]}-\hat{\boldsymbol{e}}^d)\end{aligned} \tag{34}$$

合并同类项,放缩:

$$\begin{aligned}\dot{\boldsymbol{V}}^{[*]}(z)=&-z^{\mathrm{T}}\boldsymbol{Q}z+(\boldsymbol{u}^{[*]}-\hat{\boldsymbol{u}}(\hat{z}))^{\mathrm{T}}R_u(\boldsymbol{u}^{[*]}-\hat{\boldsymbol{u}}(\hat{z}))-\\ &\hat{\boldsymbol{u}}(\hat{z})^{\mathrm{T}}R_u\hat{\boldsymbol{u}}(\hat{z})-\gamma_u^2|\boldsymbol{e}^{u[*]}-\hat{\boldsymbol{e}}^u|^2+\\ &\gamma_u^2|\hat{\boldsymbol{e}}^u|^2-\gamma_d^2|\boldsymbol{d}^{[*]}-\hat{\boldsymbol{d}}(\hat{z})|^2+\gamma_d^2|\hat{\boldsymbol{d}}(\hat{z})|^2\leqslant\\ &-z^{\mathrm{T}}\boldsymbol{Q}z+(\boldsymbol{u}^{[*]}-\hat{\boldsymbol{u}}(\hat{z}))^{\mathrm{T}}R_u(\boldsymbol{u}^{[*]}-\hat{\boldsymbol{u}}(\hat{z}))-\\ &\hat{\boldsymbol{u}}(\hat{z})^{\mathrm{T}}R_u\hat{\boldsymbol{u}}(\hat{z})+\gamma_u^2|\hat{\boldsymbol{e}}^u|^2+\gamma_d^2|\hat{\boldsymbol{d}}(\hat{z})|^2\end{aligned} \tag{35}$$

根据式(22)和 $\tilde{\boldsymbol{W}}_c=\boldsymbol{W}_c-\hat{\boldsymbol{W}}_c$,

$$\begin{aligned}|\boldsymbol{u}^{[*]}-\hat{\boldsymbol{u}}(\hat{z})|^2=&|-\frac{1}{2}R_u^{-1}\boldsymbol{G}(z)^{\mathrm{T}}(\nabla\boldsymbol{\varphi}_c(z)^{\mathrm{T}}(\tilde{\boldsymbol{W}}_c+\hat{\boldsymbol{W}}_c)+\nabla\epsilon_c)-\\ &\boldsymbol{G}(\hat{z})^{\mathrm{T}}\nabla\boldsymbol{\varphi}_c(\hat{z})^{\mathrm{T}}\hat{\boldsymbol{W}}_c|^2=\frac{1}{4}R_u^{-2}|\boldsymbol{G}(z)^{\mathrm{T}}\nabla\boldsymbol{\varphi}_c(z)^{\mathrm{T}}\tilde{\boldsymbol{W}}_c+\nabla\epsilon_c+\\ &(\boldsymbol{G}(z)^{\mathrm{T}}\nabla\boldsymbol{\varphi}_c(z)^{\mathrm{T}}-\boldsymbol{G}(\hat{z})^{\mathrm{T}}\nabla\boldsymbol{\varphi}_c(\hat{z})^{\mathrm{T}})\hat{\boldsymbol{W}}_c|^2\end{aligned} \tag{36}$$

根据假设3可推理得:

$$\begin{aligned}&|\boldsymbol{G}(z)^{\mathrm{T}}\nabla\boldsymbol{\varphi}_c(z)^{\mathrm{T}}-\boldsymbol{G}(\hat{z})^{\mathrm{T}}\nabla\boldsymbol{\varphi}_c(\hat{z})^{\mathrm{T}}|^2=\\ &\left|\nabla\boldsymbol{\varphi}_c(z)^{\mathrm{T}}(\boldsymbol{G}(z)-\boldsymbol{G}(\hat{z}))^{\mathrm{T}}+(\nabla\boldsymbol{\varphi}_c(z)-\nabla\boldsymbol{\varphi}_c(\hat{z}))^{\mathrm{T}}\boldsymbol{G}(\hat{z})^{\mathrm{T}}\right|^2\leqslant\\ &2\left(\left|\nabla\boldsymbol{\varphi}_c(z)^{\mathrm{T}}(\boldsymbol{G}(z)-\boldsymbol{G}(\hat{z}))^{\mathrm{T}}\right|^2+\left|(\nabla\boldsymbol{\varphi}_c(z)-\nabla\boldsymbol{\varphi}_c(\hat{z}))^{\mathrm{T}}\boldsymbol{G}(\hat{z})^{\mathrm{T}}\right|^2\right)\leqslant\\ &8(\nabla\boldsymbol{\varphi}_M{}^2g_M{}^2)=G_{M_1}^2\end{aligned} \tag{37}$$

其中,G_{M_1} 是正常数。由控制律(式(13))、神经网络近似关系(式(22))及假设3,可得存在正常数 f_1,f_2,使得

$$\begin{aligned}R_u|\boldsymbol{u}^{[*]}-\hat{\boldsymbol{u}}(\hat{z})|^2\leqslant&\frac{1}{2}R_u^{-1}g_M^2\nabla\varphi_M^2|\tilde{\boldsymbol{W}}_c|^2+\\ &\frac{1}{2}R_u^{-1}g_M^2\nabla\epsilon_M^2+R_u^{-1}G_{M_1}^2|\hat{\boldsymbol{W}}_c|^2=\\ &R_u^{-1}G_{M_1}^2|\hat{\boldsymbol{W}}_c|^2+f_1|\tilde{\boldsymbol{W}}_c|^2+f_2\triangleq\Phi_u\end{aligned} \tag{38}$$

以下 $L_{ij}(t)$($i,j\in\{1,2\}$)简写为 L_{ij},从而有

$$\begin{aligned}\dot{L}_{11}\leqslant&-z^{\mathrm{T}}\boldsymbol{Q}z-\hat{\boldsymbol{u}}(\hat{z})^{\mathrm{T}}R_u\hat{\boldsymbol{u}}(\hat{z})+\gamma_d^2\hat{\boldsymbol{d}}(\hat{z})^{\mathrm{T}}\hat{\boldsymbol{d}}(\hat{z})+\Phi_u+\\ &\frac{1}{4\gamma^2}\nabla\hat{\boldsymbol{V}}^{\mathrm{T}}\boldsymbol{G}(z)\boldsymbol{G}(z)^{\mathrm{T}}\nabla\hat{\boldsymbol{V}}\end{aligned} \tag{39}$$

事件没有触发时,下面对 $\dot{L}_{13}$ 进行分析,已知:

$$\dot{\hat{\boldsymbol{W}}}=-\alpha\frac{\boldsymbol{\Pi}^{\mathrm{T}}}{(\boldsymbol{\Pi}\boldsymbol{\Pi}^{\mathrm{T}}+1)^2}\epsilon_B^{k+1} \tag{40}$$

根据式(11)、式(28),$\epsilon_B^{k+1}=\boldsymbol{\Pi}^{\mathrm{T}}\hat{\boldsymbol{W}}-\Lambda$,且 $\tilde{\boldsymbol{W}}=\boldsymbol{W}-$

$\hat{\boldsymbol{W}}$ 可知

$$\epsilon_B^{k+1} = \boldsymbol{\Pi}(\boldsymbol{W}-\tilde{\boldsymbol{W}}) + \int_p^{p+\delta p} \hat{r}\mathrm{d}\nu, \hat{r} = \boldsymbol{z}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{z} + \hat{U} \tag{41}$$

$$\hat{U} = \boldsymbol{u}^{\mathrm{T}}R_u\boldsymbol{u} - \gamma^2\boldsymbol{d}^{\mathrm{T}}\boldsymbol{d} - \gamma_u^2(\hat{\boldsymbol{e}}^u)^{\mathrm{T}}\hat{\boldsymbol{e}}^u + \gamma_d^2(\hat{\boldsymbol{e}}^d)^{\mathrm{T}}\hat{\boldsymbol{e}}^d \tag{42}$$

根据式 (33):

$$\begin{aligned}\dot{L}_{13} =& \tilde{\boldsymbol{W}}^{\mathrm{T}}\dot{\tilde{\boldsymbol{W}}} = \frac{\tilde{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{\Pi}(-\tilde{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{\Pi}+\boldsymbol{\Pi}\boldsymbol{W}+\hat{r})}{(\boldsymbol{\Pi}\boldsymbol{\Pi}^{\mathrm{T}}+1)^2} \leqslant \\ & \frac{\epsilon_M^2 - \tilde{\boldsymbol{W}}^{\mathrm{T}}\boldsymbol{\Pi}\boldsymbol{\Pi}^{\mathrm{T}}\tilde{\boldsymbol{W}}}{2(1+\boldsymbol{\Pi}\boldsymbol{\Pi}^{\mathrm{T}})^2} \leqslant \frac{\epsilon_M^2 - \lambda_{\min}(\boldsymbol{\Pi}\boldsymbol{\Pi}^{\mathrm{T}})|\tilde{\boldsymbol{W}}|^2}{2(1+\lambda_{\max}(\boldsymbol{\Pi}\boldsymbol{\Pi}^{\mathrm{T}}))^2} = \\ & G_{M_2} - G_{M_3}|\tilde{\boldsymbol{W}}|^2\end{aligned} \tag{43}$$

其中, ϵ_M 为误差上界, $G_{M_2}, G_{M_3} > 0$ 为常数。所以综合上述结果, 满足式 (31) 时下式 $\dot{L}_1 \leqslant 0$:

$$\begin{aligned}\dot{L}_1 \leqslant & -\boldsymbol{z}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{z} + \gamma_d^2\hat{\boldsymbol{d}}(\hat{\boldsymbol{z}})^{\mathrm{T}}\hat{\boldsymbol{d}}(\hat{\boldsymbol{z}}) + \Phi_u - \lambda\int_p^{p+\delta p}\eta_d(\tau)\mathrm{d}\tau + \\ & \frac{1}{4\gamma^2}\nabla\hat{\boldsymbol{V}}^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{z})\boldsymbol{G}(\boldsymbol{z})^{\mathrm{T}}\nabla\hat{\boldsymbol{V}} + G_{M_2} - \boldsymbol{G}_{M_3}|\tilde{\boldsymbol{W}}|^2 \leqslant \\ & -\beta_1^2\lambda_{\min}(\boldsymbol{Q})|\boldsymbol{z}|^2 - (1-\beta_1^2)\lambda_{\min}(\boldsymbol{Q})|\boldsymbol{z}|^2 + \Upsilon_1\end{aligned} \tag{44}$$

其中, $\beta_1 > 0$, 且 $\Upsilon_1 = \Phi_u + 1/4\gamma^2\nabla\hat{\boldsymbol{V}}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{G}^{\mathrm{T}}\nabla\hat{\boldsymbol{V}}_i + \gamma_d^2\hat{\boldsymbol{d}}^{\mathrm{T}}(\hat{\boldsymbol{z}})\hat{\boldsymbol{d}}(\hat{\boldsymbol{z}}) + G_{M_2} - G_{M_3}\|\tilde{\boldsymbol{W}}\|^2$。

情况 2　在事件触发执行期间, 李雅普诺夫函数的变化量可写为:

$$\Delta L_1 = \int_p^{p+\delta p}\Delta L_{11}\mathrm{d}\nu + \int_p^{p+\delta p}\Delta L_{12}\mathrm{d}\nu + \Delta L_{13} + \int_p^{p+\delta p}\Delta\overline{L}\mathrm{d}\nu \tag{45}$$

由于事件触发机制在触发时刻之外不改变系统状态与权重, 故 $\Delta L_{11} = \Delta L_{13} = \Delta\overline{L} = 0$。而在事件触发瞬间, 状态与估计值更新满足:

$$\Delta L_{12} \leqslant -\kappa(|\hat{\boldsymbol{z}}(\tau_{j+1}) - \hat{\boldsymbol{z}}(\tau_j)|) \leqslant -\kappa(|\boldsymbol{e}^z(\tau)|) < 0 \tag{46}$$

其中 $\kappa(\cdot)$ 根据文献 [9] 表示初值为 0 的严格增的 $\mathcal{K}$ 类函数, $\boldsymbol{e}^z(\tau)$ 表示状态触发误差。满足条件, 证明结束。

引理 3[8]　对无约束系统式 (8), 具有规定的动态事件触发规则 (式 (21)), 则相邻触发时刻的最小间隔严格大于 0。

证明　根据测量状态误差定义, 我们可知: $\dot{\boldsymbol{e}}^z = \dot{\hat{\boldsymbol{z}}} - \dot{\boldsymbol{z}} = -\dot{\boldsymbol{z}}$, $\boldsymbol{e}^u = \hat{\boldsymbol{u}} - \boldsymbol{u}$, $\boldsymbol{e}^d = \hat{\boldsymbol{d}} - \boldsymbol{d}$。对等式两边取范数, 结合式 (8) 与假设 2, 可以建立与文献 [8] 的定义 (3) 一致的表达式:

$$\begin{aligned}\|\dot{\boldsymbol{e}}^z\| = \|-\dot{\boldsymbol{z}}\| = & \left\|\boldsymbol{F}(\boldsymbol{z}) + \boldsymbol{G}(\boldsymbol{z})\hat{\boldsymbol{u}} + \boldsymbol{K}(\boldsymbol{z})\hat{\boldsymbol{d}}\right\| \leqslant \\ & \|\boldsymbol{F}(\boldsymbol{z})\| + \|\boldsymbol{G}(\boldsymbol{z})\hat{\boldsymbol{u}} + \boldsymbol{K}(\boldsymbol{z})\hat{\boldsymbol{d}}\| \leqslant \\ & L_f\|\hat{\boldsymbol{z}} - \boldsymbol{e}^z\| + L_g \leqslant \\ & L_f\|\hat{\boldsymbol{z}}\| + L_f\|\boldsymbol{e}^z\| + L_g\end{aligned} \tag{47}$$

由文献 [8] 的定义 (3) 可以推知 $|\boldsymbol{e}^u| \leqslant L_e|\hat{\boldsymbol{e}}^z|$, 且 $L_f \geqslant 0$、$\boldsymbol{F}(\boldsymbol{z}) \leqslant L_f|\boldsymbol{z}|$, 后续推导一致, 直接获取结论:

$$\tau_{j+1} - \tau_j \geqslant \frac{1}{L_f}\ln\left(1 + \frac{1}{\Omega}\sqrt{\frac{\frac{1}{4\gamma_u^2}\nabla\hat{\boldsymbol{V}}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{G}\nabla\hat{\boldsymbol{V}} + \frac{\eta_d}{\beta}}{\gamma_u^2}}\right) \tag{48}$$

其中 $\Omega = L_e(L_f|\hat{\boldsymbol{z}}| + L_g)/L_f$。因此不存在 Zeno 行为, 证明结束。

3　仿真示例

在该部分, 通过一个仿真示例, 验证本文提出的预定性能约束下最优事件触发控制方法对零和博弈系统控制的有效性。考虑如下非线性仿射系统[1]:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x})\boldsymbol{u}(\boldsymbol{x}) + \boldsymbol{k}(\boldsymbol{x})\boldsymbol{d}(\boldsymbol{x}) \tag{49}$$

其中系统动态函数为:

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} -2x_1 + x_2 \\ -x_2 - 0.5x_1 + 0.25x_2(\cos(2x_1)+2)^2 \\ -0.25x_2\dfrac{1}{\gamma^2}(\sin(4x_1)^2+2)^2 \end{bmatrix} \tag{50}$$

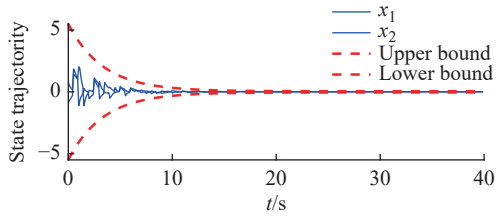
$$\boldsymbol{g}(\boldsymbol{x}) = \begin{bmatrix} 0 \\ \cos(2x_1)+2 \end{bmatrix}, \boldsymbol{k}(\boldsymbol{x}) = \begin{bmatrix} 0 \\ \sin(4x_1)^2+2 \end{bmatrix}$$

在仿真验证过程中, 式 (49) 不提供模型信息, 仅基于输入输出数据进行学习。采样周期设置为 $T = 0.1$ s, 仿真时长为 40 s, 系统初始状态设置为 $\boldsymbol{x}_0 = [1, -1]^{\mathrm{T}}$。代价函数相关参数选取为: $\boldsymbol{Q} = 2\boldsymbol{x}_0^{\mathrm{T}}\boldsymbol{x}_0$, $R_u = 1$, $\gamma = 2$, $\gamma_u = 4$, $\gamma_d = 4$。动态事件触发条件参数 $\lambda = 0.95, \beta = 0.4$。预定性能函数设计为: $\boldsymbol{\mu}(t) = 5.45\mathrm{e}^{-0.3t} + 0.05$, 即 $\boldsymbol{\mu}_0 = [5.5, 5.5]^{\mathrm{T}}$, $\boldsymbol{\mu}_\infty = [0.05, 0.05]^{\mathrm{T}}$, 衰减参数 $\beta = 0.3$, $\underline{\varsigma} = \overline{\varsigma} = 1$。

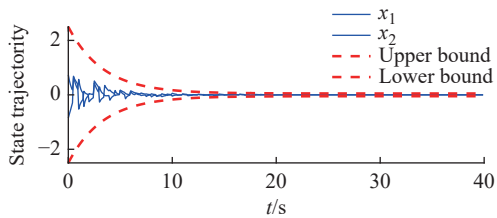
应用本文提出的预定性能最优事件触发控制方法, 评价网络激活函数选取为 $\boldsymbol{\varphi}_c = [z_1^2, z_1z_2, z_2^2]^{\mathrm{T}}$, 辅助网络激活函数选取为 $\boldsymbol{\varphi}_{lg} = \boldsymbol{\varphi}_{lk} = [z_1, z_2, z_1z_2, z_1^2, z_2^2]^{\mathrm{T}}$, wc_i $(i = 1,2,3)$、wlg_i, $wlk_i (i = 1,\cdots,5)$ 分别表示激活函数的权值。根据经验, 选择 $u_0 = z_1 + z_2 + \cos(2z_1)$, $d_0 = z_1 + z_2 + \sin(4z_1)$, 其余所有权值均在 $[0,1]$ 间进行随机初始化。每收集 $l = 20$ 个数据点后, 使用 LS 对权值更新, 并在仿真初始的 20 s 内添加幅值逐渐衰减的高斯探索噪声 $\delta\boldsymbol{x} = \boldsymbol{\mu}(t)\boldsymbol{\xi}(t)$, $\boldsymbol{\xi}(t) \in [-1,1]$ 促进系统探索。

图 1 示出了在预定性能约束下系统的状态轨迹, 并比较了不同初始边界参数 μ_0 对状态收敛行为的影响。结果显示, 经过预定性能函数约束后的系统状态始终保持在设定边界内, 验证了本文控制方

法的有效性。图 2 给出了转换后系统的最优控制策略。与连续更新策略相比，所提出的事件触发机制仅在触发条件满足时更新控制量，在确保系统稳定的同时显著降低了更新频率，减少了计算负担。图 3 示出了系统权值收敛过程，可见在迭代过程中权值逐渐趋于稳定。



(a) μ_0=5.5



(b) μ_0=2.5

图 1 预定性能约束下系统状态轨迹及参数μ_0 的影响分析

Fig. 1 Analysis of system state trajectory and the influence of parameter μ_0 under prescribed performance constraints

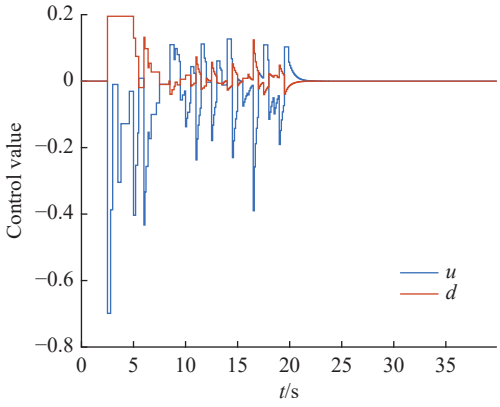


图 2 事件触发控制信号

Fig. 2 Event-triggered control signals

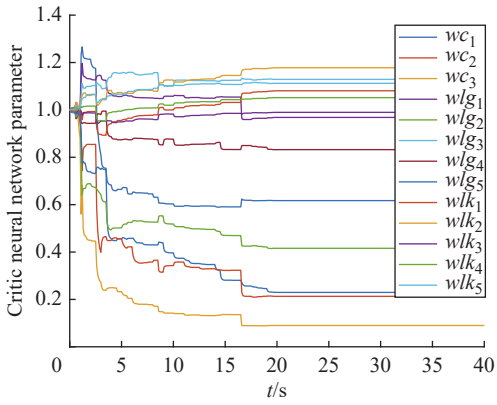
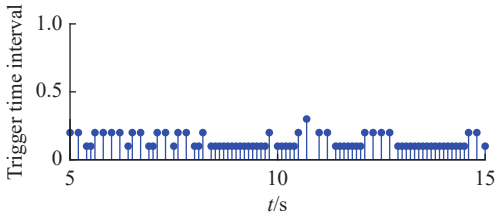


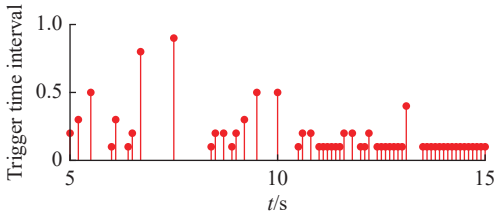
图 3 系统权值收敛过程

Fig. 3 Convergence process of system weights

图 4 示出了仿真中的事件触发间隔，结果表明系统未出现 Zeno 行为。图 5 对比了时间触发(TTC)、文献 [3] 中的静态事件触发(SETC)、文献 [9] 中的动态事件触发(DETC)以及本文最优事件触发(OS-ETC)四种机制下的采样数量。其中，为保证公平性，复现优化 SETC 中固定值 r 为 0.1，其余参数根据原文设置，而 DETC 触发条件参数 λ、β 与本文方法一致。结果显示，相比 SETC、DETC 机制，本文方法采样数量分别减少 21.7%、6.7%，明显降低了计算开销，凸显了算法的优势。此外，本文提出的方法可对博弈系统中不同参与者分别设计事件触发机制，具备较好的扩展性与普适性。



(a) Control u



(b) Control d

图 4 最优事件触发时间间隔

Fig. 4 Optimal event-triggering time intervals

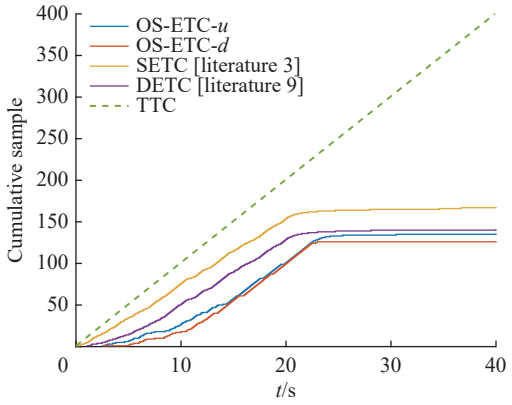


图 5 不同触发机制下的采样数量比较

Fig. 5 Comparison of sampling quantities under different triggering mechanisms

4 结论与展望

本文针对具有预定性能约束且动态未知的非线

性零和博弈系统, 提出了一种基于 off policy IRL 与事件触发机制的无模型最优控制方法。该方法通过 PPC 将约束问题转化为无约束形式, 结合 IRL 与多项式拟合实现了完全数据驱动的控制律求解, 并设计了一种基于博弈理论的事件触发机制以降低计算负担。仿真结果表明, 在不同性能约束参数下系统状态轨迹均能在 10 s 内收敛至平衡点的预设邻域内; 与 SETC 与 DETC 相比, 本文方法在 40 s 仿真时长内分别减少采样数量 21.7% 和 6.7%。需要指出的是, 本文方法主要适用于仿射非线性系统, 且神经网络的结构与参数需依据经验进行选取, 这是该工作的主要局限性。

未来的研究方向包括: (1)将所提框架拓展至非仿射或更一般的混合零和博弈系统; (2)探索神经网络参数的在线自适应调整策略, 以进一步减少人工干预并提升算法对复杂环境的适应能力。

参考文献:

[1] MU C X, WANG K. Single-network ADP for near optimal control of continuous-time zero-sum games without using initial stabilising control laws[J]. IET Control Theory & Applications, 2018, 12(18): 2449-2458.

[2] LI S B, SHAN R M, WANG J, *et al*. H∞ tracking control for de-oiling hydrocyclone systems via event-triggered reinforcement learning[J]. Journal of the Franklin Institute, 2024, 361(12): 107006.

[3] TANG Y, LV Y F, ZHAO J, *et al*. Prescribed performance event-triggered optimal control of nonlinear multi-input systems[J]. Neurocomputing, 2025, 637: 130044.

[4] HU Y Y, CHEN J Q, QIN C B. Event-triggered prescribed performance control of the multiplayer game nonlinear system via integral reinforcement learning[J]. Applied Mathematics and Computation, 2026, 511: 129716.

[5] WANG D, MU C X, ZHANG Q C, *et al*. Event-based input-constrained nonlinear H∞ state feedback with adaptive critic and neural implementation[J]. Neurocomputing, 2016, 214: 848-856.

[6] XUE S, LUO B, LIU D R, *et al*. Constrained event-triggered H∞ control based on adaptive dynamic programming with concurrent learning[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2020, 52(1): 357-369.

[7] HUO Y, WANG D, QIAO J F, *et al*. Off-policy model-free learning for multi-player non-zero-sum games with constrained inputs[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2022, 70(2): 910-920.

[8] ZHANG Y W, ZHONG W F, ZHOU G X, *et al*. Model-free game-based dynamic event-driven safety-critical control of unknown nonaffine systems[J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2024, 72(9): 5019-5032.

[9] MU C X, WANG K, QIU T. Dynamic event-triggering neural learning control for partially unknown nonlinear systems[J]. IEEE Transactions on Cybernetics, 2020, 52(4): 2200-2213.

[10] 赵刚, 王梦灵, 薛斌强, 等. 基于自适应事件触发的交通网络预测控制 [J]. 华东理工大学学报 (自然科学版), 2021, 47(3): 316-322.

[11] 陈泽彬, 余昭旭, 李树刚. 网络攻击下非线性 CPSs 的事件触发自适应控制 [J]. 华东理工大学学报 (自然科学版), 2024, 50(3): 411-417.

[12] SAHOO A, NARAYANAN V, JAGANNATHAN S. A min–max approach to event- and self-triggered sampling and regulation of linear systems[J]. IEEE Transactions on Industrial Electronics, 2018, 66(7): 5433-5440.

[13] XU W Q, WANG T, QIU J B, *et al*. A novel framework for game-based optimal event-triggered control of multi-input nonlinear systems[J]. IEEE Transactions on Automatic Control, 2024, 69(7): 4867-4874.

[14] LUO B, WU H N, HUANG T W, *et al*. Reinforcement learning solution for HJB equation arising in constrained optimal control problem[J]. Neural Networks, 2015, 71: 150-158.

[15] GIRARD A. Dynamic triggering mechanisms for event-triggered control[J]. IEEE Transactions on Automatic Control, 2014, 60(7): 1992-1997.

[16] ABU-KHALAF M, LEWIS F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach[J]. Automatica, 2005, 41(5): 779-791.

Optimal Event-Triggered Control for Nonlinear Game Systems with Prescribed Performance Constraints

ZHANG Zhixiang, WANG Mengling

(Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China)

Abstract: This paper proposes an optimal event-triggering control method based on prescribed performance constraints for nonlinear zero-sum game systems with unknown dynamics, achieving prescribed performance control for systems with unknown dynamics under reduced computational resources. First, by introducing a prescribed performance function, the constrained system control problem is transformed into an equivalent unconstrained optimal control problem. Second, leveraging an off-policy data iteration mechanism, combined with polynomial fitting techniques and the Integral Reinforcement Learning (IRL) method, a model-free iterative control strategy is derived within a purely data-driven control law framework, addressing the challenges posed by unknown system dynamics. Furthermore, an event-triggering mechanism based on game theory is designed, effectively reducing the number of samples and computational burden by modeling event-triggering sampling errors and control strategies as a zero-sum game. Then, a single critic neural network is employed to approximate the optimal control solution, with the proposed strategy used to update the network weights. Finally, theoretical analysis and simulation results demonstrate the convergence and effectiveness of the proposed method.

Key words: zero-sum game; adaptive dynamic programming; event-triggering control; prescribed performance control; integral reinforcement learning

（责任编辑：张欣）