基于自适应动态规划的移动机器人视觉伺服跟踪控制

罗 彪 [1] 欧阳志华 [1] 易昕宁 [1] 刘德荣 [2]

摘 要 针对移动机器人视觉伺服跟踪控制问题, 提出一种基于自适应动态规划 (Adaptive dynamic programming, ADP) 的控制方法. 通过移动机器人上的相机拍摄共面特征点的当前图像、期望图像以及参考图像, 利用单应性技术得到移动机器人当前的位姿信息与期望的位姿信息 (即平移量与旋转角度), 从而通过当前与期望的平移旋转之间差值得到系统的开环误差模型. 进而, 针对此系统设计最优控制器, 同时做合适的控制输入变换. 在此基础上设计一个基于 ADP 的视觉伺服控制方法以保证移动机器人完成轨迹跟踪任务. 为求出最优控制输入, 采用一个评价神经网络近似值函数, 通过不断学习逼近哈密顿−雅可比−贝尔曼 (Hamilton-Jacobi-Bellman, HJB) 方程的解. 与以往不同的是, 由于系统存在时变项, 导致 HJB 方程也含有时变项, 因此需要设计具有时变权值结构的神经网络近似值函数. 最终证明在所设计的控制方法作用下, 闭环系统是一致最终有界的.

关键词 自适应动态规划, 移动机器人, 视觉伺服, 轨迹跟踪, 神经网络控制

引用格式 罗彪, 欧阳志华, 易昕宁, 刘德荣. 基于自适应动态规划的移动机器人视觉伺服跟踪控制. 自动化学报, 2023, **49**(11): 2286−2296

DOI 10.16383/j.aas.c211230

Adaptive Dynamic Programming Based Visual Servoing Tracking Control for Mobile Robots

LUO Biao[1] OUYANG Zhi-Hua[1] YI Xin-Ning[1] LIU De-Rong[2]

Abstract In this paper, a visual servoing approach based on adaptive dynamic programming (ADP) is developed for the trajectory tracking control of mobile robots. First, according to the current image, the desired image and reference image sequence of coplanar feature points are captured by the on-board camera. The current pose information and desired pose information of the mobile robot can be reconstructed by homography technology. Then, the open-loop error model of the system is obtained by the difference between the current translation and rotation and the desired. In order to design the optimal controller for this system, the appropriate control input transformation is adopted. Therefore, a visual servoing approach based on ADP is proposed to achieve the trajectory tracking task for the mobile robot. A critic neural network structure is used to learn the time-varying solution, namely the optimal value function, of the Hamilton-Jacobi-Bellman (HJB) equation. Owing to the existence of time-varying terms, which is different from many existing works, the HJB equation is time-varying. Therefore, a neural network with a time-varying weight structure is designed to approximate the time-dependent value function of the HJB equation. Finally, it is proved that the approach proposed in this paper guarantees the closed-loop system is uniformly ultimately bounded.

Key words Adaptive dynamic programming (ADP), mobile robot, visual servoing, trajectory tracking, neural network control

Citation Luo Biao, Ouyang Zhi-Hua, Yi Xin-Ning, Liu De-Rong. Adaptive dynamic programming based visual servoing tracking control for mobile robots. *Acta Automatica Sinica*, 2023, **49**(11): 2286−2296

随着移动机器人控制技术的发展, 移动机器人如今已广泛应用于工业生产、国防军事以及生活服务等众多方面[1]. 视觉传感器近年来逐渐变得价格经济、易于获取, 加上其本身具有获取环境信息丰富、外部感知能力强和适用范围广等特点, 因而广泛装备于移动机器人和其他智能体系统[2]. 由于视

收稿日期 2021-12-24 录用日期 2022-10-17
Manuscript received December 24, 2021; accepted October 17, 2022
国家自然科学基金 (62022094, 62373375), 湖南省自然科学基金 (2020JJ2049), 之江实验室开放课题 (2021NB0AB01), 中南大学创新驱动项目 (2020CX032) 资助
Supported by National Natural Science Foundation of China (62022094, 62373375), Hunan Provincial Natural Science Foundation of China (2020JJ2049), Zhejiang Lab (2021NB0AB01), and Innovation-Driven Project of Central South University (2020CX032)
本文责任编委 刘艳军
Recommended by Associate Editor LIU Yan-Jun
1. 中南大学自动化学院 长沙 410083 2. 南方科技大学工学院 深圳 518055
1. School of Automation, Central South University, Changsha 410083 2. School of Engineering, Southern University of Science and Technology, Shenzhen 518055

觉传感器的作用, 移动机器人的灵活性和智能性得到极大提高, 可以适应更加复杂的环境, 基于视觉的机器人的定位、环境感知与控制技术也得到了很大的发展[3]. 移动机器人视觉伺服控制主要有两大控制目标, 一种是位姿校正[4], 另一种是轨迹跟踪[5]. 在实际应用中, 轨迹跟踪相比位姿校正往往更加复杂且常见, 因此基于视觉的移动机器人轨迹跟踪吸引了大量学者的关注与研究.

经典的视觉伺服控制主要分为基于图像的、基于位置的和混合视觉伺服控制. 基于图像的视觉伺服控制是通过在二维图像平面定义误差信号进行控制, 不需要利用移动机器人三维位姿信息, 且其对系统扰动具有鲁棒性, 但是很难控制移动机器人的偏转位姿[6]. 基于位置的视觉伺服控制需要在三维欧氏空间定义误差信号, 虽然需要进行三维重构, 但这样直接对三维空间的误差进行控制可以保证其收敛性. 由于移动机器人具有非完整约束特点以及单目相机深度信息的缺失造成系统带有不确定性参数, 设计移动机器人的控制器存在很多困难[7]. 根据 Brockett 定理, 连续定常的控制器无法实现对具有非完整约束的移动机器人位姿校正控制[8]. 为了克服非完整约束, 文献 [9] 提出了一种基于反步法的时变状态反馈跟踪控制方法. 然而其并未将视觉传感器引入到控制中, 并且一般假设系统状态是精确可测量的. 但在移动机器人的视觉伺服跟踪控制中, 由于视觉传感器的引入给系统带来了不确定性, 以往的设计方法并不能直接应用于视觉伺服控制中. 为了克服这些限制, 目前已提出了许多非线性控制方法. 文献 [5] 基于单应性技术设计了一种自适应控制器实现移动机器人的轨迹跟踪任务, 其视觉反馈由搭载于机器人上的相机提供. 为保持目标特征点在摄像机的视野范围内, 文献 [10] 设计了一种时变连续的混合视觉伺服控制器实现对移动机器人一致性跟踪与位姿校正任务. 为了在无需知道移动机器人位姿与速度信息下完成轨迹跟踪任务, 文献 [11] 提出了一种自适应控制方法实时估计移动机器人的位姿与速度. 近期, 也有一系列关于移动机器人的视觉伺服跟踪控制的相关成果[12−15]. 在上述针对移动机器人的视觉伺服跟踪控制中, 大多是为完成视觉跟踪任务, 但考虑最优性能指标的移动机器人视觉伺服最优跟踪控制问题仍待研究.

考虑到移动机器人的视觉伺服轨迹跟踪控制问题的系统模型是一个时变仿射非线性系统. 自适应动态规划 (Adaptive dynamic programming, ADP) 是解决非线性最优控制问题的一种有效的方法[16−21]. ADP 作为一类智能控制方法, 可以有效地解决传统动态规划中“维数灾”的问题, 同时具备处理复杂约束和不确定性的能力. 这类方法在处理具有强非线性、强耦合性的复杂非线性系统时, 具有自适应性、最优性和稳定性[22]. 求解基于 ADP 的移动机器人视觉伺服轨迹跟踪的最优控制, 需要求解哈密顿–雅可比–贝尔曼 (Hamilton-Jacobi-Bellman, HJB) 方程. 对于非线性定常系统来说, 其 HJB 方程是一个时不变的偏微分方程, 无法求得其解析解, 目前大部分求解方法都是设计合适的神经网络结构来近似未知项[23−24]. 文献 [25] 针对非线性不确定系统设计了三个神经网络分别近似系统状态、值函数和最优控制. 在系统模型已知的情况下, 文献 [26] 针对移动机器人控制问题提出了一种单网络自适应评价方法. 与上述一般非线性系统不同的是, 由于移动机器人系统模型固有的时变特性, 最优控制问题中的 HJB 方程含有两项偏微分项, 且值函数是状态与时间的函数, 这使得问题更加复杂. 在现有的研究中针对时变系统的最优控制还较少. 目前有两种神经网络的结构用于近似与时间相关的值函数: 1) 带时变权值的神经网络[27−28]; 2) 常数权值与时变激活函数的神经网络[29−30].

当前基于 ADP 的移动机器人视觉伺服控制方法及理论仍然是一个开放性问题, 有待深入研究. 本文的贡献主要体现在以下几个方面.

1) 针对移动机器人视觉伺服控制, 提出了基于 ADP 的跟踪控制方法. 与现有视觉伺服控制方法相比, 本文设计的 ADP 控制方法是基于最优控制理论, 算法收敛后, 可得到近似最优控制器.

2) 现有的大多数工作针对定常系统设计, 而移动机器人视觉伺服控制系统为时变系统, 得到的 HJB 方程是时变的. 因而, 本文的工作与现有 ADP 方法在理论分析与实现存在显著区别, 这是本文主要解决的理论难题. 同时, 本文所提的基于 ADP 的跟踪控制方法对于一般性的时变仿射非线性系统的跟踪控制问题具有普适性和通用性.

3) 针对时变 HJB 方程, 本文设计了带时变权值的神经网络结构, 给出了一种新的权值更新律, 证明了神经网络权值的收敛性和闭环系统的稳定性. 同时, 在实验仿真中验证了该类结构具有较小的误差和快速收敛性.

针对时变非线性系统的最优控制仍然与时不变系统存在不同之处, 同时带来许多挑战. 本文针对移动机器人视觉跟踪控制问题, 提出了一种新颖的基于 ADP 的跟踪控制方法. 考虑移动机器人系统模型的时变特性与非完整约束, 采用带时变权值的神经网络近似值函数, 并给出了一种新的权值更新律. 运用 Lyapunov 稳定性理论证明了权值的收敛性以及闭环系统状态是一致最终有界的.

1 问题描述

如图 1 所示, 本文考虑带有单目相机的轮式移动机器人, 假设移动机器人的坐标系与相机坐标系相同. 当前相机坐标系记为 F^c, 坐标系 F^c 的原点为移动机器人中心, x 轴方向为与移动机器人轮轴平行, z 轴方向为相机光轴方向. 由右手坐标系规则, y 轴垂直于移动机器人运动平面 xz 向下. 坐标系 F^d 与 F^* 分别表示移动机器人在期望位姿和在固定参考位姿处的坐标系, 其 x, y, z 轴的定义规则与坐标系 F^c 相同. 在移动机器人的轨迹跟踪任务中, 期望轨迹在 F^d 中由一系列关于特征点的图像描述. F^* 为固定参考坐标系, 由一张关于特征点的图像表示. θ_d 和 θ 分别表示 F^d 与 F^c 相对参考坐标系 F^* 绕 y 轴的旋转角. 基于以上的坐标系定义, 本文的目标是设计一种视觉伺服跟踪控制器以确保移动机器人完成轨迹跟踪任务, 即当 $t \to \infty$ 时, $F^c \to F^d$.

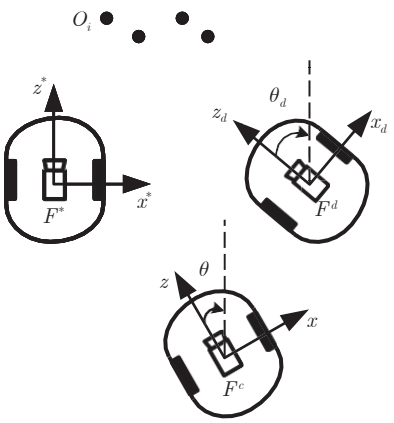

图 1 视觉伺服轨迹跟踪任务描述

Fig. 1 Visual servoing trajectory tracking task

1.1 欧氏重构

如图 1 所示, 假设移动机器人仅做平面运动, 坐标系 F^* 在坐标系 F^c 中的位姿信息可以由 $[x, z, \theta]^{\mathrm{T}}$ 表示. 其中, x 和 z 分别表示沿着 x 轴和 z 轴方向的平移, θ 表示绕 y 轴的旋转. 同样, 坐标系 F^* 在坐标系 F^d 中的位姿信息可以由 $[x_d, z_d, \theta_d]^{\mathrm{T}}$ 表示. 考虑 N 个共面的静止特征点 O_i $(i = 1, 2, \cdots, N)$, 其在坐标系 F^c, F^d, F^* 的三维坐标分别为 O_i^c, O_i^d, O_i^*, 定义为

$$O_i^c = \begin{bmatrix} X_i^c \\ Y_i^c \\ Z_i^c \end{bmatrix}, \quad O_i^d = \begin{bmatrix} X_i^d \\ Y_i^d \\ Z_i^d \end{bmatrix}, \quad O_i^* = \begin{bmatrix} X_i^* \\ Y_i^* \\ Z_i^* \end{bmatrix} \tag{1}$$

将三维坐标 O_i^c, O_i^d, O_i^* 进行归一化[5] 可得:

$$m_i^c = \frac{O_i^c}{Z_i^c}, \quad m_i^d = \frac{O_i^d}{Z_i^d}, \quad m_i^* = \frac{O_i^*}{Z_i^*} \tag{2}$$

特征点 O_i 在坐标系 F^c, F^d, F^* 下对应的可测量齐次像素坐标 p_i^c, p_i^d, $p_i^* \in \mathbf{R}^3$ 分别为

$$p_i^c = \begin{bmatrix} u_i^c \\ v_i^c \\ 1 \end{bmatrix}, \quad p_i^d = \begin{bmatrix} u_i^d \\ v_i^d \\ 1 \end{bmatrix}, \quad p_i^* = \begin{bmatrix} u_i^* \\ v_i^* \\ 1 \end{bmatrix} \tag{3}$$

由针孔相机模型可以得到特征点归一化三维欧氏坐标与可测量齐次像素坐标的关系为

$$m_i^c = K^{-1} p_i^c, \ m_i^d = K^{-1} p_i^d, \ m_i^* = K^{-1} p_i^* \tag{4}$$

其中, $K \in \mathbf{R}^{3\times3}$ 为相机标定矩阵, 为已知可逆常数矩阵. 由式 (4) 得到的 m_i^c, m_i^d, m_i^*, 它们之间的关系可以通过坐标系间的平移与旋转描述, 即

$$m_i^c = H_c m_i^* \tag{5}$$

$$m_i^d = H_d m_i^* \tag{6}$$

其中, H_c, $H_d \in \mathbf{R}^{3\times3}$ 分别为当前位置、期望位置与参考位置之间的单应性矩阵. 通过直接线性变换算法, 可以将单应性矩阵 H_c 分解得到当前相机的尺度化平移 $[\bar{x}, \bar{z}]^{\mathrm{T}}$ 和旋转角 θ; 同样地, 将单应性矩阵 H_d 分解得到期望的尺度化平移 $[\bar{x}_d, \bar{z}_d]^{\mathrm{T}}$ 和旋转角 θ_d [31]. 具体定义为

$$\begin{bmatrix} \bar{x} \\ \bar{z} \end{bmatrix} = \frac{1}{d^*} \begin{bmatrix} x \\ z \end{bmatrix}, \quad \begin{bmatrix} \bar{x}_d \\ \bar{z}_d \end{bmatrix} = \frac{1}{d^*} \begin{bmatrix} x_d \\ z_d \end{bmatrix} \tag{7}$$

其中, d^* 为深度信息常量, 表示固定参考坐标系 F^* 的原点到特征点平面的距离, 即 $d^* = Z_i^*$. 由前文的描述可知, 当 $t \to \infty$ 时, 若 $[\bar{x}, \bar{z}, \theta]^{\mathrm{T}} \to [\bar{x}_d, \bar{z}_d, \theta_d]^{\mathrm{T}}$, 则表示移动机器人完成了轨迹跟踪任务.

1.2 移动机器人运动学模型

移动机器人 F^c 的位姿信息在坐标系 F^* 中记为 $[x^*, z^*, \theta^*]$. 经典的移动机器人连续时间运动学模型可以表达为

$$\begin{cases} \dot{x}^* = v \sin(\theta^*) \\ \dot{z}^* = v \cos(\theta^*) \\ \dot{\theta}^* = \omega \end{cases} \tag{8}$$

其中, v 和 ω 分别为移动机器人的线速度与角速度. 通过坐标变换可以得到 $[x, z, \theta]^{\mathrm{T}}$ 与 $[x^*, z^*, \theta^*]^{\mathrm{T}}$ 之间的关系, 即

$$\begin{bmatrix} x \\ z \\ \theta \end{bmatrix} = \begin{bmatrix} -\cos(\theta^*) & \sin(\theta^*) & 0 \\ -\sin(\theta^*) & -\cos(\theta^*) & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x^* \\ z^* \\ \theta^* \end{bmatrix} \tag{9}$$

对式 (9) 求关于时间 t 的导数, 并将式 (8) 代入求导结果, 可得

$$\dot{x} = -z\omega, \quad \dot{z} = -v + x\omega, \quad \dot{\theta} = -\omega \tag{10}$$

同理, 期望的位姿信息在坐标系 F^* 中记为 $[x_d^*, z_d^*, \theta_d^*]$. v_d, ω_d 分别为移动机器人期望的线速度与角速度. 则期望的位姿信息对时间 t 的导数为

$$\dot{x}_d = -z_d\omega_d, \quad \dot{z}_d = -v_d + x_d\omega_d, \quad \dot{\theta}_d = -\omega_d \tag{11}$$

将式 (7) 代入式 (10) 和式 (11), 可得

$$\dot{\bar{x}} = -\bar{z}\omega, \quad \dot{\bar{z}} = -\frac{v}{d^*} + \bar{x}\omega, \quad \dot{\theta} = -\omega \tag{12}$$

$$\dot{\bar{x}}_d = -\bar{z}_d\omega_d, \quad \dot{\bar{z}}_d = -\frac{v_d}{d^*} + \bar{x}_d\omega_d, \quad \dot{\theta}_d = -\omega_d \tag{13}$$

1.3　跟踪误差系统模型

移动机器人轨迹跟踪误差定义为

$$e = \begin{bmatrix} e_x \\ e_z \\ e_\theta \end{bmatrix} = \begin{bmatrix} \bar{x} \\ \bar{z} \\ \theta \end{bmatrix} - \begin{bmatrix} \cos(e_\theta) & \sin(e_\theta) & 0 \\ -\sin(e_\theta) & \cos(e_\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \bar{x}_d \\ \bar{z}_d \\ \theta_d \end{bmatrix} \tag{14}$$

其中, e_x, e_z表示平移误差, e_θ 为旋转误差. 由式 (14) 可知, 当 $e \to 0$ 时, 则 $[\bar{x}, \bar{z}, \theta] \to [\bar{x}_d, \bar{z}_d, \theta_d]$. 因此上述问题转换为设计一反馈控制律使得 $e \to 0$, 则表示移动机器人完成了轨迹跟踪任务. 对式 (14) 求关于时间的导数, 并将式 (12) 和式 (13) 代入, 可得

$$\begin{cases} \dot{e}_x = -e_z\omega + \dfrac{v_d}{d^*}\sin(e_\theta) \\ \dot{e}_z = \dfrac{-v}{d^*} + e_x\omega + \dfrac{v_d}{d^*}\cos(e_\theta) \\ \dot{e}_\theta = -\omega + \omega_d \end{cases} \tag{15}$$

为了便于后续对跟踪控制问题控制器的设计, 受文献 [32] 启发, 使用如下输入变换:

$$\begin{cases} v = v_d\cos(e_\theta) - u_v \\ \omega = \omega_d - u_\omega \end{cases} \tag{16}$$

其中, u_v, u_ω 为新的输入控制量, 可见上述变换也是可逆的. 由此可以推导出新的误差模型, 即

$$\begin{cases} \dot{e}_x = -e_z\omega_d + \dfrac{v_d}{d^*}\sin(e_\theta) + e_z u_\omega \\ \dot{e}_z = e_x\omega_d + \dfrac{u_v}{d^*} - e_x u_\omega \\ \dot{e}_\theta = u_\omega \end{cases} \tag{17}$$

为估计深度参数信息, 深度参数 d^* 的估计更新律设计采用与文献 [33] 类似方法, 且其收敛性已在文献中得到证明.

2　控制器设计

在移动机器人跟踪控制任务中, 期望轨迹的速度 $v_d(t)$, $\omega_d(t)$ 是随时间变化的, 导致在误差模型 (17) 中含有时变项. 本节基于 ADP 方法设计最优控制器以保证移动机器人完成轨迹跟踪任务, 采用评价神经网络结构近似最优值函数, 求解最优控制输入.

2.1　最优控制设计

由误差动力学方程 (17) 可知, 该系统是一个连续时间仿射非线性系统, 可以表示为

$$\dot{e} = f(e, t) + g(e)u \tag{18}$$

其中,

$$f(e, t) = \begin{bmatrix} -e_z\omega_d + \dfrac{v_d}{d^*}\sin(e_\theta) \\ e_x\omega_d \\ 0 \end{bmatrix}$$

$$g(e) = \begin{bmatrix} 0 & e_z \\ \dfrac{1}{d^*} & -e_x \\ 0 & 1 \end{bmatrix}, \ u = [u_v, u_\omega]^{\mathrm{T}}$$

注意到系统状态方程 (18) 中控制输入为二维, 而系统状态为三维, 可见该系统是一个欠驱动系统.

假设 1. 期望的移动机器人的位姿和速度 x_d^*, z_d^*, θ_d^*, v_d, ω_d 是有界的, 且 $\|f(e, t)\| \le f_M\|e\|$, $\|g(e)\| \le g_M$.

针对此系统的最优控制问题, 本文的目标是设计一个反馈控制策略使得下列定义的值函数最小, 即

$$V(e(t), t) = \int_t^\infty L(e(\tau), u(e), \tau)\mathrm{d}\tau \tag{19}$$

其中, 效用函数取为 $L(e, u, t) = Q(e) + u^{\mathrm{T}}Ru$, $Q(e)$ 是正定的, 即对于 $\forall e \ne 0$, $Q(e) > 0$ 且 $e = 0 \Rightarrow Q(e) = 0$, $R \in \mathbf{R}^{2\times 2}$ 为对称正定矩阵. 沿着系统轨迹 (18), 对值函数求时间微分, 可得如下时变 Lyapunov 方程[34]

$$-\frac{\partial V(e, t)}{\partial t} = \frac{\partial V^{\mathrm{T}}(e, t)}{\partial e}[f(e, t) + g(e)u] + Q(e) + u^{\mathrm{T}}Ru \tag{20}$$

为了便于描述, 定义 $\nabla_e = \frac{\partial}{\partial e}$, $\nabla_t = \frac{\partial}{\partial t}$. 接下来, 定义哈密顿函数为

$$H(e, u, \nabla_e V(e, t), t) = Q(e) + u^{\mathrm{T}}Ru + \nabla_t V(e, t) + \nabla_e V^{\mathrm{T}}(e, t)[f(e, t) + g(e)u] \tag{21}$$

注意到与其他仿射非线性系统无穷时域最优控制情

况不同, 式 (21) 是与时间 t 直接相关的. 根据最优性原理, 当 $\frac{\partial H}{\partial u}=0$ 时, 可得最优控制为

$$u^*(e,t)=-\frac{1}{2}R^{-1}g^{\mathrm{T}}(e)\nabla_e V^*(e,t) \tag{22}$$

将式 (22) 代入式 (20), 可得时变 HJB 方程

$$\begin{aligned}0=&\nabla_t V^*(e,t)+\nabla_e V^{*\mathrm{T}}(e,t)[f(e,t)+g(e)u^*]+\\&Q(e)+u^{*\mathrm{T}}Ru^*=\\&\nabla_t V^*(e,t)+\nabla_e V^{*\mathrm{T}}(e,t)f(e,t)+Q(e)-\\&\frac{1}{4}\nabla_e V^{*\mathrm{T}}(e,t)g(e)R^{-1}g^{\mathrm{T}}(e)\nabla_e V^*(e,t)\end{aligned} \tag{23}$$

注意到 HJB 方程 (23) 是一个偏微分方程, 很难直接求得此方程的解析解, 因此这里使用神经网络近似估计最优值函数 $V^*(e,t)$, 以期望求得近似最优的反馈控制.

2.2 神经网络近似

与时不变仿射非线性系统最优控制问题不同的是, 这里的最优值函数 $V^*(e,t)$ 与时间变量 t 相关. 受文献 [35] 启发, 带时变权值的神经网络结构可以用来近似一致连续时变函数. 假设 $V^*(e,t)$ 是光滑的, 则最优值函数 $V^*(e,t)$ 可由下列神经网络形式表示

$$V^*(e,t)=W^{\mathrm{T}}(t)\phi(e)+\varepsilon(e,t) \tag{24}$$

其中, $W(t)\in\mathbf{R}^L$ 为真实的神经网络权值向量, $L\in\mathbf{R}$, $L>0$ 为隐含层节点数, 随着神经网络隐含层节点数增加, 逼近误差会不断减小[35], 即当 $L\to\infty$, $\varepsilon(e,t)\to 0$. $\phi(e)=[\phi_1(e),\cdots,\phi_L(e)]^{\mathrm{T}}\in\mathbf{R}^L$ 为与状态相关的连续可微的激活函数, $\varepsilon(e,t)$ 为神经网络逼近误差. 对最优值函数 $V^*(e,t)$ 求分别关于 e 和 t 的偏导

$$\nabla_e V^*(e,t)=\nabla_e^{\mathrm{T}}\phi(e)W(t)+\nabla_e\varepsilon(e,t) \tag{25}$$

$$\nabla_t V^*(e,t)=\dot{W}^{\mathrm{T}}(t)\phi(e)+\nabla_t\varepsilon(e,t) \tag{26}$$

假设 2. 神经网络逼近误差 $\varepsilon(e,t)$ 以及其分别关于状态 e 与时间 t 的偏导数 $\nabla_e\varepsilon(e,t)$, $\nabla_t\varepsilon(e,t)$ 是有界的. 真实的神经网络权值 $W(t)$ 以及其对时间的导数 $\dot{W}(t)$ 是有界的, 激活函数 $\phi(e)$ 以及其对 e 的偏导数 $\nabla_e\phi(e)$ 是有界的: $\|\varepsilon(e,t)\|\le\varepsilon_M$, $\|\nabla_e\varepsilon(e,t)\|\le\varepsilon_{eM}$, $\|\nabla_t\varepsilon(e,t)\|\le\varepsilon_{tM}$, $\|W(t)\|\le W_M$, $\|\dot{W}(t)\|\le W_{tM}$, $\|\phi(e)\|\le\phi_M$, $\|\nabla_e\phi(e)\|\le\phi_{eM}$.

将式 (25) 代入式 (22) 中, 最优控制输入可以表示为

$$\begin{aligned}u^*(e,t)=&-\frac{1}{2}R^{-1}g^{\mathrm{T}}(e)\nabla_e\phi^{\mathrm{T}}(e)W(t)-\\&\frac{1}{2}R^{-1}g^{\mathrm{T}}(e)\nabla_e\varepsilon(e,t)\end{aligned} \tag{27}$$

将式 (25) ~ (27) 代入 HJB 方程 (23) 中, 得到

$$\begin{aligned}0=&\dot{W}^{\mathrm{T}}(t)\phi(e)+W^{\mathrm{T}}(t)\nabla_e\phi(e)[f(e,t)+g(e)u^*]+\\&Q(e)+u^{*\mathrm{T}}Ru^*+\varepsilon_v(e,t)=\\&\dot{W}^{\mathrm{T}}(t)\phi(e)+W^{\mathrm{T}}(t)\nabla_e\phi(e)f(e,t)+Q(e)-\\&\frac{1}{4}W^{\mathrm{T}}(t)\nabla_e\phi(e)g(e)R^{-1}g^{\mathrm{T}}(e)\nabla_e^{\mathrm{T}}\phi(e)W(t)+\\&\varepsilon_{hjb}(e,t)\end{aligned} \tag{28}$$

其中, $\varepsilon_v(e,t)=\nabla_e^{\mathrm{T}}\varepsilon(e,t)[f(e,t)+g(e)u^*]+\nabla_t\varepsilon(e,t)$. $\varepsilon_{hjb}(e,t)$ 为 HJB 方程的残差, 即

$$\begin{aligned}\varepsilon_{hjb}(e,t)=&\nabla_t\varepsilon(e,t)+\nabla_e^{\mathrm{T}}\varepsilon(e,t)f(e,t)-\\&\frac{1}{2}W^{\mathrm{T}}(t)\nabla_e\phi(e)D_1(e)\nabla_e\varepsilon(e,t)-\\&\frac{1}{4}\nabla_e^{\mathrm{T}}\varepsilon(e,t)D_1(e)\nabla_e\varepsilon(e,t)\end{aligned} \tag{29}$$

其中, $D_1(e)=g(e)R^{-1}g^{\mathrm{T}}(e)$. 根据文献 [36], 随着神经网络隐含层节点数 $L\to\infty$, HJB 方程残差 $\varepsilon_{hjb}(e,t)\to 0$, 即对于 $\forall\varepsilon_h>0$, $\exists L:\|\varepsilon_{hjb}(e,t)\|<\varepsilon_h$.

2.3 神经网络权值更新

由于真实的神经网络权值 W 是未知的, 设计一个评价神经网络近似最优值函数 (24), 即

$$\hat{V}(e(t),t)=\hat{W}^{\mathrm{T}}(t)\phi(e(t)) \tag{30}$$

其中, $\hat{V}(e(t),t)$ 为近似最优值函数, $\hat{W}\in\mathbf{R}^L$ 为真实权值的估计值. $\phi(e(t))$满足: 若 $\|e\|=0$, 则 $\|\phi(0)\|=0$. 根据式 (27) 和式 (30), 可得近似最优控制为

$$\hat{u}(e,t)=-\frac{1}{2}R^{-1}g^{\mathrm{T}}(e)\nabla_e\phi^{\mathrm{T}}(e)\hat{W}(t) \tag{31}$$

定义真实最优控制输入与近似最优控制输入之差为

$$\tilde{u}=u^*-\hat{u} \tag{32}$$

将式 (27) 和式 (31) 代入式 (32) 中, 可得

$$\begin{aligned}\tilde{u}=&-\frac{1}{2}R^{-1}g^{\mathrm{T}}(e)\nabla_e^{\mathrm{T}}\phi(e)\tilde{W}(t)-\\&\frac{1}{2}R^{-1}g^{\mathrm{T}}(e)\nabla_e\varepsilon(e,t)\end{aligned} \tag{33}$$

将式 (30) 和式 (31) 分别作为近似最优值函数和近似最优控制代入哈密顿函数 (21) 中, 可得

$$\begin{aligned}\delta=&\ H(e,\hat{u},\nabla_e\hat{V}(e,t),t)=\\&\dot{\hat{W}}^{\mathrm{T}}(t)\phi(e)+Q(e)+\hat{u}^{\mathrm{T}}R\hat{u}+\\&\hat{W}^{\mathrm{T}}(t)\nabla_e\phi(e)[f(e,t)+g(e)\hat{u}]\end{aligned} \tag{34}$$

根据式 (34), 定义目标误差函数为

$$E=\frac{1}{2}\delta^2 \tag{35}$$

为使目标误差函数 E 不断减小, 结合梯度下降法的思想设计权值更新律为

$$\dot{\hat{W}}(t) = -\alpha\frac{\sigma[\sigma^{\mathrm{T}}\hat{W}(t) + Q(e) + \hat{u}^{\mathrm{T}}R\hat{u}]}{(\sigma^{\mathrm{T}}\sigma+1)^2 + \alpha\sigma\phi^{\mathrm{T}}(e)} \tag{36}$$

其中, $\sigma = \nabla_e\phi(e)\,[f(e,\,t) + g(e)\hat{u}] \in \mathbf{R}^L$, α 为评价网络权值的学习率. 值得一提的是, 权值更新律 (36) 的设计主要有两方面的考虑: 1) 为使式 (35) 给出的目标误差函数 E 不断减小, 因而采用梯度下降的思想; 2) 为了保证系统的稳定性, 详细的稳定性分析见第 3 节. 神经网络权值估计误差定义为

$$\tilde{W}(t) = W(t) - \hat{W}(t) \tag{37}$$

根据式 (36), 可得

$$\begin{aligned}(\sigma^{\mathrm{T}}\sigma+1)^2\dot{\tilde{W}}(t) + \alpha\sigma\phi^{\mathrm{T}}(e)\dot{\tilde{W}}(t) = \\ -\alpha\sigma\big[\hat{u}^{\mathrm{T}}R\hat{u} + \sigma^{\mathrm{T}}\hat{W}(t) + Q(e)\big]\end{aligned} \tag{38}$$

将等式 (38) 左边第 2 项移到右边, 可得

$$\begin{aligned}(\sigma^{\mathrm{T}}\sigma+1)^2\dot{\tilde{W}}(t) = -\alpha\sigma[\phi^{\mathrm{T}}(e)\dot{\tilde{W}}(t) + \sigma^{\mathrm{T}}\hat{W}(t) + \\ Q(e) + \hat{u}^{\mathrm{T}}R\hat{u}]\end{aligned} \tag{39}$$

由式 (39), 可得:

$$\dot{\tilde{W}}(t) = -\alpha\frac{\sigma[\phi^{\mathrm{T}}(e)\dot{\tilde{W}}(t) + \sigma^{\mathrm{T}}\hat{W}(t) + Q(e) + \hat{u}^{\mathrm{T}}R\hat{u}]}{(\sigma^{\mathrm{T}}\sigma+1)^2} \tag{40}$$

根据式 (31), 可得

$$\begin{aligned}\hat{u}^{\mathrm{T}}R\hat{u} = \frac{1}{4}\hat{W}^{\mathrm{T}}(t)\nabla_e\phi(e)g(e)R^{-1}g^{\mathrm{T}}(e)\nabla_e^{\mathrm{T}}\phi(e)\hat{W}(t) = \\ \frac{1}{4}\hat{W}^{\mathrm{T}}(t)D_2(e)\hat{W}(t)\end{aligned} \tag{41}$$

其中, $D_2(e) = \nabla_e\phi(e)g(e)R^{-1}g^{\mathrm{T}}(e)\nabla_e^{\mathrm{T}}\phi(e)$. 对式 (37) 求时间导数, 将式 (40) 和式 (41) 代入其中, 可得神经网络权值估计误差的动力学方程为

$$\begin{aligned}&\dot{\tilde{W}}(t) = \dot{W}(t) - \dot{\hat{W}}(t) = \dot{W}(t) + \\ &\alpha\frac{\sigma[\phi^{\mathrm{T}}(e)\dot{\tilde{W}}(t) + \sigma^{\mathrm{T}}\hat{W}(t) + Q(e) + \hat{u}^{\mathrm{T}}R\hat{u}]}{(\sigma^{\mathrm{T}}\sigma+1)^2} = \\ &\dot{W}(t) + \alpha\frac{\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(\phi^{\mathrm{T}}(e)\dot{\tilde{W}}(t) + \\ &\sigma^{\mathrm{T}}\hat{W}(t) + Q(e) + \frac{1}{4}\hat{W}^{\mathrm{T}}(t)D_2(e)\hat{W}(t)\Big)\end{aligned} \tag{42}$$

3　稳定性分析

定理 1. 针对式 (18) 描述的非线性时变仿射系统, 以式 (31) 为控制输入, 式 (36) 为评价神经网络的权值更新律, 则闭环系统的状态 e 与评价网络的权值估计误差 $\tilde{W}(t)$ 是一致最终有界的.

证明. 选择 Lyapunov 函数形式为

$$L(t) = L_w(t) + L_v(t) \tag{43}$$

其中, $L_w(t) = \frac{1}{2}\tilde{W}^{\mathrm{T}}(t)\tilde{W}(t)$, $L_v(t) = V^*(e,\,t)$ (简洁起见, 后文将 $W(t)$ 简写为 W). 对式 (43) 求关于时间的导数, 得到

$$\dot{L}(t) = \dot{L}_w(t) + \dot{L}_v(t) \tag{44}$$

根据式 (42), $\dot{L}_w(t)$ 可以表示为

$$\begin{aligned}&\dot{L}_w = \tilde{W}^{\mathrm{T}}\dot{\tilde{W}} = \\ &\quad\tilde{W}^{\mathrm{T}}\dot{W} + \tilde{W}^{\mathrm{T}}\frac{\alpha\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(\phi^{\mathrm{T}}(e)\dot{\hat{W}} + \sigma^{\mathrm{T}}\hat{W} + \\ &\quad Q(e) + \hat{u}^{\mathrm{T}}R\hat{u}\Big) = \\ &\quad\tilde{W}^{\mathrm{T}}\dot{W} + \tilde{W}^{\mathrm{T}}\frac{\alpha\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(\phi^{\mathrm{T}}(e)\dot{\hat{W}} + \sigma^{\mathrm{T}}\hat{W} + \\ &\quad Q(e) + \frac{1}{4}\hat{W}^{\mathrm{T}}D_2\hat{W} - \phi^{\mathrm{T}}(e)\dot{W} - \sigma^{\mathrm{T}}W - \\ &\quad Q(e) - \frac{1}{4}W^{\mathrm{T}}D_2W - \varepsilon_{hjb}\Big) = \\ &\quad\tilde{W}^{\mathrm{T}}\dot{W} + \tilde{W}^{\mathrm{T}}\frac{\alpha\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(\phi^{\mathrm{T}}(e)\dot{\hat{W}} - \\ &\quad\phi^{\mathrm{T}}(e)\dot{W} + \sigma^{\mathrm{T}}\hat{W} - \sigma^{\mathrm{T}}W + Q(e) + \\ &\quad\frac{1}{4}\hat{W}^{\mathrm{T}}D_2\hat{W} - Q(e) - \frac{1}{4}W^{\mathrm{T}}D_2W - \varepsilon_{hjb}\Big) = \\ &\quad\tilde{W}^{\mathrm{T}}\dot{W} + \tilde{W}^{\mathrm{T}}\frac{\alpha\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(-\phi^{\mathrm{T}}(e)\dot{\tilde{W}} - \\ &\quad(\nabla_e\phi(e)f(e,\,t))^{\mathrm{T}}\tilde{W} - \frac{1}{2}\hat{W}^{\mathrm{T}}D_2\hat{W} + \\ &\quad\frac{1}{2}W^{\mathrm{T}}D_2W + \frac{1}{4}\hat{W}D_2\hat{W} - \frac{1}{4}W^{\mathrm{T}}D_2W - \varepsilon_{hjb}\Big) = \\ &\quad\tilde{W}^{\mathrm{T}}\dot{W} + \tilde{W}^{\mathrm{T}}\frac{\alpha\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(-\phi^{\mathrm{T}}(e)\dot{\tilde{W}} - \\ &\quad(\nabla_e\phi(e)f(e,\,t))^{\mathrm{T}}\tilde{W} + \frac{1}{4}\tilde{W}^{\mathrm{T}}D_2\tilde{W} + \\ &\quad\frac{1}{2}\tilde{W}^{\mathrm{T}}D_2\hat{W} - \varepsilon_{hjb}\Big) = \\ &\quad\tilde{W}^{\mathrm{T}}\dot{W} + \tilde{W}^{\mathrm{T}}\frac{\alpha\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(-\phi^{\mathrm{T}}(e)\dot{\tilde{W}} - \sigma^{\mathrm{T}}\tilde{W} + \\ &\quad\frac{1}{4}\tilde{W}^{\mathrm{T}}D_2\tilde{W} - \varepsilon_{hjb}\Big)\end{aligned} \tag{45}$$

由式 (45), 可得:

$$\begin{aligned}\left(1 + \frac{\alpha\sigma\phi^{\mathrm{T}}(e)}{(\sigma^{\mathrm{T}}\sigma+1)^2}\right)\tilde{W}^{\mathrm{T}}\dot{\tilde{W}} = \tilde{W}^{\mathrm{T}}\dot{W} + \\ \frac{\tilde{W}^{\mathrm{T}}\alpha\sigma}{(\sigma^{\mathrm{T}}\sigma+1)^2}\Big(-\sigma^{\mathrm{T}}\tilde{W} + \frac{1}{4}\tilde{W}^{\mathrm{T}}D_2\tilde{W} - \varepsilon_{hjb}\Big)\end{aligned} \tag{46}$$

根据式 (46), 可得 $\dot{L}_w$ 为

$$\dot{L}_w = \tilde{W}^{\mathrm{T}}\dot{\tilde{W}} = \frac{(\sigma^{\mathrm{T}}\sigma+1)^2}{(\sigma^{\mathrm{T}}\sigma+1)^2+\alpha\sigma\phi^{\mathrm{T}}}\left(\tilde{W}^{\mathrm{T}}\dot{W}+\tilde{W}^{\mathrm{T}}\frac{\alpha\sigma(-\sigma^{\mathrm{T}}\tilde{W}+\frac{1}{4}\tilde{W}^{\mathrm{T}}D_2\tilde{W}-\varepsilon_{hjb})}{(\sigma^{\mathrm{T}}\sigma+1)^2}\right) = \lambda_\sigma\tilde{W}^{\mathrm{T}}\dot{W}+\lambda_\sigma\tilde{W}^{\mathrm{T}}\alpha\bar{\sigma}\left(-\bar{\sigma}^{\mathrm{T}}\tilde{W}-\frac{\varepsilon_{hjb}}{m_s}\right)+\frac{1}{4}\lambda_\sigma\alpha\tilde{W}^{\mathrm{T}}\frac{\bar{\sigma}}{m_s}\tilde{W}^{\mathrm{T}}D_2\tilde{W} \tag{47}$$

其中, $\bar{\sigma}=\sigma/(\sigma^{\mathrm{T}}\sigma+1)$, $m_s=\sigma^{\mathrm{T}}\sigma+1$, $\lambda_\sigma=(\sigma^{\mathrm{T}}\sigma+1)^2/((\sigma^{\mathrm{T}}\sigma+1)^2+\alpha\sigma\phi^{\mathrm{T}}(e))$. Lyapunov 函数第 2 项 L_v 关于时间 t 的导数为

$$\dot{L}_v=\dot{V}^*=\nabla_t V^*+\nabla_e V^*\dot{e} \tag{48}$$

将式 (18)、式 (25) 和式 (26) 代入式 (48), 可得

$$\dot{L}_v=\nabla_t V^*+\nabla_e V^*\Big(f(e,t)+g(e)\hat{u}\Big)= \dot{W}^{\mathrm{T}}(t)\phi(e)+\nabla_t\varepsilon(e,t)+\nabla_e V^*\Big(f(e,t)+g(e)\hat{u}\Big)= \dot{W}^{\mathrm{T}}(t)\phi(e)+\nabla_t\varepsilon(e,t)+\Big(\nabla_e\phi^{\mathrm{T}}(e)W(t)+\nabla_e\varepsilon(e,t)\Big)\Big(f(e,t)+g(e)\hat{u}\Big) \tag{49}$$

将式 (31) 代入式 (49), 可得

$$\dot{L}_v=\dot{W}^{\mathrm{T}}(t)\phi(e)+\nabla_e\phi^{\mathrm{T}}(e)W(t)f(e,t)-\frac{1}{2}W^{\mathrm{T}}D_2\hat{W}+\varepsilon_1 \tag{50}$$

其中, $\varepsilon_1=\nabla_e\varepsilon(e,t)(f(e,t)-\frac{1}{2}D_1\nabla_e^{\mathrm{T}}\phi(e)\hat{W})+\nabla_t\varepsilon(e,t)$. 由式 (50) 可得

$$\dot{L}_v=\dot{W}^{\mathrm{T}}(t)\phi(e)+\nabla_e^{\mathrm{T}}\phi(e)W(t)f(e,t)+\frac{1}{2}W^{\mathrm{T}}D_2(W-\hat{W})-\frac{1}{2}W^{\mathrm{T}}D_2W+\varepsilon_1= \dot{W}^{\mathrm{T}}(t)\phi(e)+\nabla_e^{\mathrm{T}}\phi(e)W(t)f(e,t)+\frac{1}{2}W^{\mathrm{T}}D_2\tilde{W}-\frac{1}{2}W^{\mathrm{T}}D_2W+\varepsilon_1 \tag{51}$$

根据 HJB 方程 (28), 可得

$$\nabla_e^{\mathrm{T}}\phi(e)W(t)f(e,t)-\frac{1}{2}W^{\mathrm{T}}D_2W= -Q(e)-\frac{1}{4}W^{\mathrm{T}}D_2W-\dot{W}^{\mathrm{T}}(t)\phi(e)-\varepsilon_{hjb} \tag{52}$$

将式 (52) 代入式 (51) 中, 可得

$$\dot{L}_v=\dot{W}^{\mathrm{T}}(t)\phi(e)+\frac{1}{2}W^{\mathrm{T}}D_2\tilde{W}-Q(e)-\frac{1}{4}W^{\mathrm{T}}D_2W-\dot{W}^{\mathrm{T}}(t)\phi(e)-\varepsilon_{hjb}+\varepsilon_1= \frac{1}{2}W^{\mathrm{T}}D_2\tilde{W}-Q(e)-\frac{1}{4}W^{\mathrm{T}}D_2W-\varepsilon_{hjb}+\varepsilon_1 \tag{53}$$

将式 (47) 和式 (53) 相加, 可得 Lyapunov 函数的导数为

$$\dot{L}=\dot{L}_w+\dot{L}_v= \lambda_\sigma\tilde{W}^{\mathrm{T}}\dot{W}+\lambda_\sigma\tilde{W}^{\mathrm{T}}\alpha\bar{\sigma}\left(-\bar{\sigma}^{\mathrm{T}}\tilde{W}-\frac{\varepsilon_{hjb}}{m_s}\right)+\frac{1}{4}\lambda_\sigma\alpha\tilde{W}^{\mathrm{T}}\frac{\bar{\sigma}}{m_s}\tilde{W}^{\mathrm{T}}D_2\tilde{W}+\frac{1}{2}W^{\mathrm{T}}D_2\tilde{W}-Q(e)-\frac{1}{4}W^{\mathrm{T}}D_2W-\varepsilon_{hjb}+\varepsilon_1 \tag{54}$$

根据假设 1 和假设 2, 可得

$$\varepsilon_1=\nabla_e\varepsilon(e,t)\Big(f(e,t)-\frac{1}{2}D_1\nabla_e^{\mathrm{T}}\phi(e)\hat{W}\Big)+\nabla_t\varepsilon(e,t)= \nabla_e\varepsilon(e,t)f(e,t)-\frac{1}{2}\nabla_e\varepsilon(e,t)D_1\nabla_e^{\mathrm{T}}\phi(e)(W-\tilde{W})+\nabla_t\varepsilon(e,t)\le\varepsilon_{eM}f_M\|e\|+\frac{1}{2}\varepsilon_{eM}\phi_{eM}\|D_1\|(\|W\|+\|\tilde{W}\|)+\varepsilon_{tM} \tag{55}$$

因为 $Q(e)$ 是正定的, 存在一个 λ_q 使得 $e^{\mathrm{T}}\lambda_q e\le Q(e)$. 根据文献 [36], 随着神经网络隐含层节点数 L 的增大, HJB 方程残差 $\varepsilon_{hjb}(e,t)$ 会逐渐收敛到零. 假设存在一正数 ε_{hM}, 选择合适的神经网络隐含层节点数 L, HJB 方程残差满足: $\varepsilon_{hjb}\le\varepsilon_{hM}$. 因而, 根据式 (54) 和式 (55), 可得

$$\dot{L}\le-\tilde{Z}^{\mathrm{T}}H_1\tilde{Z}+\tilde{Z}^{\mathrm{T}}H_2+c \tag{56}$$

其中,

$$\tilde{Z}=\left[e,\ \bar{\sigma}^{\mathrm{T}}\tilde{W},\ \tilde{W}^{\mathrm{T}}\tilde{W},\ \tilde{W}\right]^{\mathrm{T}}$$

$$H_1=\begin{bmatrix}\lambda_q I & 0 & 0 & 0\\ 0 & \alpha\lambda_\sigma & -\dfrac{\alpha\lambda_\sigma D_2^{\mathrm{T}}}{8m_s} & 0\\ 0 & -\dfrac{\alpha\lambda_\sigma D_2}{8m_s} & 0 & 0\\ 0 & 0 & 0 & 0\end{bmatrix}$$

$$H_2=\begin{bmatrix}\varepsilon_{eM}f_M\\ \dfrac{\varepsilon_{hM}}{m_s}\\ 0\\ \lambda_\sigma\|\dot{W}\|+\dfrac{1}{2}\|W\|\|D_1\|+\dfrac{1}{2}\varepsilon_{eM}\phi_{eM}\|D\|\end{bmatrix}$$

$$c=\frac{1}{2}\|D_2\|\|W\|^2+\frac{1}{2}\varepsilon_{eM}\phi_{eM}\|D_1\|\|W\|+\varepsilon_{tM}+\varepsilon_{hM}$$

式中, I 为合适维度的单位矩阵. 选择参数使 $H_1>0$, 根据式 (56), 如果

$$\|\tilde{Z}\|>\frac{\|H_2\|+\sqrt{\|H_2\|^2+4c\lambda_{\min}(H_1)}}{2\lambda_{\min}(H_1)}\equiv B_Z \quad (57)$$

则 Lyapunov 函数的导数 $\dot{L}<0$. 故闭环系统的状态 e 与评价网络的权值估计误差 $\tilde{W}$ 是一致最终有界的. □

在这里, 有必要讨论一下本文与相关工作 [16, 37−39]的区别. 在文献 [37−38] 中, 采用基于策略迭代方法设计控制器, 与此不同, 本文则是基于 ADP 设计自适应控制器, 因而在实现方法与理论分析上存在着显著的差异. 在文献 [16, 39] 中, 考虑的是非线性离散时间系统的跟踪控制问题, 与本文方法的区别主要体现在: 文献 [16, 39] 考虑的是定常系统, 也就是说, 虽然期望轨迹为时变的, 但是针对的系统是定常系统, 也即时不变系统; 与文献 [16, 39] 不同, 本文考虑移动机器人视觉伺服跟踪控制问题, 不仅期望轨迹是时变的, 而且系统也是时变的. 因此, 文献 [16, 39] 与本文所针对的问题是完全不同的.

4　仿真研究

为了验证本文提出的控制方法的有效性, 本节利用计算机进行仿真实验. 选择 4 个共面特征点作为视觉目标点, 以便能通过单应性矩阵分解成对应的平移与旋转量. 相机标定矩阵为

$$K=\begin{bmatrix}800\text{ mm} & 0 & 512\\ 0 & 800\text{ mm} & 512\\ 0 & 0 & 1\end{bmatrix}$$

移动机器人初始的位姿设置为 $[x(0),\ z(0),\ \theta(0)]^{\mathrm{T}}=[-4.5,\ -7.1,\ 0.23]^{\mathrm{T}}$, 期望轨迹的初始位姿设置为 $[x_d(0),\ z_d(0),\ \theta_d(0)]^{\mathrm{T}}=[-3.5,\ -6.5,\ 0.15]^{\mathrm{T}}$. 让移动机器人预先沿曲线运动获取一系列特征点的图像以定义期望轨迹. 初始的跟踪误差为 $e(0)=[e_x(0),\ e_z(0),\ e_\theta(0)]^{\mathrm{T}}=[-1,\ -0.6,\ -0.08]^{\mathrm{T}}$. 为求解最优控制问题中的值函数, 选择神经网络的隐含层节点数 $L=21$, 选取激活函数为 $\phi(e)=[e_x^2,\ e_z^2,\ e_\theta^2,\ e_xe_z,\ e_xe_\theta,\ e_ze_\theta,\ e_x^4,\ e_z^4,\ e_\theta^4,\ e_x^2e_z^2,\ e_x^2e_\theta^2,\ e_z^2e_\theta^2,\ e_x^2e_ze_\theta,\ e_xe_z^2e_\theta,\ e_xe_ze_\theta^2,\ e_x^3e_z,\ e_x^3e_\theta,\ e_xe_z^3,\ e_z^3e_\theta,\ e_xe_\theta^3,\ e_ze_\theta^3]^{\mathrm{T}}$. 权值更新中的学习率设置为 $\alpha=0.1$, $Q(e)=0.5e^{\mathrm{T}}e$, 正定矩阵 $R=0.1I$. 在仿真实验中, 移动机器人的初始姿态位置、期望轨迹初始姿态位置满足其能保证特征点始终在视野范围内. 初始位姿与期望的初始位姿相差较大则相应的收敛时间将变长, 而对最终的跟踪精度没有显著影响. 对于神经网络初始值的选取, 目前没有通用的方法, 通常只是基于经验和尝试. 在本文中, 仿真结果是通过多次实验尝试后, 选取了一组结果作为展示.

图 2 ~ 9 展示了实验仿真结果. 由图 2 和图 3 可以看出, 在本文设计的控制方法下系统状态最终收敛到零并且控制输入也随着时间最终趋于零. 神经网络的权值最终如期望的一样收敛于常数值如图 4 所示, 最终收敛权值为 $\hat{W}=[0.0445,\ 0.0458,\ 0.0214,\ 0.0048,\ 0.0180,\ -0.0005,\ 0.0652,\ 0.0174,\ 0.0004,\ 0.0430,\ 0.0292,\ 0.0021,\ -0.0021,\ -0.0011,\ -0.0009,\ 0.0077,\ 0.0105,\ 0.0048,\ 0.0003,\ 0.0012,\ -0.0001]^{\mathrm{T}}$. 移动机器人的线速度和角速度如图 5 所示, 可以看出, 当前线速度和角速度与期望轨迹的线速度和角速度逐渐一致.

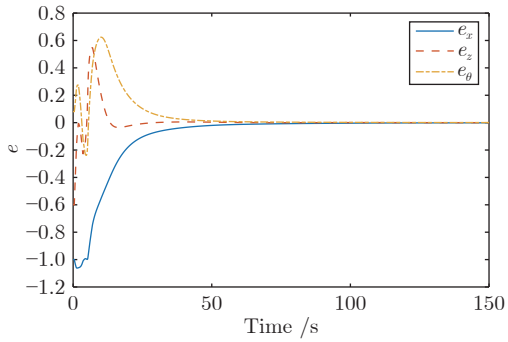

图 2　系统响应

Fig. 2　System response

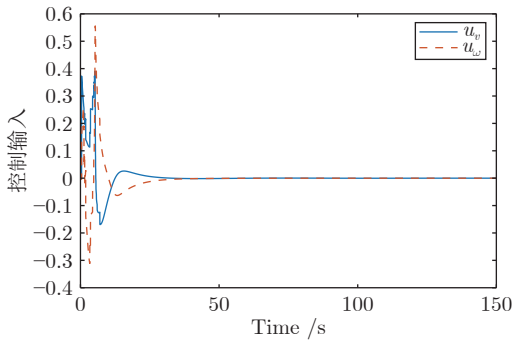

图 3　控制输入

Fig. 3　Control input

在仿真过程中, 为了验证本文所提出算法的效果, 在保证选取同样网络层数和隐含层节点个数的前提下, 将本文提出的时变权值神经网络结构与文献 [38] 中提出的时变激活函数 NN 结构进行对比. 在最优控制问题中, HJB 方程的近似误差可以用于表征控制器对性能指标优化程度的好坏. 如图 6 所示,

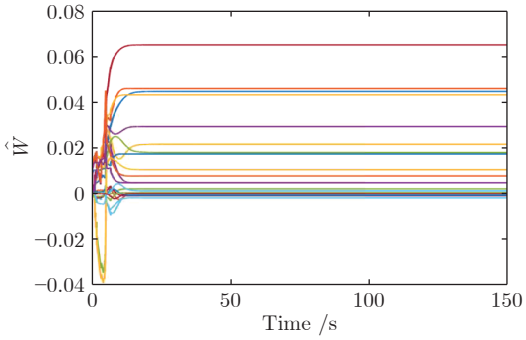

图 4 评价神经网络权值的收敛

Fig. 4 Convergence of critic neural network weights

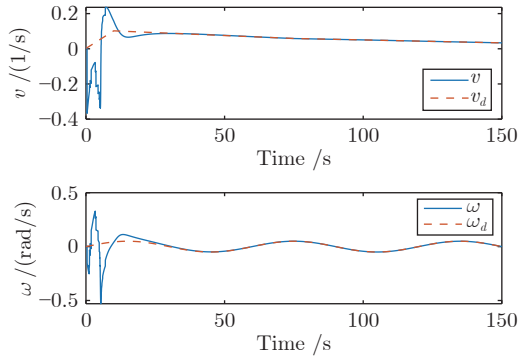

图 5 移动机器人期望轨迹速度与实际运动速度

Fig. 5 Desired and real velocities of the mobile robot

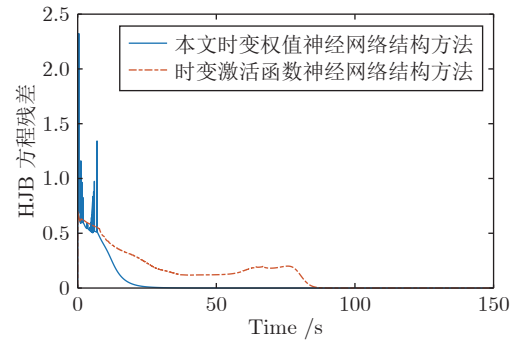

图 6 HJB 方程残差

Fig. 6 The residual error of HJB equation

在两种方法下 HJB 方程的残差最终均能收敛至零, 但本文所提方法的收敛速度相对要快很多.

在两种方法下的移动机器人期望轨迹与实际运动轨迹如图 7 和图 8 所示. 可见, 在本文所提的时变权值的神经网络结构方法下, 机器人的跟踪效果更好、跟踪误差更小. 此外, 特征点的二维图像轨迹如图 9 所示. 其中, 红色虚线表示期望图像轨迹, 实心圆点表示初始的期望图像, 五角星表示最终的期望图像; 蓝色实线表示当前的真实图像轨迹, 空心圆点表示初始的真实图像, 方形表示最终的真实图像. 由图 9 可知, 当前实际图像轨迹与期望的图像轨迹逐渐一致.

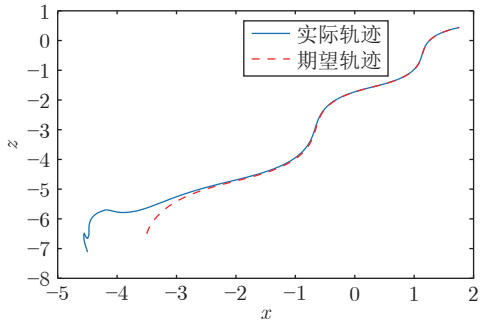

图 7 利用本文时变权值神经网络结构方法的移动机器人期望轨迹与实际运动轨迹

Fig. 7 Desired and real trajectories of the mobile robot using time-varying weights neural network structure method in this paper

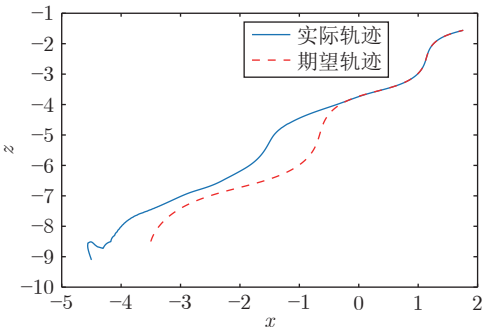

图 8 利用时变激活函数神经网络结构方法的移动机器人期望轨迹与实际运动轨迹

Fig. 8 Desired and real trajectories of the mobile robot using time-varying activation neural network structure method

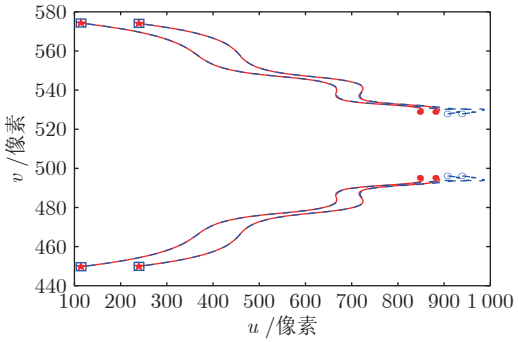

图 9 特征点二维图像轨迹

Fig. 9 2D image trajectories of the feature points

5 结束语

本文设计了一种新的基于 ADP 的跟踪控制方法来解决移动机器人视觉伺服轨迹跟踪最优控制问

题. 与以往控制对象不同的是移动机器人视觉伺服轨迹跟踪的误差系统模型是一个时变仿射非线性系统, 针对此系统的最优控制问题需要设计具有时变权值的神经网络近似值函数以求解时变的 HJB 方程. 运用 Lyapunov 稳定性理论证明了在本文提出的控制方法作用下神经网络权值的收敛性以及闭环系统的稳定性. 仿真实验结果验证了所提出方法的有效性与可行性.

References

1 Jia Bing-Xi, Liu Shan, Zhang Kai-Xiang, Chen Jian. Survey on robot visual servo control: Vision system and control strategies. *Acta Automatica Sinica*, 2015, **41**(5): 861−873
(贾丙西, 刘山, 张凯祥, 陈剑. 机器人视觉伺服研究进展: 视觉系统与控制策略. 自动化学报, 2015, **41**(5): 861−873)

2 Fomena R T, Tahri O, Chaumette F. Distance-based and orientation-based visual servoing from three points. *IEEE Transactions on Robotics*, 2011, **27**(2): 256−267

3 Yang Fang, Wang Chao-Li. Adaptive stabilization for uncertain nonholonomic dynamic mobile robots based on visual servoing feedback. *Acta Automatica Sinica*, 2011, **37**(7): 857−864
(杨芳, 王朝立. 基于视觉伺服反馈的不确定非完整动态移动机器人的自适应镇定. 自动化学报, 2011, **37**(7): 857−864)

4 Jiang Z P. Robust exponential regulation of nonholonomic systems with uncertainties. *Automatica*, 2000, **36**(2): 189−209

5 Chen J, Dixon W E, Dawson D M, McIntire M. Homography-based visual servo tracking control of a wheeled mobile robot. *IEEE Transactions on Robotics*, 2006, **22**(2): 406−415

6 Liang X, Wang H, Chen W, Guo D, Liu T. Adaptive image-based trajectory tracking control of wheeled mobile robots with an uncalibrated fixed camera. *IEEE Transactions on Control Systems Technology*, 2015, **23**(6): 2266−2282

7 Xu De. A tutorial for monocular visual servoing. *Acta Automatica Sinica*, 2018, **44**(10): 1729−1746
(徐德. 单目视觉伺服研究综述. 自动化学报, 2018, **44**(10): 1729−1746)

8 Brockett R W, Millman R S, Sussmann H J. *Differential Geometric Control Theory*. Michigan: Birkhauser Boston, 1983.

9 Jiang Z P, Nijmeijer H. Tracking control of mobile robots: A case study in backstepping. *Automatica*, 1997, **33**(7): 1393−1399

10 Li B, Fang Y, Hu G, Zhang X. Model-free unified tracking and regulation visual servoing of wheeled mobile robots. *IEEE Transactions on Control Systems Technology*, 2015, **24**(4): 1328−1339

11 Li L, Liu Y H, Jiang T, Wang K, Fang M. Adaptive trajectory tracking of nonholonomic mobile robots using vision-based position and velocity estimation. *IEEE Transactions on Cybernetics*, 2017, **48**(2): 571−582

12 Miao Z, Zhong H, Lin J, Wang Y, Chen Y, Fierro R. Vision-based bormation control of mobile robots with FOV constraints and unknown feature depth. *IEEE Transactions on Control Systems Technology*, 2021, **29**(5): 2231−2238

13 Zhang K, Chen J, Li Y, Zhang X. Visual tracking and depth estimation of mobile robots without desired velocity information. *IEEE Transactions on Cybernetics*, 2020, **50**(1): 361−373

14 Lee S, Chwa D. Dynamic image-based visual servoing of monocular camera mounted omnidirectional mobile robots considering actuators and target motion via fuzzy integral sliding mode control. *IEEE Transactions on Fuzzy Systems*, 2021, **29**(7): 2068−2076

15 Wang R, Zhang X, Fang Y, Li B. Virtual-goal-guided RRT for visual servoing of mobile robots with FOV constraint. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2022, **52**(4): 2073−2083

16 Kiumarsi B, Lewis F L. Actor-critic-based optimal tracking for partially unknown nonlinear discrete-time systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **26**(1): 140−151

17 Luo B, Wu H N, Li H X. Adaptive optimal control of highly dissipative nonlinear spatially distributed processes with neuro-dynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, **26**(4): 684−696

18 Liu D, Xue S, Zhao B, Luo B, Wei Q. Adaptive dynamic programming for control: A survey and recent advances. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, **51**(1): 142−160

19 Ming Z, Zhang H, Yan Y, Zhang J. Tracking control of discrete-time system with dynamic event-based adaptive dynamic programming. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2022, **69**(8): 3570−3574

20 Li S, Ding L, Gao H, Liu Y J, Huang L, Deng Z. ADP-based online tracking control of partially uncertain time-delayed nonlinear system and application to wheeled mobile robots. *IEEE Transactions on Cybernetics*, 2020, **50**(7): 3182−3194

21 Kong L, He W, Yang C, Sun C. Robust neurooptimal control for a robot via adaptive dynamic programming. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, **32**(6): 2584−2594

22 Zhang Hua-Guang, Zhang Xin, Luo Yan-Hong, Yang Jun. An overview of research on adaptive dynamic programming. *Acta Automatica Sinica*, 2013, **39**(4): 303−311
(张化光, 张欣, 罗艳红, 杨珺. 自适应动态规划综述. 自动化学报, 2013, **39**(4): 303−311)

23 Liu D, Wang D, Wang F Y, Li H, Yang X. Neural-network-based online HJB solution for optimal robust guaranteed cost control of continuous-time uncertain nonlinear systems. *IEEE Transactions on Cybernetics*, 2014, **44**(12): 2834−2847

24 Wang Ding, Mu Chao-Xu, Liu De-Rong. Data-driven nonlinear near-optimal regulation based on iterative neural dynamic programming. *Acta Automatica Sinica*, 2017, **43**(3): 366−375
(王鼎, 穆朝絮, 刘德荣. 基于迭代神经动态规划的数据驱动非线性近似最优调节. 自动化学报, 2017, **43**(3): 366−375)

25 Bhasin S, Kamalapurkar R, Johnson M, Vamvoudakis K G, Lewis F L, Dixon W E. A novel actor-critic-identifier architecture for approximate optimal control of uncertain nonlinear systems. *Automatica*, 2013, **49**(1): 82−92

26 Lin W S, Yang P C. Adaptive critic motion control design of autonomous wheeled mobile robot by dual heuristic programming. *Automatica*, 2008, **44**(11): 2716−2723

27 Cheng T, Lewis F L, Abu-Khalaf M. A neural network solution for fixed-final time optimal control of nonlinear systems. *Automatica*, 2007, **43**(3): 482−490

28 Heydari A, Balakrishnan S N. Fixed-final-time optimal tracking control of input-affine nonlinear systems. *Neurocomputing*, 2014, **129**(10): 528−539

29 Wang F Y, Jin N, Liu D, Wei Q. Adaptive dynamic programming for finite-horizon optimal control of discrete-time nonlinear systems with ε-error bound. *IEEE Transactions on Neural Networks*, 2010, **22**(1): 24−36

30 Zhao Q, Xu H, Jagannathan S. Neural network-based finite-horizon optimal control of uncertain affine nonlinear discrete-time systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2014, **26**(3): 486−499

31 Hartley R, Zisserman A. *Multiple View Geometry in Computer*

Vision. Cambridge: Cambridge University Press, 2003.

32 Siciliano B, Sciavicco L, Villani L, Oriolo G. *Robotics: Modelling, Planning and Control*. London: Springer-Verlag, 2009.

33 Zhang K, Chen J, Li Y, Zhang X. Visual tracking and depth estimation of mobile robots without desired velocity information. *IEEE Transactions on Cybernetics*, 2018, **50**(1): 361−373

34 Lewis F L, Vrabie D L, Syrmos V L. *Optimal Control, Third Edition*. Hoboken: John Wiley & Sons, 2012.

35 Finlayson B A. *The Method of Weighted Residuals and Variational Principles*. Philadelphia: Society for Industrial and Applied Mathematics, 2013.

36 Abu-Khalaf M, Lewis F L. Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach. *Automatica*, 2005, **41**(5): 779−791

37 Pakkhesal S, Shamaghdari S. Sum-of-squares-based policy iteration for suboptimal control of polynomial time-varying systems. *Asian Journal of Control*, 2022, **24**(6): 3022−3031

38 Wei Q, Liao Z, Yang Z, Li B, Liu D. Continuous-time time-varying policy iteration. *IEEE Transactions on Cybernetics*, 2020, **50**(12): 4958−4971

39 Zhang H, Cui L, Zhang X, Luo Y. Data-driven robust approximate optimal tracking control for unknown general nonlinear systems using adaptive dynamic programming method. *IEEE Transactions on Neural Networks*, 2011, **22**(12): 2226−2236

罗 彪 中南大学自动化学院教授. 主要研究方向为智能控制, 强化学习, 深度学习和自主决策. 本文通信作者.
E-mail: biao.luo@hotmail.com
(**LUO Biao** Professor at the School of Automation, Central South University. His research interest covers intelligent control, reinforcement learning, deep learning, and decision-making. Corresponding author of this paper.)

欧阳志华 中南大学自动化学院硕士研究生. 主要研究方向为移动机器人, 视觉伺服控制和自适应动态规划.
E-mail: ouyangzh@csu.edu.cn
(**OUYANG Zhi-Hua** Master student at the School of Automation, Central South University. His research interest covers mobile robot, visual servoing control, and adaptive dynamic programming.)

易昕宁 中南大学自动化学院硕士研究生. 主要研究方向为四旋翼视觉伺服控制, 自适应动态规划和强化学习.
E-mail: xnyi17@foxmail.com
(**YI Xin-Ning** Master student at the School of Automation, Central South University. Her research interest covers visual servoing control of quadrotor, adaptive dynamic programming, and reinforcement learning.)

刘德荣 南方科技大学工学院教授. 主要研究方向为智能控制理论及应用, 自适应动态规划, 人工神经网络和计算神经科学.
E-mail: derongliu@foxmail.com
(**LIU De-Rong** Professor at the School of Engineering, Southern University of Science and Technology. His research interest covers intelligent control theory and application, adaptive dynamic programming, artificial neural networks, and computational neuroscience.)