中国科学 : 信息科学 2026 年 第 56 卷 第 5 期: 1043–1054
SCIENTIA SINICA Informationis
安全控制理论与应用专题 • 论文

基于控制障碍函数的自适应动态规划方法及其在安全关键控制中的应用

李咸宁[1*], 王烨宾[2], 姜钟平[1]

1. Tandon School of Engineering, New York University, Brooklyn NY 11201, USA
2. Mitsubishi Electric Research Laboratories, Cambridge MA 02139, USA
* 通信作者. E-mail: xl5305@nyu.edu

收稿日期: 2025–11–04; 修回日期: 2026–01–27; 接受日期: 2026–02–28; 网络出版日期: 2026–04–17

美国国家科学基金会 (批准号: CNS-2148309, CPS-2227153)、国家自然科学基金 (批准号: 62521001) 和三菱电机研究实验室资助项目

摘要 本文针对安全关键场景下具有未知动力学的连续时间线性系统, 提出了一种结合学习型控制障碍函数安全滤波器的自适应动态规划框架. 通过自适应动态规划, 从输入 – 状态数据中学习得到一个次优反馈控制器. 针对控制障碍函数安全约束中未知的项, 利用神经网络基于在线数据进行逼近, 并将所得的学习型控制障碍函数约束融入到基于二次规划的安全滤波器中, 以修正控制输入, 从而保证系统安全约束的满足. 最后, 通过一个避障控制实例验证了所提控制方法的有效性.

关键词 基于学习的控制, 安全关键系统, 控制障碍函数, 自适应动态规划

1 引言

安全关键系统在现代工程领域中被广泛应用, 如自动驾驶和机器人等领域 [1~4]. 在这些应用中, 安全不仅是一种设计偏好, 而是一项不可或缺的要求, 因为系统失效可能导致灾难性后果, 例如自动驾驶车辆必须避免的碰撞. 为应对这些挑战, 已有大量控制方法被提出, 为安全关键控制设计奠定了基础. 模型预测控制通过反复求解带约束的优化问题来生成控制输入, 从而确保系统安全 [5]. 可达性分析通过刻画系统的所有可能轨迹集来验证是否能够避免进入不安全状态 [6]. 控制障碍函数 (control barrier function, CBF) 则提供了一种类似于李雅普诺夫函数的条件, 可以在反馈控制设计中直接施加安全约束 [7~9]. 此外, 一些混合方法结合了这些技术, 以发挥各自的互补优势 [10, 11]. 然而, 这些方法的实现通常依赖于对系统动态的精确且完整的先验知识.

随着现代控制系统的复杂性和规模不断提升, 获取精确的模型与完整的参数信息已成为一项艰巨的任务. 因此, 大量研究开始转向数据驱动的控制方法, 作为传统基于模型设计的可行替代方案. 在此背景下, 强化学习和自适应动态规划 (adaptive dynamic programming, ADP) 在自适应优化控制领域得到了广泛研究, 尤其在系统稳定化与输出调节问题中展现出显著效果 [12~14]. 然而, 现有多数 ADP 研究的一个重要局限在于未能显式考虑安全约束 [15, 16], 这限制了其在安全关键系统中的直接应用.

引用格式: 李咸宁, 王烨宾, 姜钟平. 基于控制障碍函数的自适应动态规划方法及其在安全关键控制中的应用. 中国科学: 信息科学, 2026, 56: 1043–1054, doi: 10.1360/SSI-2025-0475
Li X N, Wang Y B, Jiang Z-P. Adaptive dynamic programming with control barrier functions for safety-critical control applications. Sci Sin Inform, 2026, 56: 1043–1054, doi: 10.1360/SSI-2025-0475

近年来, 围绕学习型控制与安全约束相结合的研究逐渐受到关注, 部分工作尝试将控制障碍函数引入强化学习或自适应动态规划框架中, 以在学习过程中兼顾系统性能与安全性. 例如, 有研究通过在性能指标中引入与安全约束相关的惩罚或增广项, 使学习得到的控制策略在接近安全边界时自动趋于保守, 从而在一定程度上实现安全控制目标 [17]. 类似地, 也有工作针对离散时间非线性系统, 将安全约束以控制障碍函数的形式直接并入效用函数中, 并在此基础上构建自适应动态规划算法, 以实现含状态约束的最优控制设计 [18].

然而, 上述方法通常将安全约束与控制策略学习过程紧密耦合, 其安全性往往依赖于学习过程的收敛性质或对系统动力学的先验假设. 当系统动力学未知或处于在线学习阶段时, 这种安全 – 学习一体化的结构在安全可实施性分析和实际控制实现方面仍面临一定挑战. 针对这一问题, 也有研究提出通过额外的安全控制机制对学习策略进行屏蔽或修正, 从而在结构上将安全保障与策略学习过程分离, 并指出将安全函数本身作为学习目标可能会限制安全性的独立分析与保证 [19]. 此外, 将安全约束与控制策略综合过程直接耦合通常会形成递推时域的 OCP-CBF 问题, 其在线求解需要同时优化系统状态轨迹与控制输入序列. 在采用预测步长为 N 的情形下, 该类问题的决策变量维数通常为 $O(N(n+m))$, 其中 n 和 m 分别表示系统状态与控制输入的维数. 对于非线性系统或非线性安全约束, 该问题一般表现为非凸优化, 其计算复杂度随预测步长和系统维数的增加快速增长, 难以保证在多项式时间内实现实时求解. 相比之下, 本文方法在结构上将安全约束的执行与策略学习过程进行解耦, 在线阶段仅需求解一个决策变量维数为 $O(m)$ 且约束数量固定的凸二次规划问题, 从而具有良好的计算复杂度可预测性与实时实现能力 [20].

因此, 本文针对具有未知动态的连续时间线性系统, 采用一种分层式的控制器设计思路, 将性能优化与安全约束在结构上进行解耦. 在上层, ADP 参考控制器采用策略迭代 (policy iteration, PI) 算法, 从在线数据中直接学习次优参考反馈控制律, 而不考虑安全约束. 在下层, 基于 CBF 的安全滤波器利用神经网络和在线数据对安全约束中的未知项进行逼近, 并构建一个基于优化的最小干预控制问题. 通过求解仅含线性约束的二次规划, 对参考控制输入进行修正, 从而保证安全约束的满足.

本文的其余部分安排如下: 第 2 节针对具有安全约束的连续时间线性系统, 给出了优化控制问题的形式化描述. 第 3 节介绍了基于学习的安全控制器设计方法, 包括参考 ADP 控制器的学习过程以及应用于未知动力学线性系统的基于 CBF 的安全滤波器. 第 4 节通过包含避障的换道场景仿真验证了所提出的 ADP-CBF 方法的有效性. 最后, 第 5 节给出了本文的总结与展望.

符号说明: 对于任意向量 x, 记号 $\|x\|$ 表示其欧几里得范数; 对于任意矩阵 M, 记号 $\|M\|$ 表示由欧几里得向量范数诱导的矩阵范数, 即矩阵 M 的最大奇异值. 记 $\mathbb{R}^+$ 为正实数集合. 矩阵的克罗内克积 (Kronecker product) 用符号 $\otimes$ 表示, 常用于构造分块结构矩阵或对矩阵方程进行向量化. 给定矩阵 $A \in \mathbb{R}^{m\times n}$, 其向量化算子 $\mathrm{vec}(A)$ 通过按列堆叠 A 的所有元素生成 $\mathbb{R}^{mn}$ 中的一个向量. 若 $A = [a_1\ a_2\ \cdots\ a_n]$, 其中 $a_i \in \mathbb{R}^m$, 则 $\mathrm{vec}(A) = [a_1^\top\ a_2^\top\ \cdots\ a_n^\top]^\top$. 对于对称矩阵 $S \in \mathbb{R}^{n\times n}$, 其元素为 s_{ij}, $S \succeq 0$ 和 $S \succ 0$ 分别表示矩阵半正定和正定, 定义其对称向量化为 $\mathrm{svec}(S) = [s_{11}, 2s_{12}, \ldots, 2s_{1n},\ s_{22}, 2s_{23}, \ldots, s_{nn}]^\top$, 该操作将主对角元素与倍增后的非对角元素按顺序排列成 $\mathbb{R}^{n(n+1)/2}$ 中的向量. 对于状态向量 $x = [x_1, \ldots, x_n]^\top \in \mathbb{R}^n$, 二次特征映射 (quadratic-feature mapping) 定义为 $\psi(x) = [x_1^2,\ x_1x_2,\ \ldots,\ x_1x_n,\ x_2^2,\ x_2x_3,\ \ldots,\ x_{n-1}x_n,\ x_n^2]^\top$, 该映射将所有不重复的二次单项式堆叠成 $\mathbb{R}^{n(n+1)/2}$ 中的向量. 对于连续时间系统 $\dot{x} = Ax + Bu$, 若反馈增益 $K \in \mathbb{R}^{m\times n}$ 使得闭环矩阵 $A - BK$ 为 Hurwitz 矩阵 (即其所有特征值的实部均位于复平面的左半开区域内), 则称 K 为系统的稳定反馈增益, 从而保证原点的指数稳定性. 函数 $\alpha : \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}$ 若连续、严格单调递增且满足 $\alpha(0) = 0$, 则称其属于类 $\mathcal{K}$; 若进一步满足 $\alpha(s) \to \infty$ 当 $s \to \infty$ 时, 则称其属于类 $\mathcal{K}_\infty$.

2　问题表述

考虑如下形式的连续时间线性时不变 (linear time-invariant, LTI) 系统:

$$\dot{x} = Ax + Bu, \tag{1}$$

其中, 状态向量 $x \in \mathbb{R}^n$ 被假设为在反馈控制设计中可完全测量; 控制输入为 $u \in \mathbb{R}^m$; 系统矩阵 $A \in \mathbb{R}^{n\times n}$ 和 $B \in \mathbb{R}^{n\times m}$ 为常数. 本文所称 "未知动力学" 是指系统矩阵 A 与 B 为未知常数矩阵的线性时不变系统.

假设1　矩阵对 (A, B) 是可稳定化的.

在不失一般性的前提下, 我们考虑最小化如下线性二次型调节 (linear quadratic regulator, LQR) 代价函数:

$$J = \int_0^\infty \left(x^\top Q x + u^\top R u\right) dt, \tag{2}$$

其中, $Q = Q^\top \succeq 0$, $R = R^\top \succ 0$. 假设矩阵对 $(A, Q^{1/2})$ 是可观测的, 从而保证存在唯一的稳定化 LQR 解. 同时, 我们考虑控制输入的饱和约束:

$$u_{\min} \leqslant u \leqslant u_{\max}, \tag{3}$$

其中, $u_{\min}$ 和 $u_{\max}$ 分别表示控制输入的下限和上限. 此外, 为保证系统运行的安全性, 引入如下状态约束:

$$h_i(x) \geqslant 0, \quad i = 1, \ldots, N_{\text{safe}}, \tag{4}$$

其中, 函数 $h_i(x) : \mathbb{R}^n \to \mathbb{R}$ 表示施加于系统的状态安全约束.

假设2　每个函数 $h_i(x)$ $(i = 1, \ldots, N_{\text{safe}})$ 都是连续可微的.

为在保证实时性和理论可证性的前提下系统地施加状态安全约束, 本文采用基于 CBF 的建模方法来处理安全关键控制问题. 我们采用如下零化控制障碍函数 (zeroing CBF) 的定义 [21].

定义1 (控制障碍函数)　设 $h(x) : \mathbb{R}^n \to \mathbb{R}$ 为连续可微函数, 其定义安全集 C 为其超水平集:

$$C := \{x \in \mathbb{R}^n : h(x) \geqslant 0\}. \tag{5}$$

给定如下控制仿射系统:

$$\dot{x} = f(x) + g(x)u, \tag{6}$$

其中, $f(x)$ 与 $g(x)$ 为向量场, 且其每个分量均为局部 Lipschitz 连续函数; $x \in X \subset \mathbb{R}^n$ 与 $u \in U \subset \mathbb{R}^m$ 分别表示系统状态与控制输入. 若存在一个类 $\mathcal{K}_\infty$ 函数 $\alpha(\cdot)$, 使得对所有 $x \in C$ 均有

$$\sup_{u\in U} \left[L_f h(x) + L_g h(x) u + \alpha(h(x))\right] \geqslant 0, \tag{7}$$

则称 $h(x)$ 为系统的控制障碍函数 (CBF). 其中, $L_f h(x)$ 与 $L_g h(x)$ 分别表示 $h(x)$ 沿向量场 $f(x)$ 与 $g(x)$ 的 Lie 导数. 条件 (7) 意味着集合 C 是正向不变的, 即若初始状态满足 $x(0) \in C$, 则系统在之后的所有时刻都保持在集合 C 内.

以下定理形式化地说明: 若控制输入满足 CBF 条件, 则集合 C 在系统演化过程中保持正向不变性.

定理1 (CBF 的正向不变性)　设 $h(x)$ 为定义在式 (5) 所给安全集 C 上的控制障碍函数. 若任意 Lipschitz 连续的控制输入 $u(t)$ 在所有 $t \geqslant 0$ 时刻均满足条件 (7), 则集合 C 对于系统 (6) 是正向不变的.

假设3　本文仅关注相对阶为一的控制障碍函数, 即其导数直接包含控制输入 u.

本文将上述连续时间 LTI 系统的安全关键控制问题分解为两个子问题. 其中, 问题 1 关注在不考虑安全约束的情况下, 为系统构造一个能够最小化给定性能指标并保证闭环稳定性的标称控制输入, 作为系统的基础调节策略; 问题 2 则在此基础上, 通过引入基于优化的最小干预机制, 对标称控制输入进行必要修正, 以同时满足输入约束和安全约束要求.

问题1 (无约束优化调节问题)

$$\begin{aligned} \min_{u\in\mathbb{R}^m} \quad & \int_0^\infty \left(x^\top Q x + u^\top R u\right) dt \\ \text{s.t.} \quad & \dot{x} = Ax + Bu. \end{aligned} \tag{8}$$

设计目标是确定一个线性状态反馈控制律 $u_{\text{ref}} = -Kx$, 该控制律作为后续安全关键控制设计中的标称控制输入, 以最小化代价函数并确保闭环系统的稳定性.

问题2 (基于优化的最小干预控制)

$$\begin{aligned} u_{\text{safe}}(x, u_{\text{ref}}) = \arg\min_{u \in U} \quad & \frac{1}{2}\|u - u_{\text{ref}}\|^2 \\ \text{s.t.} \quad & L_f h_i(x) + L_g h_i(x) u + \alpha_i(h_i(x)) \geqslant 0, \\ & i = 1, \ldots, N_{\text{safe}}, \end{aligned} \tag{9}$$

其中, $U \subseteq \mathbb{R}^m$ 为在式 (3) 中定义的可行控制输入集合; 标称控制输入 u_{ref} 根据 CBF 约束进行修正, 以保证闭环系统的安全性.

采用上述分解式建模具有重要优势. 一方面, 问题 1 将闭环稳定性与性能优化问题从安全约束处理中解耦, 使得系统在标称意义下具备良好的调节性能; 另一方面, 问题 2 仅在必要时对标称控制输入进行最小幅度的调整, 从而使安全约束的引入不会过度影响系统的标称性能, 同时也降低了在线控制求解的复杂度. 若不依赖标称控制输入而直接联合考虑性能目标与安全约束构造综合控制器, 则往往需要在整个状态空间内同时权衡稳定性、性能与安全性, 不仅会显著增加控制设计与计算复杂度, 还可能导致安全约束在控制律中长期占据主导, 从而削弱标称性能表现.

3　基于学习的安全控制器设计

本节提出了一种安全控制器的学习策略. 首先, 通过基于 PI 的 ADP 方法学习参考控制器. 随后, 利用神经网络对 CBF 安全约束中依赖未知动态的项进行逼近, 并构建安全滤波器以修正控制输入, 从而保证闭环系统的稳定性与安全性.

3.1　阶段一: 参考 ADP 控制器学习

回顾问题 1, 在已知系统矩阵 A 和 B 的情况下, 参考控制器可通过求解代数 Riccati 方程 [22] 获得

$$A^\top P + PA + Q - PBR^{-1}B^\top P = 0. \tag{10}$$

根据假设 1, Riccati 方程 (10) 存在唯一的实对称正定解 P^*. 因此, 次优状态反馈增益矩阵可表示为

$$K^* = R^{-1}B^\top P^*. \tag{11}$$

由于 Riccati 方程 (10) 在 P 上是非线性的, 直接求解 P^* 对于高维系统而言较为困难. 为此, 可采用基于策略迭代的经典 Kleinman 算法 [23] 来迭代求解.

定理2 (Kleinman 算法)　设 $K_0 \in \mathbb{R}^{m\times n}$ 为初始稳定反馈增益. 对于每一个整数 $k \geqslant 0$, 定义 P_k 为以下 Lyapunov 方程的唯一对称正定解:

$$(A - BK_k)^\top P_k + P_k(A - BK_k) + Q + K_k^\top R K_k = 0, \tag{12}$$

并按如下方式更新反馈增益:

$$K_k = R^{-1}B^\top P_{k-1}. \tag{13}$$

随后, 可以证明每个闭环矩阵 $A - BK_k$ 均为 Hurwitz 矩阵, 序列 $\{P_k\}_{k=0}^{\infty}$ 满足单调性关系 $P^* \preceq P_{k+1} \preceq P_k$, 并且当 $k \to \infty$ 时, (P_k, K_k) 收敛至最优对偶 (P^*, K^*).

当系统矩阵 A 和 B 未知时, Kleinman 算法无法直接应用. 然而, 通过利用测得的状态与控制输入数据, 可以在保持收敛性的前提下迭代学习最优代价矩阵 P^* 与反馈增益 K^* [24]. 接下来, 介绍一种针对未知动力学 LTI 系统的连续时间、基于 PI 的 ADP 算法.

假设4　假设标称系统模型 (1) 是已知的, 并可据此获得系统 $\dot{x} = Ax + Bu$ 的一个稳定反馈增益矩阵 $K_0 \in \mathbb{R}^{m\times n}$.

在假设 1 与 4 下, 对于每一个 $k \in \mathbb{Z}_+$, 可计算满足式 (12) 的唯一对称正定矩阵 P_k, 并通过以下关系更新反馈增益:

$$K_{k+1} = R^{-1}B^\top P_k.$$

为便于重复利用已测得的状态与控制输入数据, 可将原系统重新表示为如下形式:

$$\dot{x} = A_k x + B\big(K_k x + u\big), \tag{14}$$

其中 $A_k = A - BK_k$. 该关系对任意 u 均成立, 因此可以通过设置 $u = -K_0 x + e$, 其中 $e(t)$ 为学习器/设计者人为注入的探索信号 (在实现中已知或可测), 用于提供持续激励以满足后续秩条件, 从而在不破坏 P_k 与 K_k 收敛性的前提下, 实现数据驱动学习所需的持续激励 [12].

随后, 对系统 (14) 的轨迹沿时间求 $x^\top P_k x$ 的导数, 结合式 (12) 与 (13), 并在区间 $[t,\, t+\delta t]$ 上积分, 可得

$$\begin{aligned} &x(t+\delta t)^\top P_k x(t+\delta t) - x(t)^\top P_k x(t) \\ &= \int_t^{t+\delta t} \Big[x^\top (A_k^\top P_k + P_k A_k)x + 2\,(u + K_k x)^\top B^\top P_k x\Big]\, d\tau \\ &= -\int_t^{t+\delta t} x^\top Q_k x\, d\tau + 2\int_t^{t+\delta t} (u + K_k x)^\top R\, K_{k+1} x\, d\tau, \end{aligned} \tag{15}$$

其中 $Q_k = Q + K_k^\top R K_k$. 基于式 (15), 可以得到如下线性矩阵方程:

$$\Lambda_k \begin{bmatrix} \mathrm{svec}\big(P_k\big) \\ \mathrm{vec}\big(K_{k+1}\big) \end{bmatrix} = \eta_k, \tag{16}$$

其中 $\Lambda_k \in \mathbb{R}^{\,l_1 \times \left(\frac{n(n+1)}{2} + mn\right)}$, $\eta_k \in \mathbb{R}^{l_1}$ 分别定义为

$$\Lambda_k = \left[\Delta_{xx},\ -2J_{xx}\left(I_n \otimes K_k^\top R\right) - 2J_{xu}\left(I_n \otimes R\right)\right],$$

$$\eta_k = -J_{xx}\,\mathrm{vec}(Q_k).$$

$\Delta_{xx} \in \mathbb{R}^{l_1 \times \frac{1}{2}n(n+1)}$, $J_{xx} \in \mathbb{R}^{l_1 \times n^2}$, $J_{xu} \in \mathbb{R}^{l_1 \times mn}$ 定义为

$$\begin{aligned} \Delta_{xx} = \big[\, &\psi(x(t_1)) - \psi(x(t_0)), \psi(x(t_2)) - \psi(x(t_1)), \ldots, \\ &\psi(x(t_{l_1})) - \psi(x(t_{l_1-1}))\,\big]^\top, \end{aligned}$$

$$J_{xx} = \left[\int_{t_0}^{t_1} x \otimes x\, d\tau,\ \int_{t_1}^{t_2} x \otimes x\, d\tau,\ \ldots,\ \int_{t_{l_1-1}}^{t_{l_1}} x \otimes x\, d\tau\right]^\top,$$

$$J_{xu} = \left[\int_{t_0}^{t_1} x \otimes u\, d\tau,\ \int_{t_1}^{t_2} x \otimes u\, d\tau,\ \ldots,\ \int_{t_{l_1-1}}^{t_{l_1}} x \otimes u\, d\tau\right]^\top,$$

其中 $0 \leqslant t_0 < t_1 < \cdots < t_{l_1}$.

对于任意稳定反馈增益矩阵 K_k, 若存在对称正定矩阵 $P_k = P_k^\top > 0$ 满足式 (12) 与 (13), 且 Λ_k 满秩, 则矩阵对 $(P_k,\, K_{k+1})$ 可以在未知 A 与 B 的情况下被唯一确定, 其所有元素组成的向量可按如下方式计算:

$$\begin{bmatrix} \mathrm{svec}\big(P_k\big) \\ \mathrm{vec}\big(K_{k+1}\big) \end{bmatrix} = (\Lambda_k^\top \Lambda_k)^{-1}\Lambda_k^\top \eta_k, \tag{17}$$

因此, PI 算法的收敛性可由引理 1 中给出的条件保证.

引理1　设 $l_0 > 0$ 为整数. 若对于所有 $l_1 \geqslant l_0$, 有

$$\operatorname{rank}([J_{xx}, J_{xu}]) = \frac{n(n+1)}{2} + mn, \tag{18}$$

则对于所有 $k \in \mathbb{Z}_+$, 矩阵 Λ_k 均具有满列秩.

定理3　假设初始增益矩阵 $K_0 \in \mathbb{R}^{m\times n}$ 使系统稳定. 在引理 1 所述条件下, 序列 $\{P_i\}_{i=0}^{\infty}$ 与 $\{K_j\}_{j=1}^{\infty}$ 均收敛至最优对偶 (P^*, K^*) [24].

注释1　需要说明的是, 尽管理论上 ADP 方法在满足相应条件时可收敛至最优解, 但在有限数据与数值计算条件下, 本文所学习得到的反馈控制器在实现层面视为次优控制器.

3.2　阶段二: 基于 CBF 的安全滤波器学习

对于具有未知动力学的系统 (1), 对控制障碍函数 $h_i(x)$ 关于时间求导, 并在时间区间 $[t, t+\delta t]$ 上积分, 可得

$$h_i(x)\Big|_t^{t+\delta t} = \int_t^{t+\delta t} \left(\nabla h_i(x)^\top Ax + \nabla h_i(x)^\top Bu\right) dt. \tag{19}$$

根据文献 [25], 在任意可行控制策略 u 下, 未知项 $\nabla h_i(x)^\top A$ 与 $\nabla h_i(x)^\top B$ 可通过神经网络逼近为

$$(\nabla h_i(x)^\top A)_{\mathrm{NN}} = \sum_{j=1}^{N_i} \phi_{i,j}(x)\, \hat{w}_{i,j}, \tag{20}$$

$$(\nabla h_i(x)^\top B)_{\mathrm{NN}} = \sum_{j=1}^{N_i} \phi_{i,j}(x)\, \hat{v}_{i,j}, \tag{21}$$

其中, $\phi_{i,j} : \mathbb{R}^n \to \mathbb{R}$ $(j = 1, 2, \ldots)$ 为一组光滑且线性无关的基函数序列; N_i 为与 $h_i(x)$ 对应的充分大的正整数; $\hat{w}_{i,j} \in \mathbb{R}^n$ 与 $\hat{v}_{i,j} \in \mathbb{R}^m$ 为待学习的参数.

将式 (20) 与 (21) 代入式 (19), 可得

$$h_i(x)\Big|_t^{t+\delta t} = \int_t^{t+\delta t} \left(\sum_{j=1}^{N_i} \phi_{i,j}(x)\, \hat{w}_{i,j}\, x + \sum_{j=1}^{N_i} \phi_{i,j}(x)\, \hat{v}_{i,j}\, u\right) dt. \tag{22}$$

进一步可得

$$\begin{aligned} h_i(x)\Big|_t^{t+\delta t} = &\int_t^{t+\delta t} \left(x^\top \otimes \phi_i(x)\right) dt \cdot \operatorname{vec}(\hat{W}_i) \\ &+ \int_t^{t+\delta t} \left(u^\top \otimes \phi_i(x)\right) dt \cdot \operatorname{vec}(\hat{V}_i), \end{aligned} \tag{23}$$

其中

$$\hat{W}_i = \left[\hat{w}_{i,1}^\top,\ \hat{w}_{i,2}^\top,\ \cdots,\ \hat{w}_{i,N_i}^\top\right]^\top,$$

$$\hat{V}_i = \left[\hat{v}_{i,1}^\top,\ \hat{v}_{i,2}^\top,\ \cdots,\ \hat{v}_{i,N_i}^\top\right]^\top,$$

$$\phi_i(x) = \left[\phi_{i,1}(x),\ \phi_{i,2}(x),\ \cdots,\ \phi_{i,N_i}(x)\right].$$

未知对偶 $(\nabla h_i(x_{\mathrm{lat}})^\top A, \nabla h_i(x_{\mathrm{lat}})^\top B)$ 可以在无需精确已知 A 与 B 的情况下, 通过求解如下线性矩阵方程确定. 该方程由式 (23) 推导而来, 利用在不同时间区间内采集的在线输入 – 状态数据建立

$$\Gamma_i \begin{bmatrix} \operatorname{vec}(\hat{W}_i) \\ \operatorname{vec}(\hat{V}_i) \end{bmatrix} = H_i, \tag{24}$$

其中 $H_i \in \mathbb{R}^{l_2}$, $\Gamma_i \in \mathbb{R}^{l_2 \times n(n+m)}$, 具体定义如下:

$$H_i = \left[h_i(x)\big|_{t_0}^{t_1} \; h_i(x)\big|_{t_1}^{t_2} \; \cdots \; h_i(x)\big|_{t_{l_2-1}}^{t_{l_2}} \right]^{\mathsf{T}},$$

$$\Gamma_i = \begin{bmatrix} \displaystyle\int_{t_0}^{t_1} x^{\top} \otimes \phi_i(x)\, dt & \displaystyle\int_{t_0}^{t_1} u^{\top} \otimes \phi_i(x)\, dt \\ \displaystyle\int_{t_1}^{t_2} x^{\top} \otimes \phi_i(x)\, dt & \displaystyle\int_{t_1}^{t_2} u^{\top} \otimes \phi_i(x)\, dt \\ \vdots & \vdots \\ \displaystyle\int_{t_{l_2-1}}^{t_{l_2}} x^{\top} \otimes \phi_i(x)\, dt & \displaystyle\int_{t_{l_2-1}}^{t_{l_2}} u^{\top} \otimes \phi_i(x)\, dt \end{bmatrix}.$$

假设5　假设 l_2 足够大, 使得矩阵 Γ_i 满秩:

$$\operatorname{rank}(\Gamma_i) = N_i(m+n). \tag{25}$$

在假设 5 下, 式 (24) 可以直接求解为

$$\begin{bmatrix} \operatorname{vec}(\hat{W}_i) \\ \operatorname{vec}(\hat{V}_i) \end{bmatrix} = (\Gamma_i^{\top}\Gamma_i)^{-1}\Gamma_i^{\top} H_i. \tag{26}$$

当 $\hat{W}_i$ 与 $\hat{V}_i$ 学习得到后, 基于 CBF 的安全约束可表示为

$$\phi_i(x)\,\hat{W}_i\,x + \phi_i(x)\,\hat{V}_i\,u + \alpha\big(h_i(x)\big) \geqslant 0. \tag{27}$$

注释2　对于 LTI 系统, 可以从 CBF 关于状态 x 的梯度分量中选取合适的基函数, 从而使 $\nabla h_i(x)^{\top}A$ 和 $\nabla h_i(x)^{\top}B$ 在无噪声的理想情况下能够被精确表示, 即基于神经网络的逼近在理论上不引入结构性逼近误差. 在实际应用中, 逼近误差主要来源于测量噪声和数值计算误差, 因此由式 (26) 求得的解可视为式 (24) 的最小二乘意义下的近似解. 当 CBF 安全约束参数选取相对保守时, 上述误差通常不会导致安全约束被违反.

注释3　秩条件 (18) 和 (25) 及相关的持续激励条件用于保证所构造数据矩阵具有足够的信息量. 在实际实现中, 该条件通常通过在闭环控制输入中叠加幅值受限的探索信号, 并在足够长的采样时间窗口内采集数据而近似满足. 已有研究和工程实践表明, 在每次迭代中使用数量不少于未知参数数量两倍的采样区间数据, 以及在周期激励情形下选取不小于激励周期的采样区间长度, 均有助于提高秩条件成立的可能性.

注释4　在参考控制器训练阶段, 采用基于标称系统模型的初始安全滤波器 $u_{\text{safe}}^{\text{nom}}(x, u_{\text{ref}})$ 可以实现安全探索, 并同时采集用于学习安全约束的数据. 这种方式使参考控制器与安全滤波器能够并行学习, 并通过在新的安全约束下复用已有数据, 从而加速安全滤波器的学习过程并提高样本效率.

注释5　本文所建立的理论结果严格针对相对阶为一的控制障碍函数情形, 即假设系统状态 x 可测, 且控制输入 u 在控制障碍函数 $h(x)$ 的一阶 Lie 导数中显式出现. 在该假设下, 所提出的基于数据驱动的学习与安全控制方法可在理论上保证 CBF 约束的可实施性. 需要指出的是, 本文在后续示例中还展示了在具有特定动力学结构的系统中, 部分具有更高相对阶的安全约束可通过结构性转化采用该方法进行处理, 但该扩展依赖于系统的具体动力学形式, 并不具有普遍性.

当所有基于 CBF 的安全约束均已学习完成后, 最初为参考控制器设计的安全滤波问题 2 可重新表述为

$$\begin{aligned} u_{\text{safe}}(x, u_{\text{ref}}) = \arg\min_{u \in U} \; & \frac{1}{2}\|u - u_{\text{ref}}\|^2 \\ \text{s.t.} \quad & \phi_i(x)\,\hat{W}_i\,x + \phi_i(x)\,\hat{V}_i\,u + \alpha\big(h_i(x)\big) \geqslant 0, \\ & i = 1, \ldots, N_{\text{safe}}. \end{aligned} \tag{28}$$

注释6　需要指出的是, 在同时考虑输入约束及多个安全约束的情形下, 由二次规划求解的控制输入在理论上可能出现不可行的情况, 例如当不同约束在当前状态下相互冲突时. 这种不可行性反映的是约束集合本身的不一致性, 而非控制律构造方法的失效. 在多约束条件下对二次规划可行性给出系统性的理论保证, 仍然是一个具有挑战性的研究问题, 超出了本文的研究范围.

最终, 整体的 ADP-CBF PI 学习框架如算法 1 所总结.

算法 1 ADP-CBF PI 学习算法.

Require: 初始稳定参考控制器 $u_{\mathrm{ref}} = -K_0 x$; 名义 CBF 安全滤波器 $u_{\mathrm{safe}}^{\mathrm{nom}}(x, u_{\mathrm{ref}})$; 探索噪声 e; 阈值 $\varepsilon > 0$; CBFs $h_i(x)$; 基函数 $\phi_{i,j}(x)$.

Ensure: 学习得到的 ADP 参考控制器 $u_{\mathrm{ref}} = -K^* x$, 以及学习得到的基于 CBF 的安全滤波器 $u_{\mathrm{safe}}(x, u_{\mathrm{ref}})$.

学习阶段 (安全探索阶段):

在时间区间 $[t_0, t_l]$ 上施加安全输入 $u_{\mathrm{safe}}^{\mathrm{nom}}(x, -K_0 x + e)$;

ADP 参考控制器学习:

计算 Δ_{xx}, J_{xx}, J_{xu}, 直至满足秩条件 (18), 并设 $k = 0$;

repeat

由式 (17) 求解 P_k 与 K_{k+1}, 然后更新 $k \leftarrow k + 1$;

until $\|P_k - P_{k-1}\| \leqslant \varepsilon$

基于 CBF 的安全约束学习 (并行执行):

计算 Γ_i 与 H_i, 直至满足秩条件 (25);

由式 (26) 求解 $\hat{W}_i$ 与 $\hat{V}_i$;

执行阶段 (学习后的安全控制器):

ADP 参考控制器由 $u_{\mathrm{ref}} = -K^* x$ 给出;

控制输入通过已学习的 CBF 安全滤波器 (28) 进行修正;

return　最终的安全控制输入 u_{safe}.

4　仿真结果

本节在一个包含障碍物避让的换道场景中对所提出的方法进行验证. 考虑一条直线路段, 其全局坐标系定义为 (X, Y), 其中 X 和 Y 分别表示车辆的纵向与横向坐标. 假设车辆的纵向速度保持恒定为 V_x, 且前轮转向角较小, 则车辆的纵向动力学可近似表示为

$$\dot{X} = V_x. \tag{29}$$

随后, 采用线性自行车模型 [26] 来描述车辆的横向动力学, 其形式如下:

$$\dot{x}_{\mathrm{lat}} = A_{\mathrm{lat}} x_{\mathrm{lat}} + B_{\mathrm{lat}} u_{\mathrm{lat}}, \tag{30}$$

其中, $x_{\mathrm{lat}} = [e_1(t),\ \dot{e}_1(t),\ e_2(t),\ \dot{e}_2(t)]^\top$, $e_1(t)$ 和 $e_2(t)$ 分别表示横向位移误差与航向角误差, u_{lat} 为前轮转向角输入. 线性系统矩阵具体形式如下:

$$A_{\mathrm{lat}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\frac{2C_{\alpha f}+2C_{\alpha r}}{mV_x} & \frac{2C_{\alpha f}+2C_{\alpha r}}{m} & -\frac{2C_{\alpha f}l_f+2C_{\alpha r}l_r}{mV_x} \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{2C_{\alpha f}l_f-2C_{\alpha r}l_r}{I_z V_x} & \frac{2C_{\alpha f}l_f-2C_{\alpha r}l_r}{I_z} & -\frac{2C_{\alpha f}l_f^2+2C_{\alpha r}l_r^2}{I_z V_x} \end{bmatrix},$$

$$B_{\mathrm{lat}} = \begin{bmatrix} 0 & \frac{2C_{\alpha f}}{m} & 0 & \frac{2C_{\alpha f}f}{I_z} \end{bmatrix}^\top,$$

其中, $C_{\alpha f}$ 与 $C_{\alpha r}$ 分别表示前后轮的侧偏刚度; l_f 与 l_r 分别表示车辆的前、后轴距; m 为车辆质量, I_z 为偏航惯性矩. 在不失一般性的情况下, 可将目标横向位置设为 Y 轴坐标原点 (即 $Y = e_1$), 从而使整个车辆横向动力学系统的跟踪目标为原点. 此时, 可将式 (8) 所示的调节问题构建出来, 以确定优化反馈控制律.

表 1　训练与控制器参数.

Table 1　Training and controller parameters.

Parameter	Value	Parameter	Value
Q	$\mathrm{diag}([100, 20, 20, 30])$	X_{ob}	20 m
R	I_1	Y_{ob}	−2.5 m
δt	0.01 s	$u_{\mathrm{lat,max}}$	0.3 rad
l	200	$u_{\mathrm{lat,min}}$	−0.3 rad
V_x	60 km/h	γ_1	100
r	3 m	γ_2	7

在该场景中, 考虑一个静止的圆形障碍物, 其中心位置为 $(X_{\mathrm{ob}}, Y_{\mathrm{ob}})$, 半径为 R_{ob}. 为了避免碰撞, 原始的安全约束可设计为

$$h_0(x_{\mathrm{veh}}) = (X - X_{\mathrm{ob}})^2 + (Y - Y_{\mathrm{ob}})^2 - r^2 \geqslant 0, \tag{31}$$

其中, $x_{\mathrm{veh}} = [X, x_{\mathrm{lat}}^{\top}]^{\top}$ 表示车辆的整体状态, 且 $r > R_{\mathrm{ob}}$ 为障碍物的膨胀半径, 用以考虑车辆自身几何尺寸与安全裕度.

从物理角度出发, 将 $h_0(x_{\mathrm{veh}})$ 设计为控制障碍函数 (CBF), 用于构建安全滤波器. 基于 CBF 的安全约束可表示为

$$\begin{aligned} &h_1(x_{\mathrm{veh}}) = \dot{h}_0(x_{\mathrm{veh}}) + \alpha_1(h_0(x_{\mathrm{veh}})) \geqslant 0 \\ &\quad \Rightarrow 2\,(X - X_{\mathrm{ob}})\,V_x + 2\,(e_1 - Y_{\mathrm{ob}})\,\dot{e}_1 + \alpha_1\left((X - X_{\mathrm{ob}})^2 + (e_1 - Y_{\mathrm{ob}})^2 - r^2\right) \geqslant 0. \end{aligned} \tag{32}$$

可以看出, 控制输入在 $h_1(x_{\mathrm{veh}})$ 中并未显式出现. 因此, $h_1(x_{\mathrm{veh}})$ 被视为高阶控制障碍函数 (high-order control barrier function, HOCBF) [27], 用于保证系统的安全性, 并且式 (32) 中的 $\mathcal{K}$ 类函数选取为具有系数 $\gamma_1 \in \mathbb{R}^+$ 的线性函数.

随后, 基于 HOCBF 的安全约束可进一步表示为

$$h_2(x_{\mathrm{veh}}) = \dot{h}_1(x_{\mathrm{veh}}) + \gamma_2(h_1(x_{\mathrm{veh}})) \geqslant 0, \tag{33}$$

其中, $\gamma_2 \in \mathbb{R}^+$ 为用于 HOCBF 安全约束中线性 $\mathcal{K}$ 类函数的系数.

注释7　在推导式 (32) 时, 利用了误差变量 e_1 与其导数 $\dot{e}_1$ 之间的已知动力学关系, 从而将原本具有较高相对阶的安全约束转化为等价的一阶约束形式, 以降低安全约束的实现复杂度. 需要指出的是, 该转化过程依赖于上述动力学关系的可获得性, 且并不依赖于具体的车辆参数取值. 因此, 本文方法在该类具有特定结构的系统中, 可扩展应用于涉及高阶控制障碍函数 (HOCBF) 的场景, 但并非对所有一般形式的 HOCBF 均适用.

随后, 应用所提出的 ADP-CBF 方法, 可以在不考虑约束的情况下学习次优参考控制器, 同时构建用于保证安全性的安全滤波器. 在该场景的仿真中, 数据采集时长为 2 s, 基函数选取为

$$\phi(x_{\mathrm{lat}}) = [e_1,\ \dot{e}_1,\ 1]\,,$$

其余训练与控制器参数汇总于表 1 中. 在学习过程中, 参考控制器的反馈增益收敛情况如图 1 所示, 其在 11 次迭代内收敛至次优反馈增益 $K_{\mathrm{lat}}^* = [10.00, 3.10, 30.26, 3.85]$.

关于基于 HOCBF 的安全约束参数选择, 低阶 CBF 的 $\mathcal{K}$ 类函数不应过于保守 (即 γ_1 不宜取值过小), 如图 2 所示. 否则, 无论高阶 CBF 的 $\mathcal{K}$ 类函数如何选取, 都无法消除由此积累的保守性. 在参考控制器与安全滤波器均学习完成后, 其闭环仿真结果如图 3 所示. 结果表明, 配备安全滤波器的参考控制器能够在换道过程中成功避开障碍物; 相比之下, 未使用安全滤波器的 ADP 控制器虽然能够稳定闭环系统, 但无法确保避障安全性.

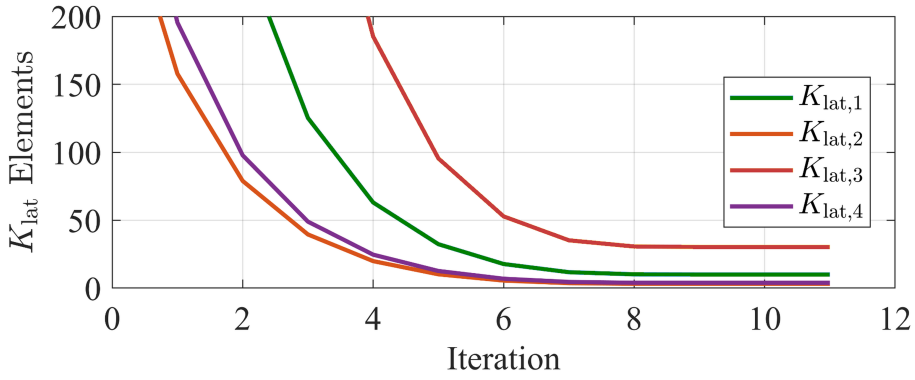

图 1　(网络版彩图) 参考控制器反馈增益的收敛过程.

Figure 1　(Color online) Convergence process of the feedback gain of the reference controller.

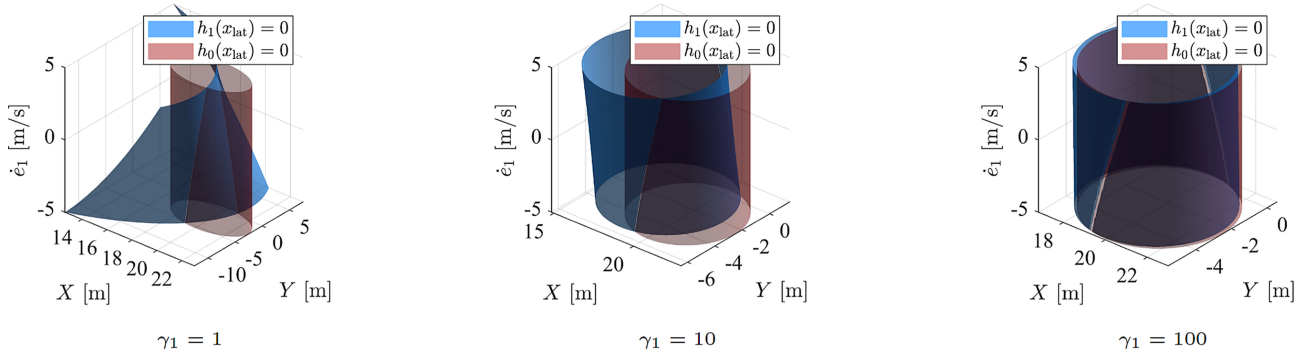

图 2　(网络版彩图) 基于 HOCBF 的安全约束参数选择.

Figure 2　(Color online) Parameter selection for the HOCBF-based safety constraints.

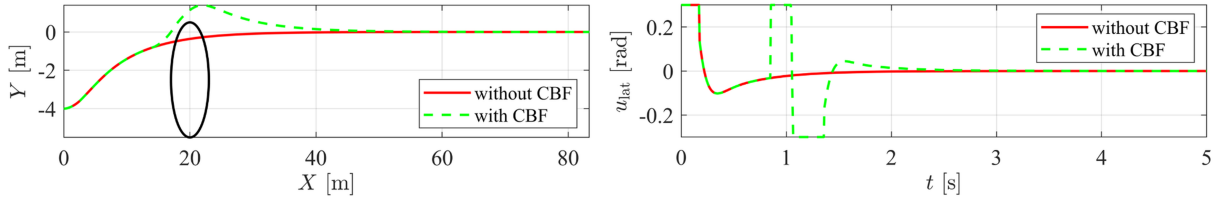

图 3　(网络版彩图) 闭环仿真结果.

Figure 3　(Color online) Closed-loop simulation results.

5　结论

本文针对具有未知动力学的连续时间 LTI 系统, 提出了一种基于学习的安全关键型控制器设计方法. 通过采用 ADP, 可以基于在线数据学习次优控制增益, 从而在不考虑安全约束的情况下保证闭环系统的稳定性. 进一步地, 通过基于 CBF 的安全约束构建的安全滤波器, 可确保闭环系统的安全性. 此外, 本文在含障碍物避让的车辆换道场景中验证了所提方法的有效性. 未来工作将进一步扩展该方法至非线性系统中, 利用神经网络对 CBF 安全约束中的未知项进行近似建模.

致谢　感谢 Kaan Ozbay 教授和 Sayan Chakraborty 在本文研究过程中给予的宝贵意见与支持.

参考文献

1 Zeng D, Jiang Y, Wang Y, et al. Robust adaptive control barrier functions for input-affine systems: application to uncertain manipulator safety constraints. IEEE Control Syst Lett, 2023, 8: 279–284

2 Chen D, Zhong R, Chen K, et al. Dynamic high-order control barrier functions with diffuser for safety-critical trajectory planning at signal-free intersections. IEEE Trans Intell Transp Syst, 2025, 26: 14011–14024

3 Hong J, Li X, Sun K, et al. Fast optimization solution and road tests for predictive cruise control of heavy-duty trucks. Proc Inst Mech Eng Part D J Automob Eng, 2025, doi:10.1177/09544070251369334

4 Li X, Wang Y, Ozbay K, et al. Physics-informed machine learning with heuristic feedback control layer for autonomous vehicle control. In: Proceedings of the IEEE Intelligent Vehicles Symposium, 2025. 2304–2309

5 Koller T, Berkenkamp F, Turchetta M, et al. Learning-based model predictive control for safe exploration. In: Proceedings of the IEEE Conference on Decision and Control, 2018. 6059–6066

6 Bajcsy A, Bansal S, Bronstein E, et al. An efficient reachability-based framework for provably safe autonomous navigation in unknown environments. In: Proceedings of the 58th IEEE Conference on Decision and Control, 2019. 1758–1765

7 Ames A D, Xu X, Grizzle J W, et al. Control barrier function based quadratic programs for safety critical systems. IEEE Trans Automat Contr, 2017, 62: 3861–3876

8 Ames A D, Coogan S, Egerstedt M, et al. Control barrier functions: theory and applications. In: Proceedings of the 18th European Control Conference, 2019. 3420–3431

9 Xiao W, Cassandras C G, Belta C A. Bridging the gap between optimal trajectory planning and safety-critical control with applications to autonomous vehicles. Automatica, 2021, 129: 109592

10 Choi J J, Lee D, Sreenath K, et al. Robust control barrier-value functions for safety-critical control. In: Proceedings of the 60th IEEE Conference on Decision and Control, 2021. 6814–6821

11 Zeng J, Zhang B, Sreenath K. Safety-critical model predictive control with discrete-time control barrier function. In: Proceedings of the American Control Conference, 2021. 3882–3889

12 Jiang Y, Jiang Z P. Robust Adaptive Dynamic Programming. Hoboken: John Wiley & Sons, 2017

13 Chakraborty S, Gao W, Vamvoudakis K G, et al. Active learning-based control for resiliency of uncertain systems under DoS attacks. IEEE Control Syst Lett, 2024, 8: 3297–3302

14 Kiran B R, Sobh I, Talpaert V, et al. Deep reinforcement learning for autonomous driving: a survey. IEEE Trans Intell Transp Syst, 2022, 23: 4909–4926

15 Chakraborty S, Cui L, Ozbay K, et al. Automated lane changing control in mixed traffic: an adaptive dynamic programming approach. Transp Res Part B-Methodological, 2024, 187: 103026

16 Cui L, Chakraborty S, Ozbay K, et al. Data-driven combined longitudinal and lateral control for the car following problem. IEEE Trans Contr Syst Technol, 2025, 33: 991–1005

17 Marvi Z, Kiumarsi B. Safe reinforcement learning: a control barrier function optimization approach. Intl J Robust Nonlinear, 2021, 31: 1923–1940

18 Xu J, Wang J, Rao J, et al. Adaptive dynamic programming for optimal control of discrete-time nonlinear system with state constraints based on control barrier function. Intl J Robust Nonlinear, 2022, 32: 3408–3424

19 Cohen M H, Belta C. Safe exploration in model-based reinforcement learning using control barrier functions. Automatica, 2023, 147: 110684

20 Xiao W, Cassandras C G, Belta C. Safe Autonomy With Control Barrier Functions: Theory and Applications. Cham: Springer, 2023

21 Ames A D, Grizzle J W, Tabuada P. Control barrier function based quadratic programs with application to adaptive cruise control. In: Proceedings of the 53rd IEEE Conference on Decision and Control, 2014. 6271–6278

22 Lewis F L, Vrabie D, Syrmos V L. Optimal Control. Hoboken: John Wiley & Sons, 2012

23 Kleinman D. On an iterative technique for Riccati equation computations. IEEE Trans Automat Contr, 1968, 13: 114–115

24 Jiang Y, Jiang Z P. Computational adaptive optimal control for continuous-time linear systems with completely unknown dynamics. Automatica, 2012, 48: 2699–2704

25 Powell M J D. Approximation Theory and Methods. Cambridge: Cambridge University Press, 1981

26 Rajamani R. Vehicle Dynamics and Control. Berlin: Springer, 2006

27 Xiao W, Belta C. Control barrier functions for systems with high relative degree. In: Proceedings of the 58th IEEE Conference on Decision and Control, 2019. 474–479

Adaptive dynamic programming with control barrier functions for safety-critical control applications

Xianning LI[1*], Yebin WANG[2] & Zhong-Ping JIANG[1]

1. *Tandon School of Engineering, New York University, Brooklyn NY 11201, USA*
2. *Mitsubishi Electric Research Laboratories, Cambridge MA 02139, USA*
* Corresponding author. E-mail: xl5305@nyu.edu

Abstract In this paper, we propose an adaptive dynamic programming (ADP) framework augmented with a learned control barrier function (CBF)-based safety filter for continuous-time linear systems with unknown dynamics in safety-critical scenarios. Using adaptive dynamic programming, a sub-optimal feedback controller is learned from input-state data. The unknown terms in the CBF-based safety constraints are approximated from online data using neural networks, and the resulting learned CBF constraints are incorporated into a quadratic program-based safety filter that modifies the control input to ensure satisfaction of the safety constraints. The effectiveness of the proposed control methodology is illustrated via an obstacle-avoidance example.

Keywords learning-based control, safety-critical systems, control barrier functions (CBFs), adaptive dynamic programming (ADP)