arXiv:2601.00476v1 [math.OC] 1 Jan 2026

基于屏障状态的安全自适应反馈控制 ⋆

Trivikram Satharasi * **Tochukwu E. Ogri** * **Muzaffar Qureshi** * **Kyle Volle** ** **Rushikesh Kamalapurkar** *

*佛罗里达大学机械与航空航天工程系，美国佛罗里达州盖恩斯维尔，邮编 *32608*（电子邮件：t.satharasi@ufl.edu, tochukwu.ogri@ufl.edu, muzaffar.qureshi@ufl.edu, rkamalapurkar@ufl.edu。）

**_Torch Technologies_，美国佛罗里达州希利马（电子邮件：Kyle.Volle@torchtechnologies.com.）

摘要： 本文提出了一种安全反馈控制框架，通过结合屏障状态增广与自适应动态规划（ADP），处理具有参数不确定性的非线性控制仿射系统。所设计的基于自适应动态规划的控制器通过优化一个显式惩罚屏障状态的价值函数来实现控制不变性，从而将安全性直接嵌入贝尔曼结构中。利用基于模型的强化学习计算得到的近似最优控制策略与并发学习估计器相结合，以识别未知参数，并在无需持续激励的情况下保证一致收敛。借助屏障状态李雅普诺夫函数，我们证明了屏障动力学的有界性，并证明了闭环稳定性和安全性。针对最优避障问题的数值仿真验证了所提出方法的有效性。

关键词：安全最优控制；自适应动态规划；屏障状态。

1. 引言

随着自主系统日益普及，安全性已成为影响其在现实世界中部署的考量因素之一。例如，飞机制导与导航、自主驾驶车辆和核反应堆等系统，都将安全性作为主要设计目标之一。安全要求推动了相关研究，旨在理解并设计能够在不确定性下保障安全的控制器。在控制系统中，安全通常被视为保持在有利条件所对应的区域或状态内，并避开不利区域和状态。这一概念由 Nagumo（1942年）和 Blanchini（1999年）进一步形式化为：相对于动力学流，集合具有正向不变性。

过去十年间，控制屏障函数（控制障碍函数，CBF）已成为保障安全的常用工具，其方法是确保指定集合具有正向不变性，参见 Wieland 和 Allgöwer（2007年）；Ames 等人（2014年）。早期关于控制屏障函数的研究以模型为基础，而近期研究则着重于拓展基于控制屏障函数的反馈控制，以确保模型不确定性下的安全性（Mahmud 等人（2021年）；Robey 等人（2021年）；Lopez 等人（2020年）；Ogri 等人（2025年））。

为保证安全性，Cohen 和 Belta (2020); Deptula 等人 (2020); Marvi 和 Kiumarsi (2021) 等基于控制障碍函数的强化学习（RL）方法将无模型策略学习与基于模型的安全过滤相结合，通过在最优控制问题的代价函数中纳入控制障碍函数（CBF），可在执行过程中确保给定安全集的正向不变性。然而，Cohen 和 Belta（2020）；Deptula 等人（2020）；Marvi 和 Kiumarsi（2021）提出的基于CBF的强化学习方法可能导致安全性与性能之间的权衡：过于严格的安全条件会妨碍系统实现其稳定性目标。为避免这种权衡，Cohen 和 Belta（2023）以及 Cohen 等人（2023）通过开发专用安全控制器，提出了一种解耦的安全感知控制器，并将其与基于模型的强化学习控制器配合使用，使系统能够在不损害稳定性的情况下维持安全（Kamalapurkar 等人（2016））。然而，Cohen 和 Belta（2023）的结果依赖于类李雅普诺夫障碍函数梯度的一个 L_∞ 界，以保证稳定性；而在安全集边界附近，这一界可能难以获得。Cohen 等人（2023）虽然使用零化障碍函数放宽了这一界限，但 Cohen 和 Belta（2023）以及 Cohen 等人（2023）中的控制器以性能为导向，却并非最优。此外，基于强化学习和二次规划的控制器（如 Cohen 等人（2023）；Krstic（2023））通过在每一步求解二次规划来确保安全，但这种安全保障可能以偏离原始基于模型的强化学习策略的方式修改控制输入，从而可能损害基于模型的强化学习控制器原本能够保证的稳定性。近期研究进一步表明，基于控制障碍函数的二次规划可能引入不理想的渐近稳定平衡点（Reis 等人（2021）），这可能导致闭环系统收敛到非预期的轨迹或平衡点，而非基于学习的控制器所指定的目标。

⋆ 本研究部分由美国空军研究实验室通过合同编号 FA8651-24-1-0019 资助。本文中的任何观点、研究结果或建议均属于作者所有，不一定反映资助机构的意见。

另一种常见的安全控制方法是屏障变换（BT）方法（Yang 等人（2019）；Greene 等人（2020）；Mahmud 等人（2021）），该方法对系统状态进行非线性坐标变换。与基于控制障碍函数的技术不同，Yang 等人（2019）的BT方法通过状态变换，将受约束的状态反馈控制问题改写为等价的无约束问题，从而能够直接设计出天然满足状态约束的控制器。随后，该方法的扩展放宽了持续激励等严格假设（Greene 等人（2020）），并引入基于学习的技术来处理参数不确定性（Mahmud 等人（2021））。尽管基于BT的方法能够得到可验证的安全状态反馈控制器，但其适用范围仅限于盒约束。

为解决这些局限，Almubarak 等（2025）引入了一种屏障算子，用于构造一种内在安全坐标，称为屏障状态（BaS）。该坐标将原始系统转换为扩展模型，使安全性表现为部分状态的有界性，并在增广动力学中形成正向不变安全锥。该方法完全绕开了合成有效CBF这一要求，而这正是传统CBF的一项关键局限。BaS直接集成到系统动力学中，确保控制决策兼顾当前安全性和近期未来的安全性。与基于BT的技术不同，基于BaS的安全控制适用于更广泛的一类约束。

BaS直接出现在系统动力学中，使得能够通过标准非线性控制工具设计安全认证控制器。BaS和CBF方法的一个根本局限在于，它们都依赖于对系统动力学的完整了解，而这在实践中很少能够实现。针对不确定系统的CBF扩展（Taylor和Ames（2020年）；Lopez等人（2020年））虽然提升了鲁棒性，但往往会施加保守控制界，限制系统性能。将BaS与参数估计结合（Aoun等人（2023）；Al-Sunni等人（2025年）），可通过自适应技术在不确定性下维持安全性；然而，Al-Sunni等人（2025年）等方法并不保证参数估计值收敛到其真实值，因此鲁棒安全与精确系统辨识之间仍存在差距。

本文将积分并发学习（ICL）、自适应动态规划（ADP）与BaS增广相结合，提出了一种适用于参数未知的非线性仿射系统的安全反馈控制框架。本文的主要贡献如下。首先，通过将安全约束嵌入增广系统状态，BaS-RL框架确保即使系统参数未知，控制器仍能保持近最优安全性。其次，通过结合行动者-评论家ADP框架与基于ICL的参数估计器，我们的方法能够学习最优价值函数，并生成兼顾安全性、稳定性和性能的控制策略。基于李雅普诺夫的分析表明，所提出的ICL-BaS-RL架构能够同时实现预期的安全性、稳定性和性能目标，标志着相较于现有安全最优控制方法的一项重大进展。

2. 问题表述

考虑如下形式的非线性控制仿射系统

$$\dot{x} = Y(x)\theta + f(x) + g(x)u, \quad x(0) = x_0, \tag{1}$$

其中，$x \in \mathbb{R}^n$ 是系统状态向量，$u \in \mathbb{R}^m$ 是控制输入，$\theta \in \mathbb{R}^p$ 是未知参数向量，$Y: \mathbb{R}^n \to \mathbb{R}^{n\times p}$、$f: \mathbb{R}^n \to \mathbb{R}^n$以及 $g: \mathbb{R}^n \to \mathbb{R}^{n\times m}$ 分别是已知的系统回归矩阵、漂移动力学和控制效能矩阵。

假设*1*。函数 f、g和 Y 是局部利普希茨连续的，并满足 $f(0) = 0$、$Y(0) = 0$以及$0 < \|g(x)\| \le G$，其中 $G > 0$，且对所有 $x \in \mathbb{R}^n$均成立。

假设*2*。存在一个已知的 $\theta > 0$，使得$||\theta|| \le \theta$。

为便于设计，定义投影算子

$$\mathrm{proj}_{\Theta}(\mu, v) \coloneqq \begin{cases} v, & \text{if } \mu \in \mathrm{int}(\Theta), \\ \left(I_p - \dfrac{\Gamma\mu\mu^\top}{\mu^\top\Gamma\mu}\right) v, & \text{if } \mu \in \partial\Theta \text{ and } \mu^\top v > 0, \\ v, & \text{otherwise}, \end{cases} \tag{2}$$

其中，$\Theta := \{\theta \in \mathbb{R}^p \mid \|\theta\| \le \theta\}$，$I_p$是 $p \times p$单位矩阵，$\Gamma \in \mathbb{R}^{p\times p}$ 是对称正定自适应增益矩阵，且 μ，$v \in \mathbb{R}^p$。

(2) 中的投影算子与 Cai 等人（2006）以及 Krstic 等人（1995）附录 E 中引入的投影算子相同。

为定义安全性，令 $\mathcal{S} \subset \mathbb{R}^n$ 表示连续可微函数 $h \in C^1(\mathbb{R}^n;\mathbb{R})$ 的零超水平集，其定义为

$$\mathcal{S} = \{x \in \mathbb{R}^n \mid h(x) \ge 0\}, \tag{3}$$

$$\partial\mathcal{S} = \{x \in \mathbb{R}^n \mid h(x) = 0\}, \tag{4}$$

$$\mathrm{int}(\mathcal{S}) = \{x \in \mathbb{R}^n \mid h(x) > 0\}, \tag{5}$$

其中，$\partial\mathcal{S}$ 和 int($\mathcal{S}$) 分别表示 $\mathcal{S}$ 的边界和内部。

假设*3.* 集合 $\mathcal{S}$ 非空、不含孤立点，且 $\nabla h(x) \neq 0$，$\forall x \in \partial\mathcal{S}$。

令 $t \mapsto x(t)$ 表示系统 (1) 在输入 $u(\cdot)$ 作用下从 x_0 出发的轨迹。本文采用的安全性概念基于条件控制不变性，其形式化定义如下。

定义*1.* 若存在可容许控制策略 $u(\cdot) \in \mathcal{U}$，使得对于任意初始条件 $x_0 \in \mathcal{S}_0$，相应的轨迹均满足 $x(t) \in \mathcal{S}$，$\forall t \ge 0$，则称集合 $\mathcal{S} \subset \mathbb{R}^n$ 关于 $\mathcal{S}_0 \subseteq \mathcal{S}$ 条件控制不变。

本文旨在估计系统 (1) 的未知参数，同时综合设计安全反馈控制策略 π，以确保满足定义1中所述的条件控制不变性。

3. 安全嵌入模型

为了以保留可微性、支持观测器和估计器设计并具备基于李雅普诺夫的保证的方式，将约束感知直接融入系统模型，可考虑一个屏障算子

$B \in \mathcal{C}^{\infty}((0,\ \infty);\mathbb{R})$，使得 $\frac{dB}{da}(a) > 0 \quad \forall a > 0$，$\lim_{a\to 0^+} B(a) = \infty$，且 $\frac{dB}{da}(a) \circ B^{-1} \in \mathcal{C}^{\infty}((0,\ \infty);\mathbb{R})$。对于定义 $\mathcal{S}$ 的连续可微函数 $h \in \mathcal{C}^1(\mathbb{R}^n;\mathbb{R})$，定义复合障碍函数

$$\beta(x) \coloneqq B\big(h(x)\big). \tag{6}$$

由于 B 在 $h(x) \to 0^+$时发散，β 满足

$$\beta(x) \geq 0,\ \forall x \in \operatorname{int}(\mathcal{S}), \qquad \lim_{x\to\partial\mathcal{S}} \beta(x) = \infty, \tag{7}$$

因此，安全条件 $h(x) > 0$等价于$\beta(x) < \infty$。下面的引理是经典结果；不过，为完整起见，我们仍予以重述。

引理*1*。（Almubarak 等（2025））设 $\mathcal{S} \subset \mathbb{R}^n$ 为式（3）中定义的安全集，并设 $\beta:\ \mathbb{R}^n \to [0,\infty]$ 为与 $\mathcal{S}$ 相关、如式（6）所规定的复合障碍函数。则对于任意 $x_0 \in \mathcal{S}_0$，可容许控制策略 $u(\cdot) \in \mathcal{U}$ 是安全的，即 $\mathcal{S}$ 相对于 $\mathcal{S}_0 \subseteq \mathcal{S}$是条件控制不变的，当且仅当 $\beta(x) < \infty \forall x \in \mathcal{S}$。

引理1促使我们采用将安全性嵌入系统动力学的安全控制策略，因为只要 $\beta(x(t))$ 保持有限，状态就会安全地保持在 $\mathcal{S}$内。

通过坐标变换，BaS 定义为$z \coloneqq \beta(x) - \beta_0$，其中 $\beta_0 \coloneqq \beta(0)$。在这些坐标下，动力学变为

$$\dot{z} = \Phi(z+\beta_0)\,\nabla h(x)\Big(Y(x)\theta + f(x) + g(x)u\Big). \tag{8}$$

其中 $\Phi = \frac{B}{dh}(b) \circ B^{-1} \in \mathcal{C}^{\infty}((0,\infty);\mathbb{R})$ 表现为由屏障算子诱导的非线性增益，且初始 BaS 为 $z_0 = \beta(x_0) - \beta_0$。

BaS 在扩展的 (x, z) 空间中诱导出一个前向不变锥，该锥将安全性编码为 z 的有界性。

引理*2*. 定义集合 $\mathcal{C} \coloneqq \{(x,z) \in \mathbb{R}^n \times \mathbb{R} \mid z = \beta(x) - \beta_0\}$。考虑在局部Lipschitz反馈 $u(x)$ 下，方程 (1) 和 (8) 的任意解 $(x(t)\ z(t))$。则对于所有 $t \geq 0$，$z(t) = \beta(x(t)) - \beta_0$ 当且仅当 $(x(0)\ z(0)) \in \mathcal{C}$。此外，对于任意 $t \geq 0$，$x(t) \in \mathcal{S}$ 当且仅当 $\beta(x(t)) < \infty$ 当且仅当 $z(t) < \infty$。

证明。对 $\beta(x(t)) - \beta_0$ 沿解求导，并与 (8) 中的 $\dot{z}$ 进行比较。根据局部Lipschitz性下解的唯一性，只要两个标量轨迹在 $t = 0$处相等，它们就始终重合。最后的等价关系由第一部分以及定义性质 $\beta(x) = \infty$ 当且仅当 $x \notin \mathcal{S}$得出。□

注记*1*。集合 $\mathcal{C} \subset \mathbb{R}^n \times \mathbb{R}$ 是映射$x \mapsto \beta(x) - \beta_0$的图。尽管 $\mathcal{C}$不是 $\mathcal{S}$的子集，但其在状态空间上的投影满足 $\pi_x(\mathcal{C}) = \{x \in \mathbb{R}^n \mid \beta(x) < \infty\} = \operatorname{int}(\mathcal{S})$。因此，$\mathcal{C}$ 在扩展的 (x, z) 坐标中，提供了安全集内部的提升表示。

引理2将安全条件形式化为与某个提升坐标的前向有界性在动态上等价，因此可以直接对z 而非对 $\beta(x)$使用Lyapunov论证。

由于 (8) 中的 θ 未知，我们引入估计BaS，记为 $\hat{z} \in \mathbb{R}$，其演化方程为

$$\dot{\hat{z}} = \Phi(z+\beta_0)\,\nabla h(x)\Big(Y(x)\hat{\theta} + f(x) + g(x)u\Big) + \gamma(z-\hat{z}), \tag{9}$$

其中，$\gamma > 0$ 是用户自定义的观测器增益。

定义 BaS估计误差为 $\tilde{z} \coloneqq z - \hat{z}$。用（8）减去（9），可得误差动态

$$\dot{\tilde{z}} = \Phi(z+\beta_0)\,\nabla h(x) Y(x)\tilde{\theta} - \gamma\tilde{z}, \tag{10}$$

其中 $\tilde{\theta} \coloneqq \theta - \hat{\theta}$ 为参数估计误差。

下一节将介绍一种针对（10）的参数估计律，该方法在确保$(\tilde{z},\ \theta)$的稳定性和有界性的同时，在较宽松的激励条件下保持引理2的正向不变锥。

4. 参数估计器设计

本节将提出一种 ICL更新律（Chowdhary and Johnson (2010)），用于估计未知参数。该参数估计器基于这样一个事实：在时刻 t和时刻 $t-T$，估计的 BaS 与真实 BaS 之间的差值可表示为参数 θ 的仿射函数，以及一个会随着 BaS估计误差减小而减小的残差。

引理*4*. 对于任意固定延迟 $T > 0$ 以及所有 $t \geq T$，系统动力学满足以下增量关系

$$\mathcal{X}(t) = \mathcal{Y}(t)\theta + \mathcal{G}_{fu}(t), \tag{11}$$

where $\mathcal{X}(t) \coloneqq x(t) - x(t-T)$, $\mathcal{Y}(t) \coloneqq \int_{t-T}^{t} Y(s(\tau))\,d\tau$, and $\mathcal{G}_{fu}(t) \coloneqq \int_{t-T}^{t} \big(f(s(\tau)) + g(s(\tau))u(\tau)\big)\,d\tau$.

证明。将微积分基本定理应用于式（1），即可直接得到该结果。□

*4.1 ICL*更新律与数据管理

引理4表明，对于任意时刻 τ，参数回归误差满足

$$\mathcal{Y}(t)\,\tilde{\theta}(\tau) = \mathcal{X}(t) - \mathcal{G}_{fu}(t) - \mathcal{Y}(t)\hat{\theta}(\tau), \tag{12}$$

由此可得以下ICL更新律

$$\dot{\hat{\theta}} = \operatorname{proj}_{\Theta}\big(\hat{\theta},\, \Gamma\phi\big), \tag{13}$$

其中，$\Gamma \in \mathbb{R}^{p\times p}$ 是时变正定最小二乘增益矩阵，且满足

$$\dot{\Gamma} = \begin{cases} \beta_\theta\Gamma - k_\theta\Gamma\sum_{i=1}^{N} \frac{\mathcal{Y}_i^{\top}\mathcal{Y}_i}{1+\kappa\|\mathcal{Y}_i\|^2}\Gamma, & \text{if } \hat{\theta} \in \operatorname{int}(\Theta) \\ & \text{or } (\hat{\theta} \in \partial\Theta \text{ and } (\Gamma\phi)^{\top}\hat{\theta} \leq 0), \\ 0 & \text{otherwise,} \end{cases} \tag{14}$$

and

$$\phi \coloneqq Y(x)^{\top}\nabla h(x)^{\top}\Phi(z+\beta_0)^{\top}\tilde{z} + k_\theta \sum_{i=1}^{N} \frac{\mathcal{Y}_i^{\top}\big(\mathcal{X}_i - \mathcal{G}_{fu,i} - \mathcal{Y}_i\hat{\theta}\big)}{1+\kappa\|\mathcal{Y}_i\|^2}, \tag{15}$$

其中，$\mathcal{X}_i \coloneqq \mathcal{X}(t_i)$，$\mathcal{Y}_i \coloneqq \mathcal{Y}(t_i)$，$\mathcal{G}_{fu,\ i} \coloneqq \mathcal{G}_{fu}(t_i)$。增益 k_θ，β_θ，$\kappa \in \mathbb{R}_{>0}$ 是用户定义的自适应增益。由于$\Gamma(t)$是正定的，因此存在常数 $0 < \underline{\Gamma} \leq \overline{\Gamma}$，使得对于所有 $t \geq 0$，

$$\underline{\Gamma}\, I_p \preceq \Gamma(t) \preceq \overline{\Gamma}\, I_p. \tag{16}$$

为便于数据重用，在一个历史栈中维护一组有限的延迟样本

$$\mathcal{H} \coloneqq \big\{\, \mathcal{X}_i,\, \mathcal{Y}_i,\, \mathcal{G}_{fu,i} \,\big\}_{i=1}^{N}, \tag{17}$$

该历史栈根据数据持久性条件在离散时刻 t_i 进行更新。

为便于分析，将参数估计误差$\tilde{\theta} := \hat{\theta} - \theta$定义为 $\theta := \theta - \hat{\theta}$。相应的动力学由式(13)得出：

$$\dot{\hat{\theta}} = \mathrm{proj}_{\Theta}\left(\hat{\theta}, \left(-k_\theta \Gamma \Sigma_{\mathcal{Y}} \tilde{\theta} - \Gamma Y(x)^\top \nabla h(x)^\top \Phi(z+\beta_0)^\top \tilde{z}\right)\right), \tag{18}$$

其中，$\Sigma_{\mathcal{Y}} \in \mathbb{R}^{p\times p}$ 表示定义为 $\Sigma_{\mathcal{Y}} := \sum_{i=1}^{N} \sigma_i \mathcal{Y}_i^\top \mathcal{Y}_i$ 的加权回归矩阵，且 $\sigma_i = \frac{1}{1+\kappa\|\mathcal{Y}_i\|^2} > 0$。假设5。存在有限个采样时刻 $\{t_i\}_{i=1}^{N}$，使得

$$\lambda_{\min}(\Sigma_{\mathcal{Y}}) =: \underline{\sigma}_\theta > 0. \tag{19}$$

注2。该假设通常用于基于ICL的参数估计框架（Chowdhary and Johnson (2010)；Parikh等人 (2019)；Ogri等人 (2023, 2025)），并意味着系统经历充分激励。然而，与经典PE条件不同，该假设仅要求在有限区间内受到激励，因此要求较弱。

满足式(19)的历史栈称为满秩。常数σ_θ 量化了存储数据的信息丰富度，并直接影响式(13)的收敛速率。

为保证充分激励并改善估计器的条件性，历史栈 $\mathcal{H} = \{(\mathcal{X}_i, \mathcal{Y}_i, \mathcal{G}_{fu,i})\}_{i=1}^{N}$ 采用最小特征值最大化策略进行更新（Chowdhary and Johnson (2010)）。若候选数据点 $(\mathcal{X}^*, \mathcal{Y}^*, \mathcal{G}^*_{fu})$ 能严格增大 $\Sigma_{\mathcal{Y}}$ 的最小特征值，则将其纳入历史栈，替换现有条目 $(\mathcal{X}_j, \mathcal{Y}_j, \mathcal{G}_{fu,j})$，即

$$\lambda_{\min}\left(\Sigma^{o}_{\mathcal{Y}}\right) < \frac{\lambda_{\min}\left(\Sigma^{*}_{\mathcal{Y}}\right)}{1+\delta}, \qquad \delta > 0, \tag{20}$$

其中 $\Sigma^{o}_{\mathcal{Y}} = \sum_{i\neq j} \sigma_i \mathcal{Y}_i^\top \mathcal{Y}_i + \sigma_j \mathcal{Y}_j^\top \mathcal{Y}_j$、$\Sigma^{*}_{\mathcal{Y}} = \sum_{i\neq j} \sigma_i \mathcal{Y}_i^\top \mathcal{Y}_i + \sigma^* \mathcal{Y}^{*\top} \mathcal{Y}^*$和 $\sigma^* = \frac{1}{1+\kappa\|\mathcal{Y}^*\|^2}$。根据构造，式（20）确保 $\Sigma_{\mathcal{Y}}$ 的最小特征值随时间非递减。这严格保证历史栈一旦达到数值满秩，便能始终保持该性质，并改善并发学习更新律的条件性（Chowdhary and Johnson (2010)）。

在下面的章节中，我们将为式(21)中的增广系统设计一个稳定控制律，使其最小化式(22)中引入的代价泛函，并满足原系统(1)的安全要求。

5. 控制设计

引理1和引理2表明，保持在安全集内的轨迹将产生有界的BaS。因此，为了设计一个能够保证BaS有界性的控制策略，我们将其增广到系统的状态空间中。

为了构造安全嵌入系统，我们定义增广状态 $s := [x^\top, z]^\top \in \mathbb{R}^{n+1}$。相应的增广动力学形式为

$$\dot{s} = A_s(s)\theta + F_s(s) + G(s)u, \quad s(0) = s_0, \; s = [x^\top \; \Phi(z+\beta_0)\nabla h(x) f(x)]^\top \quad G(s) = [g(x)^\top \; \Phi(z+\beta_0)\nabla h(x) g(x)]^\top。 \tag{21}$$

其中 $A(s) := [Y(x)^\top, \; \Phi(z+\beta_0)\nabla h(x) Y(x)]^\top$, $F(x) := [f(x)^\top, \; \Phi(z+\beta_0)\nabla h(x) f(x)]^\top$，且 $G(s) := [g(x)^\top, \; \Phi(z+\beta_0)\nabla h(x) g(x)]^\top$。 z 的构造确保，安全集 $\mathcal{S}$ 的条件控制不变性等价于 s的 $(n+1)^{\text{st}}$ 坐标的一致局部有界性，因此，约束现在被编码为动力学的一种性质，而非状态空间的一种性质。

5.1 最优控制问题

设 $Q \in \mathbb{S}^{n+1}_{+}$和 $R \in \mathbb{S}^{m}_{+}$，其中 $\mathbb{S}^{n}_{+}$ 是 n阶正定方阵。对于可测输入u：$\mathbb{R}_{\geq 0} \to \mathbb{R}^m$，定义代价泛函

$$J(s_0, u) = \int_0^\infty s(\tau)^\top Q s(\tau) + u(\tau)^\top R u(\tau)\, d\tau, \tag{22}$$

其中，$s(\cdot)$ 是从初始状态 s_0 出发并采用控制策略 $u(\cdot)$ 时方程 (21) 的解。

注 *3.* 下文假设最优控制问题存在解。为此，我们隐含地假设增广系统 (21) 是可镇定的。如果系统不可镇定，则可采用 Almubarak 等人（2025年）提出的动态补偿技术来确保其可镇定性。

将使 J 最小化的最优控制问题的价值函数定义为

$$V^*(s_0) := \inf_{u(\cdot)} J(s_0, u), \tag{23}$$

其中，下确界取遍所有使 (22) 有定义的可容许输入。

假设 *6.* 最优价值函数 $V^*: \mathbb{R}^{n+1} \to \mathbb{R}$ 在 $\mathbb{R}^{n+1}$ 上连续可微，即 $V^* \in C^1(\mathbb{R}^{n+1}; \mathbb{R})$。

根据假设 6，对于所有 $s \in \mathbb{R}^{n+1}$，梯度 $\nabla V^*(s)$ 存在且连续。因此，V^* 满足稳态 Hamilton–Jacobi–Bellman（HJB）方程（Kamalapurkar 等人，2018年，定理1.5）。

$$\inf_{u\in\mathbb{R}^m} \left\{\nabla V^*(s)^\top \big(F(s) + G(s)u\big) + s^\top Q\, s + u^\top R\, u\right\} = 0, \tag{24}$$

对所有 $s \in \mathbb{R}^{n+1}$ 均成立。与 (24) 相关的哈密顿量关于 u 严格凸，这是由于 $R \succ 0$。因此，对于每个 $s \in \mathbb{R}^{n+1}$，最小化控制由一阶最优性条件唯一确定。

$$0 = \frac{\partial}{\partial u}\Big(\nabla V^*(s)^\top G(s)u + u^\top R u\Big) = G(s)^\top \nabla V^*(s) + 2Ru, \tag{25}$$

由此得到最优反馈控制策略

$$u^*(s) = -\frac{1}{2} R^{-1} G(s)^\top \nabla V^*(s). \tag{26}$$

将 (26) 代入 (24)，得到约化形式的 HJB 方程

$$\nabla V^*(s)^\top F(s) + s^\top Q\, s - \frac{1}{4}\nabla V^*(s)^\top G(s) R^{-1} G(s)^\top \nabla V^*(s) = 0. \tag{27}$$

方程 (26) 刻画了安全嵌入动力学的最优控制器。值得注意的是，BaS坐标将状态约束最优控制问题转化为无约束最优控制问题，并通过 F 和 G的结构隐式编码安全性。

5.2 价值函数近似

对于一般的非线性控制系统，与最优控制问题 (21)–(27) 相关的 HJB方程很少存在闭式解。为此，我们采用近似动态规划（ADP）表述，在紧集 $\Omega \subset \mathbb{R}^{n+1}$ with $\Omega \supset \mathcal{C}$上进行求解，并遵循 Bhasin等人 (2012) 提出的近似架构。

设 $L \in \mathbb{N}$，并令 $\sigma\colon \mathbb{R}^{n+1} \to \mathbb{R}^L$ 表示一个由连续可微基函数组成的向量，$\sigma \in C^1(\mathbb{R}^{n+1};\mathbb{R}^L)$，且令 $W \in \mathbb{R}^L$ 为未知参数向量。若存在连续可微残差函数 $\epsilon : \Omega \to \mathbb{R}$，使得对所有 $s \in \Omega$，最优价值函数 V^* 都可在 Ω 上进行参数化近似，则称该最优价值函数可参数化近似。

$$V^*(s) = W^\top \sigma(s) + \epsilon(s). \tag{28}$$

根据通用逼近定理（Sauvigny，2012，定理 1.5），函数重构误差及其梯度有界，即 $\sup_{s\in\Omega} \|\epsilon(s)\| < \bar{\epsilon}$，且 $\sup_{s\in\Omega} \|\nabla\epsilon(s)\| < \nabla\bar{\epsilon}$。理想价值函数权重满足 $\|W\| < W$，激活函数满足 $\sup_{s\in\Omega} \|\sigma(s)\| < \bar{\sigma}$，且 $\sup_{s\in\Omega} \|\nabla\sigma(s)\| < \nabla\sigma$。

由于理想权重向量 W 未知，因此价值函数和反馈策略都用参数化函数来近似。令 $\hat{W}_c, \hat{W}_a \in \mathbb{R}^L$ 表示自适应权重估计。评价器定义为

$$\hat{V}(s,\hat{W}_c) \coloneqq \hat{W}_c^\top \sigma(s), \quad \forall s \in \Omega, \tag{29}$$

执行器定义为

$$\hat{u}(s,\hat{W}_a) \coloneqq -\frac{1}{2} R^{-1} G(s)^\top \nabla\sigma(s)^\top \hat{W}_a, \quad \forall s \in \Omega. \tag{30}$$

5.3 贝尔曼误差

利用评价器 $\hat{V}$ 和执行器 $\hat{u}$，时序差分（亦称贝尔曼误差）定义为

$$\hat{\delta}(s,\hat{\theta},\hat{W}_a,\hat{W}_c) \coloneqq s^\top Q s + \hat{u}(s,\hat{W}_a)^\top R \hat{u}(s,\hat{W}_a) + \nabla\hat{V}(s,\hat{W}_c)^\top \big(A(s)\hat{\theta} + F(s) + G(s)\hat{u}(s,\hat{W}_a)\big). \tag{31}$$

利用近似表示式 (29)–(30)，我们引入评价器权重误差和执行器权重误差，分别为 $\tilde{W}_c := W - \hat{W}_c$ 和 $\tilde{W}_a \coloneqq W - \hat{W}_a$。在后续分析中，贝尔曼误差可用 $\tilde{W}_c$ 误差 $\tilde{W}_c$ 和 $\tilde{W}_a$ 表示为

$$\hat{\delta} = -\omega^\top \tilde{W}_c + \frac{1}{4}\tilde{W}_a^\top G_\sigma \tilde{W}_a - W^\top \nabla\sigma\, A\tilde{\theta} + \Delta, \tag{32}$$

其中 $\omega \coloneqq \nabla\sigma\left(A(s)\hat{\theta} + F(s) + G(s)\hat{u}(s,\hat{W}_a)\right)$，$G_R \coloneqq GR^{-1}G^\top$，$G_\epsilon \coloneqq \nabla\epsilon G_R \nabla\epsilon^\top$，$G_\sigma \coloneqq \nabla\sigma G R^{-1} G^\top \nabla\sigma^\top$，$\Delta \coloneqq \frac{1}{2}W^\top \nabla\sigma G_R \nabla\epsilon^\top + \frac{1}{4}G_\epsilon - \nabla\epsilon F$，$A \coloneqq A(s)$，$F \coloneqq F(s)$，$\epsilon := \epsilon(s)$，以及 $\sigma \coloneqq \sigma(s)$。

在线强化学习方案通常要求满足持续激励（PE）条件以确保参数收敛；然而，这类条件无法先验地强制满足，而且通常无法通过在线数据验证。当系统模型已知时，贝尔曼误差（BE）外推可提供一种虚拟激励机制，从而产生类似 PE 的条件，其是否满足可通过最小特征值准则进行评估。由于模型已知时可在任意状态下计算 BE，因此可以在每个 t 时刻选取一组虚拟状态 $\{s_k(t)\}_{k=1}^N$，并在这些外推点处计算 BE。外推贝尔曼误差定义为

$$\hat{\delta}_k \coloneqq \hat{\delta}(s_k,\hat{\theta},\hat{W}_a,\hat{W}_c), \tag{33}$$

并且可以表示为

$$\hat{\delta}_k = -\omega_k^\top \tilde{W}_c + \frac{1}{4}\tilde{W}_a^\top G_{\sigma_k}\tilde{W}_a - W^\top \nabla\sigma_k\, A_k\tilde{\theta} + \Delta_k, \tag{34}$$

其中 $\omega_k \coloneqq \nabla\sigma_k\left(A_k\hat{\theta} + F_k + G_k\hat{u}(s_k,\hat{W}_a)\right)$，$G_{R_k} \coloneqq G_k R^{-1} G_k^\top$，$G_{\sigma_k} \coloneqq \nabla\sigma_k G R^{-1} G^\top \nabla\sigma_k^\top$，$G_{\epsilon_k} \coloneqq \nabla\epsilon_k G_{R_k} \nabla\epsilon_k^\top$，$\Delta_k \coloneqq \frac{1}{2}W^\top \nabla\sigma_k G_{R_k} \nabla\epsilon_k^\top + \frac{1}{4}G_{\epsilon_k} - \nabla\epsilon_k F_k$，$A_k \coloneqq A(s_k)$，$F_k \coloneqq F(s_k)$，$\epsilon_k \coloneqq \epsilon(s_k)$，以及 $\sigma_k \coloneqq \sigma(s_k)$。

在假设1和神经网络的通用逼近性质下，贝尔曼误差残差满足 $\sup_{s\in\Omega} \|\Delta(s)\| \le \Delta$，其中某个常数 $\Delta > 0$。同样，对于任意外推状态 $s_k \in \Omega$，相应的残差满足 $\|\Delta_k\| \le \Delta$。

5.4 Update Laws for Actor and Critic Weights

基于后续的稳定性分析，通过对平方贝尔曼误差应用梯度型自适应方法，可得到评价器更新律。由此得到

$$\dot{\hat{W}}_c = -k_{c_1}\Upsilon\frac{\omega}{\rho}\hat{\delta} - \frac{k_{c_2}}{N}\Upsilon\sum_{k=1}^{M}\frac{\omega_k}{\rho_k}\delta_k, \tag{35}$$

其中，k_{c_1}，$k_{c_2} > 0$ 为自适应增益，归一化因子 $\rho = 1 + \nu\,\omega^\top\omega$ 和 $\rho_k = 1 + \nu\,\omega_k^\top\omega_k$使用正则化常数 $\nu > 0$。增益矩阵$\Upsilon : \mathbb{R}_{\ge 0} \to \mathbb{R}^{L\times L}$ 采用标准的递归最小二乘更新，

$$\dot{\Upsilon} = \beta_c\Upsilon - k_{c_1}\Upsilon\frac{\omega\omega^\top}{\rho^2}\Upsilon - \frac{k_c}{N}\Upsilon\sum_{i=1}^{M}\frac{\omega_k\omega_k^\top}{\rho_k^2}\Upsilon, \tag{36}$$

其中 $\Upsilon(0) = \Upsilon_0$，遗忘因子为 $\beta_c > 0$。为保证参数收敛，后续的稳定性分析采用以下激励条件。

假设7。存在一个常数 $c_1 > 0$，使得轨迹 $\{s_k(\cdot)\}_{k=1}^N$ 满足

$$\underline{c}_1 \le \inf_{t\ge 0}\lambda_{\min}\left(\frac{1}{N}\sum_{k=1}^{M}\frac{\omega_k(t)\,\omega_k^\top(t)}{\rho_k^2(t)}\right). \tag{37}$$

如Kamalapurkar等人（2016）的引理1所示，在假设7成立且 $\lambda_{\min}\{\Upsilon_0^{-1}\} > 0$的条件下，式（36）中的更新律可确保最小二乘增益满足

$$\underline{\Upsilon} I_L \preceq \Upsilon(t) \preceq \overline{\Upsilon} I_L, \tag{38}$$

$\forall t \in \mathbb{R}_{\ge 0}$ 并且对于某个 $\overline{\Upsilon}, \underline{\Upsilon} > 0$。

执行器更新律通过对由相同的贝尔曼残差构成的正则化损失进行梯度下降而得到。由此得到的自适应律为

$$\dot{\hat{W}}_a = -k_{a_1}\left(\hat{W}_a - \hat{W}_c\right) - k_{a_2}\hat{W}_a + \frac{k_{c_1}G_\sigma^\top\hat{W}_a\omega^\top}{4\rho}\hat{W}_c + \sum_{k=1}^{M}\frac{k_c G_{\sigma_k}^\top\hat{W}_a\omega_k^\top}{4N\rho_k}\hat{W}_c, \tag{39}$$

其中 k_{a_1}，$k_{a_2} > 0$ 为自适应增益。

施加于系统（21）的控制输入使用由（35）生成的执行器权重，表示为

$$u(t) = \hat{u}\Big(s(t),\, \hat{W}_a(t)\Big), \qquad t \geq 0. \tag{40}$$

本节提出的自适应执行器-评价器更新律将在后续稳定性分析中使用，以证明闭环系统轨迹的一致最终有界性。

6. 稳定性分析

为便于分析，令拼接向量

$$Z \coloneqq \begin{bmatrix} s^\top & \tilde{z}^\top & \tilde{\theta}^\top & \tilde{W}_c^\top & \tilde{W}_a^\top \end{bmatrix}^\top \in \mathbb{R}^{n+p+2L+2} \tag{41}$$

表示闭环系统的状态，并令一个连续可微的候选李雅普诺夫函数$V_L:\ \mathbb{R}^{n+p+2L+2} \times \mathbb{R}_{\geq 0} \to \mathbb{R}$定义为

$$V_L\left(Z, t\right) \coloneqq V^*(s) + \frac{1}{2}\tilde{z}^\top \tilde{z} + \frac{1}{2}\tilde{\theta}^\top \Gamma^{-1}(t)\tilde{\theta} + \frac{1}{2}\tilde{W}_c^\top \Upsilon^{-1}(t)\tilde{W}_c + \frac{1}{2}\tilde{W}_a^\top \tilde{W}_a. \tag{42}$$

利用式(38)中的界，并且由于候选李雅普诺夫函数(42)是正定的，可使用(Khalil, 2002, 引理4.3)得出其有界，即

$$\underline{v}_l\left(\|Z\|\right) \leq V_L\left(Z, t\right) \leq \overline{v}_l\left(\|Z\|\right), \tag{43}$$

对所有 $t \in \mathbb{R}_{\geq 0}$ 以及所有 $Z \in \mathbb{R}^{n+p+2L+2}$均成立，其中$v_l,\ v_l:\ \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ 为class $\mathcal{K}$ 函数。利用式(36)中的最小二乘增益更新律和式(38)中的界，归一化回归向量 $\frac{\omega(t)}{\rho\, t}$ 有界为 $\left\|\frac{\omega(t)}{\rho\, t}()\right\| \leq \frac{1}{2\sqrt{\nu\Upsilon}},\ \forall t \geq 0$。对于任意连续函数 $f:\Omega \to \mathbb{R}^m$，我们将 f表示为满足$\sup_{s\in\Omega}\|f(s)\| \leq f$的任意常数。在不引起歧义时，我们取 $f = \sup_{s\in\Omega}\|f(s)\|$。为便于分析，定义常数

$$\begin{aligned}
&\varpi_1 \coloneqq (k_{c_1} + k_{c_2})\, \overline{W}\, \overline{\nabla\sigma}\, \overline{A},\\
&\varpi_2 \coloneqq (k_{c_1} + k_{c_2})\, \overline{\Delta},\\
&\varpi_3 \coloneqq \tfrac{1}{2}\overline{W}\, \overline{G_\sigma} + \tfrac{1}{2}\overline{\nabla G}\, \overline{\nabla\sigma} + k_{a_2}\, \overline{W} + \tfrac{1}{4}(k_{c_1} + k_{c_2})\, \overline{W}^2\, \overline{G_\sigma},\\
&\varpi_4 \coloneqq k_{a_1} + \tfrac{1}{4}(k_{c_1} + k_{c_2})\, \overline{W}\, \overline{G_\sigma},\\
&\varpi_5 \coloneqq \tfrac{1}{4}(k_{c_1} + k_{c_2})\, \overline{W}\, \overline{G_\sigma}.
\end{aligned}$$

此外，将正的常数残差定义为 $\iota \coloneqq \varpi_2^2 + \frac{(\varpi_3+\varpi_4)^2}{2\varepsilon_3} + \frac{1}{4}\overline{G_\epsilon}$下述定理证明了闭环系统轨迹的一致最终有界性。

定理*1*。 设 $\mathbb{B}_\chi \subset \Omega \times \mathbb{R}^{p+2L+1}$ 是一个包含原点、半径为 $\chi > 0$ 的闭球。若假设1–7成立，未知参数 θ 采用式（13）和（14）中的自适应更新律进行估计，且满足式（50）和

$$v_l^{-1}\left(\iota\right) \leq \overline{v}_l^{-1}\left(\underline{v}_l\left(\chi\right)\right), \tag{44}$$

$\hat{}\,\hat{}\,\hat{}$ 均成立，并且权重 W_c、Υ、W_a、θ和 Γ分别按照式（35）、（36）、（39）、（13）和（14）更新，则在式（40）设计的控制器作用下，拼接状态 Z局部一致最终有界。此外，令 $\mathcal{S}_0$ 取为

$$\mathcal{S}_0 \coloneqq \{\, x(0) \in \mathcal{S} \mid Z(0) \leq \overline{v}_l^{-1}\big(v_l(\chi)\big)\,\}, \tag{45}$$

那么，对于所有满足 $x(0) \in \mathcal{S}_0$ 的轨迹，均有 $x(t) \in \mathcal{S}$，对所有 $t \geq 0$ 成立。

证明。 沿闭环系统轨迹的 (42) 的轨道导数为

$$\begin{aligned}
\dot{V}_L(Z,t) &= \nabla V(s)(A(s)\theta + F(s) + G(s)\hat{u}(s,\hat{W}_a))\\
&\quad + \tilde{z}^\top \dot{\tilde{z}} + \tilde{\theta}^\top \Gamma^{-1}\dot{\tilde{\theta}} - \frac{1}{2}\tilde{\theta}^\top \Gamma^{-1}\dot{\Gamma}\Gamma^{-1}\tilde{\theta}\\
&\quad - \tilde{W}_c^\top \Gamma^{-1}\dot{\tilde{W}}_c - \frac{1}{2}\tilde{W}_c^\top \Gamma^{-1}\dot{\Gamma}\Gamma^{-1}\tilde{W}_c - \tilde{W}_a^\top \dot{\tilde{W}}_a.
\end{aligned} \tag{46}$$

代入 (10)、(18)、(35)、(14) 和 (39)，再利用投影算子的性质（Krstic 等人，1995，引理 E.1. IV），并应用柯西-施瓦茨不等式和配方法，可将 (46) 上界为

$$\begin{aligned}
\dot{V}_L(Z,t) &\leq -\lambda_{\min}(Q)\|s\|^2 - \gamma\|\tilde{z}\|^2 - k_\theta \underline{\sigma}_\theta \|\tilde{\theta}\|^2 - k_{c_2}\underline{c}\|\tilde{W}_c\|^2\\
&\quad - (k_{a_1} + k_{a_2})\|\tilde{W}_a\|^2 + \varpi_1\|\tilde{W}_c\|\,\|\tilde{\theta}\| + \varpi_3\|\tilde{W}_a\| + \varpi_2\|\tilde{W}_c\|\\
&\quad + \varpi_4\|\tilde{W}_a\| + \tfrac{1}{4}\overline{G}_\epsilon + \varpi_5\|\tilde{W}_a\|^2.
\end{aligned} \tag{47}$$

应用杨氏不等式后，导数的上界变为

$$\begin{aligned}
\dot{V}_L(Z,t) &\leq -\lambda_{\min}(Q)\|s\|^2 - \gamma\|\tilde{z}\|^2 - \Big(k_\theta \underline{\sigma}_\theta - \tfrac{\varpi_1^2}{2\varepsilon_1}\Big)\|\tilde{\theta}\|^2\\
&\quad - \Big(k_{c_2}\underline{c} - \tfrac{\varepsilon_1}{2} - \tfrac{\varepsilon_2}{2}\Big)\|\tilde{W}_c\|^2 - \Big((k_{a_1} + k_{a_2}) - \varpi_5 - \tfrac{\varepsilon_3}{2}\Big)\|\tilde{W}_a\|^2\\
&\quad + \tfrac{\varpi_2^2}{2\varepsilon_2} + \tfrac{(\varpi_3+\varpi_4)^2}{2\varepsilon_3} + \tfrac{1}{4}\overline{G}_\epsilon,
\end{aligned} \tag{48}$$

其中 $\underline{c} \coloneqq \frac{\beta}{2\Upsilon k_{c_2}} + \frac{c_1}{2}$是一个正常数，$\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ 是正标量。将有效二次系数定义为 $\alpha_s \coloneqq \lambda_{\min}(Q)$，$\alpha_z \coloneqq k_z$，$\alpha_\theta \coloneqq k_\theta \sigma_\theta - \frac{\varpi_1^2}{2\varepsilon_1}$，$\alpha_c \coloneqq k_{c_2} c - \frac{\varepsilon_1}{2} - \frac{\varepsilon_2}{2}$，以及 $\alpha_a \coloneqq (k_{a_1} + k_{a_2}) - \varpi_5 - \frac{\varepsilon_3}{2}$。则 (48) 可简洁地写为

$$\begin{aligned}
\dot{V}_L(Z,t) &\leq -\alpha_s\|s\|^2 - \alpha_z\|\tilde{z}\|^2 - \alpha_\theta\|\tilde{\theta}\|^2 - \alpha_c\|\tilde{W}_c\|^2\\
&\quad - \alpha_a\|\tilde{W}_a\|^2 + \iota.
\end{aligned} \tag{49}$$

为使二次项严格负定，选择 $\varepsilon_1, \varepsilon_2, \varepsilon_3 > 0$ 以及增益，使得

$$\alpha_\theta = (k_\theta - 1)\underline{\sigma}_\theta + \frac{\beta_\theta}{k_\theta \overline{\Gamma}} - \frac{\varpi_1^2}{2\varepsilon_1} > 0, \tag{50a}$$

$$\alpha_c = k_{c_2}\underline{c} - \frac{\varepsilon_1}{2} - \frac{\varepsilon_2}{2} > 0, \tag{50b}$$

$$\alpha_a = (k_{a_1} + k_{a_2}) - \varpi_5 - \frac{\varepsilon_3}{2} > 0. \tag{50c}$$

如果选择增益值 k_θ、k_{c_2}、k_{a_1}、k_{a_2}，使得上述等式右侧均为正，则 α_s、α_z、α_θ、α_c、α_a均为正。在这些选择下，由（49）可得

$$\dot{V}_L \leq -v_l\left(\|Z\|\right), \quad \forall \|Z\| \geq v_l^{-1}(\iota), \tag{51}$$

对于所有 $t \in \mathbb{R}_{\geq 0}$ 以及所有 $Z \in \mathbb{B}_\chi$，其中 $v_l:\ \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ 是满足以下条件的 class $\mathcal{K}$ 函数

$$v_l\left(\|Z\|\right) \leq \frac{1}{2}\min\{\alpha_s, \alpha_z, \alpha_\theta, \alpha_c, \alpha_a\}\|Z\|^2. \tag{52}$$

因此，若满足（50）中的充分条件且保持（43）中的界，则根据 Khalil（2002）定理4.18，闭环信号 Z 具有局部一致最终有界性。具体而言，每个满足初始界 $\|Z(0)\| \leq v_l^{-1}\big(v_l(\chi)\big)$ 的解都满足 $\limsup_{t\to\infty}\|Z(t)\| \leq v_l^{-1}\Big(v_l\big(v_l^{-1}(\iota)\big)\Big)$。此外，对于所有 $t \in \mathbb{R}_{\geq 0}$，拼接后的状态轨迹始终限制在球 $\mathbb{B}_\chi$ 内。由于执行器权重 $\tilde{}$ 估计误差 W_a 一致最终有界，因此得到的控制策略 $\hat{}u$ 构成对最优策略 u^* 的近似。

因此，对于任意初始条件 $x(0) \in \mathcal{S}_0$，真实的屏障状态 $z(t)$ 对所有 $t \geq 0$ 均保持有界。此外，由引理 2 可知，$z(t)$ 的有界性等价于 $x(t) \in \mathcal{S}$，因此对于所有 $t \geq 0$，$x(t)$ 均保持在安全集 $\mathcal{S}$ 内，从而确保安全性。□

定理*2.* 若假设1–5成立，且未知参数和状态估计值按照 (9) 和 (13) 更新，并且增益的选取满足

$$\frac{\beta_\theta}{k_\theta \overline{\Gamma}} > \underline{\sigma}_\theta, \tag{53}$$

则 (10) 和 (18) 中的误差系统呈指数稳定。

证明. 定义一个连续可微的候选李雅-˜普诺夫函数，V：$\mathbb{R} \times \mathbb{R}^p \times \mathbb{R}_{\geq 0} \to \mathbb{R}$，如下：

$$\tilde{V}(\tilde{z}, \tilde{\theta}, t) = \frac{1}{2}\tilde{z}^\top \tilde{z} + \frac{1}{2}\tilde{\theta}^\top \Gamma(t)^{-1}\tilde{\theta}. \tag{54}$$

存在常数 $\bar{v}, \underline{v} > 0$ 使得 $\underline{v}\|(\tilde{z}, \tilde{\theta})\|^2 \leq \tilde{V}(\tilde{z}, \tilde{\theta}, t) \leq \bar{v}\|(\tilde{z}, \tilde{\theta})\|^2$, $\forall t \geq 0$。由 V 的正定性及 $\Gamma(t)$ 的连续性，该条件在局部成立。根据 (Krstic 等人, 1995, ˜引理 E.1. IV) 中投影算子的性质，沿着轨迹 (10)、(18) 和 (14) 的 V 的轨道导数有界如下：

$$\begin{aligned}\dot{\tilde{V}}(\tilde{z}, \tilde{\theta}, t) \leq \tilde{z}^\top &\left(\Phi(z + \beta_0)\nabla h(x) Y(x)\tilde{\theta} - \gamma \tilde{z}\right) \\ &- k_\theta \tilde{\theta}^\top \Sigma_{\mathcal{Y}} \tilde{\theta} - \tilde{\theta}^\top Y(x)^\top \nabla h(x)^\top \Phi(z + \beta_0)^\top \tilde{z} \\ &- \frac{1}{2}\tilde{\theta}^\top \Gamma^{-1}\left(\beta_\theta \Gamma - k_\theta \Gamma \Sigma_{\mathcal{Y}} \Gamma\right)\Gamma^{-1}\tilde{\theta}\end{aligned} \tag{55}$$

˙当 $\Gamma = \beta_\theta \Gamma - k_\theta \Gamma \Sigma_{\mathcal{Y}} \Gamma$ 在式（14）中时。当 $\Gamma = 0$ 在式（14）中时，˜则 V 的李导数有界如下：

$$\begin{aligned}\dot{\tilde{V}}(\tilde{z}, \tilde{\theta}, t) \leq \tilde{z}^\top &\left(\Phi(z + \beta_0)\nabla h(x) Y(x)\tilde{\theta} - \gamma \tilde{z}\right) \\ &- k_\theta \tilde{\theta}^\top \Sigma_{\mathcal{Y}} \tilde{\theta} - \tilde{\theta}^\top Y(x)^\top \nabla h(x)^\top \Phi(z + \beta_0)^\top \tilde{z}.\end{aligned} \tag{56}$$

利用假设5，应用三角不等式，并且在满足式（53）中的增益条件时，对于式（14）中 Γ 的任意更新，轨道˙导数（55）可有界如下：

$$\dot{\tilde{V}}(\tilde{z}, \tilde{\theta}, t) \leq -\gamma \|\tilde{z}\|^2 - k_\theta \underline{\sigma}_\theta \|\tilde{\theta}\|^2. \tag{57}$$

因此，根据 Khalil（2002）的定理 4.8，误差动力学（10）和（18）是指数稳定的。因此，在基于 ICL 的更新律（13）下，参数估计值 $\theta(t)$ 指数收敛至真实参数θ，即 $\lim_{t\to\infty} \tilde{\theta}(t) = \lim_{t\to\infty} (\hat{\theta}(t) - \theta) = 0$。□

7. 仿真研究

本节将通过一个障碍物避让问题，考察式（30）中所提出控制策略的有效性。本研究的系统动力学是式（1）所示形式的非线性控制仿射系统，其状态为 $x = [x_1, x_2]^\top$，其中

$$Y(x) = \begin{bmatrix} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_1 + x_2 & x_1^2 x_2 \end{bmatrix}, \tag{58}$$

$g(x) = [0, \cos(2x_1) + 2]^\top$，$f(x) = [\,0 \quad 0\,]^\top$，以及 $\theta = [\theta_1 \ \theta_2 \ \theta_3 \ \theta_4]^\top = [-1 \ -1 \ -0.5 \ -0.5]^\top$。

对于该仿真，安全集由式 (3) 定义，其中约束为 $h(x) = (x_1 - 1)^2 + (x_2 - 2)^2 - 0.5^2$。

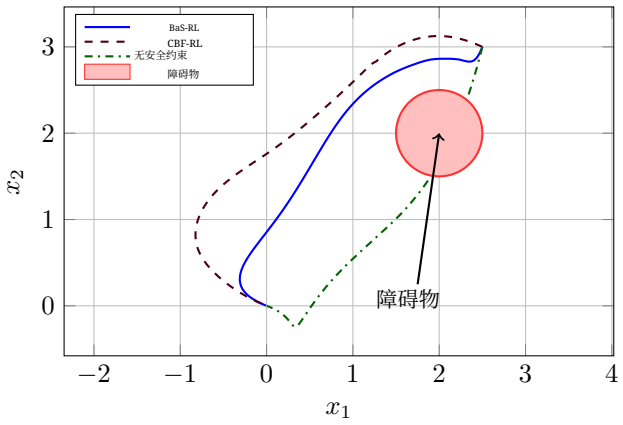

图 1. 系统的相空间轨迹

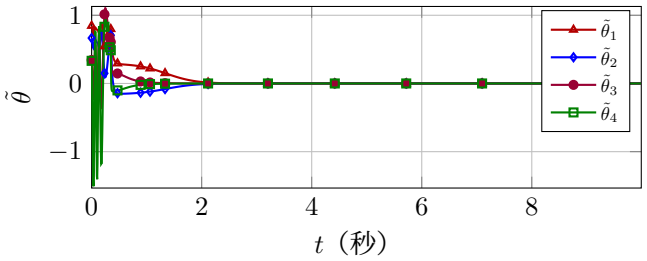

图 2. 参数估计误差随时间的变化。

控制策略的目标是确保智能体避开障碍物，从而始终处于定义 1 所规定的安全集内。式 (22) 中状态和控制作用的惩罚项或代价分别选为 $Q = I_3$ 和 $R = 1$。θ 的估计值记为 $\hat{\theta} = [\hat{\theta}_1 \ \hat{\theta}_2 \ \hat{\theta}_3 \ \hat{\theta}_4]^\top$。式 (58) 中系统的初始条件为 $x(0) = [2.5, \ 4]^\top$。为估计值、权重和增益选取的初始值为 $\hat{z}(0) = 0$, $\theta(0) = 0_{4\times 1}$, $\hat{W}_c = 0.5 1_{6\times 1}$, $\hat{W}_a = 0.5 1_{6\times 1}$, $\Gamma(0) = 10 I_4$，以及$\Upsilon(0) = 0.01 I_6$。式 (6) 中的障碍函数选为$\beta(x) = B(h(x)) = \frac{K}{h(x)}$ $\forall x \in \mathbb{R}^2$，使得 $\Phi(\beta) = -\frac{K}{\beta^2}\sigma$。观测器采用 $\gamma = 3$，障碍增益选为 $K = 0.01$。

仿真中使用的控制增益和学习增益如下。对于 ADP 框架，增益设置为$\nu = 2$、$k_{c1} = 1$、$k_{c2} = 1$、$k_{a1} = 2$、$k_{a2} = 1$ 和 $\beta_c = 0.1$，其边界为 $\Upsilon = 0$ 和 $\Upsilon = 1000$。ICL 增益选为 $k_\theta = 50$、$\overline{\kappa} = 1$和 $\beta_\theta = 1$。仿真使用 100 个固定的贝尔曼误差外推点 s_k，这些点位于以 $s = [x_1 \quad x_2 \quad z]^\top$ 坐标系原点为中心、边长为 4×4 的正方形内。价值函数的基函数选为 $\sigma(s) = \left[s_1^2 \ s_2^2 \ s_3^2 \ s_1 s_2 \ s_2 s_3 \ s_3 s_1\right]^\top$。仿真研究将考察所提出的 BaS-RL-ICL 框架、Cohen 等人（2023）提出的 CBF-RL-ICL 框架，以及不含任何安全约束的 RL-ICL 的控制策略有效性，分别记为案例 1、2 和 3。

7.1 讨论

将该框架的有效性与 Cohen 等人（2023）中描述的 CBF-RL 技术进行了比较。图

1 展示了基于BaS增强的控制策略在避开障碍物方面的有效性。此外，与基于BaS的控制策略相比，基于控制障碍函数的方法保持了更大的安全裕度。图 2 显示，系统参数误差呈指数级收敛至零，这与定理2一致。总之，这些结果凸显了BaS框架相较于现有基于控制障碍函数的技术，在效率、稳定性和学习性能方面的提升。

8. 结论

本文提出了一种基于BaS的安全控制框架，用于参数未知的非线性控制仿射系统。结果表明，该方法无需先验参数知识即可保持安全性，并在维持系统稳定运行的同时，推动参数估计值趋近真实值。该框架存在一些局限：其策略是近最优的，而非渐近最优的；此外，要实现完全收敛，还需要有限激励。此外，由于状态维数增加且模型非线性增强，用BaS动力学扩展系统动力学会提高最优控制问题的计算复杂度。尽管存在这些局限，仿真结果表明，所提出的ICL-BaS-RL框架能够保证近最优安全性和稳定性。未来的工作将把该框架扩展至更复杂的环境，包括存在时变障碍物和动态扰动的场景，并研究在限制较少的假设下实现全局安全保证的条件。

参考文献

Al-Sunni, M.F., Almubarak, H. 和 Dolan, J.M.（2025）。使用屏障状态的安全嵌入自适应控制。载于美国控制会议论文集，303–308。Almubarak, H., Sadegh, N. 和 Theodorou, E.A.（2025）。安全关键多目标控制的屏障状态理论。*IEEE* 自动控制汇刊，1–16。Ames, A.D., Grizzle, J.W. 和 Tabuada, P.（2014）。基于控制屏障函数的二次规划及其在自适应巡航控制中的应用。载于*IEEE* 决策与控制会议论文集，6271–6278。IEEE。Aoun, C.E., Zhao, P., Almubarak, H., Hovakimyan, N. 和 Theodorou, E.A.（2023）。使用屏障状态对受约束系统进行 L1自适应控制。载于*AIAA* 科技论坛论文集，1051。Bhasin, S., Kamalapurkar, R., Johnson, M., Vamvoudakis, K.G., Lewis, F.L. 和 Dixon, W.E.（2012）。用于自适应近似最优控制的演员-评论家-辨识器架构。载于 F.L. Lewis 和 D. Liu（编），用于反馈控制的强化学习与近似动态规划，IEEE 计算智能系列，258–278。Wiley 和 IEEE Press。Blanchini, F.（1999）。控制中的集合不变性。*Automatica*，35(11)，1747–1767。Cai, Z., de Queiroz, M.S. 和 Dawson, D.M.（2006）。一个充分光滑的投影算子。*IEEE* 自动控制汇刊，51(1)，135–139。Chowdhary, G. 和 Johnson, E.（2010）。无需持续激励、通过并发学习实现自适应控制收敛。载于*IEEE* 决策与控制会议论文集，3674–3679。Cohen, M.H. 和 Belta, C.（2020）。利用控制屏障函数对安全关键系统进行近似最优控制。载于*IEEE* 决策与控制会议论文集，2062–2067。Cohen, M.H. 和 Belta, C.（2023）。利用控制屏障函数进行基于模型的强化学习安全探索。*Automatica*，147，110684。Cohen, M.H., Serlin, Z., Leahy, K. 和 Belta, C.（2023）。时序逻辑引导的安全基于模型强化学习：一种混合系统方法。非线性分析：混合系统，47，101295。Deptula, P., Chen, H., Licitra, R.A., Rosenfeld, J.A. 和 Dixon, W.E.（2020）。通过近似最优运动规划避开未知移动避障区域。*IEEE* 机器人汇刊，36(2)，414–430。Greene, M.L., Deptula, P., Nivison, S. 和 Dixon, W.E.（2020）。带有屏障约束的稀疏学习型近似动态规划。*IEEE* 控制系统快报，4(3)，743–748。Kamalapurkar, R., Rosenfeld, J.A. 和 Dixon, W.E.（2016）。用于近似在线最优控制的高效基于模型的强化学习。*Automatica*，74，247–258。Kamalapurkar, R., Walters, P., Rosenfeld, J.A. 和 Dixon, W.E.（2018）。最优反馈控制的强化学习：一种基于李雅普诺夫的方法。通信与控制工程。Springer International Publishing。Khalil, H.K.（2002）。非线性系统。Prentice Hall，美国新泽西州上萨德尔河，第三版。Krstic, M.（2023）。逆最优安全滤波器。*IEEE* 自动控制汇刊，69(1)，16–31。Krstic, M., Kanellakopoulos, I. 和 Kokotovic, P.V.（1995）。非线性与自适应控制设计。John Wiley & Sons，美国纽约州纽约市。Lopez, B.T., Slotine, J.J.E. 和 How, J.P.（2020）。鲁棒自适应控制屏障函数：一种面向安全的自适应数据驱动方法。*IEEE* 控制系统快报，5(3)，1031–1036。Mahmud, S.M.N., Nivison, S.A., Bell, Z.I. 和 Kamalapurkar, R.（2021）。针对具有参数不确定性的系统的安全基于模型强化学习。机器人与人工智能前沿，8(733104)，1–13。Marvi, Z. 和 Kiumarsi, B.（2021）。安全强化学习：一种控制屏障函数优化方法。国际鲁棒与非线性控制期刊，31(6)，1923–1940。Nagumo, M.（1942年）。关于常微分方程积分曲线的位置。载于物理数学学会论文集，272–559。Ogri, T.E., Qureshi, M., Bell, Z.I., Makumi, W.A. 和 Kamalapurkar, R.（2025年）。仿射非线性系统的安全输出反馈自适应最优控制。arXiv:2510.20081。Ogri, T.E., Bell, Z.I. 和 Kamalapurkar, R.（2023）。仿射非线性系统的状态与参数估计。载于*IEEE* 决策与控制会议论文集，1517–1522。新加坡。Parikh, A., Kamalapurkar, R. 和 Dixon, W.E.（2019）。积分并发学习：利用有限激励实现参数收敛的自适应控制。国际自适应控制与信号处理期刊，33(12)，1775–1787。Reis, M.F., Aguiar, A.P. 和 Tabuada, P.（2021）。基于控制屏障函数的二次规划会引入不期望的渐近稳定平衡点。*IEEE* 控制系统快报，5(2)，731–736。Robey, A., Lindemann, L., Tu, S. 和 Matni, N.（2021）。针对不确定系统学习鲁棒混合控制屏障函数。*IFAC* 论文在线，54(5)，1–6。Sauvigny, F.（2012）。偏微分方程 *1*。Springer。Taylor, A.J. 和 Ames, A.D.（2020）。利用控制屏障函数实现自适应安全。载于美国控制会议论文集，1399–1405。Wieland, P. 和 Allgöwer, F.（2007）。利用控制屏障函数实现构造性安全。*IFAC* 会议论文集，40(12)，462–467。Yang, Y., Vamvoudakis, K.G., Modares, H., He, W., Yin, Y.X. 和 Wunsch, D.（2019）。具有行动者–评论家–屏障结构的安全感知强化学习框架。载于美国控制会议论文集，2352–2358。