arXiv:2602.05046v1 [math.OC] 4 Feb 2026

采用对数障碍有约束迭代线性二次调节器实现安全最优控制

Abhijeet 和 Suman Chakravorty

T德克萨斯农工大学，大学城，德克萨斯州， USA

摘要。 本文提出了一种用于非线性最优控制问题的有约束迭代线性二次调节器（iLQR）框架，可处理状态和控制输入上的盒约束。我们将对数障碍函数纳入阶段代价，以施加盒约束（变量的上界和下界），从而得到一种平滑的内点形式，并可无缝集成到标准 iLQR 的后向—前向过程中。对数障碍所产生的海森矩阵项为正定，既保持并增强了 iLQR 中二次近似的正定性，又带来内在的正则化效果，从而提升数值稳定性与收敛性。此外，由于负对数是凸函数，若原代价已具凸性，加入对数障碍项仍可保持凸性。我们进一步分析了增广障碍项的 iLQR 如何在约束边界附近自然调整反馈增益。特别是，在收敛时，与饱和控制通道相关的反馈项会趋于零；每当控制饱和时，系统便恢复为纯前馈行为。在有约束非线性控制问题上的数值示例表明，所提出的方法能够可靠地满足盒约束，并保持良好的收敛性质。[1]

关键词： 控制理论与优化 · 有约束优化 · 迭代线性二次调节器 · 内点法。

1 引言

最优控制理论提供了一个强大的数学框架，用于确定控制策略，以引导动力系统沿期望轨迹运行，同时最小化指定的代价函数 [6]。其应用是众多领域的基础，包括机器人学、航空航天工程、经济学和过程控制。针对一般非线性系统求解最优控制问题（OCP），推动了多种数值方法的发展。这些方法大致分为间接方法和直接法：间接方法求解由庞特里亚金极小值原理导出的两点边值问题，直接法则将最优控制问题转化为有限维非线性规划（NLP）问题 [3,4,22]。

[1] 包含代码和补充视频的项目网站：Box-iLQR。

在间接方法中，微分动态规划（DDP） [20,37,5,23,16,17], 及其变体迭代线性二次调节器（iLQR） [30], 尤为成功。这些方法利用最优控制问题的序贯结构，通过反向传递构造值函数的局部二次近似，并通过前向过程以更新后的控制策略对系统进行仿真。这种序贯分解使迭代线性二次调节极为高效，因为它将整条轨迹上的大型优化问题转化为每个时间步上一系列更小、更易处理的问题。近期工作进一步探讨了微分动态规划/迭代线性二次调节与序列二次规划之间的联系，从而对其收敛性质提供了更深入的认识 [1]。相比之下，迭代线性二次调节具有稳健的收敛性质，即使初始化远离最优解也是如此，并且对复杂控制问题非常有效 [38,24,21,28,29,11,9,10]。

尽管 iLQR 具有这些优势，但将其应用于现实世界的系统仍面临一项重大挑战：处理状态与控制约束。因此，人们开始转向使用其他优化方法来进行有约束运动规划 [42,14,27]。针对有约束 iLQR 的尝试包括控制量钳位等方法，但这类方法往往收敛较慢 [31,19]。其他方法也存在明显缺点：梯度投影技术 [31,8,41,32] 可能得出次优解，而增广拉格朗日（AL）形式通常需要复杂的参数调优，且无法保证收敛到最优解 [13,12,18]。尽管内点法（IPM）提供了一个有前景的方向 [25,26,34,33], ，但许多现有实现缺乏稳健的收敛保证。本文通过提出一种基于障碍函数的内点法的约束迭代线性二次调节算法，弥补了这一关键缺口。重要的是，我们的方法易于调优，并在一些温和假设下保证收敛。此外，该方法还具有显著优势，包括障碍项带来的固有正则化，从而提升数值稳定性。通过将这些特性融入迭代线性二次调节器框架，我们保留了其出色的全局收敛性质 [1] ，同时能够严格且高效地处理约束。

本文重点关注一类尤为重要的约束：状态和控制量上的盒约束（$x_{\min} \leq x_k \leq x_{\max}$、 $u_{\min} \leq u_k \leq u_{\max}$）。我们提出了一种新颖的迭代线性二次调节算法，通过对数障碍内点法直接处理这些约束。障碍函数法最早由 [7], 等研究提出，其通过向代价函数中添加障碍项，将约束问题转化为一系列无约束问题。该项会惩罚接近可行集边界的解，从而确保迭代点始终保持严格可行。内点法最初源于 Karmarkar 的线性规划算法 [15], ，此后已成为通用非线性优化领域的主流范式 [39,40,22,36], ，并且已有广泛可用的稳健实现 [35]。通过将这一强大的框架融入 iLQR，我们的方法使反向传递能够考虑约束并主动调整控制策略，从而实现更加稳健且高效的优化过程。

本工作的主要贡献如下：

– 在基于障碍函数的内点框架下，为 iLQR 算法量身设计了一种新的盒约束轨迹优化问题表述。– 提出了一种完整的算法，称为 Box-iLQR。该算法通过系统地减小障碍参数，使解收敛到原始约束问题的解。– 分析了使用 Box-iLQR 的各种优势（第 4 节）。– 通过多个基准最优控制问题进行数值验证，展示了所提方法的有效性。

本文其余部分安排如下：第 2 节简要介绍本文所研究的约束最优控制问题。第 3 节详细介绍我们的 Box-iLQR 算法，第 4 节随后阐述其优点。第 5 节给出数值结果，第 6 节则通过讨论未来工作总结全文。

2 背景

本节首先提出盒约束最优控制问题，随后利用对数障碍提出无约束优化问题。

2.1 有约束最优控制问题

考虑以下最优控制问题：

$$\begin{aligned}
\min_{\mathbf{u}(\cdot)} \quad & \mathcal{J} = \phi(\mathbf{x}(t_f)) + \int_0^{t_f} c(\mathbf{x}(t), \mathbf{u}(t))\, dt && \text{(1a)}\\
\text{s.t.} \quad & \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t); t) \quad \forall\, t \in [0, t_f], && \text{(1b)}\\
& \mathbf{x}(t) \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n \mid \underline{\mathbf{x}} \le \mathbf{x} \le \bar{\mathbf{x}}\}, && \text{(1c)}\\
& \mathbf{u}(t) \in \mathcal{U} := \{\mathbf{u} \in \mathbb{R}^m \mid \underline{\mathbf{u}} \le \mathbf{u} \le \bar{\mathbf{u}}\}, && \text{(1d)}\\
& \mathbf{x}(0) \ \text{ and } \ t_f \ \text{ are given.}
\end{aligned}$$

在上述方程中，$\mathbf{x}(t) \in \mathbb{R}^n$ 为状态，$\mathbf{u}(t) \in \mathbb{R}^m$ 为控制输入。函数 $\mathbf{f}: \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n$ 描述系统动力学。目标泛函 $\mathcal{J}$ 由终端代价 ϕ：$\mathbb{R}^n \to \mathbb{R}$ 与运行代价 c：$\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 组成。对于所有 $t \in [0, t_f]$，状态与控制分别有约束地位于紧集 $\mathcal{X}$ 与 $\mathcal{U}$ 内。

方程（1）中所述问题的离散形式为：

$$\begin{aligned}
\min_{\mathbf{u}_t} \quad & \mathcal{J} = \Phi(\mathbf{x}_T) + \sum_{t=0}^{T-1} C(\mathbf{x}_t, \mathbf{u}_t)\ , && \text{(2a)}\\
\text{s.t.} \quad & \mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t, \mathbf{u}_t; t) \quad \forall\, t = 0, 1 \cdots T-1, && \text{(2b)}\\
& \mathbf{x}_t \in \mathcal{X} := \{\mathbf{x} \in \mathbb{R}^n \mid \underline{\mathbf{x}} \le \mathbf{x} \le \bar{\mathbf{x}}\}, && \text{(2c)}\\
& \mathbf{u}_t \in \mathcal{U} := \{\mathbf{u} \in \mathbb{R}^m \mid \underline{\mathbf{u}} \le \mathbf{u} \le \bar{\mathbf{u}}\}, && \text{(2d)}\\
& \mathbf{x}_0\ ,\ T \text{ 已给定}
\end{aligned}$$

而 Φ, $C(\cdot)$，以及 $\mathbf{F}(\cdot,\ \cdot)$ 分别是 ϕ，c 和 $\mathbf{f}$ 的离散等价形式。

假设1 (光滑动力学)。 $\mathbf{f}(,\cdot)$ 在区间 $[0, t_f]$ 内是一个连续可微（$\mathcal{C}^1$）函数，对所有$\mathbf{x} \in \mathcal{X}$和$\mathbf{u} \in \mathcal{U}$均成立。

引理1。 根据假设*1*，式(1) 所表示的每个连续时间最优控制问题，都可以使用零阶控制保持，将其表示为式(2) 所表示的离散时间最优控制问题，其中 Δt_i、$i = 1$、$\cdots T$ 和 $\sum_i \Delta t_i = t_f$ 足够小。

证明。 假设式 (1) 所表示的最优控制问题已给定。根据假设1，存在足够小的 Δt_i ，使得$\mathbf{f}(\mathbf{x}(t),\ \mathbf{u}(t))$ 在所有区间 $[0, \Delta t_1]$，$[\Delta t_1, \Delta t_1 + \Delta t_2]$，$[\Delta t_1 + \Delta t_2$，$\Delta t_1 + \Delta t_2 + \Delta t_3]$，$\cdots \left[\sum_{i=1}^{T-1} \Delta t_i, t_f\right]$上单调。定义区间 $I_1 = [0, \Delta t_1]$，$I_2 = [\Delta t_1, \Delta t_1 + \Delta t_2]$，$\cdots I_T = \left[\sum_{i=1}^{T-1} \Delta t_i, t_f\right]$。利用这些区间和一组 T 个控制输入$\mathbf{u}_t \in \mathcal{U}$，可以将式 (1b) 中的动力学离散化为

$$\mathbf{x}_{t+1} = \int_{I_{t+1}} \mathbf{f}(\mathbf{x}(t), \mathbf{u}_t; t)dt \qquad t = 0, 1, \cdots T-1 \tag{3}$$

其中，每个积分中的状态初始条件由前一个积分推导得出，而在 $t = 0$ 处的状态初始条件为 $\mathbf{x}_0$ =$\mathbf{x}(0)$。可以看出，式 (2b) 中的 $\mathbf{F}(\cdot, \cdot)$ 与式 (3) 中的 $\int_{I_t} \mathbf{f}(\mathbf{x}(t), \mathbf{u}_t; t)dt$ 相同。此外，由于 $\mathbf{f}(\cdot, \cdot)$ 是单调的，如果 $\mathbf{x}_t \in \mathcal{X}\ \ \forall\, t$，也就是说，如果每个 $\mathbf{x}_t$ 都满足式 (2c) 中的约束，那么根据 $f(\cdot, \cdot)$ 的单调性，式 (1c) 中的约束也得到满足。此外，如果每个 $\mathbf{u}_t$ 都满足式 (2d)，那么 $\mathbf{u}(t)$ 也满足式 (1d) 中的盒约束，因为 $\mathbf{u}(t)\ = \mathbf{u}_t, t \in I_t$。此外，还可以写成

$$C(\mathbf{x}_t, \mathbf{u}_t) = \int_{I_{t+1}} c\Big(\mathbf{f}(\mathbf{x}(t), \mathbf{u}_t), \mathbf{u}_t\Big)dt \quad \text{and} \quad \Phi() = \phi(). \tag{4}$$

因此，式 (3) 和式 (4) 表示了一种变换：将式 (1) 中的连续时间最优控制问题转化为式 (2) 中的离散时间最优控制问题，同时满足约束。

□

注 *1.* 该结果可直接推广至控制的一阶保持。控制被定义为分段线性函数，在每个离散化区间内均为单调函数。根据该线性插值函数的单调性，如果区间端点处的控制值满足盒约束 (1d)，则可保证控制的所有中间值也满足这些约束。

在本研究中，我们求解离散时间最优控制问题，并在通过零阶控制保持将连续时间问题转换为离散时间问题时，将区间 Δt_i 设为等长（$\Delta t_i = \Delta t\ \ \forall\ \ i$）。

2.2 对数障碍问题

式 (2) 中的离散时间最优控制问题是一个大规模非线性规划（NLP），适合采用原始-对偶内点法 [22]求解。这种方法通过在目标函数中引入对数障碍项来处理不等式约束，并使用拉格朗日乘子处理等式约束，从而将有约束问题重构为一系列无约束问题。

为了规范不等式约束的处理方式，我们定义索引集，用以标识受约束的分量。令 $\mathcal{I}_x \subseteq \{1,\dots,n\}$ 为对应于受盒约束的状态变量的索引集。受约束状态的数量由该集合的基数给出，即 $|\mathcal{I}_x| = n_1$。因此，无约束（自由）状态变量的数量为 $n_2 = n - n_1$。类似地，令 $\mathcal{I}_u \subseteq \{1,\dots,m\}$ 为受约束控制输入的索引集，其基数为 $|\mathcal{I}_u| = m_1$。自由控制输入的数量则为 $m_2 = m - m_1$。此记法允许对状态和控制分量的任意子集施加约束。

在每个离散时间步 k，状态变量和控制变量的盒约束仅针对这些索引集所指定的分量表示如下：

$$\underline{x}_i \le x_{t,i} \le \bar{x}_i, \quad \forall i \in \mathcal{I}_x, \quad \forall t \in \{0,\dots,T\} \tag{5}$$

$$\underline{u}_j \le u_{t,j} \le \bar{u}_j, \quad \forall j \in \mathcal{I}_u, \quad \forall t \in \{0,\dots,T-1\} \tag{6}$$

其中，$x_{t,\,i}$ 是时间步 t 时状态向量的第 i 个分量，x_i 和$\bar{x}_i$ 分别是其对应的下界和上界。控制量 $u_{t,\,j}$的记法类似。接下来，我们可以使用对数障碍法将这些不等式约束纳入代价函数。通过在目标函数中加入障碍项，并用拉格朗日乘子处理动态约束，式（2）中的约束最优控制问题由此转化为无约束问题。这得到如下障碍子问题的形式，其中我们寻求增广代价函数$\mathcal{J}_{\boldsymbol{\mu},\,\boldsymbol{\sigma}}$的驻点：

$$\begin{aligned}\mathcal{J}_{\boldsymbol{\mu},\boldsymbol{\sigma}} = \Phi(\mathbf{x}_T) &- \sum_{i\in\mathcal{I}_x} \mu_i \left(\log(x_{T,i} - \underline{x}_i) + \log(\bar{x}_i - x_{T,i})\right) \\ &+ \sum_{t=0}^{T-1} \Bigg[C(\mathbf{x}_t, \mathbf{u}_t) - \sum_{i\in\mathcal{I}_x} \mu_i \left(\log(x_{t,i} - \underline{x}_i) + \log(\bar{x}_i - x_{t,i})\right) \\ &\quad - \sum_{j\in\mathcal{I}_u} \sigma_j \left(\log(u_{t,j} - \underline{u}_j) + \log(\bar{u}_j - u_{t,j})\right) + \boldsymbol{\lambda}_{t+1}^{\top}(\mathbf{F}(\mathbf{x}_t, \mathbf{u}_t) - \mathbf{x}_{t+1}) \Bigg]. \end{aligned} \tag{7}$$

在上述表达式中，变量定义如下：$\boldsymbol{\mu} = [\mu_1, \mu_2, \cdots \mu_{n_1}]$**s.t.** $\mu_i > 0 \;\; \forall i \in 1, \;\; \cdots n_1$ 和 $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \cdots \sigma_{m_1}]$ **s.t.** $\sigma_j > 0 \;\; \forall j \in 1, \;\; \cdots m_1$分别是与状态和控制不等式约束对应的障碍参数。 $\boldsymbol{\lambda}_t \in \mathbb{R}^n$ 对于 $t = 1$, ..., T，它们是拉格朗日乘子，

也称为协态，与每个时间步的离散时间动态等式约束相关联。

通过求解 $\mathcal{J}_{\boldsymbol{\mu},\boldsymbol{\sigma}}$关于所有变量 $\{\mathbf{x}_t\}$、$\{\mathbf{u}_t\}$ 以及 $\{\boldsymbol{\lambda}_t\}$ 的驻点，可以得到最优解，其中 $\boldsymbol{\mu} \to \mathbf{0}$ 和 $\boldsymbol{\sigma} \to \mathbf{0}$ 为给定序列。

3 Box-iLQR框架

在（7）中将有约束最优控制问题重新表述为一系列无约束障碍子问题后，本节将详细介绍一种基于迭代线性二次调节器（iLQR）算法的求解方法。其核心思想是利用 iLQR 在无约束轨迹优化中的计算效率——这是一种二阶方法——将其直接应用于由对数障碍函数构造的无约束问题结构。

迭代线性二次调节器（iLQR）是一种重要的轨迹优化算法，通过反复应用后向-前向传递结构 [30]来求解最优控制问题。在每次迭代中，它围绕当前名义轨迹对动力学进行线性化，并将代价函数二次化，以构造一个局部线性二次调节器（LQR）子问题。

我们调整 iLQR 算法，以求解式 (7) 中定义的障碍增广优化问题。为确保算法适定并收敛到局部最小值，我们作出以下标准假设。

假设 2 (光滑性)。 离散时间动力学函数$\mathbf{F}(\mathbf{x}_t,\mathbf{u}_t)$是 $\mathcal{C}^1$ (连续可微)的，原始代价函数 $C(\mathbf{x},\mathbf{u})$ 和$\Phi(\mathbf{x})$ 在各自的定义域上均为 $\mathcal{C}^2$ (二阶连续可微)。

假设 3 (适定子问题)。 对于任意选择的严格为正的障碍参数，$\boldsymbol{\mu} > \mathbf{0}$ 和 $\boldsymbol{\sigma} > \mathbf{0}$，式$(7)$ 定义的无约束障碍子问题均适定，且至少存在一个有限的最优解。

假设 4 (可行初始轨迹)。 输入中给定一条可行初始轨迹，$(\bar{\mathbf{x}}_0, \ldots, \bar{\mathbf{x}}_N; \bar{\mathbf{u}}_0, \ldots, \bar{\mathbf{u}}_{N-1})$。该轨迹满足所有动力学和不等式约束。

根据假设 4，算法以一条可行轨迹初始化。我们注意到，寻找这样的轨迹本身就可能是一项颇具挑战的问题，但相关方法不在本文讨论范围内。算法的核心包括后向过程和前向过程，下面将对此进行详细介绍。

障碍代价项及其导数 我们的代价函数通过添加对数障碍项来施加不等式约束。运行代价包括一个障碍项 $\boldsymbol{\omega}(\mathbf{x}_t, \mathbf{u}_t)$，终端代价包括一个障碍项$\boldsymbol{\Omega}(\mathbf{x}_T)$。

运行**障碍项**定义如下:

$$\begin{aligned}\boldsymbol{\omega}(\mathbf{x}_t, \mathbf{u}_t) = &- \sum_{i \in \mathcal{I}_x} \mu_i \left(\log(x_{t,i} - \underline{x}_i) + \log(\overline{x}_i - x_{t,i})\right) \\ &- \sum_{j \in \mathcal{I}_u} \sigma_j \left(\log(u_{t,j} - \underline{u}_j) + \log(\overline{u}_j - u_{t,j})\right)\end{aligned} \tag{8}$$

其一阶导数（梯度）包含一些零元素和一些非零元素。二阶导数（海森矩阵）是一个对角矩阵，其中包含一些零元素和一些非零元素。梯度和海森矩阵中的非零元素如下:

$$\begin{aligned}&(\boldsymbol{\omega}_{\mathbf{x}})_i = -\mu_i \left(\frac{1}{x_{t,i} - \underline{x}_i} - \frac{1}{\overline{x}_i - x_{t,i}}\right), (\boldsymbol{\omega}_{\mathbf{u}})_j = -\sigma_j \left(\frac{1}{u_{t,j} - \underline{u}_j} - \frac{1}{\overline{u}_j - u_{t,j}}\right) \\ &(\boldsymbol{\omega}_{\mathbf{xx}})_{ii} = \mu_i \left(\frac{1}{(x_{t,i} - \underline{x}_i)^2} + \frac{1}{(\overline{x}_i - x_{t,i})^2}\right), \\ &(\boldsymbol{\omega}_{\mathbf{uu}})_{jj} = \sigma_j \left(\frac{1}{(u_{t,j} - \underline{u}_j)^2} + \frac{1}{(\overline{u}_j - u_{t,j})^2}\right),\end{aligned} \tag{9}$$

其中 $i \in \mathcal{I}_x$ 和 $j \in \mathcal{I}_u$。交叉项海森矩阵 $\boldsymbol{\omega}_{\mathbf{ux}}$ 为零。
终端**终端障碍项**定义如下:
$\boldsymbol{\Omega}(\mathbf{x}_T) = -\sum_{i \in \mathcal{I}_x} \mu_i (\log(x_{T,i} - \underline{x}_i) + \log(\overline{x}_i - x_{T,i}))$。其关于终端状态$\mathbf{x}_T$ 的导数非零元素为:
$(\boldsymbol{\Omega}_{\mathbf{x}})_i = -\mu_i \left(\frac{1}{x_{T,i} - \underline{x}_i} - \frac{1}{\overline{x}_i - x_{T,i}}\right)$ and $(\boldsymbol{\Omega}_{\mathbf{xx}})_{ii} = \mu_i \left(\frac{1}{(x_{T,i} - \underline{x}_i)^2} + \frac{1}{(\overline{x}_i - x_{T,i})^2}\right)$.

反向传递 反向传递通过传播更新后的协态 $\boldsymbol{\lambda}_t = \mathbf{S}_t \delta\mathbf{x}_t + \mathbf{v}_t$的系数，计算局部最优控制策略。它从 $t = T$ 到 $t = 0$沿时间反向计算因子 $\mathbf{v}_t$ 和 $\mathbf{S}_t$。

1. 初始化（$t = T$）： 在终端时刻，使用在名义状态 $\bar{}\mathbf{x}_T$处评估的总终端代价的导数，初始化值函数导数：$\mathbf{v}_T = \Phi_{\mathbf{x}} + \boldsymbol{\Omega}_{\mathbf{x}}$ 和$\mathbf{S}_T = \Phi_{\mathbf{xx}} + \boldsymbol{\Omega}_{\mathbf{xx}}$。

2. 反向递推（$t = T-1, \ldots, 0$）： 对于每个时间步，我们计算动作价值函数 Q_t 在围绕（$\bar{}\mathbf{x}_t$， $\bar{}\mathbf{u}_t$）偏差时的导数:

$$Q_{\mathbf{x}} = (C_{\mathbf{x}} + \boldsymbol{\omega}_{\mathbf{x}}) + \mathbf{F}_{\mathbf{x}}^{\top} \mathbf{v}_{t+1} \tag{10}$$

$$Q_{\mathbf{u}} = (C_{\mathbf{u}} + \boldsymbol{\omega}_{\mathbf{u}}) + \mathbf{F}_{\mathbf{u}}^{\top} \mathbf{v}_{t+1} \tag{11}$$

$$Q_{\mathbf{xx}} = (C_{\mathbf{xx}} + \boldsymbol{\omega}_{\mathbf{xx}}) + \mathbf{F}_{\mathbf{x}}^{\top} \mathbf{S}_{t+1} \mathbf{F}_{\mathbf{x}} \tag{12}$$

$$Q_{\mathbf{uu}} = (C_{\mathbf{uu}} + \boldsymbol{\omega}_{\mathbf{uu}}) + \mathbf{F}_{\mathbf{u}}^{\top} \mathbf{S}_{t+1} \mathbf{F}_{\mathbf{u}} \tag{13}$$

$$Q_{\mathbf{ux}} = C_{\mathbf{ux}} + \mathbf{F}_{\mathbf{u}}^{\top} \mathbf{S}_{t+1} \mathbf{F}_{\mathbf{x}} \tag{14}$$

所有导数均在标称轨迹（$\bar{}\mathbf{x}_t$， $\bar{}\mathbf{u}_t$）处求值。随后计算前馈项$\mathbf{k}_t$ 和反馈增益矩阵$\mathbf{K}_t$ 。为确保其正定，向 $Q_{\mathbf{uu}}$ 添加正则化项$\zeta\mathbf{I}$ 。

$$\mathbf{k}_t = -(Q_{\mathbf{uu}} + \zeta\mathbf{I})^{-1} Q_{\mathbf{u}} \tag{15}$$

$$\mathbf{K}_t = -(Q_{\mathbf{uu}} + \zeta\mathbf{I})^{-1} Q_{\mathbf{ux}} \tag{16}$$

最后，更新$\mathbf{S}_t$ 和$\mathbf{v}_t$ 的值：

$$\mathbf{v}_t = Q_{\mathbf{x}} + \mathbf{K}_t^\top Q_{\mathbf{uu}}\mathbf{k}_t + \mathbf{K}_t^\top Q_{\mathbf{u}} + Q_{\mathbf{ux}}^\top \mathbf{k}_t \tag{17}$$

$$\mathbf{S}_t = Q_{\mathbf{xx}} + \mathbf{K}_t^\top Q_{\mathbf{uu}}\mathbf{K}_t + \mathbf{K}_t^\top Q_{\mathbf{ux}} + Q_{\mathbf{ux}}^\top \mathbf{K}_t \tag{18}$$

前向过程 在前向过程中，通过应用计算得到的控制策略生成一条新轨迹。使用参数$\alpha \in (0, 1]$ 进行线搜索，以确保代价函数得到改善并保持可行性。从 $\mathbf{x}_0^{\text{new}} = \bar{\mathbf{x}}_0$开始：

$$\mathbf{u}_t^{\text{new}} = \bar{\mathbf{u}}_t + \alpha\mathbf{k}_t + \mathbf{K}_t(\mathbf{x}_t^{\text{new}} - \bar{\mathbf{x}}_t) \qquad t \in [0, T-1] \tag{19}$$

$$\mathbf{x}_{t+1}^{\text{new}} = \mathbf{F}(\mathbf{x}_t^{\text{new}}, \mathbf{u}_t^{\text{new}}) \qquad t \in [0, T-1] \tag{20}$$

新轨迹（$\mathbf{x}^{\text{new}}$，$\mathbf{u}^{\text{new}}$）将成为下一次迭代的标称轨迹。重复这一前向—后向过程，直至收敛。对 α 执行回溯线搜索，以确保增广代价函数 $\mathcal{J}_{\boldsymbol{\mu},\,\boldsymbol{\sigma}}$在每次迭代时都下降。如果新轨迹带来了充分下降，则接受该轨迹；通常通过比较代价的实际减少量与局部二次模型预测的减少量来衡量。这可确保算法朝着局部最小值不断推进。常见的接受准则是实际减少量与预期减少量之比为正：

$$\frac{\mathcal{J}(\text{new}) - \mathcal{J}(\text{old})}{\Delta J_{\text{expected}}} < c_1 \tag{21}$$

其中，$\mathcal{J}$（旧）和 $\mathcal{J}$（新）分别是旧轨迹和新轨迹的代价，$c_1 \in (0,\ 1)$是一个较小的常数（例如 10^{-4}），而 $\Delta J_{\text{expected}} < 0$ 是二次近似预测的代价变化。

$$\Delta J_{\text{expected}} = -\Big(\alpha - \frac{\alpha^2}{2}\Big) \sum_{i=0}^{T-1} Q_{\mathbf{u}}^\top Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}}. \tag{22}$$

这里需要注意的是，我们需要 $Q_{\mathbf{uu}}$ 对于 $\Delta J_{\text{expected}} < 0$ 为正定，即预期代价会降低。

障碍松弛 上一节介绍的 Box-iLQR 方法针对固定 的一组障碍参数 $\boldsymbol{\mu}$ 和$\boldsymbol{\sigma}$求解优化问题。为了恢复原始约束问题的解，这些参数必须被驱动至零。这由一个外循环管理，该外循环迭代地减小障碍参数，并使用 iLQR 求解器求解每个新的子问题。这一系列解理想情况下沿中心路径 趋向约束最优解。我们将完整过程称为 Box-iLQR 算法，详见算法 1。该过程从保守地较大的初始障碍参数 $\boldsymbol{\mu}_0$ 和 $\boldsymbol{\sigma}_0$ 以及一条可行初始轨迹开始。随后算法进入内循环，首先求解

当前参数下的 iLQR 子问题，并使用上一次迭代的解作为热启动以加速收敛。

这一过程的一个关键方面是参数更新规则。在内循环求解成功后，$\boldsymbol{\mu}$ 和 $\boldsymbol{\sigma}$ 的每个分量都会乘以一个因子而减小。简单或过于激进的缩减可能会使热启动轨迹进入不可行区域，在该区域中对数障碍函数无定义，从而导致内循环求解器失败。为此，我们采用自适应更新计划。如果 iLQR 求解器未能收敛，说明障碍参数的缩减幅度过大。作为应对，算法会撤销参数更新，并在下一次尝试时对相应约束采用更保守的缩减因子。此过程会持续进行，直至障碍参数低于指定容差，确保它们对总成本的贡献可以忽略不计，并使解能够紧密逼近真正的约束最优解。

备注 *2*（基于模型的表述）。所提出的算法是基于模型的，因为它需要获取动力学函数的梯度，具体而言是雅可比矩阵$\mathbf{F}_{\mathbf{x}} = {}^{\partial}\mathbf{F}\,{}_{\partial\mathbf{x}}$ 和$\mathbf{F}_{\mathbf{u}} = {}^{\partial}\mathbf{F}\,{}_{\partial\mathbf{u}}$。如果动力学$\mathbf{F}$ 有解析表达式，则可以直接推导这些雅可比矩阵，也可以使用自动微分工具高效地计算它们。相反，在将$\mathbf{F}$ 视为黑箱的无模型设定中，则必须采用其他方法。例如，无模型版本的 iLQR 会在每次迭代中，根据与系统交互所收集的数据，学习动力学的局部线性模型 [24,38]。

4 算法特性与优势

接下来，我们将说明，上一节所述的算法不仅仅是为了方便，还具有显著优势。我们将展示，对数障碍项的海森矩阵可作为一种天然的正则化项，相较于基础表述，能够增强数值稳定性。此外，我们还将展示，在约束边界附近获得的反馈增益矩阵具有一些显著特性，我们称之为“约束感知反馈”。对数障碍表述会自然地将约束信息编码到反馈增益中，使其在状态或控制接近约束边界时逐渐减小。这实际上产生了一种遵守约束的反馈律，也适用于存在不确定性的情况。为证实这些论断，以下小节将介绍相关优势，并给出相应的数学证明与分析。

4.1 固有正则化与稳定性

iLQR反向传递的稳定性取决于矩阵 $Q_{\mathbf{uu}}$ 的正定性，因为这可确保每次迭代都能按照式（22）降低代价。标准 iLQR 通过添加显式正则化项（例如 $\zeta\mathbf{I}$ ，参见式（15）和式（16））来强制满足这一条件。我们的对数障碍表述通过两种机制引入固有正则化，从而增强数值稳定性。

算法 1 Box-iLQR

要求： 初始可行轨迹 $(\bar{\mathbf{x}}\ \bar{\mathbf{u}})$
要求： 初始障碍参数 $\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0$
要求： 初始缩减因子 $\mathbf{r}_{\mu} \in (0,1)^{n_1}, \mathbf{r}_{\sigma} \in (0,1)^{m_1}$
要求： 缩减因子更新率 $\beta_r > 1$（例如，1.5）
要求： 障碍终止容差 $\epsilon_{\text{barrier}}$
初始化 $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}_0$, $\boldsymbol{\sigma} \leftarrow \boldsymbol{\sigma}_0$
初始化 $\boldsymbol{\mu}_{\text{prev}} \leftarrow \boldsymbol{\mu}$, $\boldsymbol{\sigma}_{\text{prev}} \leftarrow \boldsymbol{\sigma}$
当 $\|\boldsymbol{\mu}\| > \epsilon_{\text{barrier}}$ 或 $\|\boldsymbol{\sigma}\| > \epsilon_{\text{barrier}}$ **时**
    ▷ 使用热启动求解内循环问题。
    $(\bar{\mathbf{x}}\ \bar{\mathbf{u}}, \text{成功}, \text{失败_索引}) \leftarrow$ 迭代线性二次调节_求解$(\bar{\mathbf{x}}\ \bar{\mathbf{u}}\ \boldsymbol{\mu}, \boldsymbol{\sigma})$
    如果未成功 **则**
        ▷ 缩减过于激进；撤销并调整。
        $\boldsymbol{\mu} \leftarrow \boldsymbol{\mu}_{\text{prev}}$, $\boldsymbol{\sigma} \leftarrow \boldsymbol{\sigma}_{\text{prev}}$
        如果 约束 i在 $\boldsymbol{\mu}$ 中失败（来自失败的_索引）**则**
            $(\mathbf{r}_{\mu})_i \leftarrow \min(1,\ (\mathbf{r}_{\mu})_i \cdot \beta_r)$ ▷ 降低缩减的激进程度。
        否则如果 约束 j 在 $\boldsymbol{\sigma}$ 失败**则**
            $(\mathbf{r}_{\sigma})_j \leftarrow \min(1, (\mathbf{r}_{\sigma})_j \cdot \beta_r)$
        如果结束
        继续 ▷使用旧参数和新速率重新求解。
    else
        ▷ 求解成功；准备进行下一次缩减。
        $\boldsymbol{\mu}_{\text{prev}} \leftarrow \boldsymbol{\mu}$, $\boldsymbol{\sigma}_{\text{prev}} \leftarrow \boldsymbol{\sigma}$
        $\boldsymbol{\mu} \leftarrow \mathbf{r}_{\mu} \odot \boldsymbol{\mu}$ ▷ 逐元素乘法。
        $\boldsymbol{\sigma} \leftarrow \mathbf{r}_{\sigma} \odot \boldsymbol{\sigma}$
    结束 if
结束 while
返回最优轨迹（$\bar{\mathbf{x}}\ \bar{\mathbf{u}}$）

首先，将式（9）中的项 $\boldsymbol{\omega}_{\mathbf{uu}}$ （代表控制障碍函数的二阶导数）直接加到 $Q_{\mathbf{uu}}$ 表达式中。作为凸函数的海森矩阵， $\boldsymbol{\omega}_{\mathbf{uu}}$ 是正定的，可作为自适应正则化项，其作用类似于手动调节的 $\varsigma\mathbf{I}$。

其次，状态对数障碍函数会产生间接的正则化效果。项 $\boldsymbol{\omega}_{\mathbf{xx}}$ 通过直接加到 $Q_{\mathbf{xx}}$ 项中（式（12））， 改善了下一时间步的价值函数海森矩阵 $\mathbf{S}_{t+1}$的条件。$\mathbf{S}_{t+1}$ 条件的改善会通过式（13）中的 $Q_{\mathbf{uu}}$ 更新传播，从而进一步确保其正定性。总体而言，这种双重正则化效果能够稳定反向传递，尤其是在未正则化的 $Q_{\mathbf{uu}}$可能条件不良或不定的区域。

4.2 收敛保证

假设5 （控制海森矩阵的正定性）。 式（13）中的控制海森矩阵项 $Q_{\mathbf{uu}}$ 在所有时间步均为正定。

定理1（收敛到最优解）。在假设*2-5*下，对于每个$(\boldsymbol{\mu}, \boldsymbol{\sigma})$ 对，*Box-iLQR*算法*1* 总会收敛到一个最优解，并在$(\boldsymbol{\mu}, \boldsymbol{\sigma}) \rightarrow \mathbf{0}$时达到式$(2)$中最优控制问题的最优解。

证明。让我们分析由控制更新引起的总成本 $\mathcal{J}$的变化。我们首先考虑由式(22)中的二次模型所预测的期望代价变化：

$$\Delta J_{\text{expected}} = -\Big(\alpha - \frac{\alpha^2}{2}\Big) \sum_{i=0}^{T-1} Q_{\mathbf{u}}^\top Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}}.$$

在假设5（$Q_{\mathbf{uu}}$ 为正定）下，且对于线搜索参数$\alpha \in (0,1)$，项$(\alpha - \alpha^2/2)$为正。因此， $\Delta J_{\text{expected}} \leq 0$。因而，除非解已收敛（$Q_{\mathbf{u}} = \mathbf{0}$），iLQR反向传递总是预测成本降低。要使真实的非线性成本 $\mathcal{J}$下降，二次近似必须足够有效。这可通过适当选择 α来保证。

为说明这一点，我们首先分析状态和控制扰动的量级。式 (19) 中的控制更新规则为$\delta\mathbf{u}_t = \alpha\mathbf{k}_t + \mathbf{K}_t\delta\mathbf{x}_t$。我们分析从固定的初始状态 $\mathbf{x}_0$, 开始的轨迹展开，这意味着 $\delta\mathbf{x}_0 = \mathbf{0}$。

在时间步 $t = 0$:

$$\delta\mathbf{u}_0 = \alpha\mathbf{k}_0 + \mathbf{K}_0\delta\mathbf{x}_0 = \alpha\mathbf{k}_0 = \mathcal{O}(\alpha).$$

下一步的状态扰动 $\delta\mathbf{x}_1$，可通过展开动力学方程得到：

$$\begin{aligned}\delta\mathbf{x}_1 =& \mathbf{F}_{\mathbf{x}}^0\delta\mathbf{x}_0 + \mathbf{F}_{\mathbf{u}}^0\delta\mathbf{u}_0 \\ &+ \frac{1}{2}\delta\mathbf{x}_0^\top\mathbf{F}_{\mathbf{xx}}^0\delta\mathbf{x}_0 + \delta\mathbf{u}_0\mathbf{F}_{\mathbf{ux}}^0\delta\mathbf{x}_0 + \frac{1}{2}\delta\mathbf{u}_0^\top\mathbf{F}_{\mathbf{uu}}^0\delta\mathbf{u}_0 + \mathcal{O}(\|\delta\mathbf{x}_0\|^3, \|\delta\mathbf{u}_0\|^3).\end{aligned}$$

利用 $\delta\mathbf{x}_0 = \mathbf{0}$ 和 $\delta\mathbf{u}_0 = \mathcal{O}(\alpha)$，可得

$$\delta\mathbf{x}_1 = \mathbf{F}^0_{\mathbf{u}}\delta\mathbf{u}_0 + \frac{1}{2}\delta\mathbf{u}_0^\top\mathbf{F}^0_{\mathbf{uu}}\delta\mathbf{u}_0 + \mathcal{O}(\alpha^3),$$

由此可得

$$\delta\mathbf{x}_1 = \alpha\mathbf{F}^0_{\mathbf{u}}\mathbf{k}_0 + \frac{\alpha^2}{2}\mathbf{k}_0^\top\mathbf{F}_{\mathbf{uu}}\mathbf{k}_0 + \mathcal{O}(\alpha^3).$$

类似地，在下一个时间步，我们有

$$\begin{aligned}\delta\mathbf{x}_2 =&\mathbf{F}^1_{\mathbf{x}}\delta\mathbf{x}_1 + \mathbf{F}^1_{\mathbf{u}}\delta\mathbf{u}_1\\ &+ \frac{1}{2}\delta\mathbf{x}_1^\top\mathbf{F}^1_{\mathbf{xx}}\delta\mathbf{x}_1 + \delta\mathbf{u}_1\mathbf{F}^1_{\mathbf{ux}}\delta\mathbf{x}_1 + \frac{1}{2}\delta\mathbf{u}_1^\top\mathbf{F}^1_{\mathbf{uu}}\delta\mathbf{u}_1 + \mathcal{O}(\|\delta\mathbf{x}_1\|^3, \|\delta\mathbf{u}_1\|^3).\end{aligned}$$

将 $\delta\mathbf{x}_1$,代入，我们得到

$$\begin{aligned}\delta\mathbf{x}_2 =&\mathbf{F}^1_{\mathbf{x}}\big(\alpha\mathbf{F}^0_{\mathbf{u}}\mathbf{k}_0 + \frac{\alpha^2}{2}\mathbf{k}_0^\top\mathbf{F}_{\mathbf{uu}}\mathbf{k}_0 + \mathcal{O}(\alpha^3)\big)\\ &+ \alpha\mathbf{F}^1_{\mathbf{u}}\mathbf{k}_1 + \mathbf{F}^1_{\mathbf{u}}\mathbf{K}_1\big(\alpha\mathbf{F}^0_{\mathbf{u}}\mathbf{k}_0 + \frac{\alpha^2}{2}\mathbf{k}_0^\top\mathbf{F}_{\mathbf{uu}}\mathbf{k}_0 + \mathcal{O}(\alpha^3)\big) + \mathcal{O}(\alpha^2),\end{aligned}$$

其可简化为

$$\delta\mathbf{x}_2 = \alpha\mathbf{F}^1_{\mathbf{u}}\mathbf{k}_1 + \alpha\big(\mathbf{F}^1_{\mathbf{x}} + \mathbf{F}^1_{\mathbf{u}}\mathbf{K}_1\big)\mathbf{F}^0_{\mathbf{u}}\mathbf{k}_0 + \mathcal{O}(\alpha^2).$$

因此，对于给定的时间步 t， $\delta\mathbf{x}_t$ 为:

$$\begin{aligned}\delta\mathbf{x}_t =&\alpha\mathbf{F}^{t-1}_{\mathbf{u}}\mathbf{k}_{t-1} + \alpha\big(\mathbf{F}^{t-1}_{\mathbf{x}} + \mathbf{F}^{t-1}_{\mathbf{u}}\mathbf{K}_{t-1}\big)\mathbf{F}^{t-2}_{\mathbf{u}}\mathbf{k}_{t-2}\\ &+ \alpha\big(\mathbf{F}^{t-1}_{\mathbf{x}} + \mathbf{F}^{t-1}_{\mathbf{u}}\mathbf{K}_{t-1}\big)\big(\mathbf{F}^{t-2}_{\mathbf{x}} + \mathbf{F}^{t-2}_{\mathbf{u}}\mathbf{K}_{t-2}\big)\mathbf{F}^{t-3}_{\mathbf{u}}\mathbf{k}_{t-3} + \cdots\\ &+ \alpha\big(\mathbf{F}^{t-1}_{\mathbf{x}} + \mathbf{F}^{t-1}_{\mathbf{u}}\mathbf{K}_{t-1}\big)\cdots\cdots\mathbf{F}^0_{\mathbf{u}}\mathbf{k}_0 + \alpha^2(\cdots) + \alpha^3(\cdots) + \cdots.\end{aligned}$$

根据上述方程可知，动力学中的一阶项系数为 α，二阶项系数为 α^2，依此类推。这表明，选择较小的 α 可以将扰动限制在整个轨迹上均较小的范围内，这是确保局部线性近似有效的关键。动力学中的高阶项（例如二阶项）为 $\mathcal{O}(\alpha^2)$阶，其衰减速度更快。

现在，我们考虑代价的实际变化量， $\Delta\mathcal{J} = \mathcal{J}_{\text{new}} - \mathcal{J}_{\text{old}}$。泰勒级数展开表明，实际代价变化量与二次预测变化量 $\Delta J_{\text{expected}}$相关。差异来源于两个方面: **(i)** 代价函数中的高阶项（三阶及以上），以及 **(ii)** 反向传递中使用的非线性动力学与线性模型之间的误差。高阶代价项的形式为 $\mathcal{O}(\|\delta\mathbf{x}\|^3,\ \|\delta\mathbf{u}\|^3)$，根据我们的分析，这些项为 $\mathcal{O}(\alpha^3)$。动力学非线性的影响为 $\mathcal{O}(\alpha^2)$阶，它们沿轨迹积分后再乘以协态 $\boldsymbol{\lambda}$。其余动力学非线性导致的误差为 $\mathcal{O}(\alpha^3)$阶。因此，最终误差为 $\mathcal{O}(\alpha^2)$阶。由此，代价的实际变化量可表示为:

$$\Delta\mathcal{J} = \Delta J_{\text{expected}} + \mathcal{O}(\alpha^2).$$

式 (21) 中的线搜索接受准则要求比值 $\rho = \Delta\mathcal{J}/\Delta J_{\text{expected}}$ 大于某个参数 $c_1 \in (0, 1)$。下面考察当 $\alpha \to 0$ 时该比值的极限：

$$\begin{aligned}\lim_{\alpha\to 0} \frac{\Delta\mathcal{J}}{\Delta J_{\text{expected}}} &= \lim_{\alpha\to 0} \frac{-\left(\alpha - \frac{\alpha^2}{2}\right)\sum Q_{\mathbf{u}}^\top Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}} + \mathcal{O}(\alpha^2)}{-\left(\alpha - \frac{\alpha^2}{2}\right)\sum Q_{\mathbf{u}}^\top Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}}} \\ &= \lim_{\alpha\to 0} \frac{-\alpha\sum Q_{\mathbf{u}}^\top Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}} + \mathcal{O}(\alpha^2)}{-\alpha\sum Q_{\mathbf{u}}^\top Q_{\mathbf{uu}}^{-1} Q_{\mathbf{u}} + \mathcal{O}(\alpha^2)} = 1.\end{aligned}$$

由于实际代价降低量与预期代价减少量之比在 $\alpha \to 0$ 时趋近于 1，因此对于任意 $c_1 \in (0,1)$，必然存在一个足够小的 $\alpha > 0$，使得 $\Delta\mathcal{J}/\Delta J_{\text{expected}} > c_1$。这保证了只要算法尚未收敛（$Q_{\mathbf{u}} \neq \mathbf{0}$），线搜索最终就会找到一个可接受的步长，使真实代价 $\mathcal{J}$ 降低。因此，可以保证代价降低，且 $\mathcal{J}_{\text{new}} < \mathcal{J}_{\text{old}}$。等式 $\mathcal{J}_{\text{new}} = \mathcal{J}_{\text{old}}$ 仅在收敛时成立，此时 $\Delta J_{\text{expected}} = 0$。在已证明对于一组固定参数（$\boldsymbol{\mu}, \boldsymbol{\sigma}$），Box-iLQR 会收敛到一个最优解之后，我们现在将证明，在算法的外循环中逐渐减小这些参数时，仍能保持约束满足。

在已证明对于一组固定参数（$\boldsymbol{\mu}, \boldsymbol{\sigma}$），内层 iLQR 循环会收敛到一个局部极小值之后，我们现在证明，在算法的外循环中减小这些参数时，仍能保持严格的约束满足。

我们首先定义算法某次迭代中的相关量。令 $\boldsymbol{\Gamma} = [\boldsymbol{\mu}, \boldsymbol{\sigma}]^\top$ 表示所有障碍参数组成的向量。在每个时间步 t，最优状态和控制变量的拼接向量表示为 $\mathbf{z}_t = [\mathbf{x}_t, \mathbf{u}_t]^\top$。

$\mathbf{z}_t$ 相对于其下界和上界 $\mathbf{z}_t$ 和 $\bar{\mathbf{z}}_t$ 的松弛度分别由向量 $\boldsymbol{\varepsilon}_t$ 和 $\bar{\boldsymbol{\varepsilon}}_t$ 给出。这些量定义如下：

$$\mathbf{z}_t - \underline{\mathbf{z}}_t = \underline{\boldsymbol{\varepsilon}}_t, \tag{23}$$

$$\bar{\mathbf{z}}_t - \mathbf{z}_t = \bar{\boldsymbol{\varepsilon}}_t. \tag{24}$$

对于 $\mathbf{z}_t$ 的任意无约束分量，相应的边界为无穷大（即 $z_{t,i} = -\infty$ 或 $\bar{z}_{t,i} = \infty$），相应的松弛度也为无穷大（$\varepsilon_{t,i} = \infty$ 或 $\bar{\varepsilon}_{t,i} = \infty$）。

为确保更新后的变量在后续迭代中保持严格可行，变化量 $\delta\mathbf{z}_t$ 必须满足：

$$\underline{\mathbf{z}}_t < \mathbf{z}_t + \delta\mathbf{z}_t < \bar{\mathbf{z}}_t. \tag{25}$$

更新量 $\delta\mathbf{z}_t$ 是在障碍参数从 $\boldsymbol{\Gamma}^{\text{iter}}$ 变为 $\boldsymbol{\Gamma}^{\text{iter}+1}$ 后，根据式（7）中的增广代价函数 $\mathcal{J}$ 进行二次近似而得到的。使用线搜索参数 α，该更新量可近似表示为：

$$\delta\mathbf{z}_t \approx -\alpha\mathcal{J}_{\mathbf{z}_t,\mathbf{z}_t}^{-1}(\Delta\mathcal{J}_{\mathbf{z}_t}),$$

其中， $\Delta\mathcal{J}_{\mathbf{z}_t} = \mathcal{J}_{\text{iter}_t}{}^{+1} - \mathcal{J}^{\text{iter}}_{\mathbf{z}_t}{}_{\mathbf{z}}$表示连续迭代之间成本梯度的差值，而$\mathcal{J}_{\mathbf{z}_t,\mathbf{z}_t}$是当前迭代的海森矩阵，并假设其可逆。

为保持相对于下界的可行性，我们要求 $\mathbf{z}_t + \delta\mathbf{z}_t > \mathbf{z}_t$。代入 $\mathbf{z}_t$ 和 $\delta\mathbf{z}_t$ 的表达式可得：

$$\begin{aligned}(\underline{\mathbf{z}}_t + \underline{\varepsilon}_t) - \alpha\mathcal{J}^{-1}_{\mathbf{z}_t,\mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}) &> \underline{\mathbf{z}}_t \\ \underline{\varepsilon}_t - \alpha\mathcal{J}^{-1}_{\mathbf{z}_t,\mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}) &> \mathbf{0} \\ \alpha\mathcal{J}^{-1}_{\mathbf{z}_t,\mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}) &< \underline{\varepsilon}_t.\end{aligned}$$

该不等式规定了 α 的最大允许值。对于无约束变量，由于其对应的松弛度 $\varepsilon_{t,s}$ 为无穷大，约束显然得到满足。此外，对于任意分量 s，若项 $\left(\mathcal{J}^{-1}_{\mathbf{z}_t,\ \mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t})\right)_s$ 为负，则对于任意正的 α，该不等式均成立。

因此， α 的取值仅受满足 $\left(\mathcal{J}^{-1}_{t\ \mathbf{z},\ \mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t})\right)_s > 0$ 的分量约束。令 $(\mathcal{J}^{-1}_{\mathbf{z}_t,\ \mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}))^+$ 表示一个仅包含受约束变量中这些正元素的向量，并令 ε^+_t 表示相应松弛变量组成的向量。为确保不违反任何约束， α 必须同时满足所有此类分量的条件。由此得到界：

$$\alpha < \min_s \left[\frac{\underline{\varepsilon}^+_{t,s}}{\left((\mathcal{J}^{-1}_{\mathbf{z}_t,\mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}))^+\right)_s} \right], \tag{26}$$

其中 $(\cdot)_s$ 表示对应向量的第 s个元素。

对上界可行性条件进行类似推导，$\mathbf{z}_t + \delta\mathbf{z}_t < \bar{}\mathbf{z}_t$，可得到关于 α的类似约束：

$$\alpha < \min_s \left[\frac{\bar{\varepsilon}^+_{t,s}}{\left((\mathcal{J}^{-1}_{\mathbf{z}_t,\mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}))^+\right)_s} \right]. \tag{27}$$

为确保在每一个时间步 t都满足所有约束，并同时满足式(21)中的充分下降条件，最终的线搜索参数必须选取为：

$$\alpha < \min_{t,s} \left[\frac{\underline{\varepsilon}^+_{t,s}}{\left((\mathcal{J}^{-1}_{\mathbf{z}_t,\mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}))^+\right)_s}, \ \frac{\bar{\varepsilon}^+_{t,s}}{\left((\mathcal{J}^{-1}_{\mathbf{z}_t,\mathbf{z}_t}(\Delta\mathcal{J}_{\mathbf{z}_t}))^+\right)_s}, \ \alpha_{\text{LQ}} \right], \tag{28}$$

其中 α_{LQ} 表示使式(21)中的不等式成立的线搜索参数取值。代价梯度的变化 $\Delta\mathcal{J}_{\mathbf{z}_t}$由障碍参数的变化所驱动。对于第 p个分量，该关系由下式给出：

$$(\Delta\mathcal{J}_{\mathbf{z}_t})_s = \left(\gamma^{\text{iter}}_s - \gamma^{\text{iter}+1}_s\right)\left[\frac{1}{\underline{\varepsilon}_{t,s}} - \frac{1}{\bar{\varepsilon}_{t,s}}\right],$$

其中， γ^p_s 是 $\mathbf{\Gamma}$ 在第 p次迭代时的第 s个分量。如果$\varepsilon_{t,\ s} << 1$ 或$\bar{\varepsilon}_{t,\ \underline{s}} << 1$，变量$z_{t,\ s}$ 非常接近其下界或上界

界，$\mathcal{J}_{\mathbf{z}_t}^{-1},\ \mathbf{z}_t(\Delta\mathcal{J}_{\mathbf{z}_t})$由障碍项主导，因此可以近似为

$$(\mathcal{J}_{\mathbf{z}_t,\mathbf{z}_t}^{-1}(\Delta\mathcal{J}_{\mathbf{z}_t}))_s \approx \left(\frac{\gamma_s^{\text{iter}+1}}{\varepsilon_{t,s}^2}\right)^{-1}\left(\frac{\left(\gamma_s^{\text{iter}}-\gamma_s^{\text{iter}+1}\right)}{\varepsilon_{t,s}}\right)=\frac{\gamma_s^{\text{iter}}-\gamma_s^{\text{iter}+1}}{\gamma_s^{\text{iter}+1}}\varepsilon_{t,s}. \quad (29)$$

如果将其代入式 (28)，上式中的$\varepsilon_{t,\ s}$会与分子相消，因此对 α 的相应要求变为

$$\alpha < \frac{\gamma_s^{\text{iter}+1}}{\gamma_s^{\text{iter}}-\gamma_s^{\text{iter}+1}}. \quad (30)$$

因此， α 的界限永远不会变为 0，而是保持为正值。这意味着存在一个 $\alpha\in$（0，1），使得当障碍参数（$\boldsymbol{\mu}$，$\boldsymbol{\sigma}$）$\to$ 0 时，式 (2c) 和式 (2d) 中的盒约束不会被违反。

式 (28) 揭示了步长 α 与障碍参数变化幅度（$\gamma^{\text{iter}}-\gamma^{\text{iter}+1}$）之间的反向关系，因为该项通过 $\Delta\mathcal{J}_{\mathbf{z}_t}$ 出现在分母中。这表明，若要大幅缩减障碍参数，就必须采用更小的步长 α 来保持可行性，这可能会减慢收敛。反之，该方程表明，α 与松弛度 ε 成正比，因此当变量远离其界限时，可以采用更大的步长。

因此，通过选择合适的线搜索参数 α，我们可以说，对于给定的一对 ($\boldsymbol{\mu}$, $\boldsymbol{\sigma}$)，Box-iLQR 会收敛到最优解，并且在减小障碍参数时满足约束。因此，当 ($\boldsymbol{\mu}$，$\boldsymbol{\sigma}$) $\to$ 0 时，Box-iLQR 会收敛到最优解。

□

根据定理1，可以推知，如果式 (2) 中的问题本身是凸的，那么 Box-iLQR 会保持凸性。

注 *3.* 在哈密顿量 [2] 具有凸性且满足假设 2–4 的条件下，当参数 $\boldsymbol{\mu}$，$\boldsymbol{\sigma}\to$ 0 时，Box-iLQR 算法生成的解序列会收敛至全局最优解。

这种正则化和结构保持相较于其他方法 [31,16,25,12],具有显著优势，因为其他方法可能需要临时采用正则化措施，而且还可能收敛到次优解。

4.3 约束感知反馈控制

迭代线性二次调节算法通过前馈项（$\mathbf{k}_t$）和反馈增益矩阵（$\mathbf{K}_t$）更新控制。本节重点介绍反馈项，它具有两个关键功能。首先，在优化过程中，它确保迭代更新保持可行，防止违反约束。其次，在收敛时，

最优反馈增益$\mathbf{K}_t^*$ 构成一个局部线性时变（LTV）反馈控制器，用于在扰动[38,21]下跟踪标称轨迹。反馈律如下：

$$\mathbf{u}_t = \mathbf{u}_t^* + \mathbf{K}_t^*(\mathbf{x}_t - \mathbf{x}_t^*) \tag{31}$$

其中，根据与最优状态 $\mathbf{x}_t^*$ 的偏差施加纠正动作。

我们将证明，由我们的 Box-iLQR 导出的反馈控制器在以下两方面本质上具有“约束感知”特性：1. 当控制输入 $u_{t,j}$ 处于其边界（式 (2d)）时，反馈矩阵的对应行 $(\mathbf{K}_t)_j$ 变为零向量。这可防止反馈动作指令产生不可行变化。以及 2. 障碍项会在局部代价到达函数 $(Q_{\mathbf{uu}}, Q_{\mathbf{ux}})$ 中诱导出陡峭曲率，从而使反馈增益积极抵消任何朝向不可行区域的状态偏差（式 (2c)）。

我们现在着手将这些论断形式化并加以证明。

定理 2（控制边界处的反馈增益结构）。 式(31) 中的反馈控制器本质上遵守来自式(6)的控制饱和限制。具体而言，若最优控制输入 $u_{t,j}$ 在其边界处饱和，则反馈矩阵 $\mathbf{K}_t$ 的对应行为零向量。

证明。 首先，让我们重新审视反馈方程：

$$\mathbf{K}_t = -(Q_{\mathbf{uu}} + \zeta\mathbf{I})^{-1}Q_{\mathbf{ux}}, \tag{32}$$

并记作

$$\mathbf{M} = Q_{\mathbf{uu}} + \zeta\mathbf{I} = C_{\mathbf{uu}} + \boldsymbol{\omega}_{\mathbf{uu}} + \mathbf{F}_{\mathbf{u}}^\top \mathbf{S}_{t+1}\mathbf{F}_{\mathbf{u}} + \zeta\mathbf{I}, \tag{33}$$

其中（$\boldsymbol{\omega}_{\mathbf{uu}}$）$_{jj}$ 由公式（9）给出。容易看出，每当控制输入接近饱和时，$(\boldsymbol{\omega}_{\mathbf{uu}})$的相应对角项就趋于 ∞。令$\mathcal{I}_{u,\ c} \subseteq \mathcal{I}_u$ 表示与饱和控制输入对应的索引集合，并令$|\mathcal{I}_{u,\ c}| = m_{1,\ c}$表示其基数，其中$1 \leq m_{1,\ c} \leq m_1$。现在，我们可以说（$\mathbf{M}$）$_{kk} \to \infty \quad \forall\, k \in \mathcal{I}_{u,\ c}$，即$\mathbf{M}$ 的 k阶对角元素 $\to \infty$。现在可以写成$\mathbf{M}^{-1}\mathbf{M} = \mathbf{I}$，或者用指标记号表示为

$$\sum_{b=1}^{m} (\mathbf{M}^{-1})_{ab}(\mathbf{M})_{bc} = \boldsymbol{\delta}_{ac},$$

其中 $\boldsymbol{\delta}_{ac}$ 是克罗内克δ函数。在上式中，$\forall k \in \mathcal{I}_{u,\ c}$，我们将得到

$$\begin{aligned}
&\sum_{b=1}^{m} (\mathbf{M})_{kb}(\mathbf{M}^{-1})_{bc} = \boldsymbol{\delta}_{kc} \\
\Rightarrow &\sum_{b=1}^{m} (\mathbf{M})_{kb}(\mathbf{M}^{-1})_{bc} \approx (\mathbf{M})_{kk}(\mathbf{M}^{-1})_{kc} = \boldsymbol{\delta}_{kc} \quad \text{as} \quad (\mathbf{M})_{kk} >> (\mathbf{M})_{bk} (b \neq k).
\end{aligned}$$

当 $k \neq c$ 时，我们有 $\boldsymbol{\delta}_{kc} = 0$，因此 $(\mathbf{M}^{-1})_{kc} = 0$，若 $k \neq c$。此外，$(\mathbf{M}^{-1})_{kk} = 1/(\mathbf{M})_{kk} \to 0$，因为 $(\mathbf{M})_{kk} \to \infty$。所以，$\mathbf{M}^{-1}$ 的第 k 行元素全为零。

将此结果代入不等式 (32)，可知 k 行的 $\mathbf{K}_t^*$ 元素也全为零。因此，控制饱和时没有反馈。

□

注 *4.* 在实际应用中，与饱和控制量对应的$\mathbf{K}_t$ 行不会严格变为零，而是收敛到较小但非零的值。这是使用对数障碍内点法的直接结果，该方法可确保解始终严格位于可行集内。只有当障碍参数趋于零时，控制值才会趋近其限值。尽管这种方法在理论上是可靠的，但可能导致系统矩阵的对角元素非常大，从而引发数值病态。为缓解这一问题，常见的策略有两种：

1. 当解达到距边界所需容差范围内时终止算法，并留出较小的裕度。换言之，在障碍参数达到某些足够小的值时终止算法。2. 将安全裕度 $\varepsilon > 0$ 直接纳入约束，通过施加 $[u_j + \varepsilon, \bar{u}_j - \varepsilon]$控制限幅。

这也意味着，在算法收敛时得到的反馈$\mathbf{K}_t^*$不会影响饱和控制；若将其用于围绕不确定性下的最优解 [21,24]设计闭环系统，则也是如此。

假设6 （线性化动力学的有效性）。 假设从一次迭代到下一次迭代，状态扰动的演化可由以下线性模型精确描述： $\delta \mathbf{x}_{t+1} = \mathbf{F}_{\mathbf{x}} \delta \mathbf{x}_t + \mathbf{F}_{\mathbf{u}} \delta \mathbf{u}_t$.

定理3（状态边界处的反馈增益结构）。在假设*6*下，考虑式（19）中的反馈律。若最优状态轨迹的第 i个分量 $x_{t+1,i}$处于式（5）定义的约束边界上，则反馈增益矩阵$\mathbf{K}_t$ 具有这样的结构：控制修正量 $\delta \mathbf{u}_t = \alpha \mathbf{k}_t + \mathbf{K}_t \delta \mathbf{x}_t$ 对下一时间步的状态偏差第 i个分量$\delta x_{t+1,i}$没有净影响。从数学上说，这意味着状态偏差的第 i个分量完全由开环动力学决定：$(\delta \mathbf{x}^{t+1})_i = (\mathbf{F}_{\mathbf{x}} \delta \mathbf{x}_t)_i$.

证明。 通过在代价函数中加入惩罚项来强制执行 $x_{t+1,\ i}$ 上的激活状态约束，导致状态惩罚海森矩阵 $(\boldsymbol{\omega}_{\mathbf{xx}})_{ii}$ 的相应对角元素趋于无穷大。该惩罚会通过黎卡提递推向后传播。

$\mathbf{S}_{t+1}$ 在反向传递过程中计算。其依赖于状态的分量由惩罚项主导：

$$\mathbf{S}_{t+1} \approx \boldsymbol{\omega}_{\mathbf{xx}} + \dots$$

当 $(\boldsymbol{\omega}_{\mathbf{xx}})_{ii} \to \infty$ 时， i 阶对角元素（$\mathbf{S}_{t+1}$）$_{ii}$ 也趋于无穷大。对于对称矩阵，这意味着其最大特征值 η_i 趋于无穷大，而对应的特征向量 $\boldsymbol{v}_i$ 收敛于第 i 个标准基向量 $\mathbf{e}_i$。此外，由于对称

矩阵的特征向量彼此正交，所有其他特征向量 $\boldsymbol{v}_k$ （对于 $k \neq i$）都将与 $\mathbf{e}_i$ 正交，这意味着它们的第 i 个分量为零。

在时间步 t，动作价值函数的海森矩阵$\mathbf{Q_{uu}}$为：

$$\mathbf{Q_{uu}} = C_{\mathbf{uu}} + \boldsymbol{\omega}_{\mathbf{uu}} + \mathbf{F_u}^\top \mathbf{S}_{t+1} \mathbf{F_u}.$$

令$\mathbf{P} = \mathbf{F_u}^\top \mathbf{S}_{t+1} \mathbf{F_u}$。使用谱分解$\mathbf{S}_{t+1} = \sum_{k=1}^{n} \eta_k \boldsymbol{v}_k \boldsymbol{v}_k^\top$，$\mathbf{P}$ 的第（a, b）个元素为：

$$(\mathbf{P})_{ab} = (\mathbf{F_u}^a)^\top \mathbf{S}_{t+1} \mathbf{F_u}^b = \sum_{k=1}^{n} \eta_k \left(\boldsymbol{v}_k^\top (\mathbf{F_u}^a)\right) \left((\mathbf{F_u}^b)^\top \boldsymbol{v}_k\right),$$

其中$\mathbf{F_u}^a$ 和$\mathbf{F_u}^b$ 分别是$\mathbf{F_u}$的第 a列和第 b列。当 $\eta_i \to \infty$且 $\boldsymbol{v}_i \to \mathbf{e}_i$时，该求和由第 i项主导：

$$(\mathbf{P})_{ab} \approx \eta_i \left(\mathbf{e}_i^\top (\mathbf{F_u}^a)\right) \left((\mathbf{F_u}^b)^\top \mathbf{e}_i\right) = \eta_i (\mathbf{F_u})_{ia} (\mathbf{F_u})_{ib}.$$

这表明，如果控制$\mathbf{u}_t$ 的第 a个和第 b个分量会影响状态分量（$\mathbf{x}_t$）$_i$（即（$\mathbf{F_u}$）$_{ia} \neq 0$ 和（$\mathbf{F_u}$）$_{ib} \neq 0$），那么（$\mathbf{P}$）$_{ij} \to \infty \; \forall i, j \in a, b$中的元素，以及由此得到的（$\mathbf{Q_{uu}}$）$_{ij} \to \infty \; \forall i, j \in a, b$。

反馈增益为 $\mathbf{K}_t^* = -(Q_{\mathbf{uu}})^{-1} Q_{\mathbf{ux}}$。令 $\mathcal{I}_i = \{a \mid (\mathbf{F_u})_{ia} \neq 0\}$ 为影响第 i 个状态的控制量索引集合。对于任意 $a \in \mathcal{I}_i$，相应的对角项 $(\mathbf{Q_{uu}})_{aa}$ 趋于无穷大。根据定理 2 的结果，矩阵逆中与无穷大对角项对应的行趋于零。因此，对于所有 $a \in \mathcal{I}_i$，$(\mathbf{Q_{uu}})^{-1}$ 的第 a 行都趋于零向量。由此，对于所有 $a \in \mathcal{I}_i$，$\mathbf{K}_t^*$的第 a 行也必须为零。即使一行中的多个元素都趋于 ∞，这一结论仍然成立。

控制修正量为 $\delta \mathbf{u}_t = -\mathbf{K}_t^* \delta \mathbf{x}_t$。如果 $a \in \mathcal{I}_i$，第 a 个分量 $(\delta \mathbf{u}_t)_a$ 为零。现在考虑控制量对下一状态偏差的影响中的第 i 个分量，即 $(\mathbf{F_u} \delta \mathbf{u}_t)_i$：

$$(\mathbf{F_u} \delta \mathbf{u}_t)_i = \sum_{a=1}^{m} (\mathbf{F_u})_{ia} (\delta \mathbf{u}_t)_a.$$

此和恒等于零。每一项都为零，因为要么（$\mathbf{F_u}$）$_{ia} = 0$ （若$a \notin \mathcal{I}_i$），要么（$\delta \mathbf{u}_t$）$_a = 0$ （若 $a \in \mathcal{I}_i$）。因此，完整状态偏差的第 i个分量可简化为：

$$(\delta \mathbf{x}_{t+1})_i = (\mathbf{F_x} \delta \mathbf{x}_t)_i + (\mathbf{F_u} \delta \mathbf{u}_t)_i = (\mathbf{F_x} \delta \mathbf{x}_t)_i.$$

这表明，反馈控制器有效地“切断”了其对饱和状态分量的影响，确保反馈控制动作不会违反约束。

□

根据定理2和定理3，可以说，从 Box-iLQR 获得的反馈增益是“约束感知的”。

注*5*。 在收敛时获得的反馈律 $\mathbf{K}_t^*$ 可用于围绕不确定性下的最优解设计闭环控制器。一种实用策略是保留障碍参数 $\boldsymbol{\mu}$ 的某个有限值，而不是让其减小到非常小的值。这样可以确保标称轨迹始终严格位于边界 $[x_i+\varepsilon, \bar{x}_i-\varepsilon]$ 以内。在随机情形下，当下一状态受到前一状态影响时，这一裕度有助于维持约束满足。此外，方程（31）中的反馈律能够感知先前时间步即将出现的约束，因为约束信息通过方程（17）和方程（18）中的反向传递得以传播。

5 结果与讨论

我们通过三个基准摆起任务的数值仿真验证理论结果：摆、小车倒立摆和 Acrobot。所有仿真均涉及具有控制约束和/或状态约束的系统。每个问题都采用二次代价函数，并使用引理1中的方法进行离散化。各系统的仿真参数列于表1。表2总结了 Box-iLQR 算法的超参数。

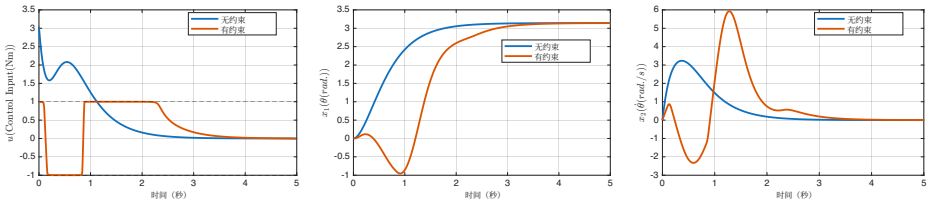
系统	Q	R	Q_f	t_f	Δt	状态约束	控制约束
摆	$3\mathbf{I}_2$	3	$30\mathbf{I}_2$	5	0.01	—	$-1 \le u \le 1$
小车倒立摆	$10\mathbf{I}_4$	10	$10^4\mathbf{I}_4$	10	0.01	$-0.2 \le x_1 \le 0.2$	$-2 \le u \le 2$
Acrobot	$500\mathbf{I}_4$	10	$5 \times 10^4\mathbf{I}_4$	10	0.01	—	$-5 \le u \le 5$

表1：最优控制问题的参数。这里，Q 和 R 分别是二次状态与控制阶段代价的权重矩阵，而 Q_f 是终端状态代价的权重矩阵。

5.1 摆

图1展示了摆锤摆起任务中状态与控制轨迹的对比，比较了无约束控制器与有约束控制器。两种情况下的物理系统参数详见表1，且完全相同；唯一的区别是，有约束情况下的控制力矩限制为± 1 牛顿·米。

蓝色曲线所示的无约束系统采用直接摆起策略。如图1a所示，控制器持续施加幅值较大的正力矩。这使摆锤角度 $x_1(t)$ 单调增加，并平稳收敛至 π rad 的直立位置（图1b）。与此同时，角速度 $x_2(t)$ 收敛至零，表明系统成功稳定在目标状态（图1c）。图2中的轨迹可视化进一步印证了这种直接、单向的运动



（a）控制输入 u（t）。 (b) x_1（t）。 (c) $x_2(t)$。

图1：摆锤摆起任务中无约束控制与有约束控制的比较。图中展示了随时间变化的 (a) 控制输入 $u(t)$、(b) 摆的角度（x_1）$= \theta(t)$ 和 (c) 角速度 $x_2(t) = \dot{\theta}(t)$。无约束情况以蓝色表示，有约束情况以红色表示。

（浅蓝色），显示摆在 $t = 2.22$ s 时到达直立构型。

与此形成鲜明对比的是，受约束控制器的控制能力有限，无法直接实现摆起。因此，系统必须执行能量泵送动作，以积累足够的动量。该策略先沿正 θ 方向小幅摆动，随后朝相反方向大幅摆动，以积累动能。这一行为在状态轨迹中清晰可见（图1b，红色曲线），其驱动力是控制输入饱和于下界（图1a）。图2中 $t = 0.56$ 秒和 $t = 1.11$ 秒时的快照（红色）直观呈现了这一关键的反向摆动。获得足够能量后，控制器再次饱和于其正向极限，驱动摆朝直立位置运动，并在此处随着控制输入减小而逐渐稳定。

这项分析突显了执行器限制所决定的控制策略之间的根本差异。无约束系统可以依靠蛮力直接完成动作，而受约束系统则必须有策略地利用摆的自然动力学，才能实现相同的摆起目标。

5.2 小车倒立摆

如表1所述，我们分别在控制约束和状态约束下分析小车倒立摆系统，约束值分别为 $|u| \leq 2$ 牛顿和 $|x_1| \leq 0.2$ 米。我们在三种不同情形下，将该方法与无约束基线进行性能比较：(a) 仅有控制约束的小车倒立摆；(b) 仅有状态约束的小车倒立摆；(c) 同时具有状态与控制约束的小车倒立摆。

系统采用的不同控制策略如图3所示。无约束控制器施加较大的控制输入，峰值接近 t =2s （约 10 牛顿），以快速摆起。相比之下，存在控制约束时，系统必须采用能量泵送策略。这一点从控制约束情形（图3a，红色）以及完全

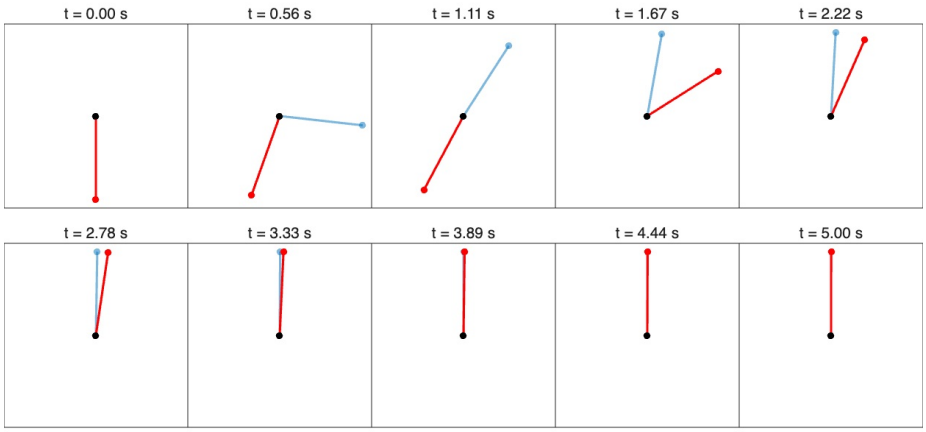
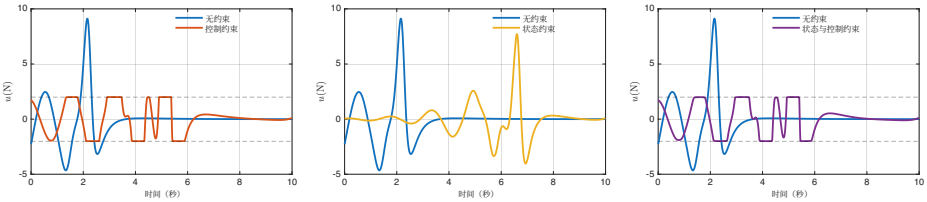


图2：摆杆摆起轨迹的连续快照。无约束控制器产生的运动以浅蓝色显示，而受约束控制器下的轨迹以红色显示。

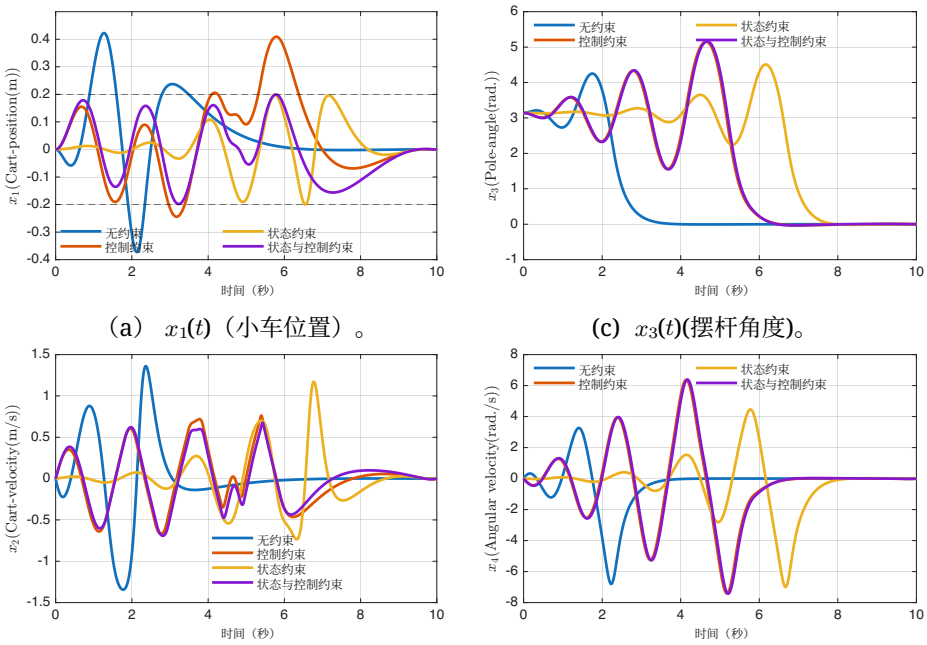
超参数	$\boldsymbol{\mu}_0$		$\boldsymbol{\sigma}_0$	$\epsilon_{\text{barrier}}$	$\mathbf{r}_\mu$	$\mathbf{r}_\sigma$	β_r
值	$1e8\mathbf{I}$	$1e8\mathbf{I}$	0.01	0.5	0.5	(1/0.95)	

表2：Box-iLQR 使用的超参数。



（a）控制约束。 （b）状态约束。 （c）两者均受约束。

图3：小车倒立摆摆起过程中控制输入 $u(t)$ 的比较。每个子图都将无约束控制轨迹（蓝色）与不同的有约束情形进行对比：（a）控制约束，（b）状态约束，以及（c）两者结合。控制约束边界 $-2 \leq u \leq 2$ 在（a）和（c）中以黑色虚线表示。



（a） $x_1(t)$（小车位置）。 (c) $x_3(t)$(摆杆角度)。

（b） $x_2(t)$（小车速度）。 (d) $x_4(t)$(摆杆角速度)。

图4：对比分析无约束条件和三种有约束条件下倒立摆摆起的状态演化。

有约束情形（图3c，紫色）在 ± 2 牛顿的饱和边界之间表现出更多振荡。状态约束情形（图3b，黄色）则呈现出第三种更为谨慎的策略：控制器起初施加较小的输入，仅在操作过程的后期逐渐增大输入，以防止小车违反其位置约束。

图4所示的最终状态轨迹揭示了这些控制策略所产生的后果。正如预期，图4a中的小车位置（x_1）表明，状态约束和全约束情形均遵守位于 ± 0.2 米处的位置边界，分别以黄色和紫色表示。比较控制约束（红色）与全约束（紫色）情形，可以得到一个关键洞见。尽管两者的摆杆动力学（x_3 和 x_4）几乎无法区分（图4c和4d），但在小车的速度剖面（x_2，图4b）中出现了一个细微却至关重要的差异。速度轨迹的这一微小调整，正是全约束系统能够满足位置约束，而纯控制约束系统却无法做到这一点的原因。

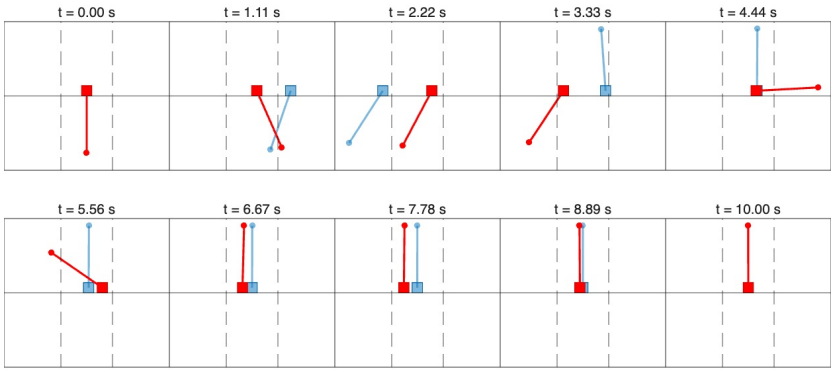
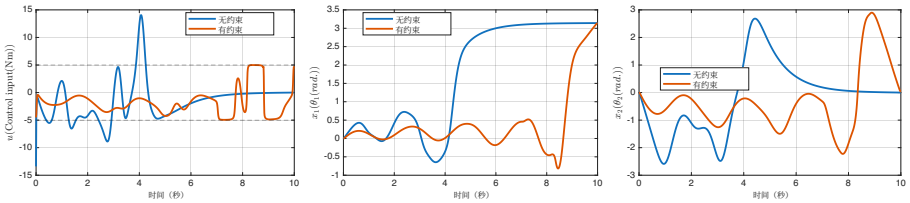


图 5：倒立摆摆起任务的延时可视化，对比无约束轨迹（浅蓝色）与同时满足状态和控制约束的轨迹（红色）。黑色虚线表示小车的位置约束，$-0.2 \leq x_1 \leq 0.2$。

图 5 通过对比无约束和全约束两种情况，进一步以延时可视化的方式展示了这些发现。无约束系统（浅蓝色）在多个时刻（例如 $t = 1.11$ s、2.22 s 和 3.33 s）被显示为违反了状态边界，同时快速完成摆起，并在约 $t = 5.56$ s 时稳定下来。相反，受约束系统（红色）必须通过更复杂的振荡机动来积累动量，并且在观测时间段结束时仍在运动，这展示了系统在受约束情形下的表现。

5.3 Acrobot

图 6 展示了 Acrobot 的状态与控制轨迹，对比了无约束（蓝色）和控制约束（红色）的摆起策略。无约束控制器采用幅度较大、较为激进的策略，控制输入范围约为 -13 牛顿·米至 $+15$ 牛顿·米（图 6a）。这种



(a) $u(t)$. (b) $x_1(t)$. (c) $x_2(t)$。

(d) $x_3(t)$。 (e) $x_4(t)$。

图6：双连杆机器人摆起任务的状态与控制轨迹，对比了无约束（蓝色）和控制约束（红色）两种情况。子图显示了角度 θ_1 (b)和 θ_2 (c)随时间的变化、对应的角速度 $\dot{\theta}_1$ (d)和 $\dot{\theta}_2$ (e)，以及施加的控制力矩 $u(t)$ (a)。控制输入受限于 $u = \pm 5$，如(a)中的虚线所示。

直接的方法能够快速完成摆起，并在约 $t = 6 - 7$ 秒时稳定在直立位置；控制输入和所有状态轨迹均收敛至零，证实了这一点（图 6a–6e）。

相比之下，控制约束系统采用了更为稳健、以能量积累为主的机动方式，类似于摆和小车-倒立摆所采用的策略。系统在较长时间内积累动能，然后执行最后一次强力摆动，以达到直立构型。这一策略反映在状态轨迹中：在机动后期进行最后摆动时，轨迹会出现剧烈而快速的变化。

图 7 中的轨迹快照直观地印证了这两种不同策略。无约束系统（浅蓝色）在约 $t = 3.39$ 秒时开始主要摆动，在 $t = 5.60$ 秒前到达直立位置，并在 $t = 7.80$ 秒时看起来已完全稳定。相比之下，有约束 Acrobot（红色）直到接近 $t = 7.80$ 秒时才开始最后一次决定性的摆动，并且直到 $t = 10$ 秒的最后一帧才达到直立构型，这清楚地展现了在驱动限制下运行所固有的性能权衡。

5.4 反馈项分析

如第4.3小节所述，Box-iLQR生成的反馈律是“考虑约束的”，这意味着随着控制输入接近其饱和边界，反馈增益会逐渐减小。本节使用摆、小车倒立摆和双连杆机器人摆起任务，对这一特性进行数值验证。所有仿真

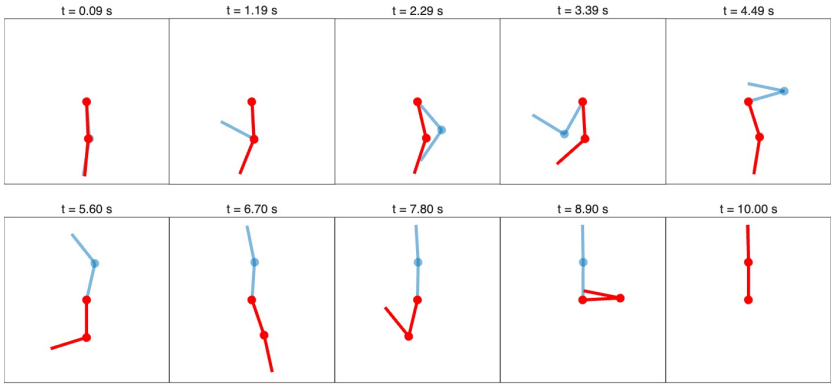


图7：Acrobot摆起轨迹对比，将无约束（浅蓝色）情形与控制约束（红色）情形进行对照。

均在障碍参数设为 $\mu = \sigma = 0.01$ 的情况下运行至收敛（见表1）。正如具有非零障碍参数的内点法所具有的典型特征，控制输入始终严格处于可行集内，虽会接近饱和边界，但永远不会完全达到该边界。

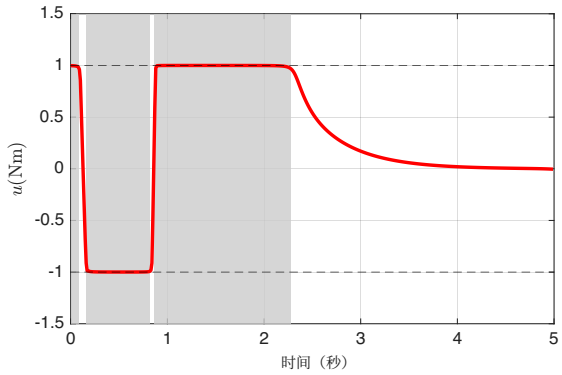
在全部三个系统中，结果与定理2所给出的理论一致。在相应图中，控制输入接近饱和的区域以灰色标出。对于摆（图8），可观察到图8b和图8c中的反馈增益在这些灰色区域内趋近于零。倒立摆系统也表现出类似行为。当控制输入接近其边界时（图9a），反馈增益的全部四个分量 K_1至 K_4均显著减小，如图9b–9e所示。

Acrobot系统在其近饱和阶段（图10a）也呈现出这一趋势。尽管大多数反馈分量明显接近于零（图10b–10e），第四个增益 K_4 （图10e）因图中纵轴缩放而看似非零。然而，其数值仍远小于峰值幅度。这与理论预测一致：对于有限的障碍参数 $\epsilon_{\text{barrier}}$，增益会变得很小，但并不恒等于零。

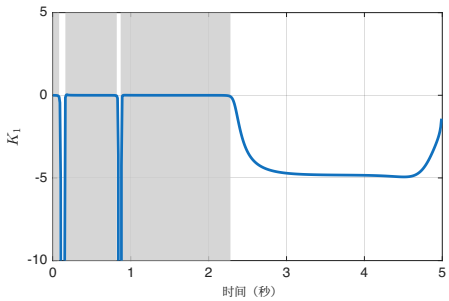
此外，在所有仿真中还一致观察到第二个值得注意的现象：在接近饱和阶段的前后，反馈增益都会出现尖峰。这种行为表明存在一种补偿机制。当控制器预判即将进入或退出因约束而导致其纠正能力受限的阶段时，便会准备施加更大的反馈修正。

最后，考察小车倒立摆在最终饱和阶段（约 $t > 6$ 秒）之后的反馈曲线（图 9b–9e），我们发现，增益在轨迹末端附近出现尖峰之前始终平滑。这种终端尖峰是一种独特现象，由表1中指定的高终端代价所致，

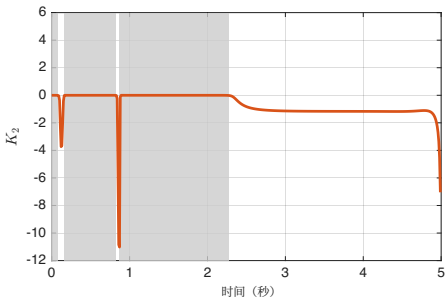
因为控制器正准备进行大幅修正，以确保精确满足终端状态约束。



(a) 控制输入 - $u(t)$。



(b) 反馈增益 - $K_1(t)$。



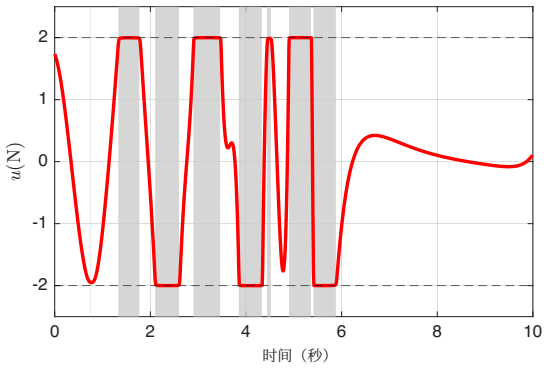
(c) 反馈增益 - $K_2(t)$。

图 8：摆的上摆过程中，控制量及反馈分量 $\mathbf{K}_t^*$ 的曲线。图中的阴影区域表示控制量接近饱和、反馈项趋于零的区域。

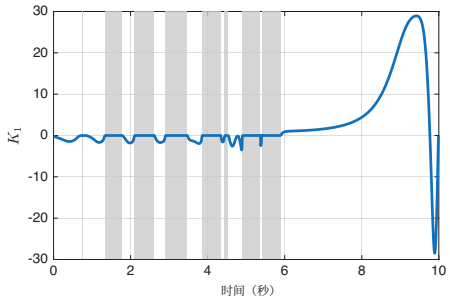
6 结论与未来工作

本文介绍了 Box-iLQR，这是一种将对数障碍内点法融入迭代线性二次调节器框架的算法，用于求解状态和控制量受盒约束的轨迹优化问题。通过系统地减小障碍参数，该方法收敛至最优约束解；我们在多个问题上验证了这一结果。该方法的一个主要优点是，障碍项可作为天然的正则化项，确保正定性并保持凸性。

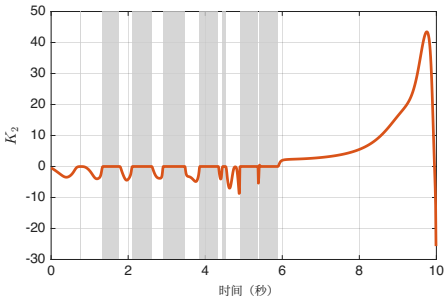
未来工作将沿两个主要方向展开。首先，我们将探索障碍参数的自适应更新策略，通过避免



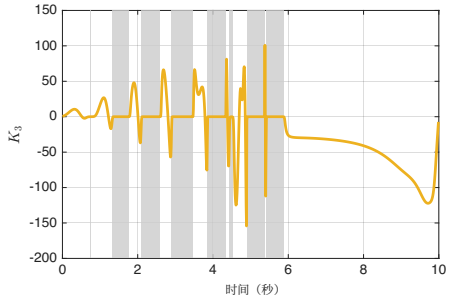
(a) 控制输入 - $u(t)$。



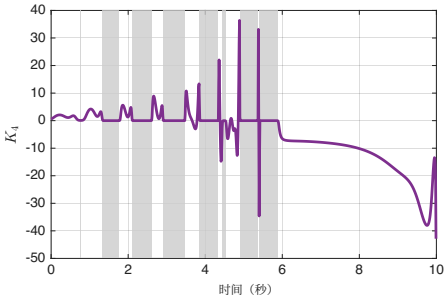
(b) $K_1(t)$。



(c) $K_2(t)$。

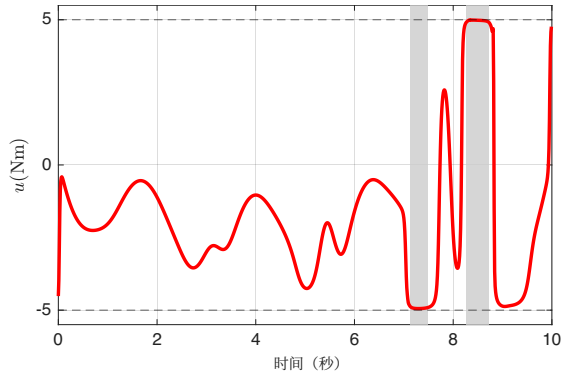


(d) $K_3(t)$。

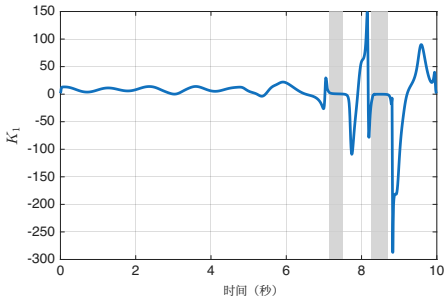


(e) $K_4(t)$。

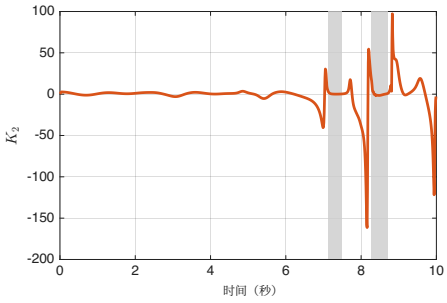
图9：小车倒立摆摆起的控制输入和反馈分量$\mathbf{K}_t^*$图。图中阴影区域表示控制接近饱和、反馈项趋于零的区域。



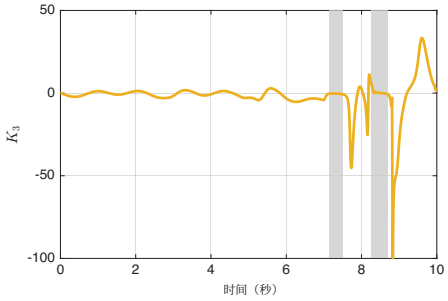
(a) $u(t)$.



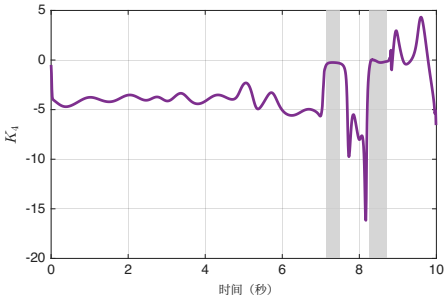
(b) $K_1(t)$.



(c) $K_2(t)$.



(d) $K_3(t)$.



(e) $K_4(t)$.

图 10：双连杆机器人摆起的控制输入和反馈项分量$\mathbf{K}_t^*$的曲线图。图中的阴影区域表示控制输入接近饱和且反馈项趋近于零的范围。

完全求解每个中间子问题的必要性。其次，我们将分析所得“约束感知”反馈律的性质，该反馈律有望在模型预测控制应用中实现更稳健的性能。总体而言，Box-iLQR 朝着面向约束最优控制的快速、可靠且通用的 iLQR 求解器迈出了重要一步。

参考文献

1. Abhijeet, Chakravorty, S.：从序列二次规划视角看最优控制（2025 03475 2 2510），https://arxiv.org/abs/。Abhijeet, Mohamed, M.N.G., Sharma, A., Chakravorty, S.：最优控制问题中的凸性。载于：2024年IEEE第63届决策与控制会议（CDC）。第261页 10886652 3 266 2024 10 1109 56724 2024 –（）。https://doi.org/。/CDC。。。Betts, J.T.：轨迹优化数值方法综述。《制导、控制与动力学杂志》**21**（2），193–207（1998） 4. Betts, J.T.：《使用非线性规划进行最优控制与估计的实用方法》，《设计与控制进展》，第19卷。SIAM，宾夕法尼亚州费城（2010） 5. Bhatia, T., Wang, A., Wensing, P.M.：针对滑移和力可控接触的足式机器人轨迹优化。载于：2021年IEEE机器人与自动化国际会议（ICRA）。第12674–12680页。IEEE（2021） 6. Bryson, Jr., A.E., Ho, Y.C.：《应用最优控制：优化、估计与控制》。Blaisdell Publishing Company，马萨诸塞州沃尔瑟姆（1969） 7. Fiacco, A.V., McCormick, G.P.：《非线性规划：序列无约束最小化技术》。Wiley，纽约（1968），1990年由SIAM重印 8. Giftthaler, M., Buchli, J.：一种用于等式约束迭代线性二次最优控制的投影方法。2017年IEEE-RAS第17届人形机器人国际会议（Humanoids），第61–66页（2017），https://api.semanticscholar.org/CorpusID:8565564 9. Goyal, R., Gul Mohamed, M.N.G., Wang, R., Sharma, A., Chakravorty, S.：基于信息状态的方法用于非线性系统的最优输出反馈控制。载于：2025年美国控制会议（ACC）。第4011–4016页（2025）。https://doi.org/10 11107601 10 23919 63710 2025。/ACC。。。Goyal, R., Naveed Gul Mohamed, M., Wang, R., Sharma, A., Chakravorty, S.：基于信息状态的强化学习用于部分可观测非线性系统的控制。《IEEE神经网络与学习系统汇刊》**36**3593259 11 12 20386 20400 2025 10 1109 2025（），–（）。https://doi.org/。/TNNLS。。。Gul Mohamed, M.N.G., Goyal, R., Chakravorty, S.：无限时域非线性控制问题的最优解。载于：2023年第62届IEEE决策与控制会议（CDC）。第1643–1648页（2023）。https://doi.org/10.1109/CDC49753.2023.10384307 12. Howell, T.A., Jackson, B.E., Manchester, Z.：ALTRO：一种快速的有约束轨迹优化求解器。载于：2019年IEEE/RSJ智能机器人与系统国际会议（IROS）。第7666–7673页。中国澳门（2019）。https://doi.org/10 8967839 13 1109 40897 2019 2019。/IROS。。。Jackson, B.E.：Al-ilqr教程（），https://api.semanticscholar.org/CorpusID:209455095

14. Kalakrishnan, M., Chitta, S., Todorov, E., Righetti, L., Schaal, S.：STORC：用于运动规划的随机轨迹优化。载于：2011年IEEE机器人与自动化国际会议。第5551–5556页。IEEE（2011） 15. Karmarkar, N.：一种用于线性规划的多项式时间算法。《组合学》**4**（4），373–395（1984） 16. Lantoine, G., Russell, R.P.：用于约束最优控制问题的混合微分动态规划算法。第1部分：理论。《优化理论与应用杂志》**154**（2），382–417（2012）。https://doi.org/10.1007/s10957 0 17 012 0039 0, 10 1007 10957 012 0039 - - - https://doi.org/。/s - - -。Lantoine, G., Russell, R.P.：用于约束最优控制问题的混合微分动态规划算法。第2部分：应用。《优化理论与应用杂志》**154**（2），418–442（2012）。https://doi.org/10.1007 1 18 10957 012 0038 1, 10 1007 10957 012 0038 /s - - - https://doi.org/。/s - - -。Lin, T.C., Arora, J.S.：用于约束最优控制的微分动态规划技术。《计算力学》**9**，41–53（1991）。https://doi.org/10.1007/BF00369914 19. Mastalli, C., Havoutis, I., Focchi, M., Caldwell, D.G., Semini, C.：可行性驱动微分动态规划。《IEEE机器人学汇刊》**36**（5），1364–1378（2020） 20. Mayne, D.Q.：确定非线性离散时间系统最优轨迹的二阶梯度法。《国际控制杂志》**3**（1），85–95（1965） 21. Mohamed, M.N.G., Chakravorty, S., Goyal, R., Wang, R.：关于随机最优非线性控制中的反馈律。载于：2022年美国控制会议（ACC）。第970 9867673 22 975 2022 10 23919 53348 2022 –（）。https://doi.org/。/ACC。。。Nocedal, J., Wright, S.J.：《数值优化》。施普林格，纽约，第二版（2006） 23. Ortiz, L.C., Castro, G., Wensing, P.M.：基于DDP的控制器在动态腿式机器人运动中的生存能力与性能。载于：2020年IEEE-RAS第20届人形机器人国际会议（Humanoids）。第186–193页。IEEE（2020） 24. Parunandi, K.S., Sharma, A., Chakravorty, S., Kalathil, D.：D2c 2.0：一种解耦数据驱动方法，通过无模型迭代线性二次调节学习控制随机非线性系统（2020 07368 25 2002），https://arxiv.org/。。Prabhu, S., Rangarajan, S., Kothare, M.：使用内点法处理逐阶段等式与不等式约束的微分动态规划。载于：2025年美国控制会议（ACC）。第2255–2261页（2025）。https://doi.org/10.23919/ACC63710.2025.11108083 26. Rao, C.V., Wright, S.J., Rawlings, J.B.：内点法在模型预测控制中的应用。《优化理论与应用杂志》**99**（3），723 1021711402723 27 757 1998 10 1023 –（）。https://doi.org/。/A:。Schulman, J., Ho, J., Lee, A., Awwal, I., Bradlow, H., Abbeel, P.：通过序贯凸优化寻找局部最优的无碰撞轨迹。载于：《机器人学：科学与系统》。第9卷（2013）。https://doi.org/10.15607/RSS.2013.IX.033 28. Sharma, A., Chakravorty, S.：一种基于降阶模型的强化学习方法，用于控制非线性偏微分方程。《动态系统、测量与控制杂志》**148**（2）（2026年3月）。https://doi.org/10 4070654 29 1115 1。/。Sharma, A., Chakravorty, S.：非线性最优控制的数据驱动建模。载于：2025年美国控制会议（ACC）。第4017–4022页（2025）。https://doi.org/10.23919/ACC63710.2025.11107931

30. Tassa, Y., Erez, T., Todorov, E.：通过在线轨迹优化合成并稳定复杂行为。载于：2012 年 IEEE/RSJ智能机器人与系统国际会议。第 4906–4913 页（2012）。https: // doi.org/10 6386025 31 1109 2012 . /IROS. . . Tassa, Y., Mansard, N., Todorov, E.：控制受限的微分动态规划。载于：2014 年 IEEE国际机器人与自动化会议（ICRA）。第 1168 6907001 32 1175 2014 10 1109 2014 – ( ) 页。https://doi.org/ . /ICRA. . . Todorov, E., Li, W.：一种用于约束非线性随机系统局部最优反馈控制的广义迭代 LQG 方法。载于：2005 年美国控制会议论文集，2005。第 300–306 页，第 1 卷（2005）。https://doi.org/ 10.1109/ACC.2005.1469949 33. Vanroye, L., Sathya, A., De Sc hutter, J., Decré, W.：Fatrop：一种用于机器人轨迹优化与控制的快速有约束最优控制问题求解器。载于：2023 年 IEEE/RSJ智能机器人与系统国际会议（IROS）。第 10036 10342336 34 10043 2023 10 1109 55552 2023 – ( ) 页。https://doi.org/ . /IROS . . . Vanroye, L., Sathya, A., Schutter, J.D., Decré, W.：Fatrop：一种用于机器人轨迹优化与控制的快速有约束最优控制问题求解器（2023），https://arxiv.org/abs/2303 16746 35 . . Wächter, A., Biegler, L.T.：关于用于大规模非线性规划的原始-对偶内点滤子线搜索算法的实现。《数学规划》**106**(1), 25–57 (2006) 36. Waltz, R.A., Morales, J.L., Nocedal, J., Orban, D.：一种结合线搜索与信赖域步长的非线性优化内点算法。《数学规划》**107**(3), 391–408 (2006)。https://doi.org/10.1007/ s10107-004-0560-5 37. Wang, A., Wensing, P.M.：用于快速、稳健机器人机动的基于腿式的 S-CON-MIP。载于：2022 年国际机器人与自动化会议（ICRA）。第 273–279 页。IEEE（2022）38. Wang, R., Sharma, A., Parunandi, K.S., Goyal, R., Mohamed, M.N.G., Chakravorty, S.：强化学习中的反馈探索。《动态系统、测量与控制杂志》**147**(6), 061002（2025 年 6 月）。https: //doi.org/10 4068705 39 1115 1 4068705, 10 1115 1 . / . https://doi.org/ . / . . Wright, S.J.：约束优化的内点法。Acta Numerica **1**, 341–407（1992）40. Wright, S.J.：《原始-对偶内点法》。SIAM，宾夕法尼亚州费城（1997）41. Xie, Z., Liu, C.K., Hauser, K.K.：带非线性约束的微分动态规划。2017 年 IEEE国际机器人与自动化会议（ICRA），第 695–702 页（2017），https://api.semanticscholar.org/CorpusID: 7648776 42. Zucker, M., Ratliff, N., Dragan, A.D., Pivtoraiko, M., Kuffner, J.J., Dellin, C.M., Srinivasa, S.S.：Chomp：用于高效运动规划的梯度优化技术。《国际机器人研究杂志》**32**(9-10), 1145–1167（2013）