arXiv:2304.00375v1 [math.OC] 1 Apr 2023

无限时域非线性控制问题的最优解

Mohamed Naveed Gul Mohamed、Raman Goyal、Suman Chakravorty

摘要——**本文研究非线性系统的无限时域最优控制问题。在原点附近线性化系统满足可控性，且系统具有到达包含原点的终端集的非线性可控性的条件下，我们建立了一种近似正则化求解方法。该方法由一个到达终端集的“有限自由终端时间”最优转移问题，以及终端集内的一个无限时域线性调节问题组成；后者可使原点全局渐近稳定。此外，我们证明，随着终端集的规模趋于零，这些近似解会收敛至真实最优代价函数。我们在摆和小车倒立摆摆起问题上对该方法进行了实证评估，结果表明，与不采用所提出正则化方法求解无限时域问题所需的有效时域相比，有限时间转移所需时间短得多。**

索引词——**非线性控制、无限时域最优控制、控制李雅普诺夫函数**

I. 引言

最优控制问题的目标是在满足系统动力学约束的条件下，找到使给定代价函数最小化的控制输入。此外，还希望最优控制问题能够得到全局渐近稳定的闭环系统。为满足全局渐近稳定性要求，必须构造一个无限时域控制问题，即代价为无穷和的问题。尽管针对特定系统（大多为线性系统）存在无限时域最优控制律，但由于无限时域 [1]，求解无限时域非线性控制问题的最优解是一项极具挑战性的任务。本文提出一种近似解，将该问题转化为等价的有限时域问题。具体而言，我们注意到，若无限时域代价有限，则代价的尾和必须趋于零，这意味着系统在无限时间内的大部分时间都必须处于原点附近。因此，假设系统线性化具有可控性，我们可以用求解定常黎卡提方程得到的最优线性代价函数，近似表示经过某个有限时间后的尾部代价。由此，无限时域问题便转化为：寻找进入包含原点的最优线性代价函数某一水平集的最优插入时间及相应的控制输入。随着这些水平集的规模缩小，所得到的近似解已证明会收敛到真实最优代价函数。此外，尽管对于有限水平集，该解可能不是最优解，但它仍能提供一个控制李雅普诺夫函数；在一个较弱的非线性可控性假设下，即任意状态都可以被控制到一个包含原点的终端集内，该函数能够使底层系统全局渐近稳定。

无限时域最优控制问题可以等价地表述为离散时间系统的定常动态规划（DP）问题，或连续时间系统的定常 Hamilton-Jacobi-Bellman（HJB）方程 [1, 2]；只要已知最优代价函数，就能将序贯决策问题转化为单步决策问题。众所周知，定常 HJB 方程的最优解是一个能够实现全局渐近稳定的控制李雅普诺夫函数 [3, 4, 5]。然而，众所周知，由于 Bellman 所说的可怕“维数灾难” [1, 2]，求解该问题在计算上难以实现。因此，大量研究致力于近似动态规划（ADP）和强化学习，以缓解维数灾难。近似动态规划方法 [6, 7] 通常采用聚合方法 [8]–[9]或各种函数逼近方法 [10] [11]，以高置信度获得近似最优的策略/价值函数。强化学习 [12] 和近似动态规划通常被视为同义词，不过“强化学习”这一术语往往与无模型方法相关：这类方法不具备模型的“显式”知识，而是通过与环境反复交互、观察系统的响应来改进控制策略。反复交互，即学习试验，使这些算法无需显式构建未知动力系统的模型，便能求得动态规划问题的解（最优价值/Q值函数或最优策略）。标准强化学习算法大体分为基于价值的方法（如 Q学习 [13],）和基于策略的方法（如策略梯度算法 [14]）。近年来，利用深度神经网络进行函数逼近，显著提升了强化学习算法的性能，推动了“深度强化学习” [15, 16, 17, 18]相关研究的不断发展。尽管这些方法已成功应用于维度相对以往更高的问题，但其解固有的方差 [19, 20, 21] 使其可靠性不足，而且所需训练时间仍然难以承受。

另一种求解 HJB 的“直接”方法，是针对给定的初始状态求解其 underlying 无限时域最优控制问题

作者来自德克萨斯农工大学航空航天工程系，美国德克萨斯州大学城，邮编 77843。 {naveed, ramaniitrgoyal92, schakrav}@tamu.edu

。模型预测控制（MPC）领域采用这种方法来求解无限时域问题；然而，由于所涉及的最优控制问题具有无限时域性质，MPC 转而求解“固定终止时间”的有限时域问题，采取第一个控制动作，并在下一个状态处重复这一过程 [22, 23]。MPC 方法实现渐近稳定性的“稳定化要素”，是在优化问题中使用合适的终端代价函数；该函数在某个包含原点的终端集内是系统的控制李雅普诺夫函数 [22]。然而，MPC控制律的吸引域可能小得不理想，因此人们提出了不同的方法来扩大吸引域 [24, 25, 26]。另一种方法是不使用终端代价函数，而是通过采用合适的长时域来设定 [27]，但由于预测时域过长，这通常会导致问题难以处理[28]。我们求解无限时域问题的方法与 MPC 类似，即“直接”求解最优控制问题；但关键区别在于，对于给定状态，我们求解一个“自由终止时间”问题，使系统进入终端集。在该终端集中，最优线性反馈律具有渐近稳定性，且/或能够较好地逼近真实最优代价函数。随后，在温和的非线性可控性假设下，我们可以证明所得反馈律的全局渐近稳定性；还可以证明，当终端集的大小减小至零时，该近似将收敛于真实最优代价函数。最后，该方法无需重新规划。其局限在于，我们未在问题中考虑状态约束或控制约束。

本文的主要贡献是提出一种易于处理的直接方法，用于求解无限时域最优控制问题。在系统原点附近的线性化系统满足可控性，以及满足关于进入包含原点的终端集的非线性可控性假设的条件下，该方法能够使非线性系统实现全局渐近稳定。本文其余部分的组织如下：第二节介绍问题，第三节详细阐述求解方法，包括构造背后的启发式思想，第四节在若干非线性系统上对该方法进行实证测试。

II. 预备知识

考虑下面我们希望求解的问题：

$$J^*_\infty(x) = \min_{\{u_t\}} \sum_{t=0}^{\infty} c(x_t, u_t); \quad \text{given } x_0 = x \tag{1a}$$

$$\text{subject to the dynamics: } x_{t+1} = f(x_t, u_t), \tag{1b}$$

其中，$x_t \in \mathbb{R}^n$ 表示动力系统的状态，$u_t \in \mathbb{R}^p$ 表示动力系统的控制输入，而 $c(x_t, u_t)$表示在状态 x_t下采取控制动作 u_t 所产生的增量代价。上述问题是一个无限时域最优控制问题（IH-OCP），因此，由于问题具有无限时域，求解该问题通常难以实现。

本研究旨在通过将上述问题转化为合适的有限时域问题，提出一种可行的方法来求解该问题。

鉴于我们可以获得 IH-OCP 式 (1) 的解，众所周知，无限时域代价到达函数 $J_\infty(\cdot)$ 满足贝尔曼方程 [1, Ch.7]:

$$J^*_\infty(x) = \min_u \{c(x,u) + J^*_\infty(f(x,u))\}. \tag{2}$$

鉴于 $c(x,u) > 0 \ \forall (x,u) \neq (0,0)$，可以看出，$J^*_\infty(\cdot)$ 是式 (1b) 中动力系统的控制李雅普诺夫函数（CLF），因此，由最优代价到达函数 $J^*_\infty(\cdot)$ 隐式定义的控制反馈策略能够使式 (1b) 中的动力系统实现全局渐近稳定。证明是直接的，如下所示。

推论*1*：若 $J^*_\infty(x)$ 满足贝尔曼方程（式(2)），则它是式(1b)中系统的控制李雅普诺夫函数，可使原点全局渐近稳定。

证明：令 u^*_t 为根据贝尔曼方程在状态 x_t 下的控制动作，其中 (x_t, u^*_t) 是给定 $x_0 = x$ 时的系统轨迹。于是：

$$\begin{aligned} & J^*_\infty(x_t) = c(x_t, u^*_t) + J^*_\infty(f(x_t, u^*_t)). \\ \Rightarrow\ & J^*_\infty(x_t) > J^*_\infty(x_{t+1}) \ \ (\text{since } c(x_t, u^*_t) > 0). \\ \Rightarrow\ & J^*_\infty(x_t) \to 0 \text{ as } t \to \infty \ \ , \end{aligned}$$

其中，最后一行由 $J^*_\infty(x_t)$ 下界为零这一事实得出，因此，序列 $\{J^*_\infty(x_t)\}$ 严格单调，意味着该序列必收敛至零。这对任意 x 均成立。由于 $J^*_\infty(x') = 0$ 仅在原点处成立，这意味着轨迹 $x_t \to 0$ 在 $t \to \infty$ 时趋于原点，结论得证。 ■

此外，假设存在一个 $J_\infty(\cdot)$ 满足贝尔曼方程（不一定是最优解）

$$J_\infty(x) = \min_u \{c(x,u) + J_\infty(f(x,u))\}, \tag{3}$$

则 $J_\infty(\cdot)$也是一个控制李雅普诺夫函数，可使原点全局渐近稳定(GAS)。

因此，我们求解式(1)的另一个目标是构造如式(2)/式(3)所示的控制李雅普诺夫函数，使原点达到全局渐近稳定。

III. 无限时域最优控制问题的解

定义以下有限时域最优控制问题(FH-OCP)：

$$J^T_\infty(x) = \min_{\{u_t\}} \sum_{t=0}^{T-1} c(x_t, u_t) + \bar{J}_\infty(x_T), \tag{4}$$

$$\text{subject to: } x_{t+1} = f(x_t, u_t)$$

其中，$\bar{J}_\infty(\cdot)$是一个将在下文定义的终端代价函数。

本文余下部分将采用以下假设。

假设*1*：我们假设代价函数 $c(x,\ u)$ 在 $(x,\ u)=(0,0)$ 处取得全局最小值，即 $\left.\frac{\partial c}{\partial x}\right|_{x=0,u=0}=0$ 且 $\left.\frac{\partial c}{\partial u}\right|_{x=0,u=0}>0$，$\forall\ x\ u\ \neq\ 0,0$，$(\ )$ 和 $(,\ )$ 是 $(\ )$。

假设*2*：我们假设，对于任意 $x\in\mathcal{X}$ 和任意包含原点的 $\Omega\subset\mathcal{X}$，都存在一个控制序列 $\exists_{\{u_t\}^{T(x)}t=0}$，在上述动力学（式（1b））下，可确保在某个 $T(x)<\infty$ 时进入 $x_{T(x)}\in\Omega$。

假设2是一个可控性假设，它确保任何状态都能在有限时间内通过控制进入区域 Ω。

假设*3*：我们假设，动力系统式（1b）在 (0,0) 附近的线性化是可控的。

根据假设1和假设3，我们可以定义最优“线性”无限时域问题：

$$\bar{J}_\infty(x)=\min_{\{u_t\}}\sum_{t=0}^{\infty}(x_t'Qx_t+u_t'Ru_t),\tag{5a}$$

$$\text{subject to: } x_{t+1}=Ax_t+Bu_t,\tag{5b}$$

其中，(Q,R) 和 (A,B) 分别通过对 $c(x,u)$ 进行二次展开，以及在原点 $(x,u)=(0,0)$ 附近对式（1b）中的动力学进行线性展开得到。

注意，由于上述线性可控性假设3，$J_\infty(\cdot)$可通过求解定常代数黎卡提方程得到，结果为

$$\bar{J}_\infty(x)=x'P_\infty x,\tag{6}$$

其中，P_∞ 是定常黎卡提方程的解。

定义*1*：上述有限时域最优控制问题（方程（4））中的终端代价 $J_\infty(\cdot)$定义为 $J_\infty(x)=x'P_\infty x$，其中 P_∞ 是定常黎卡提方程的解。

A. 启发式思想

无限时域最优控制问题的解背后的启发式思想如下。设最优轨迹为$(x_t^*,\ u_t^*)$，其起始状态为 x。为了使最优代价定义良好，需要满足 $\sum_{t=0}^{\infty}c(x_t^*,\ u_t^*)<\infty$。然而，要使这个无限和定义良好，尾和 $\sum_{t>T}^{\infty}c(x_t^*,\ u_t^*)\to 0$ 趋于$T\to\infty$。因此，经过有限时间后，成本 $c(x_t^*,\ u_t^*)$ 必然很小。鉴于根据假设1，成本 $c(x,\ u)$仅在原点处 $=0$，可知系统会在原点附近停留很长时间。接下来，根据假设3，原点附近的线性化是可控的，因此，成本的尾和可由通过求解定常黎卡提方程得到的最优线性代价函数很好地近似，如上文所定义的 J_∞。因此，无限时域代价可以分为两部分：“转移成本”，即使系统进入线性近似成立的区域Ω；以及线性调节成本，即系统进入Ω后将其调节至原点的成本（见图1）。由此，基本思想是将无限时域问题

通过近似尾和将其转化为有限时域问题

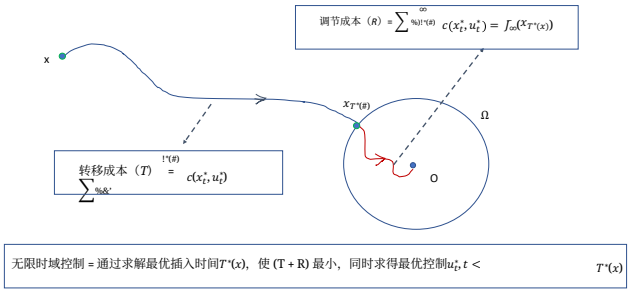


图 1：示意图，说明求解无限时域最优控制问题的策略

用最优线性代价函数近似代价，并从状态 x 找到从状态 x 插入区域 Ω 的最优插入时间 $T^*(x)$ 以及相关的最优控制。随着区域 Ω 变小，我们会得到更好的最优代价函数近似，并在极限情况下获得最优解。然而，即使对于有限的 Ω，该过程也能构造出一个使原点全局渐近稳定的控制李雅普诺夫函数。事实上，根据我们对增量代价函数 $c(x,\ u)$ 的选择，可以得到不可数多个这样的控制李雅普诺夫函数。

B. 有限时域最优控制问题的最优性

接下来，我们将证明有限时域最优控制问题方程（4）的最优性，特别是对所有$x\in\mathcal{X}$，$J_\infty^T(x)\to J_\infty^*(x)$。

为此，定义集合：

$\Omega_\varepsilon=\{x:\ |J_\infty(x)-J_\infty^*(x)|<\varepsilon$，并且 $J_\infty(x)=x'P_\infty x\leq M_\varepsilon\}$，即，$M_\epsilon$ 的选取使得在水平集 Ω_ε中，线性最优代价函数与真实最优代价函数之间的误差小于容差 ϵ。

上述集合定义了最优线性代价函数的一个次水平集，在该集合中，代价函数 ε–接近真实（未知）最优值 $J_\infty^*(x)$。

为简单起见，以下我们将代价 $c(x,\ u)$ 视为二次型。

引理*1*：存在有限时间 $T_\varepsilon<\infty$，使得对于 $T=T_\varepsilon$，有限时域最优控制问题方程（4）的解，记为 $(x_t^\varepsilon,\ u_t^\varepsilon)$，满足 $\bar{J}_\infty(x_T)=x_T^{\varepsilon'}P_\infty x_T^\varepsilon\leq M_\varepsilon$ 首次成立，即对于任意 $T<T_\varepsilon$，$\bar{J}_\infty(x_T)=x_T^{\varepsilon'}P_\infty x_T^\varepsilon>M_\varepsilon$。

证明：若 T_ε 有限，则根据定义，$J_\infty(x_{T_\varepsilon}^\varepsilon)\leq M_\varepsilon$首次在 T_ε 处成立。由于假设2，存在一个控制序列 $\{u_t\}_{t=0}^{\bar{T}}$，使得经过有限的 时间 T后，$\bar{x}_{\bar{T}}\in\Omega_\varepsilon$。将此成本记为 J。由于方程（4）的解对于任意 T都不会进入Ω_ε，且成本 $c(x,u)\geq\Delta>0$，若 $x\notin\Omega_\varepsilon$（由假设1可知），则 $J_\infty^T\to\infty$，其中 $T\to\infty$。然而，根据上面的轨迹 $(\bar{x}_t,\bar{u}_t)$：$\bar{J}(x)=\sum_{t=0}^{\bar{T}-1}c(\bar{x}_t,\bar{u}_t)+\sum_{t\geq\bar{T}}c(\bar{x}_t,\bar{u}_t)$，其中 $\bar{T}$是轨迹到达Ω_ε 的时刻，此后 $\bar{u}_t$ 遵循线性控制。选择一个 T，使得 $\forall\ T>T'$，$J^T(x)>\bar{J}(x)+\varepsilon$（这总是可行的，因为 $J^T(x)\to\infty$）。现在，选择一个 T，使得 $T>T$且 $T>T'$。这意味着：$J_\infty^T(x)=\sum_{t=0}^{T-1}c(x_t,u_t)+\bar{J}_\infty(x_T)>\sum_{t=0}^{T'-1}c(\bar{x}_t,\bar{u}_t)+$

$\sum_{t\geq T'} c\ \bar{x}_t\ \bar{u}_t\ + \varepsilon \qquad \sum_{t\geq T} c\ \bar{x}_t\ \bar{u}_t\ > J^*_\infty\ \bar{x}_T\ \geq$（，）。但是（，）（）$J_\infty(\bar{x}_T) - \varepsilon.$

$\Rightarrow J^T_\infty(x) = \sum_{t=0}^{T-1} c(x_t,\ u_t) + \bar{J}_\infty(x_T) > \sum_{t=0}^{T-1} c(\bar{x}_t,\ \bar{u}_t) + \bar{J}_\infty(\bar{x}_T)$，这与 $J^T_\infty(x)$ 是时间 T 的最优代价这一事实相矛盾。因此，存在 $\exists$ 一个有限时间 $T_\varepsilon < \infty$，使得根据方程（4）的解得到的轨迹到达集合 Ω_ε.

■

上述证明的基本思路是：如果最优轨迹从 x 出发后始终未进入 Ω_ε，$J^T \to \infty$然而，由于存在一条满足 $J < \infty$，的可行轨迹，我们总能找到一个合适的有限时间 T，在该时间内¯最优控制的成本高于 J，从而与该时间 T 的最优性相矛盾。

引理*2*：给定解首次到达集合 Ω_ε 的时间 T_ε，最优解 T_ε 满足：$|J^{T_\varepsilon}_\infty(x) - J^*_\infty(x)| \leq \varepsilon$。

证明：令 (x^*_t, u^*_t) 表示最优轨迹，并令 T^*_ε 为其到达 Ω_ε 的时间。有三种可能：$T^*_\varepsilon > T_\varepsilon$, $T^*_\varepsilon < T_\varepsilon$，以及 $T^*_\varepsilon = T_\varepsilon$。

(a) $T^*_\varepsilon > T_\varepsilon$: 令 $J^*_\infty(x) = \sum_{t=0}^{T^*_\varepsilon-1} c(x^*_t, u^*_t) + J^*_\infty(x^*_{T^*_\varepsilon})$。$J^{T_\varepsilon}_\infty(x) = \sum_{t=0}^{T_\varepsilon-1} c(x^\varepsilon_t, u^\varepsilon_t) + \bar{J}_\infty(x^\varepsilon_{T_\varepsilon})$，其中 $(x^\varepsilon_t, u^\varepsilon_t)$ 是求解方程（4）所得、时长为 T_ε 的最优轨迹；最后令 $\bar{J}^*_\infty(x) = \sum_{t=0}^{T^*_\varepsilon-1} c(x^*_t, u^*_t) + \bar{J}^*_\infty(x^*_{T^*_\varepsilon})$。

但由于 Ω_ε，的定义 $J^*(x) - J^{T_\varepsilon}(x) \leq \varepsilon$. 但 $\bar{J}^*_\infty(x) = \sum_{t=0}^{T^*_\varepsilon-1} c(x^*_t,\ u^*_t) + \sum_{t>T_\varepsilon}^{T^*_\varepsilon} c(x^*_t,\ u^*_t) + \bar{J}_\infty(x^*_{T^*_\varepsilon})$。注意到 $x^*_{T^*_\varepsilon}$和 $x^\varepsilon_{T_\varepsilon}$ 都位于 $\partial\Omega_\varepsilon$（集合 Ω_ε 的边界）上，并且注意到 $J^{T_\varepsilon}_\infty(x)$ 是方程（4）中 T_ε 的最优解，因此可得 $J^{T_\varepsilon}_\infty(x) < \bar{J}^*_\infty(x) \Rightarrow |J^{T_\varepsilon}_\infty(x) - \bar{J}^*_\infty(x)| < \varepsilon.$

(b) $T^*_\varepsilon < T_\varepsilon$: 令 T^*_ε 的有限时域最优控制问题的解为 $(\tilde{x}_t,\ \tilde{u}_t)$，并定义：

$$J^*_\infty(x) = \sum_{t=0}^{T^*_\varepsilon-1} c(x^*_t, u^*_t) + J^*_\infty(x^*_{T^*_\varepsilon}).$$

$$J^{T_\varepsilon}_\infty(x) = \sum_{t=0}^{T^*_\varepsilon-1} c(x^\varepsilon_t, u^\varepsilon_t) + \sum_{t\geq T^*_\varepsilon}^{T_\varepsilon-1} c(x^\varepsilon_t, u^\varepsilon_t) + \bar{J}_\infty(x^\varepsilon_{T_\varepsilon}).$$

$$J^{T^*_\varepsilon}_\infty(x) = \sum_{t=0}^{T^*_\varepsilon-1} c(\tilde{x}_t, \tilde{u}_t) + \bar{J}_\infty(\tilde{x}_{T^*_\varepsilon}).$$

$$\bar{J}^*_\infty(x) = \sum_{t=0}^{T^*_\varepsilon-1} c(x^*_t, u^*_t) + \bar{J}_\infty(x^*_{T^*_\varepsilon})$$

由于 Ω_ε 的定义：

$$|\bar{J}^*_\infty(x) - J^*_\infty(x)| < \varepsilon \tag{7}$$

（注意 $x^*_{T^*_\varepsilon} \in \partial\Omega_\varepsilon$）。但 $J^{T^*_\varepsilon}_\infty(x)$ 是 $T^*_\varepsilon \Rightarrow$ 下有限时域最优控制问题方程（4）的最优解

$$J^{T^*_\varepsilon}_\infty(x) < \bar{J}^*_\infty(x). \tag{8}$$

最后：

$$J^{T^*_\varepsilon}_\infty(x) = \sum_{t=0}^{T_\varepsilon-1} c(\tilde{x}_t, \tilde{u}_t) + \bar{J}_\infty(\tilde{x}_{T_\varepsilon}), \tag{9}$$

其中 $\tilde{u}_t = u_l(\tilde{x}_t)\ \forall\ t > T^*_\varepsilon$，其中 $u_l(\cdot)$ 表示根据 $\bar{J}_\infty(\cdot)$ 得到的线性控制。将其与 $J^{T_\varepsilon}_\infty(x)$ 进行比较，并注意到 $J^{T_\varepsilon}_\infty(x)$ 对于 $T_\varepsilon \Rightarrow$ 是最优的

$$J^{T_\varepsilon}_\infty(x) \leq J^{T^*_\varepsilon}_\infty(x). \tag{10}$$

结合方程（7）至方程（10），$\Rightarrow |J^*_\infty - J^{T_\varepsilon}_\infty(x)| \leq \varepsilon$。

（c）$T_\varepsilon = T^*_\varepsilon$ 的情形显然成立，因此结论得证。

■

推论 *2*：令 $T' < T$，则 T 下方程（4）的解，T'如下：$J^{T'}_\infty(x) \geq J^T_\infty(x)$。证明可由上述引理 2 证明的最后一部分得出。

定理 *1*：给定任意 $x \in \mathcal{X}$，$\lim_{T\to\infty} J^T_\infty(x) \to J^*_\infty(x)$。

证明：我们刚刚在引理 2 中证明，对于任意 $\varepsilon > 0$，都可以找到 $T_\varepsilon(x) < \infty$，使得：

$$|J^{T_\varepsilon}_\infty(x) - J^*_\infty(x)| \leq \varepsilon. \tag{11}$$

此外，根据推论 2，对于任意 $T > T_\varepsilon$，$J^T_\infty(x) \leq J^{T_\varepsilon}_\infty(x)$。注意到 $J^{T_\varepsilon}_\infty(x) > J^*_\infty(x)$，由此可得 $|J^T_\infty(x) - J^*_\infty(x)| \leq \varepsilon\ \forall\ T > T_\varepsilon(x)$，这证明了对于任意 $x \in \mathcal{X}$，$J^T_\infty(x) \to J^*_\infty(x)$。

■

C. 一种替代构造

下面我们定义一种次优构造，尽管如此，它仍能为动力系统式 (1b) 提供一个控制李雅普诺夫函数（CLF），使原点全局渐近稳定。

设线性¯控制器对应终端成本函数 J_∞ 的吸引域（ROA）为 $\mathcal{D}$。假设我们定义集合 $\Omega \subset \mathcal{D}$ 如下：

$$\Omega = \{x : \bar{J}_\infty(x) = x' P_\infty x \leq M\},$$

¯它是受限于 ROA$\mathcal{D}$ 内的 $J_\infty(\cdot)$ 的一个水平集。

固定 $M < \infty$，并考虑以下问题：

$$J_\infty(x) = \min_{u_t, T} \sum_{t=0}^{T-1} c(x_t, u_t) + \max\left(\bar{J}_\infty(x_T), M\right),\ x \notin \Omega,$$
$$J_\infty(x) = \bar{J}_\infty(x) = x' P_\infty x,\ x \in \Omega. \tag{12}$$

请注意，上述优化问题包含一个自由终端时间 T，需要与控制动作 u_t 一起进行优化。提出上述问题的理由是，最优控制会将系统从任意状态 x 转移到集合 Ω 的边界 $\partial\Omega$；从该处开始，线性控制器接管并将系统调节至原点。由于 Ω 位于吸引域 $\mathcal{D}$ 内，线性控制器无论从边界上的哪个点接管，都能将系统调节至原点。

在下一个结果中，我们将证明，对于上述问题，最优时间就是该问题的解首次到达集合 Ω 边界的时间。

引理 *3*：最优控制问题（方程（12））的最优时间 $T^*(\Omega)$ 有限，且是解首次到达集合Ω的时间 T。

证明：注意，只要 $x_T \notin \Omega$，方程（12）就与方程（4）完全相同。因此，引理1表明，Ω 的首次到达时间（记为 $T^*(\Omega)$）是有限的。此外，推论 2 表明，对于任意 $T > T(\Omega)$，$J_\infty^T(x) < J_\infty^{T(\Omega)}(x)$，其中：

$$J_\infty^T(x) = \min_{u_t} \sum_{t=0}^{T-1} c(x_t, u_t) + \bar{J}_\infty(x_T). \tag{13}$$

设 $T > T(\Omega)$，并设根据式 (12) 中的最优控制问题得到的最优轨迹为 $\{\tilde{x}_t, \tilde{u}_t\}$，对应的时间为 T。那么：

$$J_\infty^T(x) = \sum_{t=0}^{T-1} c(\tilde{x}_t, \tilde{u}_t) + \max\left(\bar{J}_\infty(\tilde{x}_T), M\right). \tag{14}$$

由于 $T > T(\Omega)$，$J_\infty(\tilde{x}_T) < M$，因此，$\max(J_\infty(\tilde{x}_T), M) = M$。注意到 $T(\Omega)$ 是式 (12) 的解首次到达 $\partial\Omega$ 的时刻，$J_\infty^T(x) > J_\infty^{T(\Omega)}(x)$，结论得证。 ■

接下来，我们证明由此构造的次优代价函数满足贝尔曼方程。

引理 *4*：与最优控制问题式 (12) 对应的最优代价 $J_\infty(x)$ 对任意 $x \notin \Omega$ 均满足贝尔曼方程：

$$J_\infty(x) = \min_u [c(x,u) + J_\infty(f(x,u))].$$

证明：将最优控制问题式 (12) 中的终端代价函数记为 $\Phi(x_T) = \max\left(\bar{J}_\infty(x_T), M\right)$。那么对任意 $x \notin \Omega$：

$$\begin{aligned} J_\infty(x) &= \min_{u_t, T} \left[\sum_{t=0}^{T-1} c(x_t, u_t) + \Phi(x_T) \right], \qquad (15) \\ &= \min_{u_0} \min_{\{u_t\}_{t=1}^{T-1}, T} \left[c(x, u_0) + \sum_{t=1}^{T-1} c(x_t, u_t) + \Phi(x_T) \right], \\ &= \min_{u_0} \left[c(x, u_0) + \min_{\{u_t\}_{t=1}^{T-1}, T} \sum_{t=1}^{T-1} \left[c(x_t, u_t) + \Phi(x_T) \right] \right]. \end{aligned}$$

注意到 $x_1 = f(x, u_0)$，并且

$$J_\infty(f(x,u_0)) = \min_{\{u_t\}_{t=1}^{T}, T} \sum_{t=1}^{T-1} \left[c(x_t, u_t) + \Phi(x_T) \right],$$

由此可得：

$$J_\infty(x) = \min_u [c(x,u) + J_\infty(f(x,u))],$$

从而证明了该结果。 ■

注意，$J_\infty(\cdot)$ 满足定常贝尔曼方程的原因在于问题构造中采用了自由终端时间 T；否则，该总和将随时间变化。又由于 $J_\infty(\cdot)$ 满足贝尔曼方程，因此有：

推论 *3*：最优控制问题（OCP）式 (12) 的最优代价函数 $J_\infty(\cdot)$ 是一个控制李雅普诺夫函数，可使系统式 (1b) 的原点全局渐近稳定。

证明：显然，对于任意 $x \notin \Omega$，由于 $J_\infty(.)$ 在 Ω 外满足贝尔曼方程，沿着最优控制运动的轨迹必然会到达 Ω。现在，由于 Ω 位于线性系统的吸引域 $\mathcal{D}$ 内

控制器，J_∞ 是系统关于状态 $x_t \to 0$ 的控制李雅普诺夫函数，其中 $t \to \infty$。因此，最优代价函数 $J_\infty(.)$ 是系统的控制李雅普诺夫函数。 ■

这里注意，所定义的控制李雅普诺夫函数 $J_\infty(。)$ 是连续的，但不一定光滑；尤其是，对于任意有限的 M，在 Ω 的边界处存在连续但非光滑的过渡。

此外，还可以建立以下收敛结果。给定 $M < \infty$，定义：

$$J_\infty^M(x) = \min_{u_t, T} \left[\sum_{t=0}^{T-1} c(x_t, u_t) + \max\left(\bar{J}_\infty(x_T), M\right) \right], \tag{16}$$

其中回顾一下，$\Omega = \{x : J_\infty(x) = x' P_\infty x \le M\}$。

定理 *2*：给定最优控制问题（式 (16)），其解满足：$\lim_{M\to 0} J_\infty^M(x) = J_\infty^*(x)$，对所有 $x \in \mathcal{X}$ 均成立。

证明：证明与引理 2 和定理 1 的证明基本相同。 ■

最后，注意对于任意有限的 $M < \infty$，$J_\infty^M(\cdot) \neq J_\infty^*(\cdot)$。然而，$J_\infty^M(\cdot)$ 仍满足贝尔曼方程；这是一个非平凡情形，在此情形下，满足贝尔曼方程并不足以保证最优性。

D. 讨论

接下来，我们讨论所提出的正则化方法的几个关键要点。

计算效率：随着时间范围T 增大，所提出的方法会收敛到最优代价函数；等价地说，以 M 表征的终端集大小会变小。一个合理的问题是：如果不设置终端代价函数，只让时间范围不断增大，那么这与所提出的方法有何不同？在这种情况下，有限时域问题的最优代价函数最终会收敛到无限时域最优代价。为了看出所提出方法的优势，设从状态 x 转移到集合$\Omega_M = \{J_\infty \le M\}$所需的最优转移时间为 $T_M^*(x)$。此外，将在集合Ω_M内所需的有效调节时间记为 $T_M^R(x)$。那么，很容易看出，对于任意有限的 M，$T_M^*(x) << T_M^R(x)$，。因此，与不含任何终端代价的有限时域问题相比，我们的方法所对应的优化问题具有短得多的时间范围。下一节中的实验清楚地证明了这一点：我们需要求解的问题的有效时域，仅为不含任何终端代价的有限时域问题所需时间范围的一小部分。

自由终端时间与全局渐近稳定性：我们方法的另一个关键方面，是从给定状态 $x \in \mathcal{X}$ 出发的优化问题具有自由终端时间。事实上，我们已经指出，自由终端时间使近似代价函数 $J_\infty^M(。)$ 满足定常贝尔曼方程，而非时变贝尔曼

方程，从而建立全局渐近稳定性。若固定时间范围，我们就会回到模型预测控制中的情形，需要借助终端代价/集合约束来实现稳定，并且只能像 [22] 中那样得出局部渐近稳定性的结论。从物理角度看，这很合理，因为状态空间中不同点对应的最优转移时间各不相同，因此最优控制问题中的有限时域需要是可变的。自由终端时间恰好刻画了问题的这一特点，从而使闭环系统实现全局渐近稳定。通过采用一个“足够大”的时间范围，而非遍历时间，可以高效求解自由终端时间问题。基于上述论证，本问题中的“足够大”远小于无终端约束问题所需的“足够大”时间。

最优控制解：我们的方法需要求解一个最优控制问题，而通常来说，满足贝尔曼方程要求我们找到非线性最优控制问题的全局最小值。在我们最近的工作中 [29, 30], ，我们证明了在动力学关于控制为仿射、代价关于控制为二次型（可放宽为控制凸代价）的条件下，满足极小值原理 [31] 也足以保证得到全局最小值。尽管该问题名义上是非凸的，但利用特征线法，我们可以证明，满足极小值原理的轨迹必然是唯一的。此外，我们还证明了迭代线性二次调节器（ILQR）算法在相对宽松的条件下能够保证满足极小值原理，从而确保我们能够找到全局最小值[21, 20]。因此，在第四节的实验中，我们采用ILQR方法求解最优控制问题。

IV. 实证结果

通过对两个非线性系统——倒立摆和摆——进行仿真，验证了所提出的理论。仿真在 MATLAB 上完成，使用 ode45 模拟这两个系统的非线性动力学。所进行的实验列表见表 I。仿真旨在从不同的初始条件出发，使系统直立，并让摆/杆在顶端保持平衡。仿真总共运行 150 个时间步，时间离散间隔为 $0.1s$。我们将代价函数设为二次型：$c(x,\ u) = x'Qx + u'Ru$。

实验编号	系统	初始状态	终端状态
1.	倒立摆 $(x,\ \theta,\ \dot{x},\ \dot{\theta})$	$(0, 0, 0, 0)$	$(0, \pi, 0, 0)$
2.	倒立摆 $(x, \theta, \dot{x}, \dot{\theta})$	$(0, 3\pi/4, 0, 0)$	$(0, \pi, 0, 0)$
3.	摆 $(\theta, \dot{\theta})$	$(0,\ 0)$	$(\pi,\ 0)$
4.	摆 $(\theta,\ \dot{\theta})$	$(5\pi/12,\ 0)$	$(\pi,\ 0)$

表 I：所进行的实验

这些实验旨在表明，我们所构建的有限时域最优控制问题中的成本收敛于无限时域最优代价：在非线性动力学约束下，$J_\infty^T(x) = \min_{u_t} \sum_{t=0}^{T-1} c(x_t, u_t) + J_\infty(x_T)$，收敛于

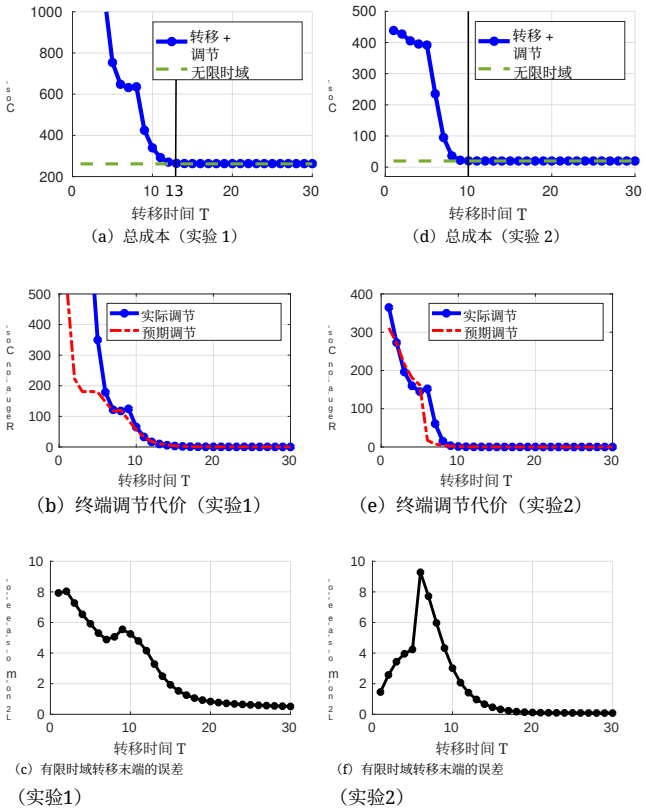
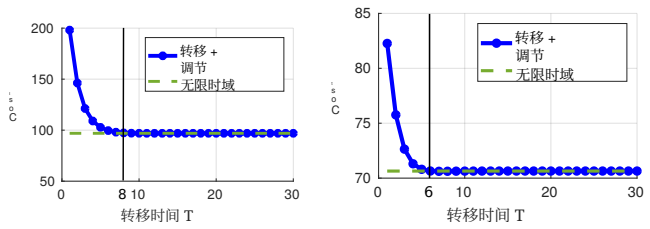


图 2：倒立摆摆起。请注意，标记为 (a)–(c) 的图对应表 I 中的实验 1，(d)–(f) 对应实验 2。图 (a) 和 (d) 中的结果表明，经过一段时间 T^*（以黑色竖线标出）后，有限时域转移+ 调节成本收敛至无限时域代价。还可以注意到，图 (a) 和 (d) 中的 T^* 数值不同，因为系统的初始条件不同。

最优代价 $J_\infty^*(x)$，随着 T 增大，即对于 $T > T^*$，$|J_\infty^T(x) - J_\infty^*| < \varepsilon$。这些实验还将表明，时间 T^* 也取决于系统的初始条件。

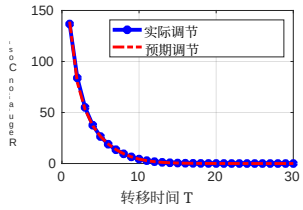
为了求出 $J_\infty(x)$，我们在原点附近对系统进行线性化。由于代价是二次型，我们可以利用稳态黎卡提方程计算 P_∞ 在原点处的值。因此，$J_\infty(x) = x'P_\infty x$。我们将 $J_\infty(x)$ 用作有限时域最优控制问题的终端代价，并使用迭代线性二次调节器（iLQR）算法，对不同的 T 求解优化问题 [32]。已有研究表明，对于具有控制仿射动力学和二次控制代价的光滑非线性问题，只要满足最优控制必要条件 [33]，(iLQR 能够做到这一点)，就可以收敛到唯一的全局最优解。因此，我们的实验避免了局部最优解问题。在有限时域或“转移时间” T之后，我们将控制切换为在原点处设计的 LQR 调节器，直至 $t = 150$。由于无法求解真正的无限时域问题，我们在没有终端代价的情况下求解 $T = 150$，将其作为无限时域代价的替代，并在图中标记为“无限时域”。我们注意到，采用较短时域进行求解无法使系统稳定在直立位置。

在我们的图中，标注为“转移 + 调节”的成本是有限时域控制器和调节器的总成本。在全部四项实验中，我们将观察到，

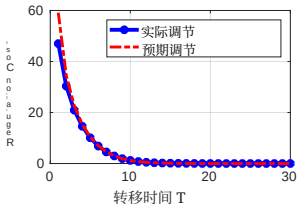


(a) 总成本（实验3）

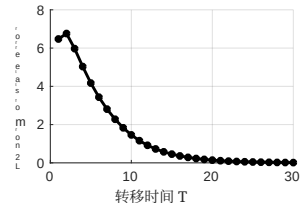
(d) 总成本（实验4）



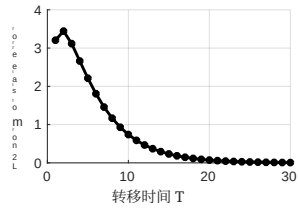
(b) 终端调节代价（实验3）



(e) 终端调节代价（实验4）



(c) 结束时的误差
有限时域转移（实验3）



（f）结束时的误差
有限时域转移（实验4）

图3：摆杆起立。标记为（a）至（c）的图对应表 I 中的实验3，（d）至（f）对应实验4。

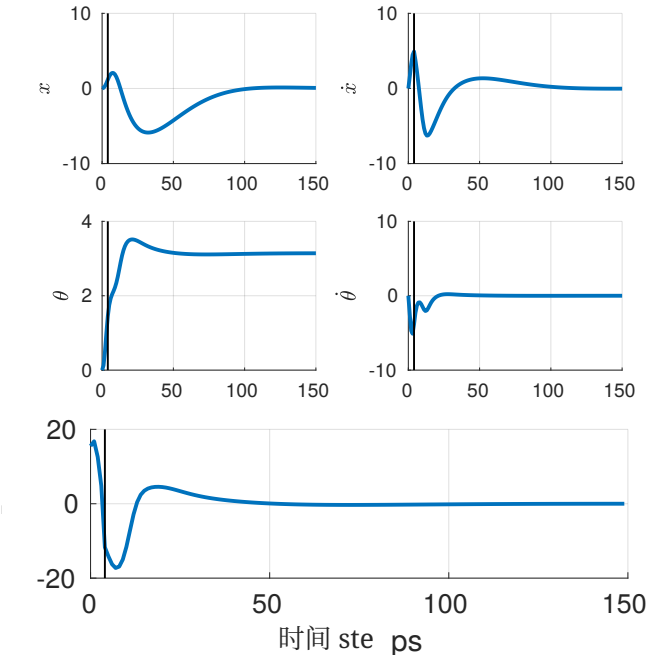


图4：时域为 $T = 4$ 时倒立摆的响应。黑色竖线表示从有限时域切换到调节控制。请注意，切换过程中状态和控制量出现了跳变。从有限时域控制器切换到调节器时，控制量从 4.8 跳变至 -11.7。

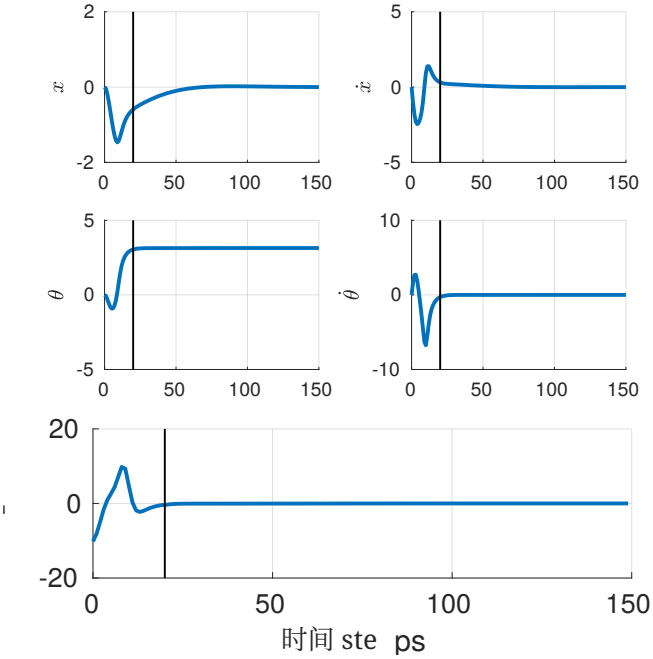


图5：时域为 $T = 20$ 时倒立摆的响应。黑色竖线表示从有限时域切换到调节控制。与图4不同，此处的切换过程是光滑的。

总成本会在一段时间后收敛 $T = T^*$。还可以看出，收敛后的成本与无限时域成本相符。实验还表明，数值 T^* 取决于所选的初始条件。预期调节成本使用 $x_T' P_\infty x_T$计算，其中 x_T 是有限时域最优控制问题的终端状态。实际调节根据终端调节器的轨迹计算。预期调节成本与实际调节成本之间的差异表明，线性终端调节器对于状态 x_T并非最优。

在实践中，通常会将无限时域最优控制问题表述为时域 T 很大的有限时域最优控制问题来求解，这会带来极高的计算成本。在这种情况下，它是 $T = 150$。但是，按照第三节中的定义，利用终端代价 $J_\infty(\cdot)$ 对有限时域最优控制问题进行正则化后，就可以用小得多的时域 T^* 求解。请注意，在我们的实验中，对于所考虑的两种情况，$T^* < 15 << 150 = T$，并且如结果所示，我们实现了近似最优性。求解较小的时域在计算上成本低得多，也为随机环境下需要频繁重新规划的应用提供了可能。此外，由于我们构造的代价到达函数是一个控制李雅普诺夫函数，因此可以保证原点的全局渐近稳定性。

V. 结论

本文提出了一种易于求解的方法，用于求解非线性无限时域最优控制问题的近似解。该方法具有全局渐近稳定性，并且随着终端集趋于零，其解在极限情况下收敛于真正的最优解。实证结果表明，与求解无限时域问题所需的有效时域相比，该方法能在很短时间内实现实际收敛；而求解

无限时域问题时，该方法不采用自由终端时间和终端集正则化。未来的工作将纳入状态约束和控制约束，并在一系列复杂程度各异的非线性问题上测试该方法；这将需要结合我们此前的工作[20, 21]，实现基于数据的泛化。我们还将考虑把该方法扩展到最优非线性输出反馈控制问题，并结合适当的基于数据的泛化。

参考文献

[1] D. P. Bertsekas。《动态规划与最优控制》。第2版，第I卷。Athena Scientific，2000年。[2] R. Bellman。《动态规划》。普林斯顿大学出版社，1957年。[3] D. S. Bernstein。“非二次型代价与非线性反馈控制”。《国际鲁棒与非线性控制期刊》3（1993）。[4] C.-J. Wan和D. S. Bernstein。“一类能够全局稳定角速度的最优非线性反馈控制器”。*[1992]*《第31届IEEE决策与控制会议论文集》。IEEE。1992年。[5] C.-J. Wan和D. S. Bernstein。“具有全局稳定化特性的非线性反馈控制”。《动力学与控制》5（1995）。[6] J. Si、A. G. Barto、W. B. Powell和D. Wunsch。《学习与近似动态规划手册》。第2卷。John Wiley & Sons，2004年。[7] F. L. Lewis和D. Liu。《用于反馈控制的强化学习与近似动态规划》。第17卷。John Wiley & Sons，2013年。[8] D. P. Bertsekas和D. A. Castanon。“用于无限时域动态规划的自适应聚合方法”。《IEEE自动控制汇刊》34（1989）。[9] W. B. Haskell、R. Jain和D. Kalathil。“经验动态规划”。《运筹学数学》41（2016）。[10] A. Geramifard、T. J. Walsh、S. Tellex、G. Chowdhary、N. Roy和J. P. How。“动态规划与强化学习中线性函数逼近器教程”。《机器学习基础与趋势》6（2013）。[11] R. Munos和C. Szepesv´ari。“拟合值迭代的有限时间界”。《机器学习研究期刊》9（2008）。[12] R. S. Sutton和A. G. Barto。《强化学习：导论》。MIT出版社，2018年。[13] C. J. Watkins和P. Dayan。“Q学习”。《机器学习》8（1992）。[14] R. S. Sutton、D. A. McAllester、S. P. Singh和Y. Mansour。“结合函数逼近的强化学习策略梯度方法”。《神经信息处理系统进展》。2000年。[15] D. Silver、A. Huang、C. J. Maddison、A. Guez、L. Sifre、G. Van Den Driessche、J. Schrittwieser、I. Antonoglou、V. Panneershelvam和M. Lanctot。“利用深度神经网络和树搜索精通围棋”。《自然》529（2016）。[16] J. Schulman、S. Levine、P. Abbeel、M. Jordan和P. Moritz。“信赖域策略优化”。《国际机器学习会议》。2015年。[17] T. Haarnoja、A. Zhou、P. Abbeel和S. Levine。“软演员-评论家：基于随机演员的离策略最大熵深度强化学习”。ICML。2018年。[18] S. Fujimoto、H. van Hoof和D. Meger。“解决演员-评论家方法中的函数逼近误差”。ArXiv abs/1802.09477（2018）。[19] P. Henderson、R. Islam、P. Bachman、J. Pineau、D. Precup和D. Meger。“真正重要的深度强化学习”。第三十二届AAAI人工智能会议。2018年。[20] R. Wang、K. S. Parunandi、A. Sharma、R. Goyal和S. Chakravorty。“强化学习中的反馈搜索”。2021年第60届IEEE决策与控制会议（CDC）。2021年。[21] R. Wang、K. S. Parunandi、A. Sharma、R. Goyal和S. Chakravorty。《强化学习中的反馈搜索》。2020年。[22] D. Q. Mayne。“模型预测控制：最新进展与未来展望”。Automatica 50（2014）。

[23] D. Q. Mayne、J. B. Rawlings、C. V. Rao和P. O. Scokaert。“约束模型预测控制：稳定性与最优性”。*Automatica* 36（2000）。[24] D. Limon、T. Alamo和E. Camacho。“扩大模型预测控制器的吸引域”。*Automatica* 41（2005）。[25] P. Falugi和D. Q. Mayne。“用于跟踪随机参考信号的模型预测控制”。*2013*年欧洲控制会议（*ECC*）。2013年。[26] L. Fagiano和A. R. Teel。“用于模型预测控制的广义终端状态约束”。*Automatica* 49（2013年9月）。[27] L. Gr¨une、J. Pannek、L. Gr¨une和J. Pannek。《非线性模型预测控制》。Springer，2017年。[28] J. K¨ohler和F. Allg¨ower。“采用有限尾部代价的模型预测控制中的稳定性与性能”。*IFAC-PapersOnLine* 54（2021）。第7届IFAC非线性模型预测控制会议（NMPC 2021）。[29] M. N. G. Mohamed、S. Chakravorty、R. Goyal和R. Wang。随机最优非线性控制中的最优反馈律。2020年。[30] M. N. G. Mohamed、S. Chakravorty、R. Goyal和R. Wang。“随机最优非线性控制中的反馈律”。*2022*年美国控制会议（*ACC*）。2022年。[31] A. Bryson和H. Y.-C.《应用最优控制：优化、估计与控制》。华盛顿：Hemisphere Pub. Corp.，1975年。[32] Y. Tassa、T. Erez和E. Todorov。“通过在线轨迹优化综合并稳定复杂行为”。*2012*年*IEEE/RSJ*智能机器人与系统国际会议。IEEE。2012年。[33] M. N. G. Mohamed、S. Chakravorty、R. Goyal和R. Wang。“随机最优非线性控制中的反馈律”。美国控制会议。2022年。