arXiv:2601.03016v1 [math.OC] 6 Jan 2026

基于策略梯度方法的未知线性切换系统自适应控制

Felix Laurent、Feiran Zhao、Jaap Eising、Florian D¨orfler

摘要——**我们研究策略梯度自适应控制（PGAC）框架，该框架通过对线性二次型调节器代价执行基于数据的梯度下降步，实时自适应地更新控制策略。已有经验表明，该方法能够高效地响应模型参数变化等情形。为将这一观察形式化，我们设计了一种 PGAC 方法，用于稳定线性切换系统，其中模型参数和切换时间均未知。我们使用滑动窗口数据进行策略梯度估计，并证明在满足驻留时间条件且动态变化量较小时，策略能够跟踪切换动力学并确保闭环稳定性。我们通过仿真验证了理论结果。**

I. 引言

线性二次调节器（LQR）是控制理论、强化学习及其应用中的一项标准基准。传统上，LQR 控制器的设计流程是间接的：首先确定待控系统的模型，然后设计控制器。随着人们对数据驱动控制的兴趣日益浓厚，与之相对的直接 方法也应运而生：旨在基于系统测量数据学习控制器。

这一问题曾从多个不同角度进行探讨。继基于 Willems 基本引理的方法之后 [1], , 有限时域情形在滚动时域形式中受到了广泛关注，相关研究始于 [2], [3]。与这些方法不同， [4],[5]研究了无限时域线性二次调节器的控制器一次性设计。同时，研究者也借助强化学习工具研究了学习型控制器的收敛性 [6]–[8]。这些方法通常处理按回合采集的数据，甚至是离线数据：控制器通过数据批次进行学习，在此期间策略保持不变，因此学习与控制彼此分离。由此，上述方法无法高效地实时改进策略。

相比之下，自适应控制能够监测自身的性能，并朝着提升性能的方向调整参数 [9], [10]。经典自适应控制侧重于鲁棒稳定性，基于李雅普诺夫理论，利用在线闭环数据设计控制输入 [11]。具有代表性的方法包括模型参考自适应控制

本研究得到了苏黎世联邦理工学院以及瑞士国家科学基金会通过国家自动化研究能力中心提供的支持。F. Laurent 任职于瑞士苏黎世 8092 苏黎世联邦理工学院机械与过程工程系。（电子邮箱：flaurent@ethz.ch）F. Zhao、J. Eising 和 F. D¨orfler 任职于瑞士苏黎世 8092 苏黎世联邦理工学院信息技术与电气工程系。（电子邮箱：zhaofe@control.ee.ethz、jeising@control.ee.ethz、dor-fler@ethz.ch）

（MRAC）、自整定调节器和无模型自适应控制 [11], [12]。此外，还有一些自适应控制方法通过确定性等价线性二次调节器 [13]–[16]，同时追求稳定性和策略收敛。尽管这些进展是在针对线性时不变（LTI）系统的研究中取得的，但当系统不是线性时不变系统时，控制策略的自适应仍发挥着重要作用 [17]–[19]。特别是当未知动力学存在切换时，必须利用在线数据实时更新策略，才能实现稳定性。近期，行为系统理论和数据驱动控制中的数据信息性方法已扩展至线性切换系统的情形 [20]–[22]。然而，这些方法依赖于在每个时间步求解优化问题，这可能导致较高的计算开销，并且容易受到噪声影响。

实现自适应的另一种方法是基于梯度的方法，即在实时运行中通过梯度下降迭代更新策略。最近，我们提出了策略梯度自适应控制（PGAC）框架 [23], ，该框架通过高效计算由数据得到的 LQR 代价梯度，并沿梯度下降方向在反馈中自适应地更新策略。该框架包括直接和间接梯度计算方法：间接方法首先估计模型，再基于该模型计算策略梯度；直接方法采用 LQR 的样本协方差参数化，并直接从数据中估计梯度 [24]。对于线性系统，在温和假设下，所提出的策略梯度方法表现出闭环稳定性，并且策略能够收敛至最优线性二次调节增益。此外，尽管理论保证仅针对线性时不变系统给出，PGAC 框架已成功应用于非线性或时变情形，例如功率变换器系统 [25], 、航空航天控制 [26], 以及自主自行车控制 [27]。

本文重点研究一种用于稳定线性切换系统的 PGAC 方法。不同于相关文献 [28],，我们假设模型参数和切换时刻均未知。我们采用无限时域线性二次调节代价作为目标函数，因为最优线性二次调节增益能够稳定系统，并具有一定程度的稳定裕度。随后，我们采用间接 PGAC方法，其中代价梯度基于辨识得到的动力学模型进行计算。为了使该梯度能够较好地反映时变动力学，我们使用滑动窗口数据进行模型辨识。我们证明，在满足驻留时间条件且动力学具有有界变化的情况下，该方法能够跟踪切换动力学并确保闭环稳定性。最后，通过仿真验证了我们的理论结果。

本文其余部分的组织如下。

第二节回顾间接PGAC方法，并 formulates 线性切换系统控制问题。第三节证明关于系统稳定性的主要结果。第四节进行数值案例研究。第五节给出结论。所有证明均推迟至附录。

II. 预备知识

本节将回顾策略梯度自适应控制 （PGAC）算法 [23]。

A. 线性二次调节器的策略梯度方法考虑线性时不变（ LTI）系统

$$x_{t+1} = Ax_t + Bu_t + w_t, \tag{1}$$

其中， $t \in \mathbb{N}$是状态， $u_t \in \mathbb{R}^m$ 是控制输入， $w_t \in \mathbb{N}$ 是噪声。我们假设系统（1）是可控的。众所周知，线性二次调节器（LQR）代价可以（例如参见 [6]）视为状态反馈增益 $K \in \mathbb{R}^{m\times n}$的函数:

$$C(K) := \mathrm{Tr}\left((Q + K^\top RK)\Sigma_K\right), \tag{2}$$

其中，权重矩阵（Q， R）是正定的， Σ_K 是通过以下李雅普诺夫方程的正定解得到的闭环状态协方差矩阵

$$\Sigma_K = I_n + (A + BK)\Sigma_K(A + BK)^\top. \tag{3}$$

策略梯度方法并不最小化该代价，而是使用梯度下降更新反馈增益，即策略:

$$K_{t+1} = K_t - \eta\nabla C(K_t), \tag{4}$$

其中， $\nabla C(K_t)$是 C 关于 K_t的梯度。定义 $\mathcal{S}$: $= \{K \in \mathbb{R}^{m\times n} | \rho(A + BK) < 1\}$ 为稳定化增益集合。于是，梯度 $\nabla C(K)$可表示为闭式解。

引理 *1*（ *[6,* 引理 *1]*）：对于任意 $K \in \mathcal{S}$， $C(K)$的梯度为

$$\nabla C(K) = 2\left(\left(R + B^\top PB\right)K + B^\top PA\right)\Sigma_K, \tag{5}$$

其中， Σ_K 满足式（3）， P是李雅普诺夫方程的正定解

$$P = Q + K^\top RK + (A + BK)^\top P(A + BK). \tag{6}$$

本文考虑，）未知、且无法直接计算梯度的情况。在这种情况下，PGAC 使用模型的在线辨识来进行梯度计算，概述如下。

B. 策略梯度自适应控制

由于模型（A， B）未知，我们从测量数据 [29] 中对其进行辨识。考虑长度为 t 的状态、输入、噪声和后继状态时间序列

$$\begin{aligned} X_{0,t} &:= [x_0, x_1, \ldots, x_{t-1}] &&\in \mathbb{R}^{n\times t}, \\ U_{0,t} &:= [u_0, u_1, \ldots, u_{t-1}] &&\in \mathbb{R}^{m\times t}, \\ W_{0,t} &:= [w_0, w_1, \ldots, w_{t-1}] &&\in \mathbb{R}^{n\times t}, \\ X_{1,t} &:= [x_1, x_2, \ldots, x_t] &&\in \mathbb{R}^{n\times t}, \end{aligned} \tag{7}$$

它们满足线性动力学

$$X_1 = AX_{0,t} + BU_{0,t} + W_{0,t}. \tag{8}$$

假设数据矩阵$D_{0,\ t} = [U_{0,t}^\top, X_{0,t}^\top]$ 满行秩。于是，估计模型可由普通最小二乘问题的唯一解得到

$$[\hat{B}_t, \hat{A}_t] = \arg\min_{[B,A]} \|X_{1,t} - [B, A]D_{0,t}\|_F = X_{1,t}D_{0,t}^\dagger, \tag{9}$$

其中，$(\cdot)^\dagger$ 表示摩尔－彭若斯伪逆。随后，可将模型估计值 $(\hat{}A,\ \hat{}B)$ 作为真实值，计算代价函数 $C(K)$ 的梯度估计值。

通过迭代辨识动力学、对反馈增益执行梯度下降步骤，并利用所得策略在线收集闭环数据，该方法能够实现自适应线性二次调节器学习，并提供最优的非渐近性能保证 [14], [30]。该方法总结于算法 1。它需要一个初始稳定化增益和一批离线持续激励数据。在在线阶段，控制输入中包含探测噪声 e_t ，以确保数据具有信息性。更多详情请参阅 [23], [24]。

算法 1 线性时不变系统的策略梯度自适应控制

1: **初始化：** 离线数据 $(X_{0,\ t_0},\ U_{0,\ t_0},\ X_{1,\ t_0})$，一个初始稳定化策略 K_{t_0}，以及步长 η。
2: **对于** $t = t_0, t_0 + 1, \ldots$ **do**
3: 应用 $u_t = K_t x_t + e_t$ 并观察 x_{t+1}。
4: 通过普通最小二乘估计模型参数

$$[\hat{B}_{t+1}, \hat{A}_{t+1}] = \arg\min_{B,A} \|X_{0,t+1} - [B, A]D_{0,t+1}\|_F.$$

5: 执行一次策略梯度下降

$$K_{t+1} = K_t - \eta\nabla\hat{C}_{t+1}(K_t),$$

其中 $\nabla\hat{}C_{t+1}(K_t)$ 是引理 1 中的策略梯度 $\hat{}\ \hat{}$，其中采用估计模型 $(A_{t+1},\ B_{t+1})$。6: **endfor**

III. 线性切换系统的策略梯度自适应控制

算法 1 中的 PGAC 方法目前仅在线性时不变系统 [23], [24] 上得到了理论认证。本节将 PGAC 方法扩展至切换系统的稳定化。

A. 线性切换系统的问题表述与 *PGAC* 方法

考虑以下切换系统

$$x_{t+1} = A_i x_t + B_i u_t, \quad T_{i-1} \le t < T_i \tag{10}$$

对于 $i \in \mathbb{N}$，其中 T_i 是 i^{th} 与 $(i + 1)^{th}$ 模式之间的切换时间。我们假设两种动力学 $(A_i,\ B_i)$ 以及 T_i、$\forall i \in \mathbb{N}$ 均未知。我们的目标是开发一种 PGAC 方法来稳定(10)。显然，仅在每个时间 t 稳定系统是不够的，因此我们采用序贯强稳定性的概念 [31]。

定义 *1*（序贯强稳定性）：对于系统 (10)，策略序列 K_t 是（κ，α）-强稳定的（对于$\kappa \geq 1$ 和 $0 < \alpha \leq 1$），如果存在矩阵 $H_t \succ 0$ 和L_t 使得

$$A_t + B_t K_t = H_t L_t H_t^{-1} \quad \text{for all } t,$$

and

(i) $\|L_t\| \leq 1 - \alpha$ 和 $\|K_t\| \leq \kappa$;
(ii) $\|H_t\| \leq \kappa$ 和 $\|H_t^{-1}\| \leq 1$;
(iii) $\|H_{t+1}^{-1} H_t\| \leq 1 + \alpha/2$。

特别是，这意味着每个策略 K_t 都必须各自稳定（满足 (i) 和 (ii)），并且两个连续闭环系统之间的变化必须很小（满足 (iii)）。序贯强稳定性给出了稳定性的定量定义，并意味着状态渐近稳定。现在，我们的目标是调整 PGAC，使系统 (10) 满足序贯强稳定性。

为避免误差累积，我们使用滑动窗口数据进行辨识，也就是说，我们只使用最近的数据。定义窗口大小固定为 L 的数据序列如下:

$$\begin{aligned} X_t &:= [x_{t-L}, x_{t-L+1}, ..., x_{t-1}] \in \mathbb{R}^{n \times L}, \\ U_t &:= [u_{t-L}, u_{t-L+1}, ..., u_{t-1}] \in \mathbb{R}^{m \times L}, \\ X_{t+1} &:= [x_{t-L+1}, x_{t-L+2}, ..., x_t] \in \mathbb{R}^{n \times L}. \end{aligned} \tag{11}$$

为确保 $D_t = [U_t^\top, X_t^\top]$满足满行秩条件，我们要求 $L \geq n + m$。由此得到的 PGAC 方法与算法 1 类似；参见算法 2。

算法 2基于滑动窗口数据的线性切换系统策略梯度自适应控制

1: **初始化：** 离线数据（X_{t_0}，U_{t_0}，X_{t_0+1}），一个初始稳态-定策略 K_{t_0}，以及一个步长 η。
2: **对于** $t = t_0, t_0 + 1, \ldots$ **do**
3: 应用 $u_t = K_t x_t + e_t$，然后观测 x_{t+1}。
4: 使用（X_{t+1}，U_{t+1}，X_{t+2}）估计模型并且或-普通最小二乘法

$$[\hat{B}_{t+1}, \hat{A}_{t+1}] = \arg\min_{B,A} \|X_{t+2} - [B, A] D_{t+1}\|_F.$$

5: 执行一次策略梯度下降

$$K_{t+1} = K_t - \eta \nabla \hat{C}_{t+1}(K_t), \tag{12}$$

其中 $\nabla \hat{C}_{t+1}(K_t)$ 是基于估计模型的策略梯度 ^ ^ 模型 (A_{t+1}, B_{t+1})。6: **结束循环**

一个主要挑战是，当动力学随时间变化时，最小二乘辨识 (9) 会变得脆弱。更准确地说，切换后收集的数据包含由不同的 (A_i，B_i) 生成的样本，而对这种混合模式数据执行普通最小二乘法，得到的估计模型可能既不对应当前的动力学，也不对应之前的动力学。

B. 闭环稳定性分析

在本文余下部分中，我们需要一些假设。首先，我们要求状态和输入信号具有持续激励性。

假设 *1*：滑动窗口数据矩阵满行秩，即 rank(D_t) $= m + n$ 对所有 $t \geq t_0$均成立。

通过令 $L \geq n + m$ 并对控制输入施加随机探测噪声，可确保该假设以概率 1 成立

$$u_t = K x_t + e_t. \tag{13}$$

我们还假设驻留时间存在一个下界，即系统在切换到另一模式之前保持当前模式的时长。

假设 *2*：驻留时间以下限滑动窗口长度为界，即对于任意 $i \in \mathbb{N}$，$T_{i+1} - T_i \geq L$。

该假设确保数据序列 X_t，U_t，X_{t+1}至多包含两个连续动力学的数据，即（A_i，B_i）和（A_{i+1}，B_{i+1}），从而简化后续分析。

我们首先利用普通最小二乘和滑动窗口数据量化辨识误差。令第 i 次切换处动力学的变化量为

$$\Delta_i = [B_{i+1}, A_{i+1}] - [B_i, A_i].$$

然后，我们得到以下结果。

引理 *2*：考虑使用滑动窗口数据的普通最小二乘问题

$$[\hat{B}_t, \hat{A}_t] = \arg\min_{[B,A]} \|X_{t+1} - [B, A] D_t\|_F. \tag{14}$$

因此，对于任意 $i \in \mathbb{N}$，我们有

$$\begin{aligned} \|[\hat{B}_t, \hat{A}_t] - [B_i, A_i]\| &\leq \|\Delta_i\|, \quad T_i \leq t < T_i + L, \\ \|[\hat{B}_t, \hat{A}_t] - [B_{i+1}, A_{i+1}]\| &\leq \|\Delta_i\|, \quad T_i \leq t < T_i + L, \\ [\hat{B}_t, \hat{A}_t] &= [B_{i+1}, A_{i+1}], \quad T_i + L \leq t < T_{i+1}. \end{aligned}$$

引理 2 表明，在 $T_i \leq t < T_i + L$ 中，i^{th} 时刻的 $(i+1)^{th}$ 模式下，辨识误差的上界由动力学变化量 $\|\Delta_i\|$ 给出。对于时间 $T_i + L \leq t < T_{i+1}$，滑动窗口数据仅由 (A_{i+1}，B_{i+1}) 生成，因此辨识误差为零。利用引理 2，我们可以量化在策略更新 (12) 中，使用辨识模型而非真实模型进行策略梯度所带来的影响。

我们的主要结果给出了相应条件，在这些条件下，我们可以保证系统(10)与 Algorithm2 之间闭环的序贯强稳定性。

定理 *1*（切换中的 N 稳定性）：设假设 1 和 2 成立。考虑在 $N + 1$ 个不同系统之间进行 N 次切换。存在常数 ν_1 和 $\nu_{2,\,i}$，其取决于（A_i，B_i，Q，R），使得，如果 $\eta \leq \nu_1$，如果初始策略 K_{t_0}使初始系统（A_0，B_0）稳定，并且如果

$$\|\Delta_i\| \leq \nu_{2,i}, \quad i \in \{1, ..., N\}, \tag{15}$$

那么算法 2 是可行的，并且序列 K_t 是（κ，α）-强稳定的，其中常数为（κ，α）。此外，状态由以下式子有界：

$$\|x_t\| \le \overline{\kappa}(1-\frac{\overline{\alpha}}{2})^{t-t_0}\|x_{t_0}\| + \frac{2\overline{\kappa}}{\overline{\alpha}} \max_{t_0 \le j < t} \|B_j e_j\|. \quad (16)$$

定理 1 表明，若步长和动力学变化量足够小，则闭环系统是序贯强稳定的，且状态序列有上界。请注意，状态界包含两项：第一项在幅度为 $\bar{\kappa}$ 的超调后，随初始状态呈指数衰减；第二项则对探测噪声引起的偏差提供一致的界。

回顾假设 2，它要求对于每个 i，都有 $T_{i+1} - T_i \ge L$，以确保滑动窗口数据至多由两个系统生成。因此，条件 (15) 隐含地给出了系统矩阵变化量的上界。

现在可以反复应用定理 1，确保算法在发生切换之前有足够的时间收敛至模态特定最优性。因此，系统的代价函数始终有界。进而，我们可以给出更严格的驻留时间条件，在该条件下，算法 2 能够保证在无限多次切换下的序贯稳定性。

推论 *1*（无限次切换）：设假设 1 成立。假设存在变量 $\nu_{1、i}$ 和 $\nu_{2、i}$，它们取决于（A_i、B_i、Q、R），使得初始策略 K_{t_0} 能够稳定初始系统（A_0、B_0）、$\eta \le \nu_{1、i}$，条件（15）成立，并且对于所有 i

$$T_{i+1} - T_i \ge \frac{2\mu_i}{\eta} \log\left(\frac{1}{\overline{C}_i - C_i^*}\right) + L \quad (17)$$

此时，算法 2 对无限次切换仍然可行。此外，状态满足有界性

$$\|x_t\| \le \nu_{3,i}(1-\frac{\nu_{4,i}}{2})^{t-t_0}\|x_{t_0}\| + \frac{2\nu_{5,i}}{\nu_{4,i}} \max_{t_0 \le j < t} \|B_j e_j\|, \quad (18)$$

其中 $\nu_{3、i}$、$\nu_{5_i} > 1$ 和 $0 < \nu_{4、i} < 1$。

下面我们讨论该结果的局限性和扩展。

一致界。注意，条件（15）和（17）是模态相关的，因此依赖于时间。具体而言，每个变化量 Δ_i 和驻留时间由相应的模态参数（A_i、B_i、C_i^*、C_i）决定。为了使推论 1 得到一个与时间无关的状态上界，我们可以假设对所有 i 都存在一个共同的上界 $C \ge C_i$（K_t）。进一步地，对系统矩阵施加额外的有界性假设，可以推导出代价的一致上界。

含过程噪声的切换系统。虽然我们的分析未明确考虑过程噪声，但将时不变情形下的结果[23] 推广到这一问题并不困难。具体而言，需要将 (15) 中的 $\nu_{2,i}$改为表示稳定性的鲁棒裕度，并在 (18) 中增加一项以考虑噪声。

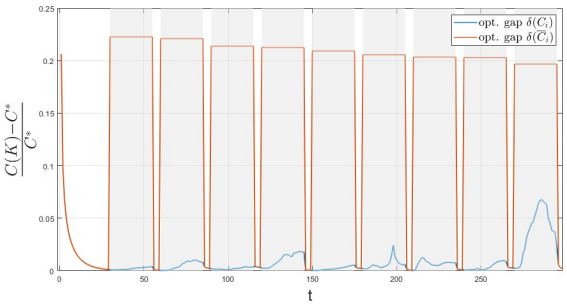


图 1. 真实代价函数 $C_i(K_t)$ 与其上界 C_i 的比较

通过多步策略梯度缩短驻留时间。在算法 2 中，我们在每个时刻只执行一次梯度下降步。如果在线计算允许，我们也可以在每个时刻执行多步梯度下降，从而提高收敛速度。这意味着算法能够快速收敛到适用于时变动力学的最优策略，并缩短驻留时间。

放宽驻留时间条件。为便于阐述，我们采用假设 2，并考虑驻留时间有界的切换系统 (10)。不过，对于切换更频繁的系统，也可以推导出与引理 2 类似的结果。假设平均驻留时间，则可以推导出期望意义下的序贯强稳定性。在这种情况下，与使用滑动窗口数据 (11) 相比，考虑带有遗忘因子的数据可能更为自然。

IV. 仿真

我们进行了一项仿真，以展示 LQR 代价的上界以及状态稳定性。我们选择以下参数：

$$A_0 = \begin{bmatrix} -0.13 & 0.14 & -0.29 & 0.28 \\ 0.48 & 0.09 & 0.41 & 0.30 \\ -0.01 & 0.04 & 0.17 & 0.43 \\ 0.14 & 0.31 & -0.29 & -0.10 \end{bmatrix}, B_0 = \begin{bmatrix} 1.63 & 0.93 \\ 0.26 & 1.79 \\ 1.46 & 1.18 \\ 0.77 & 0.11 \end{bmatrix}.$$

此外，我们假设动力学遵循随机游走，即对于每个 i，取 A_i，B_i 矩阵，其中每个元素均从标准正态分布中抽取，令 $A_{i+1} = A_i + 0.1A_i$，并令 $B_{i+1} = B_i + 0.1B_i$.。最后，对于 LQR 和算法参数，我们取 Q 和 R 为单位矩阵，并取 $L = 25$，$T_{i+1} - T_i = 30$，$\Delta t_i = 5$.

A. LQR 代价的上界性

如第三节所述，在切换后的过渡期内，代价满足不等式 $C_{i+1}(K_t) \le C_{i+1}$，即 $t \in [T_i + 1, T_i + L]$。在仿真中，我们通过最优性间隙来展示代价函数及其对应上界的演变，最优性间隙定义为

$$\delta(C(K)) = \frac{C(K) - C^*}{C^*}. \quad (19)$$

在图 1 中，代价的上界按照附录中引理 11 所述进行计算。灰色区域

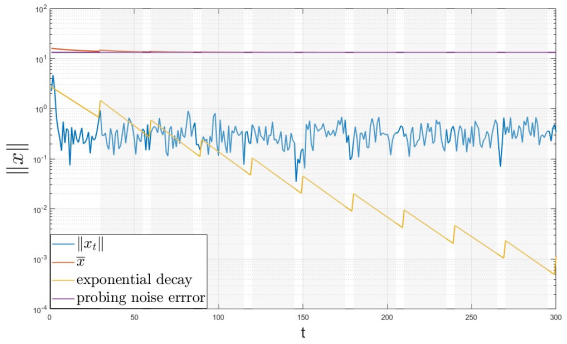


图 2. 状态 $\|x_t\|$ 范数与其上界的比较。

表示过渡时段 $t \in [T_i+1, T_i+L]$，其中数据序列跨越了两个连续的模式。图 1 显示，在每个过渡时段内，成本上界始终得到满足。这还表明，每次过渡后，代价函数都会迅速恢复到接近最优的值，这意味着采用显著更短的驻留时间可能是可行的。

B. 状态范数

我们还展示了在所设定条件下状态范数的上界。如图 2 所示，在多个切换时刻之间，状态范数始终保持较小且有界。每次切换时，上界都会发生变化，这反映了推论 1 中变量ν_3、 ν_4和 ν_5 的时变特性。然而，探测噪声诱导的误差项（紫色曲线）很快便超过指数衰减（黄色曲线），使得我们的上界过于保守。

V. 结论

本文提出了一种策略梯度自适应控制方法，用于稳定未知线性切换系统，其中系统在经过一段驻留时间后发生变化。我们证明了所提方法的闭环稳定性，并通过仿真展示了相关结果。

未来，研究不含驻留时间的连续变化系统将很有意义。此外，采用折扣线性二次调节器代价作为性能指标并分析其性能，也值得进一步研究。

附录I 证明

A. 引理 2（辨识误差）的证明

定义 $W_t = \Delta_i D_t S$，其中 $S = \mathrm{diag}(1,\dots,1,0,\dots,0)$ 选取切换前的列。然后，我们将子空间关系表述为

$$X_{t+1} = A_i X_t + B_i U_t + W_t. \tag{20}$$

对于普通最小二乘问题（14），有 $\|[\hat{}B_{t+1}, \hat{} A_{t+1}] - [B_i, A_i]\| = \|\Delta_i D_t S D_t^\dagger\| \le \|\Delta_i\| \, \|D_t S D_t^\dagger\|$。根据诱导矩阵范数的定义，并令 $M = D_t^\top D_t \succ 0$，可得

$$\|D_t S D_t^\dagger\|^2 \le \sup_{z\neq 0} \frac{z^\top S M S z}{z^\top M z} \le 1,$$

其中，最后一个不等式由 $SMS \preceq M$ 得出。因此，有 $\|D_t S D_t^\dagger\| \le 1$，从而完成证明。

B. LQR 代价的上界性的证明

不同于 [23] 中针对线性时不变系统的情形，显然，分析代价函数 $C(K)$ 的收敛性并不相关，因为我们假设动力学中存在未知切换，这将导致代价函数的非最优性。一种更为合理的方法是保证代价在变化过程中始终具有一个上界。

引理 3：令 $A_0 + B_0 K_{t_0}$ 稳定。假设

$$\|\Delta_i\| \le \nu_{1,i} \quad \eta \le \nu_2, \tag{21}$$

其中， η 为梯度下降步的步长，而ν_1， ν_2 是依赖于（A_i, B_i， Q， R）的变量。因此，代价函数 $C_t(K_t)$ 有上界。

该引理保证切换系统的代价函数在每个时间步都有上界。

如假设 2 所述，连续两次切换之间的时间间隔足够长，因此数据序列至多跨越两个连续模式（i， $i+1$）。因此，我们可以将每次切换独立处理。我们重点关注数据序列跨越两个模式的时间段，即 $t \in]T_i,\ T_i + L]$。首先，我们证明与切换前系统（A_i， B_i）相关的代价函数 $C_i(K)$ 在切换后仍有上界。然后，我们证明与切换后系统（A_{i+1}， B_{i+1}）相关的代价函数 $C_{i+1}(K)$ 有上界。

引理4（引理 11, [23]）：对于 $K \in \mathcal{S}_i$，有 (i)$\|\Sigma\| \le \frac{C(K)}{\sigma(Q)}$，$\|P\| \le C(K)$，$\|K\|_F \le \sqrt{\frac{C(K)}{\sigma(R)}}$，(ii) $\left\|\begin{pmatrix} C & K \\ \sigma & R \end{pmatrix}^{1/2}\right\| \le C(K)$，(iii) $\|K\| \le C(K)$。

引理 5（引理 12, [23]）：设 $A \in \mathbb{R}^{n\times n}$ 稳定，且 $\Sigma(A) = I + A\Sigma(A)A^\top$。若 $\|A' - A\| \le (4\|\Sigma(A)\|(1+\|A\|))^{-1}$，则 A' 稳定，且 $\|\Sigma(A') - \Sigma(A)\| \le 4\|\Sigma(A)\|^2(1+\|A\|)\|A' - A\|$。

引理 6：定义

$$p_1(a) = \frac{\underline{\sigma}(Q)}{4a\left(1+\frac{a}{\underline{\sigma}(Q)}\right)\left(1+\sqrt{\frac{a}{\underline{\sigma}(R)}}\right)}.$$

若 $\|\Delta_i\| \le p_1(C_i)$，则 $\hat{} A_{t+1} + \hat{} B_{t+1} K_t$ 稳定，且 $\hat{} \hat{} C_i(K_t) \nabla C_i(K_t) < \infty$。

证明：结合引理4和引理2，可得 $\hat{} \hat{}$ $\| A_{t+1} + B_{t+1}K_t - A_i + B_i K_t \| \le \|\Delta_i\|\ 1 + \|K_t\| \le \|\Delta_i\| \left(1 + \sqrt{\frac{C_i(K_t)}{\sigma(R)}}\right)$ 5 $\| \| \|$。根据引理并应用引理4中的界，必有

$$\|\Delta_i\|(1+\sqrt{\frac{C_i(K_t)}{\underline{\sigma}(R)}}) \le \frac{\underline{\sigma}(Q)}{4C_i(K_t)(1+\frac{C_i(K_t)}{\underline{\sigma}(Q)})} \tag{22}$$

假设 $C_i(K_t) \leq C_{i0}$。因此，$p_1(C_i) \leq p_1(C_i(K_t))$。

■

现在我们证明 $C_i(K_t) \leq C_{i0}$。我们将$C_i(K)$记为系统 $[A_i, B_i]$ 的真实成本函数，并将 $\hat{C}_t$ 记为估计模型 $[A_t, B_t]$的代价函数，定义 K' 和 K'' 为

$$K' = K - \eta \nabla \hat{C}_t(K) \tag{23}$$

$$K'' = K - \eta \nabla C_i(K) \tag{24}$$

根据 [23, 引理 2 和 3]，，若 $\eta \leq \frac{1}{l_i}$，梯度下降得到

$$C_i(K'') - C_i^* \leq (1 - \frac{\eta}{2\mu_0})(C_i(K) - C_i^*) \tag{25}$$

引理 7（引理 *14, [23]*）：设 $K \in \mathcal{S}$。对于任意 K，若 $\|K - K\| \leq p_4$，则 $K \in \mathcal{S}$ 且

$$\|\tilde{\Sigma} - \Sigma\| \leq p_5 \|\tilde{K} - K\| \tag{26}$$

$$|C(\tilde{K}) - C(K)| \leq p_6 \|\tilde{K} - K\| \tag{27}$$

where $p_4 = \frac{\sigma_{\min}(Q)\mu}{4C(K)\|B\|(\|A-BK\|+1)}$, $p_5 = 4\left(\frac{C(K)}{\sigma_{\min}(Q)}\right)^2 \frac{\|B\|(\|A-BK\|+1)}{\mu}$ and $p_6 = p_5 C(K)$.

引理 *8*：设 $K \in \mathcal{S}_{i0}$。若 $\|\Delta_i\| \leq p_1$，则

$$\|\nabla \hat{C}_i(K) - \nabla C_i(K)\| \leq p_3 \|\Delta_i\| \tag{28}$$

where $p_3 = \text{poly}\left(\frac{C_i(K)}{\sigma(Q)}, \|A_i\|, \|B_i\|, \|R\|, \frac{1}{\sigma(R)}\right)$.

证明：由 [23], $\nabla C_K = 2E\Sigma_{RK+B^TPA+BK}$ $E =$可得 $\nabla C_K - \nabla C_K = 2E\Sigma - 2\hat{E}\hat{\Sigma} = 2(E-\hat{E})\Sigma - 2\hat{E}(\Sigma - \hat{\Sigma})$。首先，我们关注第二项 $2\hat{E}(\Sigma - \hat{\Sigma})$。根据 [6, 引理 11]，，可得

$$\text{Tr}(E^T E) \leq \|R + B^T P B\|(C(K) - C^*). \tag{29}$$

结合引理 5 和引理 6，并定义 $p_2 = \frac{C_{(K)}}{\sigma(Q)p_1}$，可得

$$\|\hat{\Sigma} - \Sigma\| \leq p_2 \|[\hat{B}, \hat{A}] - [B, A]\| \leq p_2 \|\Delta\|. \tag{30}$$

将 (29) 和 (30) 合并，可得

$$\|2E(\Sigma - \hat{\Sigma})\| \leq 2\|R + B^T P B\|(C(K) - C^*)p_2 \|\Delta\|. \tag{31}$$

为界定项 $2(\hat{E} - E)\hat{\Sigma}$，注意到 $E - \hat{E} = B^T P((A + BK) - (\hat{A} + \hat{B}K)) + (B^T P - \hat{B}^T \hat{P})(\hat{A} + \hat{B}K)$。对各项分别取界，可得 $\|B^T P((A+BK) - (\hat{A}+\hat{B}K))\| \leq \|B^T P\|(1+\|K\|)\|\Delta\|$，以及 $\|A + BK\| \leq \|A + BK\| + (1 + \|K\|)p_1$。对于项 $B^T P - \hat{B}^T \hat{P}$，将其展开为 $B^T P - \hat{B}^T \hat{P} = \hat{B}^T(P - \hat{P}) + (B^T - \hat{B}^T)P$。注意到 $\|\hat{B}\| \leq \|B\| + p_1$，$\|B^T - \hat{B}^T\| \leq \|\Delta\|$，$\|P - \hat{P}\| \leq \|Q + K^T R K\|\|\Sigma - \hat{\Sigma}\| \leq C_K p_2 \|\Delta\|$，$\|\Sigma\| \leq 2\|\Sigma\|$，$\|B^T P - \hat{B}^T \hat{P}\| \leq$。因此，$(\|B\| + p_1)\|Q + K^T R K\| p_2 \|\Delta\| + \|\Delta\|\|P\|$。

此外，可得 $\|E - \hat{E}\| \leq \|B^T P\|(1 + \|K\|)\|\Delta\| + ((\|B\| + p_1)\|Q + K^T R K\| p_2 + \|P\|) \times (\|A + BK\| + (1 + \|K\|)p_1)\|\Delta\|$。注意到$\|P\|$，$\|\Sigma\|$，$\|K\|$ 可由一个多项式作为上界关于 $(C(K), \sigma(Q), \frac{1}{\sigma(R)}(), \|R\|, \|A\|, \|B\|)$，证明完毕。

■

这表明，由模型失配导致的梯度误差是有界的，并且与变化量的大小成正比。为了界定梯度不准确性所导致的代价函数误差，我们使用以下引理。

引理 *9*：设 $K \in \mathcal{S}_{i0}$。存在一个关于 $\left(\frac{\sigma(Q)}{C_i}(K), \frac{1}{\|A_i\|}, \frac{1}{\|B_i\|}, \frac{1}{\|R\|}, \sigma(R)\right)$ 的多项式 p_7，使得，如果

$$\|\Delta\| \leq \bar{p}_1 \quad \text{and} \quad \eta \leq \min\left\{\frac{p_4}{p_3\|\Delta\|}, p_7\right\}, \tag{32}$$

则有 $|C_i(K'') - C_i(K')| \leq \eta p_3 p_6 \|\Delta_i\|$.

引理 9 改编自 [23, 引理 15]，它给出了梯度下降步不准确所导致的代价函数误差的上界。

证明：设 [6,定理 7]中步长的多项式为 p_7。由 [6, 定理 7] 可知，该更新返回一个稳定化策略 K''。此外，还有 $\|K' - K''\| \leq \eta\|\nabla \hat{C}(K) - \nabla C(K)\| \leq \eta p_3 \|\Delta\|$。根据引理 6 $\eta \leq \frac{p_4}{p_3\|\Delta\|}$ $|C K'' - C K'| \leq$，并且若，$\eta p_3 p_6 \|\Delta\|$.

■

我们得到 ( ) ( )

回到式 (25)，并注意引理 9，可得

$$C_i(K') - C_i^* \leq (1 - \frac{\eta}{2\mu_i})(C_i(K) - C_i^*) + \eta p_3 p_6 \|\Delta_i\|. \tag{33}$$

设 $p_{\mathcal{T}}$ 为与 C_i 相关的多项式，p_j^* 为与 C_i^* 相关的多项式，则；

引理 *10*：设 $K_T \in \mathcal{S}_{i0}$。若

$$\|\Delta_i\| \leq \min\left\{\frac{1}{2\mu_i \bar{p}_3 \bar{p}_6}, \bar{p}_1\right\}, \eta \leq \min\left\{\frac{1}{\bar{l}_i}, \frac{p_4^*}{\bar{p}_3\|\Delta\|}, \bar{p}_7\right\},$$

则有

$$C_i(K_t) \leq \max\{C_i(K_{T_i}), C_i^* + 1\}. \tag{34}$$

根据收敛结果 n，emma 3 可得 iC_i(的一个上界，前提是 $\|\Delta_i\|$ 和 η 足够小。

证明：由式 (25) 和引理 9，C_i 的收敛性可写为

$$C_i(K_{t+1}) - C_i^* \leq (1 - \frac{\eta}{2\mu_i})(C_i(K_t) - C_i^*) + \eta \bar{p}_3 \bar{p}_6 \|\Delta_i\|.$$

根据条件 $\|\Delta_i\| \leq \frac{1}{2\mu_i p_3 p_6}$，由 (I-B) 可得

$$C_i(K_{t+1}) - C_i^* \leq (1 - \frac{\eta}{2\mu_i})(C_i(K_t) - C_i^*) + \frac{\eta}{2\mu_i}. \tag{35}$$

假设 $K_t \in \mathcal{S}_{i0}$。若 $C_i(K_t) \geq C_i^* + 1$，则 (35) 变为

$$\begin{aligned} C_i(K_{t+1}) - C_i(K_t) &\leq -\frac{\eta}{2\mu_i}(C_i(K_t) - C_i^*) + \frac{\eta}{2\mu_i} \\ &\leq -\frac{\eta}{2\mu_i} + \frac{\eta}{2\mu_i} = 0 \end{aligned} \tag{36}$$

若 $C_i(K_t) \leq C_i^* + 1$，则 (35) 变为

$$\begin{aligned} C_i(K_{t+1}) - C_i^* &\leq (1 - \frac{\eta}{2\mu_i})(C_i(K_t) - C_i^*) + \frac{\eta}{2\mu_i} \\ &\leq 1 - \frac{\eta}{2\mu_i} + \frac{\eta}{2\mu_i} = 1 \end{aligned} \tag{37}$$

由于 $C_i(K_{T_i}) \in \mathcal{S}_i$，结合 (36) 和 (37)，通过归纳可得 $C_i(K_t) \le \max\{C_i(K_{T_i})\ C_i^* + 1\} =: C_i$，从而完成证明。 ■

现在，我们证明 $C_{i+1}(K_t)$ 对于$t \in]T_i,\ T_i + L]$有上界。

引理 *11*：如果引理 3 的条件成立，那么

$$\overline{C}_{i+1} \le \overline{C}_i(1 + \overline{p}_2\|\Delta_i\|). \tag{38}$$

引理 11 给出了新系统 C_{i+1} 的代价上界，该上界与 C_i 及其差值 $\|\Delta_i\|$ 成正比。

证明：：根据定义，

$$|C_{i+1}(K) - C_i(K)| = \left|\mathrm{Tr}\left((Q + K^\top RK)(\Sigma_{i+1} - \Sigma_i)\right)\right|.$$

根据引理 7，我们知道 $\|\Sigma_{i+1} - \Sigma_i\| \le p_2\|\Delta_i\|$.因此，

$$\left|\ (\ Q + K^\top RK\ \ \Sigma_{i+1} - \Sigma_i\ )\right| \le\ \ Q\ + \mathrm{Tr}\ (\ )\ K^\top RK\ \|\Sigma_{i+1} - \Sigma_i\| \le\ \ Q + K^\top RK\ p_2\|\Delta_i\|$$

$(\ )\ \mathrm{Tr}(\ )$，$\mathrm{Tr}(\ )$。根据 Σ 的定义，$\sigma(\Sigma) \ge 1$，这意味着

$$\begin{aligned} C(K) &\ge \mathrm{Tr}(Q + K^\top RK)\underline{\sigma}(\Sigma), \\ \frac{C(K)}{\underline{\sigma}(\Sigma)} &\ge \mathrm{Tr}(Q + K^\top RK), \\ C(K) &\ge \mathrm{Tr}(Q + K^\top RK). \end{aligned} \tag{39}$$

由于 C_i 是任意 K_t 的上界，因此 $\mathrm{Tr}(Q + K_t^\top RK_t) \le C_i$. 综合以上结果，我们得到

$$\begin{aligned} |C_{i+1}(K_t) - C_i(K_t)| &\le \mathrm{Tr}(Q + K_t^\top RK_t)\|\Sigma_{i+1} - \Sigma_i\| \\ &\le \overline{C}_i\overline{p}_2\|\Delta_i\|. \end{aligned} \tag{40}$$

我们得出 $C_{i+1} \le \overline{C}_i(1 + \overline{p_2}\|\Delta_i\|)$，它等于 (38)。引理 11 的证明至此完成。 ■

C. 定理 *1* 的证明（状态稳定性）

我们证明在区间$t \in]T_i,\ T_i + L]$内具有序贯强稳定性。

引理 *12*（引理 *18, [23]*）：策略 $K \in \mathcal{S}$ 是 $(\kappa,\ \alpha)$-强稳定的，且

$$\kappa = \sqrt{\frac{C(K)}{\min\{\sigma(R), \sigma(Q)\}}}, \quad \alpha = 1 - \sqrt{1 - \frac{1}{\kappa^2}}.$$

利用引理 12 和 C_{i+1}，注意到每个 $K_t,\ t \in]T_i,\ T_i + L]$ 都是 $(\kappa_{i+1},\ \alpha_{i+1})$-强稳定的。

回顾 $p_i, i \in \{1, 2, .., 7\}$ 的定义，我们有引理 *13*：如果

$$\|\Delta_i\| \le \left\{\frac{1}{2\mu_i\overline{p}_3\overline{p}_6}, \overline{p}_1\right\}, \eta = \min\left\{\frac{\overline{\alpha}_1}{2\overline{\kappa}_1^2\overline{p}_5\overline{p}_8}, \frac{1}{\overline{l}_i}, \frac{p_4^*}{\overline{p}_8}, \overline{p}_7\right\}, \tag{41}$$

那么，对于 $t \in]T_i, T_i + L]$，策略序列 K_t 是 $(\kappa_{i+1}, \alpha_{i+1})$-强稳定的。

证明：证明过程与 [23] 中的引理 17 相同。 ■

D. 状态收敛

引理 *14*：如果满足条件 (41)，则可保证序贯强稳定性，且

$$\|x_t\| \le \overline{\kappa}_{i+1}(1 - \frac{\overline{\alpha}_{i+1}}{2})^{t-T_i-1}\|x_{T_i}\| + \frac{2\overline{\kappa}_{i+1}}{\overline{\alpha}_{i+1}}\max\|B_{i+1}e_t\|$$

对于 $t \in]T_i,\ T_i + L]$。

证明：证明过程与 [23] 中的引理 18 相同。 ■

引理 *15*：存在依赖于 (κ_i, α_i) 的常数 ν_1, ν_2, ν_3，对于 $\tau = [0, \ldots, n]$，使得当 $\kappa > 1$、$0 < \alpha < 1$ 和 $t \in [t_0, T_{i+1}]$ 为常数时，序列 K_t 是 (κ, α)-强稳定的，且状态有界

$$\|x_t\| \le \nu_1(1 - \frac{\nu_2}{2})^{t-t_0-i}\|x_{t_0}\| + \frac{2\nu_3}{\nu_2}\max_{T_{i-1} \le j < t}\|B_j e_j\|.$$

证明：由 [23], 可知，对于 $t \in [T_{i-1} + L,\ T_i]$，序列 K_t 是（$\bar{\kappa}_i,\ \bar{\alpha}_i$）-强稳定的（与（$\bar{\kappa}_i,\ \bar{\alpha}_i$）相关联）。进一步，对于 $t \in [T_i + L, T_{i+1}]$，序列 K_t 是（$\bar{\kappa}_{i+1},\ \bar{\alpha}_{i+1}$）-强稳定的（与（$\bar{\kappa}_{i+1},\ \bar{\alpha}_{i+1}$）相关联）。

$$\begin{aligned} &\|x_t\| \le \bar{\kappa}_i(1 - \bar{\alpha}_i/2)^{t-T_{i-1}-1}\|x_{T_{i-1}}\| \\ &+ \bar{\kappa}_i \sum_{j=T_{i-1}}^{t-1}(1 - \bar{\alpha}_i/2)^{t-j-1}\|B_i e_j\|, \quad t \le T_i \\ &\|x_t\| \le \bar{\kappa}_{i+1}(1 - \bar{\alpha}_{i+1}/2)^{t-(T_i+L)-1}\|x_{T_i+L}\| \\ &+ \bar{\kappa}_{i+1} \sum_{j=T_i+L}^{t-1}(1 - \bar{\alpha}_{i+1}/2)^{t-j-1}\|B_{i+1}e_j\|, t > T_i + L \end{aligned} \tag{42}$$

结合式 (42) 和引理 14，我们得到

$$\begin{aligned} \|x_t\| &\le \overline{\kappa}_i(1 - \frac{\overline{\alpha}_i}{2})^{t-1}\|x_{T_{i-1}}\| \\ &\quad + \frac{2\overline{\kappa}_i}{\overline{\alpha}_i}\max_{T_{i-1} \le j < T_i}\|B_i e_i\|, t \in]T_{i-1}, T_i] \\ \|x_t\| &\le \overline{\kappa}_i\overline{\kappa}_{i+1}(1 - \frac{\overline{\alpha}_i}{2})^{T_i-1}(1 - \frac{\overline{\alpha}_1}{2})^{t-T_i-1}\|x_{T_{i-1}}\| \\ &\quad + \bar{\kappa}_i \sum_{j=T_{i-1}}^{T_i-1}(1 - \bar{\alpha}_i/2)^{T_i-j-1}\|B_i e_j\| \\ &\quad + \bar{\kappa}_{i+1} \sum_{j=T_i}^{t-1}(1 - \bar{\alpha}_{i+1}/2)^{t-j-1}\|B_{i+1}e_j\| \\ &\quad \le \overline{\kappa}_i\overline{\kappa}_{i+1}(1 - \frac{\overline{\alpha}_i}{2})^{T_i-1}(1 - \frac{\overline{\alpha}_{i+1}}{2})^{t-T_i-1}\|x_{T_{i-1}}\| \\ &\quad + \max\{\overline{\kappa}_i, \overline{\kappa}_{i+1}\} \sum_{j=Ti-1}^{t-1}(1 - \min\{\bar{\alpha}_i, \bar{\alpha}_{i+1}\}/2)^{t-j-1} \\ &\quad \cdot \max\|B_{\{i,i+1\}}e_j\| \\ &\le \overline{\kappa}_i\overline{\kappa}_{i+1}(1 - \frac{\overline{\alpha}_i}{2})^{T_i-1}(1 - \frac{\overline{\alpha}_1}{2})^{t-T_i-1}\|x_{T_{i-1}}\| \\ &\quad + \frac{2\max\{\overline{\kappa}_i, \overline{\kappa}_{i+1}\}}{\min\{\bar{\alpha}_i, \bar{\alpha}_{i+1}\}}\max_{T_{i-1} \le j < t}\|B_{\{i,i+1\}}e_j\|, t \in]T_i, T_{i+1}] \end{aligned}$$

取 $\nu_1 = \kappa_i\kappa_{i+1}$，$\nu_2 =$ 的最小值 $\{\bar{\alpha}_1, \bar{\alpha}_{i+1}\}$ 和 $\nu_3 =$ 的最大值 $\{\kappa_i, \kappa_{i+1}\}$，可得

$$\begin{aligned} \|x_t\| &\le \nu_1(1 - \frac{\nu_2}{2})^{t-T_{i-1}-2}\|x_{T_{i-1}}\| \\ &\quad + \frac{2\nu_3}{\nu_2}\max_{T_{i-1} \le j < t}\|B_{\{i,i+1\}}e_j\|. \end{aligned}$$

我们可以将 ν_1, ν_2 扩展为 $\nu_1 = \prod_0^i \kappa_j$，$\nu_2$ =最小值$\{\alpha_0, \ldots, \alpha_i\}$和 ν_3 =最大值$\{\kappa_0, \ldots, \kappa_i\}$，用于 $i =$

$[0,\dots,n]$。因此，状态范数变为：

$$\|x_t\| \le \nu_1(1-\frac{\nu_2}{2})^{t-t_0-i}\|x_{t_0}\| + \frac{2\nu_3}{\nu_2}\max_{0\le j<t}\|B_j e_j\|. \quad (43)$$

这证明了引理 15。 ■

代入 $\kappa = \prod_0^n \kappa_j(1-\frac{\alpha_j}{2})^{-1}$和 $\alpha = \min\{\alpha_0,\dots,\overline{\alpha}_n\}$，定理 1 的证明得以完成。

E. 推论 *1* 的证明（驻留时间条件）

我们证明了，对于两个连续的模式，它们对应的代价函数之间满足如下关系

$$\overline{C}_{i+1} \le \overline{C}_i(1+\overline{p}_2\|\Delta_i\|). \quad (44)$$

我们还知道，无噪声线性时不变系统（出现在 $\forall t \in]T_{i-1}+L,\ T_i]$ 中）的收敛速率为

$$C_i(K_{t+1}) - C_i^* \le (1-\frac{\eta}{2\mu_i})(C_i(K_t)-C_i^*), \quad (45)$$

并且，由于辨识误差，我们的代价函数有界于

$$C_i \le \min\{C_i(K_{T_i}), C_i^*+1\}. \quad (46)$$

在 $t \in]T_i, T_i+L]$ 期间。为确保无限时域情形下的有界性，我们要求代价函数C_i 在发生新的切换之前至少下降 C_i^*+1 。将此驻留时间条件应用于我们的收敛方程，可得：

$$\begin{aligned} C_i(K_{T_i}) - C_i^* &\le (1-\frac{\eta}{2\mu_i})^{\Delta t_i}(C_i(K_{T_{i-1}+L}) - C_i^*) \\ C_i(K_{T_i}) - C_i^* &\le (1-\frac{\eta}{2\mu_i})^{\Delta t_i}(\overline{C}_i - C_i^*) \le 1, \end{aligned} \quad (47)$$

由此可得

$$\Delta t_i \ge \frac{\log\left(\frac{1}{\overline{C}_i - C_i^*}\right)}{-\log\left(1-\frac{\eta}{2\mu_i}\right)} \approx \frac{2\mu_i}{\eta}\log\left(\frac{1}{\overline{C}_i - C_i^*}\right) \quad (48)$$

对于$\frac{\eta}{2\mu_i} \ll 1$。

参考文献

[1] J. C. Willems、P. Rapisarda、I. Markovsky 和 B. L. De Moor，《关于持续激励的一则说明》，《系统与控制快报》，第54卷第4期，页码325–329，2005年。[2] J. Coulson、J. Lygeros 和 F. D¨orfler，《数据赋能预测控制：浅谈 DeePC》，载于第*18*届欧洲控制会议（*ECC*），2019年，页码307–312。[3] J. Berberich、J. K¨ohler、M. A. M¨uller 和 F. Allg¨ower，《具有稳定性与鲁棒性保证的数据驱动模型预测控制》，*IEEE* 自动控制汇刊，第66卷第4期，页码1702–1717，2020年。[4] C. De Persis 和 P. Tesi，《数据驱动控制公式：稳定化、最优性与鲁棒性》，*IEEE* 自动控制汇刊，第65卷第3期，页码909–924，2019年。[5] H. J. Van Waarde、J. Eising、H. L. Trentelman 和 M. K. Camlibel，《数据信息性：数据驱动分析与控制的新视角》，*IEEE* 自动控制汇刊，第65卷第11期，页码4753–4768，2020年。[6] M. Fazel、R. Ge、S. Kakade 和 M. Mesbahi，《线性二次调节器策略梯度方法的全局收敛性》，载于国际机器学习会议，2018年，页码1467–1476。[7] H. Mohammadi、A. Zare、M. Soltanolkotabi 和 M. R. Jovanovi´c，《无模型线性二次调节器问题中梯度方法的收敛性与样本复杂度》，*IEEE* 自动控制汇刊，第67卷第5期，页码2435–2450，2022年。

[8] A. Tsiamis、I. Ziemann、N. Matni 和 G. J. Pappas，《控制的统计学习理论：有限样本视角》，IEEE 控制系统杂志，第43卷第6期，页码67–97，2023年。[9] R. Drenick 和 R. Shahbender，《自适应伺服机构》，《美国电气工程师学会汇刊》，第二部分：应用与工业，第76卷第5期，页码286–292，1957年。[10] B. D. Anderson，《自适应控制理论的失效及其解决》，Commun. Inf. Syst.，第5卷第1期，页码1–20，2005年。[11] P. A. Ioannou 和 J. Sun，《鲁棒自适应控制》。PTR Prentice-Hall，上萨德尔河，新泽西州，1996年，第1卷。[12] Z. Hou 和 S. Xiong，《论无模型自适应控制及其稳定性分析》，IEEE 自动控制汇刊，第64卷第11期，页码4555–4569，2019年。[13] H. Mania、S. Tu 和 B. Recht，《确定性等价对于线性二次控制是高效的》，载于《神经信息处理系统进展》，第32卷。Curran Associates, Inc.，2019年。[14] Y. Lu 和 Y. Mo，《自适应线性二次调节器的几乎必然遗憾界》，arXiv 预印本 arXiv:2301.05537，2023年。[15] M. Simchowitz 和 D. Foster，《朴素探索对在线线性二次调节器是最优的》，载于《国际机器学习会议》。PMLR，2020年，页码8937–8948。[16] A. Cohen、T. Koren 和 Y. Mansour，《仅凭遗憾即可高效学习线性二次调节器》，载于《国际机器学习会议》。PMLR，2019年，页码1300–1309。[17] X. Dai、C. De Persis 和 N. Monshizadeh，《未知非线性系统数据驱动控制器的在线自适应》，《国际鲁棒与非线性控制》，2025年。[18] J. Teutsch、S. Ellmaier、S. Kerz、D. Wollherr 和 M. Leibold，《直接数据驱动控制的在线自适应策略》，IFAC-PapersOnLine，第56卷第2期，页码644–649，2023年。[19] Z. Wang、G. O. Berger 和 R. M. Jungers，《利用情景优化对未知切换线性系统进行数据驱动控制》，IEEE 自动控制汇刊，第69卷，7325，2024年11月，7310页。[20] M. Rotulo、C. De Persis 和 P. Tesi，《未知切换线性系统数据驱动控制器的在线学习》，Automatica，第145卷，文章编号110519，2022年。[21] S. Liu、K. Chen 和 J. Eising，《针对未知线性时变系统的在线数据驱动自适应控制》，载于2023年第62届IEEE决策与控制会议（CDC）。IEEE，2023年，页码8775–8780。[22] M. Bianchi、S. Grammatico 和 J. Cort´es，《切换与受约束线性系统的数据驱动稳定化》，Automatica，第171卷，文章编号111974，2025年。[23] F. Zhao、A. Chiuso 和 F. D¨orfler，《线性二次调节器的策略梯度自适应控制：间接与直接方法》，arXiv 预印本 arXiv:2505.03706，2025年。[24] F. Zhao、F. D¨orfler、A. Chiuso 和 K. You，《用于线性二次调节器直接自适应学习的数据赋能策略优化》，IEEE 自动控制汇刊，页码1–16，2025年（提前出版）。[25] F. Zhao、R. Leng、L. Huang、H. Xin、K. You 和 F. D¨orfler，《通过输出反馈数据赋能策略优化实现并网变换器的直接自适应控制》，载于2025年欧洲控制会议（ECC），2025年，页码2563–2568。[26] X. Wang、F. Zhao、A. J¨urisson、F. D¨orfler 和 R. S. Smith，《通过数据赋能策略优化统一实现气动弹性颤振与载荷控制》，IEEE 航空航天与电子系统汇刊，第61卷第5期，页码11 437–11 449，2025年。[27] N. Persson、F. Zhao、M. Kaheni、F. D¨orfler 和 A. V. Papadopou-los，《一种用于自主自行车控制的自适应数据赋能策略优化方法》，arXiv 预印本 arXiv:2502.13676，2025年。[28] Z. Sun 和 S. S. Ge，《切换线性系统：控制与设计》。Springer，2005年。[29] F. D¨orfler、P. Tesi 和 C. De Persis，《论直接数据驱动线性二次调节器设计的确定性等价方法》，IEEE 自动控制汇刊，第68卷第12期，页码7989–7996，2023年。[30] F. Wang 和 L. Janson，《线性二次自适应控制的精确渐近》，《机器学习研究杂志》，第22卷第1期，页码12 136–12 247，2021年。[31] A. Cohen、A. Hasidim、T. Koren、N. Lazic、Y. Mansour 和 K. Talwar，《在线线性二次控制》，载于《国际机器学习会议》。PMLR，2018年，页码1029–1038。