

具有能量约束的线性二次高斯系统的最优控制与能量调度

杨修远

上海理工大学理学院, 上海

收稿日期: 2024年3月7日; 录用日期: 2024年3月28日; 发布日期: 2024年4月30日

摘要

本文考虑了一种带有能量收集约束的能量传输反馈线性二次高斯(LQG)系统的最优控制器合成。该系统面临着选择传输能量给控制器以及每个能量操作成本和能量收集约束的选择。目标是共同选择传输能量和控制器, 以在控制性能和成本之间保持最佳平衡。在一定的假设下, 这个问题可以分解为两个优化问题: 一个用于最优控制器合成, 另一个用于最优传输能量选择。最优控制器合成子问题由Riccati方程描述, 而最优传输能量选择策略则通过解决某个马尔可夫决策过程(MDP)来找到。最后, 我们通过仿真验证了上述方法的有效性。

关键词

LQG最优控制, 最优能量传输, 马尔可夫 - 决策过程

Optimal Control and Energy Scheduling for Linear-Quadratic-Gaussian System with Energy Harvesting Constraints

Xiuyuan Yang

College of Science, University of Shanghai for Science and Technology, Shanghai

Received: Mar. 7th, 2024; accepted: Mar. 28th, 2024; published: Apr. 30th, 2024

Abstract

In this paper, we consider the optimal controller synthesis of an energy-transmission-feedback

Linear-Quadratic Gaussian (LQG) system with energy harvesting constraints. The system is presented with the choice of transmission energy to the controller along with the cost of operating each energy and energy harvesting constraints. The objective is to jointly select the transmission energy and the controller that would maintain an optimal balance between the control performance and the cost. Under certain assumptions, this problem can be solved in two optimization problems: one for optimal controller synthesis and the other for optimal transmission energy selection. The optimal controller synthesis subproblem is characterized by Riccati equations and the optimal transmission energy selection policy is found by solving a certain Markov-Decision-Process (MDP). In the end, we verify the effectiveness of the above method through a simulation.

Keywords

LQG Optimal Control, Optimal Energy Transmission, Markov-Decision-Process

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

由于网络、通信技术、计算和控制的迅速发展, 这些领域在各种系统中将网络与实际世界紧密结合起来。对这些系统的控制需要在共享通信网络上交换不受压抑且可靠的信号。从[1] [2] [3]可以看出, 在一些资源约束下, 提高系统性能和可靠性将带来挑战。

LQG 系统一直具有广泛的应用, 在经典的线性二次高斯(LQG)系统中, 噪声是不可避免, 一般而言, 噪声被假设服从高斯分布。正如[4]所示, 在无限时域的 LQG 系统中, 作者展示了控制成本和通信数据速率之间的权衡。关于在通信约束下控制 LQG 系统的问题已经在多个研究中进行了探讨, 如[5] [6]。例如, [6]表明最优控制器并不表现出分离原则; 但在[5]中, 作者提供了控制器表现出分离原则的必要条件, 但是控制器和能量选择器的最优结构尚未得到充分研究, 因而, 系统控制策略和能量传输功率选择策略的共同设计一直是研究的热点。

与此同时, 能量收集这一研究领域也引起了学者的广泛兴趣[7] [8]。Li 等人[9]的工作探讨了使用能量收集传感器进行远程状态估计的传感器传输能量控制, 采用连续时间方法和扰动分析。在此基础上, 彭等人[10]表明了存在最优的确信性和稳态功率传输分配策略。在实践中, 由于实际电路的限制, 同时进行能量收集和数据包传输可能无法实现[11]。能量收集和信息传输的方法通常分为两个主要的常见类别: 一种是功率分割, 即使用接收到的部分能量进行能量收集, 其余部分用于信息处理; 另一种是时间切换方案, 接收器在传输信息和获取能量之间切换时间[12]。

基于[5]中提出的分离原则, 作者在[13]中提出了一个新的框架, 不再设计能量选择器, 而是是否可以从给定的能量选择器集合中找到最优的传输能量水平。在本文中, 我们遵循这一框架, 分别设计控制器和能量选择器。最后, 我们考虑了一个场景, 即控制系统可以从给定的能量选择器中选择最优的能量水平来传输其测量结果。

本文的贡献是: 1) 通过考虑能量成本和能量收集约束, 研究了具有一组能量传输水平的 LQG 最优控制问题; 2) 利用后向归纳法研究出最优控制器结构, 将控制器合成问题与能量选择问题分离; 3) 将最优功率选择问题表示为约束 MDP, 并证明了存在最优确定性功率传输选择策略。

2. 问题定式化

让我们考虑一个定常的 LQG 系统:

$$X_{t+1} = AX_t + BU_t + W_t \tag{1}$$

$X_t \in \mathbb{R}^n$ 是系统状态, $U_t \in \mathbb{R}^m$ 是控制信息, A 和 B 是相容维数的常数矩阵。 $\{W_t\}_{t \in \mathbb{N}_0}$ 是一个具有零均值高斯分布的独立同分布噪声序列 $W_t \sim \mathcal{N}(0, \mathcal{W})$ 。初始状态 X_0 是一个符合 $\mathcal{N}(\mu_0, \Sigma)$ 的高斯随机变量分布。

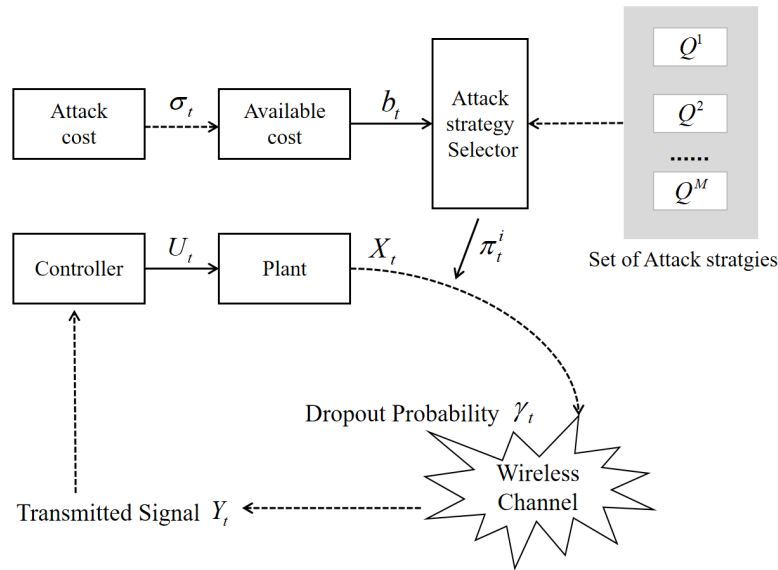


Figure 1. Control system flow chart
图 1. 控制系统流程图

如图 1 所示, 我们将传输能量分为 M 个级别, 并从 M 个能量级别中选择传输能量 p_t 。第 i 个传输能量级别提供的传输能量用 $Q^i (i \in \{1, 2, \dots, M\})$ 表示, 并且让 $\lambda(Q^i) = \lambda_i \in \mathbb{R}_+$ 表示与之相关的成本。在一个无线数据包丢失通信信道上, 系统进行状态估计并将估计值发送到控制器。假设无线信道是一个加性白噪声高斯信道(AWGN), 其比特错误率(BER)与传输能量 p_t 的关系由[10]所描述:

$$\text{BER} = 2\Phi\left(\sqrt{(\beta p_t)/(S_0 K)}\right)$$

其中 $\Phi(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty \exp(-t^2/2) dt$, $\beta > 0$ 是一个常数, S_0 为噪声功率的频谱密度, K 为无线信道带宽。

在本文中, 我们考虑一个能量受限、所需传输信息量通常较小的场景。为方便起见, 我们假设每个数据包只包含一个比特(例如用于奇偶校验)。在传输过程中, 如果数据包丢失, 则接收端将无法正确验证奇偶校验, 即数据包丢失等于比特错误率(BER)。数据检查方法可以检测数据包错误, 这些方法在[9]中有描述。我们可以利用一个二元随机过程 $\gamma(t)$ 来描述传输过程: 如果控制器接收到的信号无误, 则 $\gamma(t) = 1$, 否则 $\gamma(t) = 0$ (此情况可视为丢失)。传输信号的丢失概率由 γ_t 表示, 定义为

$$\gamma_t \triangleq \Pr[\gamma(t) = 0] = 2\Phi\left(\sqrt{\beta \frac{p_t}{S_0 K}}\right), \text{ 其中控制器可用的测量值表示为:}$$

$$Y_t = \gamma(t) X_t + (1 - \gamma(t))(AX_{t-1} + BU_{t-1}). \tag{2}$$

对于环境无线射频信号,无线信息发射器和能量接收器分别具有两种明显不同的功率灵敏度。此外,系统需要在传输信息和收集能量之间进行时间切换。因此,在商业可用的电路中,时间切换方案优于功率分割方案,因此我们采用时间切换方案。

将从起始到步骤 t 结束的收集到的能量表示为 σ_t 。假定能量收集过程 σ_t 是一个离散时间稳态一阶齐次马尔可夫过程。将电池的最大能量存储容量表示为 $C \geq p_t$ 。选择器不能同时收集能量和传输数据,我们有

$$b_t = \begin{cases} \min\{b_t + \sigma_t, C\}, & p_t = 0, \\ b_t - p_t, & 0 < p_t \leq b_t. \end{cases}$$

用于传输能量选择器的新决策变量 π_t^i 表示如下:如果在时间 t 使用第 i 个传输能量,则 $\pi_t^i = 1$, 否则 $\pi_t^i = 0$ 。因此,向量 $\pi_t \triangleq [\pi_t^1, \pi_t^2, \dots, \pi_t^M] \in \{0,1\}^M$, 描述了在时间 t 的传输能量切换方案的决策。因为传输能量选择器在每个时间步只选择一个传输能量,所以我们有 $\sum_{i=1}^M \pi_t^i = 1$ 对于所有 $t \in \mathbb{N}_0$ 。

将集合定义为 $\mathcal{X}_t \triangleq \{X_0, X_1, \dots, X_t\}$, $\mathcal{Y}_t \triangleq \{Y_0, Y_1, \dots, Y_t\}$, $\mathcal{U}_t \triangleq \{U_0, U_1, \dots, U_t\}$ 和 $\Pi_t \triangleq \{\pi_0, \pi_1, \dots, \pi_t\}$, 分别表示状态历史、测量历史、控制历史和传输选择历史。

控制器在时间 t 可以接收到的信息是 $\mathcal{F}_t^c = \{\mathcal{Y}_t, \mathcal{U}_{t-1}, \Pi_t\} = \mathcal{F}_{t-1}^c \cup \{Y_t, U_{t-1}, \pi_t\}$, 其中 $\mathcal{F}_0^c = \{Y_0, \pi_0\}$ 。根据先前的定义,可以发现接纳控制策略可以视为一个映射: $\mathcal{F}_t^c \rightarrow \mathbb{R}^m$ 。我们用 ξ_t^u 来表示这样的策略。另一方面,在时间 t 可用于能量传输选择器的信息是 $\mathcal{F}_t^e = \{\mathcal{X}_t, \mathcal{Y}_{t-1}, \mathcal{U}_{t-1}, \Pi_{t-1}\} = \mathcal{F}_{t-1}^e \cup \{X_t, Y_{t-1}, U_{t-1}, \pi_{t-1}\}$, 其中 $\mathcal{F}_0^e = \{X_0\}$ 。用于选择发送器的接纳策略也可以视为一个映射: $\mathcal{F}_t^e \rightarrow \{0,1\}^M$ 。我们用 ξ_t^π 来表示这样的策略。因此,在一个时间步内的决策过程如下:

$$\dots \rightarrow \mathcal{F}_t^e \xrightarrow{\xi_t^\pi} \pi_t \rightarrow Y_t \rightarrow \mathcal{F}_t^c \xrightarrow{\xi_t^u} U_t \rightarrow X_{t+1} \rightarrow \mathcal{F}_{t+1}^e \rightarrow \dots$$

由控制器和能量传输选择器共同最小化的成本函数是一个有限时间段的二次准则,给出为如下:

$$J(\xi^u, \xi^\pi) = E \left[\sum_{t=0}^{T-1} (X_t^\top Q_1 X_t + U_t^\top R U_t + \pi_t^\top \wedge) + X_T^\top Q_2 X_T \mid U_t = \xi_t^u(\mathcal{F}_t^c), \pi_t = \xi_t^\pi(\mathcal{F}_t^e) \right] \quad (3)$$

其中, $\wedge = [\lambda_1, \lambda_2, \dots, \lambda_M]^\top$, $R \succ 0$, $Q_1 \succeq 0$, $Q_2 \succeq 0$, ξ^π 表示整个序列 $\{\xi_0^\pi, \xi_1^\pi, \dots, \xi_{T-1}^\pi\}$, 类似地, ξ^u 也类似定义。

3. 最优控制和能量选择

3.1. 基于分离原理的最优控制

为了解决前一节提出的问题,我们需要找到最小化可接受策略中成本函数(3)的最优映射 ξ^{u^*} 和 ξ^{π^*} :

$$\left(\xi^{u^*}, \xi^{\pi^*} \right) = \arg \min_{\xi^u, \xi^\pi} J(\xi^u, \xi^\pi). \quad (4)$$

与[6] [14]中的基于传输反馈的控制方法不同,我们将在时间 t 传输 W_{t-1} 而不是 X_t 。容易发现 W_{t-1} 可以很容易地从 X_t 、 X_{t-1} 和 U_{t-1} 的值计算得出。

现在让我们定义 $\hat{X}_t = E[X_t \mid \mathcal{F}_{t-1}^c]$, 称为 X_t 的预测; $\tilde{X}_t = E[X_t \mid \mathcal{F}_t^c]$, 称为 X_t 的更新。因此,我们记 $\hat{\omega}_t(\pi_{t+1}) = E[W_t \mid \mathcal{F}_{t+1}^c]$ 。

利用(1), 并且 U_t 是 \mathcal{F}_t^c 可测的, 所以可以得到

$$\hat{X}_{t+1} = A\tilde{X}_t + BU_t \quad (5)$$

$$\tilde{X}_t = E[X_t | \mathcal{F}_t^c] = E[AX_{t-1} + BU_{t-1} + W_{t-1} | \mathcal{F}_t^c] = A\tilde{X}_{t-1} + BU_{t-1} + \hat{w}_{t-1}(\pi_t) = \hat{X}_t + \hat{w}_{t-1}(\pi_t). \quad (6)$$

让我们定义误差 $\Delta_t = X_t - \tilde{X}_t$ ，则有

$$\Delta_{t+1} = A\Delta_t + W_t - \hat{w}_t = \dots = A^{t+1}\Delta_0 + \sum_{k=0}^t A^{t-k}(W_k - \hat{w}_k) \quad (7)$$

其中 $\Delta_0 = W_{-1} - \hat{w}_{-1}(\pi_0)$ 。状态估计误差 Δ_t 通过变量 $\{\hat{w}_{-1}, \dots, \hat{w}_{t-1}\}$ 依赖于序列 $\{\pi_0, \dots, \pi_t\}$ 。它不依赖于控制策略 ξ^u 。

然而，在传输过程中发生数据包丢失时，控制器无法接收到 \hat{w}_t ，我们有以下函数：

$$\tilde{X}_t = \begin{cases} g(\tilde{X}_{t-1}), & \gamma(t) = 0 \\ \tilde{X}_t, & \gamma(t) = 1 \end{cases}$$

其中 $g(X) \triangleq AX + BU$ 。显然， \tilde{X}_t 的值包含在一个可数无限集合中： $\{\tilde{X}_t, g(\tilde{X}_{t-1}), g^2(\tilde{X}_{t-2}), \dots\}$ 。我们可以采用随机变量 τ_t 来表示从上次成功传输到时间 t 的持续时间，即

$$\tau_t = t - \max\{t^* : \gamma(t^*) = 1, 0 \leq t^* \leq t\}, \quad (8)$$

我们有 $\tilde{X}_t = g^{\tau_t}(\tilde{X}_{t-\tau_t})$ 。

上述分析暗示了控制器和发送器选择之间的分离结构。接下来，我们将正式展示出这个问题的分离原理的出现。与成本函数(3)相关联，让我们定义值函数：

$$V_k(\mathcal{F}_k^c, \mathcal{F}_k^e) = \min_{\xi_k^u, \xi_k^\pi} E \left[\sum_{t=k}^{T-1} (X_t^\top Q_1 X_t + U_t^\top R U_t + \pi_t^\top \wedge) + X_T^\top Q_2 X_T \mid U_t = \xi_t^u, \pi_t = \xi_t^\pi, t = k, \dots, T-1 \right]. \quad (9)$$

通过使用动态规划原理，表达式(9)可以重写为：

$$V_k(\mathcal{F}_k^c, \mathcal{F}_k^e) = \min_{\xi_k^u, \xi_k^\pi} E \left[\sum_{t=k}^{T-1} (X_t^\top Q_1 X_t + U_t^\top R U_t + \pi_t^\top \wedge) + V_{k+1}(\mathcal{F}_{k+1}^c, \mathcal{F}_{k+1}^e) \mid U_k = \xi_k^u, \pi_k = \xi_k^\pi \right]. \quad (10)$$

如果 ξ_k^u 和 ξ_k^π 最小化(10)的右侧，那么可得 $U_k^* = \xi_k^{u*}(\mathcal{F}_k^c)$ 、 $\pi_k^* = \xi_k^{\pi*}(\mathcal{F}_k^e)$ 。从(9)中，我们也得到：

$$\min_{\xi_k^u, \xi_k^\pi} J(\xi^u, \xi^\pi) = E[V_0(\mathcal{F}_0^c, \mathcal{F}_0^e)] \quad (11)$$

在(11)中的期望是针对随机变量 \mathcal{F}_0^c 和 \mathcal{F}_0^e 进行的。为了保持后续分析中符号的简洁性，我们将 $V_k(\mathcal{F}_k^c, \mathcal{F}_k^e)$ 写为：

$$V_k(\mathcal{F}_k^c, \mathcal{F}_k^e) = \min_{\xi_k^u, \xi_k^\pi} E_{\mathcal{F}_k} \left[(X_k^\top Q_1 g^{\tau_k} X_k + U_k^\top R U_k + \pi_k^\top \wedge) + V_{k+1}(\mathcal{F}_{k+1}^c, \mathcal{F}_{k+1}^e) \right]$$

在这种情况下， $E_{\mathcal{F}_k}[\cdot]$ 表示给定组合信息集 $\mathcal{F}_k = \{\mathcal{F}_k^c, \mathcal{F}_k^e\}$ 的条件期望。请注意，信息集 \mathcal{F}_k^e 包含状态 X_k 的实现 x_k 。以下定理表征了所有 $k = 0, 1, \dots, T-1$ 的最优策略 ξ_k^{u*} 。

定理 1: 在时间 k 给控制器的信息为 \mathcal{F}_k^c 的条件下，能最小化(10)右侧的最优控制策略为 $\xi_k^{u*} : \mathcal{F}_k^c \rightarrow \mathbb{R}^m$ ，其具有以下结构：

$$U_k^* = \xi_k^{u*}(\mathcal{F}_k^c) = -G_k g^{\tau_k} \left(E[X_{k-\tau_k} | \mathcal{F}_k^c] \right). \quad (12)$$

更具体的表达形式如下：

$$U_k^* = - \left[(I - A)^{-1} (I - A^{\tau_k}) B + G_k^{-1} \right] A^{\tau_k} E[X_k | \mathcal{F}_k^c] \quad (13)$$

对于所有的 $k = 0, 1, \dots, T-1$,

$$\begin{aligned} G_k &= (R + B^\top P_{k+1} B)^{-1} B^\top P_{k+1} A, \\ P_k &= Q_1 + A^\top P_{k+1} A - G_k^\top (R + B^\top P_{k+1} B) G_k, \\ P_T &= Q_2. \end{aligned} \tag{14}$$

证明：为了简洁起见，我们将使用 g^{τ_k} 来代替 $g^{\tau_k}(x_{k-\tau_k})$ ，并且使用 $g_\Delta^{\tau_k}$ 来代替 $g^{\tau_k}(\Delta_{k-\tau_k})$ 。该定理的证明基于动态规划原理。思路是验证与最优控制问题相关的值函数是

$$V_k(\mathcal{F}_k^c, \mathcal{F}_k^e) = g^{\tau_k \top} P_k g^{\tau_k} + C_k + o_k \tag{15}$$

其中， x_k 是状态 X_k 的实现， P_k 如(14)所示，对于所有的 $k=0,1,\dots,T-1$ ，

$$C_k = \min_{\{\xi_t^u\}_{t=k}^{T-1}} E_{\mathcal{F}_k^e} \left[\sum_{t=k}^{T-1} g_\Delta^{\tau_t \top} H_t g_\Delta^{\tau_t} + \pi_t^\top \wedge \right]. \tag{16}$$

矩阵 $H_k \in \mathbb{R}^{n \times m}$ 和标量 o_k 由以下给出：

$$\begin{aligned} H_k &= G_k^\top (R + B^\top P_{k+1} B) G_k, \\ o_k &= o_{k+1} + \text{tr}(P_{k+1} \mathcal{W}), \\ o_T &= 0. \end{aligned} \tag{17}$$

可以验证， $V_{T-1}(\mathcal{F}_{T-1}^c, \mathcal{F}_{T-1}^e)$ 确实具有形式(15)。为了验证(15)在时间 k 也成立，我们使用反向归纳，并假设(15)对某个时间 $k+1$ 成立。为此我们得到：

$$V_k(\mathcal{F}_k^c, \mathcal{F}_k^e) = \min_{\xi_k^u, \xi_k^z} E_{\mathcal{F}_k} \left[(g^{\tau_k \top} Q_1 g^{\tau_k} + U_k^\top R U_k + \pi_k^\top \wedge) + g^{\tau_{k+1} \top} P_{k+1} g^{\tau_{k+1}} + C_{k+1} + o_{k+1} \right].$$

利用方程(1)和(14)，我们可以得到：

$$V_k(\mathcal{F}_k^c, \mathcal{F}_k^e) = \min_{\xi_k^u, \xi_k^z} E_{\mathcal{F}_k} \left[\|U_k + G_k g^{\tau_k}\|_{(R+B^\top P_{k+1} B)}^2 + g^{\tau_k \top} P_k g^{\tau_k} + \pi_k^\top \wedge + \text{tr}(P_{k+1} \mathcal{W}) + C_{k+1} + o_{k+1} \right]. \tag{18}$$

最小化(18)的最优控制 U_k^* 是：

$$U_k^* = \xi_k^{u^*}(\mathcal{F}_k^c) = -G_k g^{\tau_k} \left(E[X_{k-\tau_k} | \mathcal{F}_k^c] \right). \tag{19}$$

在(18)中代入来自(19)的最优控制后，我们可以得到：

$$\begin{aligned} &V_k(\mathcal{F}_k^c, \mathcal{F}_k^e) \\ &= \min_{\xi_k^z} E_{\mathcal{F}_k^e} \left[g_\Delta^{\tau_k \top} (G_k^\top (R + B^\top P_{k+1} B) G_k) g_\Delta^{\tau_k} + \pi_k^\top \wedge + C_{k+1} \right] + g^{\tau_k \top} P_k g^{\tau_k} + \text{tr}(P_{k+1} \mathcal{W}) + o_{k+1} \\ &= \min_{\xi_k^z} E_{\mathcal{F}_k^e} \left[g_\Delta^{\tau_k \top} H_k g_\Delta^{\tau_k} + \pi_k^\top \wedge + C_{k+1} \right] + g^{\tau_k \top} P_k g^{\tau_k} + o_k \\ &= g^{\tau_k \top} P_k g^{\tau_k} + C_k + o_k. \end{aligned}$$

因此，值函数确实具有形式(15)，而时间 $k=0,1,\dots,T-1$ 处的最优控制由(19)给出。

[注] 1：分离控制问题的最优反馈控制器由[5]中的定理 1 给出，它表明如果策略满足[5]中方程(4)的结构，则存在一个分离原理。在本文中，方程(14)与[5]中方程(4)的结构完全相同，这表明我们提出的最优控制策略满足分离原理。

最优控制在 \tilde{X}_k 中是线性的，最优增益是 $-G_k$ ，可以在不知道 ξ^{Π^*} 的情况下离线计算。与[13]不同，本文考虑了延迟： $\tau_t = t - \max\{t^* : \gamma(t^*) = 1, 0 \leq t^* \leq t\}$ 。在延迟存在的情况下，控制器可用的信息将受到影响，因为某些测量到达将被延迟，从而状态估计将受到影响。定理 1 考虑了延迟，并为这种情况提供了

最优控制器结构: $U_k^* = -\left[(I-A)^{-1}(I-A^{\tau_k})B + G_k^{-1}\right]A^{\tau_k}E\left[X_k | \mathcal{F}_k^c\right]$ 。

3.2. 最佳能量选择和 MDP

上述解决过程让人联想到马尔可夫决策过程(MDP)。接下来,我们将构建一个新的状态空间 \mathcal{S} 和一个新的动作空间 \mathcal{A} , 以创建一个标准的 MDP 问题。在每个时间 t , MDP 的状态被表示为 $s_t = (\tau_{t-1}, b_t)$, 其取值范围是状态集合 $\mathcal{S} = \mathbb{Z}_+ \times [0, C]$ 。将 $\mathbb{A} = \{0, p_t(1), \dots, p_t(M)\}$ 表示为传输能量动作集, 将 $\mathbb{A}(s)$ 表示为状态 $s \in \mathcal{S}$ 的可接受动作集。

定义 $Z_t \triangleq X_t X_t^\top$, 如果发生数据包丢失, 则 $\tilde{Z}_t = \tilde{X}_t \tilde{X}_t^\top$ 、 $Z_t = A\tilde{Z}_t A^\top + \mathcal{W}$ 。因此, 一个阶段的奖励是:

$$r(s_t, p_t) = \text{Tr}\left[f^{\tau_{t-1}}(Z_{t-1-\tau_{t-1}})\right] \quad (20)$$

其中 $f(Z_t) \triangleq AZ_t A^\top + \mathcal{W}$, 而(20)与传输能量动作 p_t 无关。对于当前状态 s_t , 选择器在时间 t 选择动作 $p_t = p_{s_t}(i) \leq b_t$, 并获得奖励 $r(s_t, p_t)$, 然后进入时间 $t+1$ 。根据(2), 如果时间 t 的状态之一是 τ_{t-1} , 那么时间 $t+1$ 的状态 τ_t 的值只能在两个值之间选择: 以概率 $1-\gamma_t$ 选择 0; 以概率 γ_t 选择 $\tau_{t-1}+1$ 。因此, MDP 的状态转移概率可以写成:

$$t(s_{t+1} | s_t, p_t) = t(\tau_t | s_t, p_t) \times t(b_{t+1} | s_t, p_t), \quad (21)$$

其中 $t(\tau_t | s_t, p_t)$ 和 $t(b_{t+1} | s_t, p_t)$ 的定义与[10]中相同。

利用马尔可夫决策, 前面的成本函数可以重写为:

$$J(\mathcal{U}, \Pi) = J(s, \xi^\pi) = E\left[\sum_{t=0}^T r(s_t, p_t)\right] \quad (22)$$

其最优值为 $J(s, \xi^{\pi^*}) = \inf_{\xi^\pi \in \Pi} J(s, \xi^\pi)$ 。

寻找最优策略 ξ^{π^*} 的问题可以通过一个 MDP 来解决。将使用策略 $\xi^\pi \in \Pi$ 的每个状态在状态空间 \mathcal{S} 中的期望成本表示为一个函数 $V_{\xi^\pi}: \mathcal{S} \rightarrow \mathbb{R}$ 。因此, 最优值可以通过贝尔曼方程([15], Sec. 8.4.1)解决, 如下所示:

$$J(s, \xi^{\pi^*}) + V_{\xi^{\pi^*}}(s, \xi^{\pi^*}) = \min_{p_t \in \mathbb{A}(s)} \left\{ r(s, p_t) + \sum_{s' \in \mathcal{S}} t(s' | s, p_t) V_{\xi^{\pi^*}}(s') \right\}. \quad (23)$$

因此, 最优选择策略 ξ^{π^*} 可以通过以下方式获得:

$$\xi^{\pi^*} = \arg \min_{p_t \in \mathbb{A}(s)} \left\{ r(s, p_t) + \sum_{s' \in \mathcal{S}} t(s' | s, p_t) J_{\pi^*}(s') \right\},$$

接下来, 我们将证明这个策略的存在。

定理 2: 存在最优确定性稳态策略 $\xi^{\pi^*} \in \Pi$, 满足:

$$J(s, \xi^{\pi^*}) \leq J(s, \xi^\pi), \forall s \in \mathcal{S}, \forall \xi^\pi \in \Pi, \quad (24)$$

通过解决最优贝尔曼方程(23), 我们可以获得最优策略 ξ^{π^*} 。

证明: 让我们考虑相同的策略 ξ^π , 让 ν 是一个正常数, 使得 $Z \leq \nu I$ 和 $\mathcal{W} \leq \nu I$ 。定义一个函数

$\varphi(X) = AXA^\top + \nu I$ 。因此, $f^{\tau_t}(Z_{t-\tau_t}) \leq \varphi^{\tau_t}(\nu I) \leq \nu \sum_{i=0}^{\tau_t} A^i (A^\top)^i$, 由此得到 $\text{Tr}(f^{\tau_t}(Z_{t-\tau_t})) \leq \nu \sum_{i=0}^{\tau_t} (\rho(A))^{2i}$,

其中 ρ 表示矩阵的谱半径。此外, 让我们定义:

$$M(s) = M(\tau_{t-1}, b_t) = \eta \sum_{i=0}^{\tau_{t-1}} (\rho(A))^{2i},$$

其中 $\eta = \nu / (1 - \gamma_{\max} \rho^2(A)) > 0$, 并且 $\gamma_{\max} = \max_{k \in \mathbb{Z}} \{\gamma_k\}$ 。存在非负整数 N_1 和 N_2 , 使得 $\rho(A)^{2(N_1+1)} \geq 1 / (\gamma_{\max} - \gamma_{N_1})$ 且 $N_2 \geq p_t$ 。

对于 $\tau_{t-1} > N_1, b_t > N_2$, 动作是 $0 < p_t \leq b_t$, 那么可得:

$$\begin{aligned} & \sum_{j \in \mathbb{S}} t(j|s, \pi(s)) M(j) - M(s) + r(s, \pi(s)) \\ & \leq \eta + (\gamma_t \eta \rho^2(A) - \eta) \sum_{i=0}^{\tau_{t-1}} (\rho(A))^{2i} + \nu \sum_{i=0}^{\tau_{t-1}} (\rho(A))^{2i} \\ & = \eta + (\gamma_t \eta \rho^2(A) - \eta + \nu) \sum_{i=0}^{\tau_{t-1}} (\rho(A))^{2i}. \end{aligned}$$

由于 $\eta = \nu / (1 - \gamma_{\max} \rho^2(A))$, 我们有 $\gamma_t \eta \rho^2(A) - \eta + \nu \leq 0$ 。从 $\rho(A)^{2(N_1+1)} \geq 1 / (\gamma_{\max} - \gamma_{N_1})$, 可得 $\eta + (\gamma_{N_1} \eta \rho^2(A) - \eta + \nu) (\rho(A))^{2N_1} \leq 0$ 。因此, 根据[16]中的 Assumptions 和 Theorem, 上述定理成立。

4. 总结

在这项工作中, 我们考虑了一个经典的基于能量传输成本的 LQG 问题, 在该模型中需要选择一个最优的能量传输选择器, 以最小化传输和控制性能的综合成本。通过求解与 LQG 问题相关的经典黎卡蒂方程, 我们计算了最优控制增益, 并证明了最优控制器具有分离原理。通过建立一个马尔科夫决策过程, 我们也证明了通过求解一个贝尔曼方程可以得到最优选择器。在今后的研究中, 我们将考虑从能量选择器到无线的数据包丢失现象, 这在实际问题中也具有重要意义。

参考文献

- [1] Cao, X.H., Zhou, X.W., Liu, L. and Cheng, Y. (2015) Energy-Efficient Spectrum Sensing for Cognitive Radio Enabled Remote State Estimation over Wireless Channels. *IEEE Transactions on Wireless Communications*, **14**, 2058-2071. <https://doi.org/10.1109/TWC.2014.2379642>
- [2] Peng, L.H., Cao, X.H., Sun, C.Y., Cheng, Y. and Jin, S. (2018) Energy Efficient Jamming Attack Schedule against Remote State Estimation in Wireless Cyber-Physical Systems. *Neurocomputing*, **272**, 571-583. <https://doi.org/10.1016/j.neucom.2017.07.036>
- [3] Luo, W.C., Lu, P.L., Du, C.K. and Liu, H.K. (2022) Cooperative Output Tracking Control of Heterogeneous Multi-Agent Systems with Random Communication Constraints: An Observer-Based Predictive Control Approach. *IEEE Transactions on Circuits and Systems II: Express Briefs*, **69**, 1139-1143. <https://doi.org/10.1109/TCSII.2021.3107845>
- [4] Kostina, V. and Hassibi, B. (2019) Rate-Cost Tradeoffs in Control. *IEEE Transactions on Automatic Control*, **64**, 4525-4540. <https://doi.org/10.1109/TAC.2019.2912256>
- [5] Borkar, V.S. and Mitter, S.K. (1997) LQG Control with Communication Constraints. In: *Communications, Computation, Control, and Signal Processing* (pp. 365-373), Springer, Berlin. https://doi.org/10.1007/978-1-4615-6281-8_21
- [6] Liu, K., Skelton, R.E. and Grigoriadis, K. (1992) Optimal Controllers for Finite Wordlength Implementation. *IEEE Transactions on Automatic Control*, **37**, 1294-1304. <https://doi.org/10.1109/9.159567>
- [7] Weddell, A.S., Merrett, G.V., Kazmierski, T.J. and Al-Hashimi, B.M. (2011) Accurate Supercapacitor Modeling for Energy Harvesting Wireless Sensor Nodes. *IEEE Transactions on Circuits and Systems II: Express Briefs*, **58**, 911-915. <https://doi.org/10.1109/TCSII.2011.2172712>
- [8] Kang, T., Kim, S., Hyong, C., Kang, S. and Park, K. (2015) An Energy Combiner for a Multi-Input Energy-Harvesting System. *IEEE Transactions on Circuits and Systems II: Express Briefs*, **62**, 911-915. <https://doi.org/10.1109/TCSII.2015.2435231>
- [9] Li, Y.Z., Zhang, F., Quevedo, D.E., Lau, V., Dey, S. and Shi, L. (2017) Power Control of an Energy Harvesting Sensor for Remote State Estimation *IEEE Transactions on Automatic Control*, **62**, 277-290. <https://doi.org/10.1109/TAC.2016.2553128>

-
- [10] Peng, L.H., Cao, X.H. and Sun, C.Y. (2021) Optimal Transmit Power Allocation for an Energy-Harvesting Sensor in Wireless Cyber-Physical Systems. *IEEE Transactions on Cybernetics*, **51**, 779-788. <https://doi.org/10.1109/TCYB.2019.2939420>
- [11] Zhang, R. and Ho, C.K. (2013) MIMO Broadcasting for Simultaneous Wireless Information and Power Transfer. *IEEE Transactions on Wireless Communications*, **12**, 1989-2001. <https://doi.org/10.1109/TWC.2013.031813.120224>
- [12] Song, X., Xu, S.Y., Xie, Z.G. and Han, X.W. (2019) Joint Optimal Power Allocation and Relay Selection Scheme in Energy Harvesting Two-Way Relaying Network. *Future Internet*, **11**, 47. <https://doi.org/10.3390/fi11020047>
- [13] Maity, D. and Tsiotras, P. (2022) Optimal Controller Synthesis and Dynamic Quantizer Switching for Linear-Quadratic-Gaussian Systems. *IEEE Transactions on Automatic Control*, **67**, 382-389. <https://doi.org/10.1109/TAC.2021.3057338>
- [14] Williamson, E. and Kadiman, K. (1989) Optimal Finite Wordlength Linear Quadratic Regulation. *IEEE Transactions on Automatic Control*, **34**, 1218-1228. <https://doi.org/10.1109/9.40767>
- [15] Puterman, M.L. (2005) Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, New York.
- [16] Sennott, L.I. (1989) Average Cost Optimal Stationary Policies in Infinite State Markov Decision Processes with Unbounded Costs. *Operations Research*, **37**, 626-633. <https://doi.org/10.1287/opre.37.4.626>