

The background of the slide features a complex, abstract pattern of thin, dark grey lines. These lines form a dense, overlapping web of circles and arcs, creating a sense of movement and complexity. The pattern is most prominent in the top and bottom sections of the slide, framing a central dark green horizontal band.

随机最优控制理论

Prof. Yuan-Li Cai

Spring 2024

0. Outline

- 1 最优性原理 / 4
- 2 确定性最优控制基础 / 6
- 3 随机动态规划技术 / 54
- 4 连续时间线性二次型高斯问题 / 100
- 5 离散时间线性二次型高斯问题 / 151
- 6 非线性随机系统次优控制方法 / 164

7 本章小结 / 169

随机最优控制理论研究存在随机性或不确定性时如何找到最优控制策略，在航空、航天、航海、金融、经济、机器人以及过程控制等领域有广泛应用。基于最优性原理的动态规划技术是随机最优控制理论中最主要的方法，也是本章重点介绍的内容。

随机最优控制根植于确定性最优控制理论，我们首先介绍确定性最优控制的基本原理。

1. 最优性原理

动态规划是最优控制理论中的重要组成部分，其中核心思想是最优性原理 (principle of optimality)，它指导我们如何找到问题的最优解。通过将复杂问题分解为一系列子问题，并保证每个子问题的解都是最优的，从而最终得到整个问题的最优解。

最优性原理是 Bellman 教授 1957 年提出的，它指出：在任何一个状态下，无论过去的历史如何，做出的选择都应该从当前状态到最终状态的最优选择。(An optimal policy has the property that no matter what the previous

decisions (i.e. controls) have been, the remaining decisions must constitute an optimal policy with regards to the state resulting from those previous decisions.)

最优性原理可以非常直观地用几何曲线来表达。如图1所示，假设 C 点是由 A 点到 B 点的最优路径（黑线）上的任一点，那么由 C 点到 B 点的最优路径仍然是黑线上的 CB 部分。



Figure 1: 最优性原理示意图

2. 确定性最优控制基础

2.1 连续时间系统

考虑如下非线性微分方程描述的系统：

$$\dot{\mathbf{x}} = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t], \quad \mathbf{x}(t_0) = \mathbf{x}_0, \quad t_0 \leq t \leq t_f \quad (1)$$

式中， $\mathbf{x}(t) \in R^{n_x}$ 表示系统的状态向量， $\mathbf{u}(t) \in R^{n_u}$ 表示系统的控制信号。
 t_0 、 t_f 分别表示初始时间和终了时间。

考虑如下形式的（标量）性能指标：

$$J = \varphi[\mathbf{x}(t_f), t_f] + \int_{t_0}^{t_f} L[\mathbf{x}(t), \mathbf{u}(t), t] dt \quad (2)$$

其中， $\varphi(\cdot)$ 、 $L(\cdot)$ 是给定的标量函数。

最优控制问题可描述为：求解容许的时间函数 $\mathbf{u}(t)$ 以最小化（或最大化）性能指标 J 。如果设控制信号的容许集为 \mathcal{U} ，最优控制 $\mathbf{u}^*(t) \in \mathcal{U}$ 对应的性能指标为 J^* ，任意的 $\mathbf{u}(t) \in \mathcal{U}$ 对应的性能指标为 J ，对于最小化问题有

$$J^* \leq J$$

习惯上，我们用上标“*”表示“最优”值。

2.1.1 动态规划方法

一般地, 令

$$J[\mathbf{x}(t), t] \triangleq \varphi[\mathbf{x}(t_f), t_f] + \int_t^{t_f} L[\mathbf{x}(\tau), \mathbf{u}(\tau), \tau] d\tau \quad (3)$$

表示从 t 时刻到最终时刻的性能指标, 则对应的最优指标为

$$J^*[\mathbf{x}(t), t] \triangleq \min_{\substack{\mathbf{u}(\tau) \\ t \leq \tau \leq t_f}} J[\mathbf{x}(t), t] \quad (4)$$

由于

$$\begin{aligned} J^*[\mathbf{x}(t), t] &= \min_{\substack{\mathbf{u}(\tau) \\ t \leq \tau \leq t_f}} \left\{ \varphi[\mathbf{x}(t_f), t_f] + \int_t^{t_f} L[\mathbf{x}(\tau), \mathbf{u}(\tau), \tau] d\tau \right\} \\ &= \min_{\substack{\mathbf{u}(\tau) \\ t \leq \tau \leq t_f}} \left\{ \int_t^{t+\Delta t} L[\mathbf{x}(\tau), \mathbf{u}(\tau), \tau] d\tau + J[\mathbf{x}(t+\Delta t), t+\Delta t] \right\} \end{aligned}$$

根据最优性原理, 有

$$J^*[\mathbf{x}(t), t] = \min_{\substack{\mathbf{u}(\tau) \\ t \leq \tau \leq t+\Delta t}} \left\{ \int_t^{t+\Delta t} L[\mathbf{x}(\tau), \mathbf{u}(\tau), \tau] d\tau + J^*[\mathbf{x}(t+\Delta t), t+\Delta t] \right\} \quad (5)$$

将式中 $J^*[\mathbf{x}(t + \Delta t), t + \Delta t]$ 泰勒级数展开, 可得

$$\begin{aligned} J^*[\mathbf{x}(t + \Delta t), t + \Delta t] = & J^*[\mathbf{x}(t), t] + J_t^*[\mathbf{x}(t), t] \Delta t \\ & + J_x^*[\mathbf{x}(t), t] \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \Delta t + o(\Delta t) \end{aligned}$$

将其代入最优指标 (5), 当 Δt 为小量时, 忽略高阶小量可得

$$\begin{aligned} J^*[\mathbf{x}(t), t] = \min_{\mathbf{u}(t)} \{ & L[\mathbf{x}(t), \mathbf{u}(t), t] \Delta t + J^*[\mathbf{x}(t), t] \\ & + J_t^*[\mathbf{x}(t), t] \Delta t + J_x^*[\mathbf{x}(t), t] \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \Delta t \} \end{aligned}$$

由此可得

$$-J_t^*[\mathbf{x}(t), t] = \min_{\mathbf{u}(t)} \left\{ L[\mathbf{x}(t), \mathbf{u}(t), t] + J_x^*[\mathbf{x}(t), t]^T \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \right\} \quad (6)$$

引入哈密顿函数

$$H[\mathbf{x}(t), \mathbf{u}(t), J_x^*, t] = L[\mathbf{x}(t), \mathbf{u}(t), t] + J_x^*[\mathbf{x}(t), t]^T \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] \quad (7)$$

则得到如下关于 J^* 的偏微分方程:

$$-J_t^*[\mathbf{x}(t), t] = \min_{\mathbf{u}(t)} H[\mathbf{x}(t), \mathbf{u}(t), J_x^*, t] \quad (8)$$

该式即为著名的 哈密顿-雅可比-贝尔曼 (HJB) 方程。

根据 (3)、(4), 不难发现 HJB 方程的边界条件为

$$J^*[\mathbf{x}(t_f), t_f] = \varphi[\mathbf{x}(t_f), t_f] \quad (9)$$

HJB 方程与边界条件共同构成了最优解的充分条件。

2.1.2 变分法

在传统的优化问题中，变分法是最早出现的数学方法，也可以用来解决连续时间系统的最优控制问题。

对于更一般的最优控制问题，除了状态方程 (1) 表示的动态约束外，还可以考虑终端状态约束

$$\psi[x(t_f), t_f] = 0 \quad (10)$$

式中， $\psi(\cdot, \cdot) \in R^{n_\psi}$ 。通常，终了时间 t_f 未必是固定的，这里假设是自由的，需要根据最优性来确定具体取值。

根据拉格朗日乘子法, 可以分别通过乘子 ξ 和 $\lambda(t)$ (也称为协态变量) 将终端约束 (10) 与系统状态方程约束 (1) 引入优化指标 (3), 从而得到扩展的性能指标

$$\begin{aligned} J = & \varphi[\mathbf{x}(t_f), t_f] + \xi^T \psi[\mathbf{x}(t_f), t_f] \\ & + \int_{t_0}^{t_f} \{L[\mathbf{x}, \mathbf{u}, t] + \lambda^T(t) [\mathbf{f}(\mathbf{x}, \mathbf{u}, t) - \dot{\mathbf{x}}]\} dt \end{aligned} \quad (11)$$

定义哈密顿函数

$$H[\mathbf{x}, \mathbf{u}, \lambda, t] = L[\mathbf{x}, \mathbf{u}, t] + \lambda^T(t) \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \quad (12)$$

并记 $\phi = \varphi + \boldsymbol{\xi}^T \boldsymbol{\psi}$, 且对 (11) 右端积分号内最后一项分部积分, 可得

$$\begin{aligned} J = & \phi[\boldsymbol{x}(t_f), t_f] - \boldsymbol{\lambda}^T(t_f) \boldsymbol{x}(t_f) + \boldsymbol{\lambda}^T(t_0) \boldsymbol{x}(t_0) \\ & + \int_{t_0}^{t_f} \left\{ H[\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{\lambda}(t), t] + \dot{\boldsymbol{\lambda}}^T(t) \boldsymbol{x}(t) \right\} dt \end{aligned}$$

则由控制 $\mathbf{u}(t)$ 以及终端时刻 t_f 和初始状态的变分引起的 J 的变分为

$$\begin{aligned}\delta J &= \frac{\partial \phi}{\partial t_f} \delta t_f + \frac{\partial \phi}{\partial \mathbf{x}^T(t_f)} \delta \mathbf{x}(t_f) - \dot{\boldsymbol{\lambda}}^T(t_f) \delta t_f \mathbf{x}(t_f) - \boldsymbol{\lambda}^T(t_f) \delta \mathbf{x}(t_f) \\ &\quad + \boldsymbol{\lambda}^T(t_0) \delta \mathbf{x}(t_0) + \int_{t_0}^{t_f} \left\{ \frac{\partial H}{\partial \mathbf{x}^T} \delta \mathbf{x}(t) + \frac{\partial H}{\partial \mathbf{u}^T} \delta \mathbf{u}(t) + \dot{\boldsymbol{\lambda}}^T(t) \delta \mathbf{x}(t) \right\} dt \\ &\quad + \left\{ H[\mathbf{x}(t_f), \mathbf{u}(t_f), \boldsymbol{\lambda}(t_f), t_f] + \dot{\boldsymbol{\lambda}}^T(t_f) \mathbf{x}(t_f) \right\} \delta t_f \\ &= \left\{ \frac{\partial \phi}{\partial t_f} + H[\mathbf{x}(t_f), \mathbf{u}(t_f), \boldsymbol{\lambda}(t_f), t_f] \right\} \delta t_f \\ &\quad + \left[\frac{\partial \phi}{\partial \mathbf{x}^T(t_f)} - \boldsymbol{\lambda}^T(t_f) \right] \delta \mathbf{x}(t_f) + \boldsymbol{\lambda}^T(t_0) \delta \mathbf{x}(t_0) \\ &\quad + \int_{t_0}^{t_f} \left\{ \left[\frac{\partial H}{\partial \mathbf{x}^T} + \dot{\boldsymbol{\lambda}}^T(t) \right] \delta \mathbf{x}(t) + \frac{\partial H}{\partial \mathbf{u}^T} \delta \mathbf{u}(t) \right\} dt\end{aligned}$$

考虑到 $\delta t_f, \delta[\mathbf{x}(t_f)]$, $\delta \mathbf{x}(t)$ 可以取任意值, $\delta \mathbf{x}(t_0) = 0$ (对应初始状态

给定), 由 $\delta J = 0$ (泛函优化的必要条件), 可导出

$$\dot{\lambda}(t) = -\frac{\partial H}{\partial \mathbf{x}} \quad (13)$$

$$\lambda(t_f) = \frac{\partial \phi}{\partial \mathbf{x}(t_f)} \quad (14)$$

$$\frac{\partial \phi}{\partial t_f} = -H[\mathbf{x}(t_f), \mathbf{u}(t_f), \lambda(t_f), t_f] \quad (15)$$

对于无控制约束情况, $\delta \mathbf{u}(t)$ 也可以取任意值, 因此还有

$$\frac{\partial H}{\partial \mathbf{u}} = \mathbf{0} \quad (16)$$

该式即为最优控制的必要 (驻点) 条件。

上述方程可以汇总为

$$\frac{\partial H}{\partial \mathbf{u}} = \mathbf{0} \quad (17)$$

$$\dot{\boldsymbol{\lambda}}(t) = -\frac{\partial H}{\partial \mathbf{x}}, \quad \boldsymbol{\lambda}(t_f) = \frac{\partial \varphi}{\partial \mathbf{x}(t_f)} + \boldsymbol{\xi}^T \frac{\partial \psi}{\partial \mathbf{x}(t_f)} \quad (18)$$

$$\frac{\partial \phi}{\partial t_f} = -H[\mathbf{x}(t_f), \mathbf{u}(t_f), \boldsymbol{\lambda}(t_f), t_f] \quad (19)$$

上述关系式在变分理论中称为欧拉-拉格朗日方程 (Euler-Lagrange equation)。

将欧拉-拉格朗日方程与状态方程及其初始条件 (1)、终端约束方程 (10) 联立, 可解出最优控制 $\mathbf{u}^*(t)$ (n_u 维)、最优状态轨迹 $\mathbf{x}^*(t)$ (n_x 维)、协态变量 $\boldsymbol{\lambda}(t)$ (n_x 维)、终端约束乘子 $\boldsymbol{\xi}$ (n_ψ 维) 以及终端时间 t_f 。对比状态

变量方程 (1) 和协态变量方程 (18), 不难发现, 前者给定初始条件, 后者给定终端条件, 所以本质上我们获得了一个两点边值问题。另外, 对于 t_f 给定情况, $\delta t_f = 0$, 方程 (15) 及 (19) 自动消失, 相应的未知数减少 1 个。

欧拉-拉格朗日方程给出的是最优解的必要条件, 对应的充分条件需要计算二次变分。对于极小值问题, 即为 $\delta^2 J > 0$ 。

2.1.3 极小值原理

在上述经典变分法中，没有考虑控制 $\mathbf{u}(t)$ 有约束的情况。在贝尔曼创建动态规划理论的同时，前苏联科学家庞特里亚金发展了经典变分理论，使得可以处理控制有约束情况，提出了著名的极小值原理。求极大值问题可以简单地化为求极小值问题，所以极小值原理有时也称为极大值原理。

显然当 $\mathbf{u}(t)$ 在容许集 \mathcal{U} (约束集) 边界取值时， $\frac{\partial H}{\partial \mathbf{u}} = 0$ 不再成立，庞特里亚金给出了如下替代方程：

$$\mathbf{u}^*(t) = \arg \min_{\mathbf{u}(t) \in \mathcal{U}} H[\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\lambda}(t), t] \quad (20)$$

因此, 对于我们上述最优控制问题, 庞特里亚金极小值原理可以描述为联立求解下述方程组:

$$\mathbf{u}^*(t) = \arg \min_{\mathbf{u}(t) \in \mathcal{U}} H[\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\lambda}(t), t] \quad (21)$$

$$\dot{\mathbf{x}} = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t], \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (22)$$

$$\dot{\boldsymbol{\lambda}} = -\frac{\partial H}{\partial \mathbf{x}}, \quad \boldsymbol{\lambda}(t_f) = \frac{\partial \phi}{\partial \mathbf{x}(t_f)} + \boldsymbol{\xi}^T \frac{\partial \psi}{\partial \mathbf{x}(t_f)} \quad (23)$$

$$\frac{\partial \phi}{\partial t_f} = -H[\mathbf{x}(t_f), \mathbf{u}(t_f), \boldsymbol{\lambda}(t_f), t_f] \quad (24)$$

显然这仍然是一个两点边值问题。其中, 哈密尔顿函数定义同前, 即

$$H[\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}, t] = L[\mathbf{x}, \mathbf{u}, t] + \boldsymbol{\lambda}^T(t) \mathbf{f}(\mathbf{x}, \mathbf{u}, t)$$

Example 2.1 (线性二次型调节 (LQR) 问题)

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (25)$$

$$\min_{\mathbf{u}(t)} J = \frac{1}{2} \mathbf{x}^T(t_f) \mathbf{S}_f \mathbf{x}(t_f) + \frac{1}{2} \int_{t_0}^{t_f} [\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}] dt \quad (26)$$

其中, $\mathbf{S}_f \geq 0, \mathbf{Q} \geq 0, \mathbf{R} > 0$ 为对称加权矩阵。

[解] 根据式 (7), 此时的哈密顿函数可表示为

$$H[\mathbf{x}(t), \mathbf{u}(t), J_x^*, t] = \frac{1}{2} [\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}] + J_x^{*T} [\mathbf{A} \mathbf{x} + \mathbf{B} \mathbf{u}] \quad (27)$$

根据 HJB 方程 (8), 可导出

$$\frac{\partial H}{\partial \mathbf{u}} = H_u = \mathbf{R} \mathbf{u} + \mathbf{B}^T J_x^* = 0$$

由此可解出

$$\mathbf{u}^*(t) = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{J}_x^* \quad (28)$$

将对应的 H^* 代入 HJB 方程 (8) 得到

$$-J_t^* = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \frac{1}{2}\mathbf{J}_x^{*T}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{J}_x^* + \mathbf{J}_x^{*T}\mathbf{A}\mathbf{x} \quad (29)$$

边界条件为

$$\mathbf{J}^*[\mathbf{x}(t_f), t_f] = \frac{1}{2}\mathbf{x}^T(t_f)\mathbf{S}_f\mathbf{x}(t_f) \quad (30)$$

根据上述边界条件, 假设

$$\mathbf{J}^*[\mathbf{x}(t), t] = \frac{1}{2}\mathbf{x}^T(t)\mathbf{P}(t)\mathbf{x}(t) \quad (31)$$

其中 $\mathbf{P} \geq 0$ 为对称矩阵, 则有 $J_t^* = \frac{1}{2} \mathbf{x}^T(t) \dot{\mathbf{P}}(t) \mathbf{x}(t)$ 、 $J_x^* = \mathbf{P}(t) \mathbf{x}(t)$, 代入 (29), 考虑到 $2\mathbf{x}^T \mathbf{P} \mathbf{A} \mathbf{x} = \mathbf{x}^T (\mathbf{P} \mathbf{A} + \mathbf{A}^T \mathbf{P}) \mathbf{x}$, 可整理得到

$$\dot{\mathbf{P}} + \mathbf{P} \mathbf{A} + \mathbf{A}^T \mathbf{P} - \mathbf{P} \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{P} + \mathbf{Q} = 0 \quad (32)$$

该式称为矩阵 Riccati 微分方程, 其边界条件为 $\mathbf{P}(t_f) = \mathbf{S}_f$ 。

最后可得 LQR 问题的最优解为

$$\mathbf{u}^*(t) = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{P} \mathbf{x}(t) \quad (33)$$

$$J^*[\mathbf{x}_0, t_0] = \frac{1}{2} \mathbf{x}_0^T \mathbf{P}(t_0) \mathbf{x}_0 \quad (34)$$

以上是采用动态规划方法求解的，我们当然也可以根据极小值原理来解决这个问题。根据极小值原理，此时的哈密尔顿函数为

$$H[x, u, \lambda, t] = \frac{1}{2} [x^T Q x + u^T R u] + \lambda^T [A x + B u] \quad (35)$$

由此可知

$$u^*(t) = -R^{-1} B^T \lambda(t) \quad (36)$$

并得到如下两点边值问题：

$$\dot{x}(t) = A x - B R^{-1} B^T \lambda, \quad x(t_0) = x_0 \quad (37)$$

$$\dot{\lambda}(t) = -A^T \lambda - Q x, \quad \lambda(t_f) = S_f x(t_f) \quad (38)$$

通过设 $\lambda(t) = P(t)x(t)$ ，则可得到与动态规划方法相同的结果。 \square

Example 2.2 (简单的拦截与交会问题) 如图2所示, 以初始视线为参考基准, 考察飞行器偏离初始视线的相对运动, 空间拦截与交会问题可以简单地描述为

$$\dot{v} = a(t), \quad \dot{y} = v$$

$$J = \frac{1}{2} \left(\begin{bmatrix} v & y \end{bmatrix} \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix} \begin{bmatrix} v \\ y \end{bmatrix} \right)_{t=t_f} + \frac{1}{2} \int_{t_0}^{t_f} a^2 dt$$

其中 c_1, c_2, t_f 为常数, 求 $a(v, y, t)$ 使 J 最小。

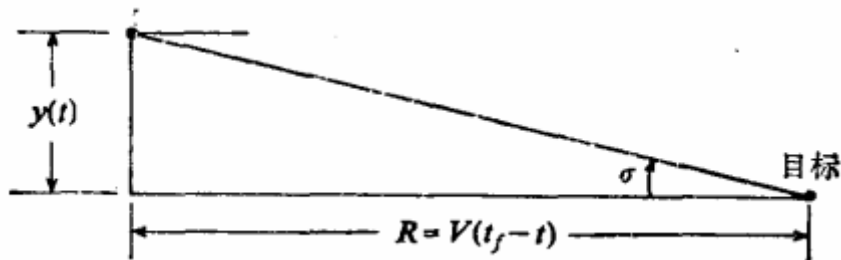


Figure 2: 拦截或交会问题示意图

[解] 将问题写为 LQR 标准形式, 有

$$\mathbf{x} = \begin{bmatrix} v \\ y \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$
$$\mathbf{S}_f = \begin{bmatrix} c_1 & 0 \\ 0 & c_2 \end{bmatrix}, \quad \mathbf{Q} = 0, \quad \mathbf{R} = 1$$

则由二次型问题的结果可得

$$a(t) = -\Lambda_v(t)v(t) - \Lambda_y(t)y(t) \quad (39)$$

其中

$$\Lambda_v(t) = \frac{\left(\frac{1}{c_2}\right) + \left(\frac{1}{c_1}\right)(t_f - t)^2 + \left(\frac{1}{3}\right)(t_f - t)^3}{D}$$

$$\Lambda_y(t) = \frac{\left(\frac{1}{c_1}\right)(t_f - t) + \left(\frac{1}{2}\right)(t_f - t)^2}{D}$$

$$D = \left[\left(\frac{1}{c_2}\right) + \left(\frac{1}{3}\right)(t_f - t)^3 \right] \left[\left(\frac{1}{c_1}\right) + (t_f - t) \right] - \left(\frac{1}{4}\right)(t_f - t)^4$$

在上述公式中，通常记 $t_{go} \triangleq t_f - t$ ，习惯上称为“剩余飞行时间 (time-to-go)”。下面针对 c_1, c_2 的几种可能的取值，进行简单讨论。

(1) 若 $c_2 \rightarrow 0$, 意味着对 $y(t_f)$ 不作要求, 相当于对末端相对位置不限制, 于是可导出

$$\Lambda_v(t) = \frac{1}{\left(\frac{1}{c_1}\right) + t_f - t}, \quad \Lambda_y(t) = 0$$

对应的制导加速度指令为

$$a(t) = -\frac{v(t)}{\left(\frac{1}{c_1}\right) + t_f - t}$$

此时即为“速度增量”控制。

(2) 若 $c_1 \rightarrow 0$, 意味着对 $v(t_f)$ 不作要求, 相当于对末端相对速度不限制,

可导出

$$\Lambda_v(t) = \frac{(t_f - t)^2}{\left(\frac{1}{c_2}\right) + \left(\frac{1}{3}\right)(t_f - t)^3}, \quad \Lambda_y(t) = \frac{(t_f - t)}{\left(\frac{1}{c_2}\right) + \left(\frac{1}{3}\right)(t_f - t)^3}$$

如果进一步取 $c_2 \rightarrow \infty$, 意味着 $y(t_f) \rightarrow 0$, 则有

$$a(t) = -\frac{3}{t_f - t}v(t) - \frac{3}{(t_f - t)^2}y(t)$$

设飞行器之间的接近速度为 V , 如图2所示, 视线角可近似表达为 $\sigma \cong \frac{y(t)}{V(t_f - t)}$, 于是上式可写为

$$a(t) = -3V\dot{\sigma} \quad (40)$$

此即为拦截问题中著名的“比例导引 (PNG)”。上式表明, 比例导引中最优的导航系数为 3, 工程中一般取 $3 \sim 5$ 。

(3) 若同时取 $c_1, c_2 \rightarrow \infty$, 意味着 $v(t_f) \rightarrow 0, y(t_f) \rightarrow 0$, 可得

$$a(t) = -\frac{4}{t_f - t}v(t) - \frac{6}{(t_f - t)^2}y(t)$$

也可以表达为

$$a(t) = -V \left(4\dot{\sigma} + \frac{2\sigma}{t_f - t} \right)$$

此即为交会问题（要求末端的位置和速度均一致）的一种变形的比例导引。

注意，在该问题中，可以将标量 v, y, a 替换为向量，从而得到三维拦截律。□

传统的最优控制理论只有一个操控方，可以认为是单边最优控制。其实，可以稍微进行一些扩展，相关理论就可以处理具有相互对抗的 双边最

优控制问题，相关理论被称为 微分对策理论，也称为零和微分对策。下面通过一个空间追逃博弈问题进行简单说明。

Example 2.3 (空间最优追逃问题) 假设在空间有两个飞行器，分别称为拦截方 p 与目标飞行器 e ，根据牛顿定理，它们的空间运动可以分别描述为

$$\dot{\boldsymbol{v}}_p = \boldsymbol{g}_p(t) + \boldsymbol{a}_p(t), \quad \dot{\boldsymbol{r}}_p = \boldsymbol{v}_p$$

$$\dot{\boldsymbol{v}}_e = \boldsymbol{g}_e(t) + \boldsymbol{a}_e(t), \quad \dot{\boldsymbol{r}}_e = \boldsymbol{v}_e$$

其中 r, v, g, a 分别为双方的位置、速度、地球引力加速度、控制加速度向量。通常两者之间的距离远小于地心距，双方的引力差可以忽略，此时只需考虑双方控制力的影响。对于追逃问题，性能指标可定义为

$$J = \frac{b}{2} [\mathbf{r}_p(t_f) - \mathbf{r}_e(t_f)]^T [\mathbf{r}_p(t_f) - \mathbf{r}_e(t_f)] + \frac{1}{2} \int_0^{t_f} (c_p^{-1} \mathbf{a}_p^T \mathbf{a}_p - c_e^{-1} \mathbf{a}_e^T \mathbf{a}_e) dt$$

其中 c_p, c_e 分别为拦截方（追逐方）和目标（逃逸方）关于控制能量的加权系数。显然，追逐方希望极小化上述 J ，而逃逸方希望极大化上述 J ，由此构成了一个 $\min\text{-max}$ 问题。

[解] 根据 LQR 理论, 可导出双方的最优控制策略分别为

$$\mathbf{a}_p = \frac{-c_p (t_f - t) \{ \mathbf{r}_p(t) - \mathbf{r}_e(t) + [\mathbf{v}_p(t) - \mathbf{v}_e(t)] (t_f - t) \}}{\left(\frac{1}{b}\right) + [(c_p - c_e)/3] (t_f - t)^3}$$

$$\mathbf{a}_e = \frac{c_e}{c_p} \mathbf{a}_p$$

当 $b \rightarrow \infty$, 追逐方的最优控制策略简化为

$$\mathbf{a}_p = \frac{-3 \{ \mathbf{r}_p(t) - \mathbf{r}_e(t) + [\mathbf{v}_p(t) - \mathbf{v}_e(t)] (t_f - t) \}}{\left(1 - \frac{c_e}{c_p}\right) (t_f - t)^2}$$

参考上例中的场景和符号定义, 追逐方的侧向最优控制策略可写为

$$a_p^\perp = \frac{3}{1 - \frac{c_e}{c_p}} V \dot{\sigma}$$

此即为有效导引系数 $N = \frac{3}{1 - \frac{c_e}{c_p}}$ 的比例导引。如前面所述, 实际经验表明 N 的最佳值为 3~5。若只按本例中的模型可以看到, $N = 3$ 对应目标不机动 ($c_e = 0$) 情况, $N = 5$ 对应 $\frac{c_e}{c_p} = 2/5$ 情况。□

Example 2.4 (加速度有界时的极小极大脱靶量问题) 在追逐-逃逸问题中, 设双方的控制量都是垂直于初始视线 (ILOS) 的加速度, 分别记为 $a_p(t)$ 和 $a_e(t)$ 。沿 ILOS 双方最为接近时的时间记为 t_f , 双方垂直于 ILOS 的相对速度和相对位移分别记为 $v(t)$ 和 $r(t)$, 则运动方程为

$$\dot{v} = a_p - a_e, \quad v(t_0) = v_0$$

$$\dot{y} = v, \quad y(t_0) = 0$$

追逐方希望最小化脱靶量 $|y(t_f)|$ ，而逃逸方希望使其最大化，故性能指标可取为

$$J = \frac{1}{2} y(t_f)^2$$

另外，考虑双方的加速度都存在取值限制

$$|a_p| \leq a_{pm}, \quad |a_e| \leq a_{em}$$

一般地， $a_{pm} > a_{em}$ 。

[解] 构造哈密顿函数

$$H = \lambda_v (a_p - a_e) + \lambda_y v$$

可导出对应的协态方程为

$$\dot{\lambda}_v = -\lambda_y, \quad \lambda_v(t_f) = 0$$

$$\dot{\lambda}_y = 0, \quad \lambda_y(t_f) = y(t_f)$$

由此可解算得到

$$\lambda_v(t) = (t_f - t) y(t_f)$$

$$\lambda_y(t) = y(t_f) = \text{Const.}$$

根据极大（小）值原理，可知双方的最优控制为

$$a_p = -a_{pm} \text{sgn}(\lambda_v)$$

$$a_e = -a_{em} \text{sgn}(\lambda_v)$$

考虑到

$$\operatorname{sgn}(\lambda_v) = \operatorname{sgn}[y(t_f)] = \text{Const.}$$

将最优控制代入状态方程并积分，易得终端状态

$$y(t_f) = v_0(t_f - t_0) - \frac{1}{2}(a_{pm} - a_{em})(t_f - t_0)^2 \operatorname{sgn}[y(t_f)]$$

对符号函数分段讨论可得

$$y(t_f) = \begin{cases} \frac{1}{2} \left[\frac{2v_0}{t_f - t_0} - (a_{pm} - a_{em}) \right] (t_f - t_0)^2, & \frac{2v_0}{(t_f - t_0)(a_{pm} - a_{em})} > 1 \\ \frac{1}{2} \left[\frac{2v_0}{t_f - t_0} + (a_{pm} - a_{em}) \right] (t_f - t_0)^2, & \frac{2v_0}{(t_f - t_0)(a_{pm} - a_{em})} < -1 \end{cases}$$

当

$$-1 < \frac{2v_0}{(t_f - t_0)(a_{pm} - a_{em})} < 1$$

$y(t_f)$ 无解。事实上, 对于该初始条件, 追逐方总能使脱靶量为零, 即 $y(t_f) = 0$ 。例如选择 $a_p = a_e + \frac{2v}{(t_f - t)}$ 。对于追逐方, 上式定义了可捕获区。这是微分对策研究中一个重要的主题。 \square

2.2 离散时间系统

考虑如下非线性差分方程描述的多级系统:

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad i = 0, \dots, N-1 \quad (41)$$

给定初始状态 \mathbf{x}_0 。 $\mathbf{x}_k \in R^{n_x}$ 和 $\mathbf{u}_k \in R^{n_u}$ 分别表示 k 时刻系统的状态与控制。

考虑性能指标

$$J_0(\mathbf{x}_0) = \varphi(\mathbf{x}_N, N) + \sum_{k=0}^{N-1} L_k(\mathbf{x}_k, \mathbf{u}_k) \quad (42)$$

此时，最优控制问题可描述为：求解控制序列 $\{\mathbf{u}_k^*\} \in \mathcal{U}$ (\mathcal{U} 为容许控制集) 以最小化 J ，即使得 $J_0^*(\mathbf{x}_0) \leq J_0(\mathbf{x}_0)$ 。

2.2.1 动态规划方法

一般地, 令

$$J_k^*(\mathbf{x}_k) = \min_{\substack{\mathbf{u}_i \\ k \leq i < N}} \left[\varphi(\mathbf{x}_N, N) + \sum_{i=k}^{N-1} L_i(\mathbf{x}_i, \mathbf{u}_i) \right] \quad (43)$$

根据最优性原理, 可知

$$J_k^*(\mathbf{x}_k) = \min_{\mathbf{u}_k} [L_k(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1}^*(\mathbf{x}_{k+1})] \quad (44)$$

该式称为贝尔曼 (Bellman) 方程。对应的边界条件为 $J_N^*(\mathbf{x}_N) = \varphi(\mathbf{x}_N, N)$ 。

2.2.2 极小值原理

定义哈密尔顿函数

$$H_k = L_k(\mathbf{x}_k, \mathbf{u}_k) + \boldsymbol{\lambda}_{k+1}^T \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k) \quad (45)$$

其中, $\boldsymbol{\lambda}_{k+1}$ 为与状态同维的协态变量。

由此可导出最优控制的必要条件为

$$\mathbf{u}_k^* = \arg \min_{\mathbf{u}_k \in \mathcal{U}} H_k(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\lambda}_{k+1}) \quad (46)$$

$$\mathbf{x}_{k+1} = \mathbf{f}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 \text{ 给定} \quad (47)$$

$$\boldsymbol{\lambda}_k = \frac{\partial H_k}{\partial \mathbf{x}_k}, \quad \boldsymbol{\lambda}_N = \frac{\partial \varphi}{\partial \mathbf{x}_N} \quad (48)$$

可见, 这是一个离散时间的两点边值问题。

Example 2.5 (线性二次型调节 (LQR) 问题)

$$\mathbf{x}_{k+1} = \mathbf{A}_k \mathbf{x}_k + \mathbf{B}_k \mathbf{u}_k, \quad \mathbf{x}_0 \text{ 给定} \quad (49)$$

$$\min_{\mathbf{u}_k} J_0(\mathbf{x}_0) = \frac{1}{2} \mathbf{x}_N^T \mathbf{S}_N \mathbf{x}_N + \frac{1}{2} \sum_{k=0}^{N-1} [\mathbf{x}_k^T \mathbf{Q}_k \mathbf{x}_k + \mathbf{u}_k^T \mathbf{R}_k \mathbf{u}_k] \quad (50)$$

其中, $\mathbf{S}_N \geq 0, \mathbf{Q} \geq 0, \mathbf{R} > 0$ 为对称加权矩阵。

[解] $k = N$ 时, 由贝尔曼方程的边界条件可知

$$J_N^*(\mathbf{x}_N) = \frac{1}{2} \mathbf{x}_N^T \mathbf{S}_N \mathbf{x}_N \triangleq \frac{1}{2} \mathbf{x}_N^T \mathbf{P}_N \mathbf{x}_N \quad (51)$$

$k = N - 1$ 时, 贝尔曼方程为

$$\begin{aligned} J_{N-1}^*(\mathbf{x}_{N-1}) &= \min_{\mathbf{u}_{N-1}} \left[\frac{1}{2} (\mathbf{x}_{N-1}^T \mathbf{Q}_{N-1} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T \mathbf{R}_{N-1} \mathbf{u}_{N-1}) + \frac{1}{2} \mathbf{x}_N^T \mathbf{P}_N \mathbf{x}_N \right] \\ &= \min_{\mathbf{u}_{N-1}} \left[\frac{1}{2} (\mathbf{x}_{N-1}^T \mathbf{Q}_{N-1} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T \mathbf{R}_{N-1} \mathbf{u}_{N-1}) \right. \\ &\quad \left. + \frac{1}{2} (\mathbf{A}_{N-1} \mathbf{x}_{N-1} + \mathbf{B}_{N-1} \mathbf{u}_{N-1})^T \mathbf{P}_N (\mathbf{A}_{N-1} \mathbf{x}_{N-1} + \mathbf{B}_{N-1} \mathbf{u}_{N-1}) \right] \end{aligned}$$

由此可导出

$$\begin{aligned} \mathbf{u}_{N-1}^* &= -(\mathbf{R}_{N-1} + \mathbf{B}_{N-1}^T \mathbf{P}_N \mathbf{B}_{N-1})^{-1} \mathbf{B}_{N-1}^T \mathbf{P}_N \mathbf{A}_{N-1} \mathbf{x}_{N-1} \\ &\triangleq -\mathbf{K}_{N-1} \mathbf{x}_{N-1} \end{aligned} \quad (52)$$

将其代入 $J_{N-1}^*(\mathbf{x}_{N-1})$ 可得

$$\begin{aligned} J_{N-1}^*(\mathbf{x}_{N-1}) &= \frac{1}{2} \mathbf{x}_{N-1}^T [\mathbf{Q}_{N-1} + \mathbf{K}_{N-1}^T \mathbf{R}_{N-1} \mathbf{K}_{N-1} \\ &\quad + (\mathbf{A}_{N-1} + \mathbf{B}_{N-1} \mathbf{K}_{N-1})^T \mathbf{P}_N (\mathbf{A}_{N-1} + \mathbf{B}_{N-1} \mathbf{K}_{N-1})] \mathbf{x}_{N-1} \\ &\triangleq \frac{1}{2} \mathbf{x}_{N-1}^T \mathbf{P}_{N-1} \mathbf{x}_{N-1} \end{aligned} \quad (53)$$

继续时间倒推求解, $k = N-2, N-3, \dots$ 。一般地, 对 $\forall 0 \leq k < N$, 有

$$\mathbf{u}_k^* = -\mathbf{K}_k \mathbf{x}_k, \quad J_k^*(\mathbf{x}_k) = \frac{1}{2} \mathbf{x}_k^T \mathbf{P}_k \mathbf{x}_k \quad (54)$$

其中

$$\mathbf{K}_k = (\mathbf{R}_k + \mathbf{B}_k^T \mathbf{P}_{k+1} \mathbf{B}_k)^{-1} \mathbf{B}_k^T \mathbf{P}_{k+1} \mathbf{A}_k \quad (55)$$

$$\mathbf{P}_k = \mathbf{Q}_k + \mathbf{K}_k^T \mathbf{R}_k \mathbf{K}_k + (\mathbf{A}_k + \mathbf{B}_k \mathbf{K}_k)^T \mathbf{P}_{k+1} (\mathbf{A}_k + \mathbf{B}_k \mathbf{K}_k), \quad \mathbf{P}_N = \mathbf{S}_N \quad (56)$$

上式称为矩阵黎卡提 (Riccati) 方程。

可见, LQR 问题归结为黎卡提方程的求解。

□

Example 2.6 已知

$$x_{k+1} = 2x_k + u_k, \quad x_0 = 1$$

求 u_0, u_1, u_2 , 最小化如下性能指标:

$$J = \sum_{k=0}^2 (x_k^2 + u_k^2)$$

[解] (1) $k = 2$ 时, 由

$$J_2^*(x_2) = \min_{u_2} (x_2^2 + u_2^2)$$

可得

$$u_2^* = 0, \quad J_2^*(x_2) = x_2^2$$

(2) $k = 1$ 时, 由

$$\begin{aligned} J_1^*(x_1) &= \min_{u_1} [x_1^2 + u_1^2 + J_2^*(x_2)] \\ &= \min_{u_1} (x_1^2 + u_1^2 + x_2^2) \\ &= \min_{u_1} [x_1^2 + u_1^2 + (2x_1 + u_1)^2] \end{aligned}$$

可得

$$u_1^* = -x_1, \quad J_1^*(x_1) = 3x_1^2$$

(3) $k = 0$ 时, 由

$$\begin{aligned} J_0^*(x_0) &= \min_{u_0} [x_0^2 + u_0^2 + J_1^*(x_1)] \\ &= \min_{u_0} (x_0^2 + u_0^2 + 3x_1^2) \\ &= \min_{u_0} [x_0^2 + u_0^2 + 3(2x_0 + u_0)^2] \end{aligned}$$

可得

$$u_0^* = -\frac{3}{2}x_0 = -\frac{3}{2}, \quad J_0^*(x_0) = 4$$

综上, 最优开环控制策略为

$$u_0^* = -\frac{3}{2}, \quad u_1^* = -\frac{1}{2}, \quad u_2^* = 0$$

最优闭环（反馈）控制策略为

$$u_0^* = -\frac{3}{2}x_0, \quad u_1^* = -x_1, \quad u_2^* = 0x_2$$

最优指标为 $J^* = J_0^*(x_0) = 4$ 。

2.3 最优控制的数值解

对于一般的非线性系统的最优控制问题，通常无法获得解析解，需要借助计算机获取数值解。典型的应用包括飞行器轨道优化、机器人及自动驾驶车辆航迹规划等。非线性系统最优控制的数值解法可以大致分为间接法、直接法和其他方法三类，都得到了大力发展和广泛应用。

2.3.1 间接法

间接方法主要应用庞特里亚金极小值原理或经典变分法，把含有性能指标的泛函最优化问题转换为求解函数极值问题，由此可以得到一个两点边值问题。主要有打靶法、拟线性化法、多重打靶法、间接配置方法等。

间接方法的主要特点是显示求解协态方程、极小值原理方程和贯截条件方程。在实际应用中，该方法存在几个缺点：

- 间接法必须根据特定条件推导解析表达式，比较麻烦；
- 协态变量没有物理意义，其初值猜测非常困难，而且其收敛域一般非常小；

- 对于有路径约束的优化问题，在迭代前必须猜测受约束的子弧段，这也是一个很困难的工作。

2.3.2 直接法

与间接方法不同，直接方法是把泛函优化问题转化为数学规划问题，不需要没有明显物理含义的协态方程（伴随方程），避免了繁杂的梯度计算，更易于计算机实现。

直接方法也很多，其中非线性规划方法是一种比较常用的方法。这种方法的中心思想是将轨迹分成若干段，每一段根据优化性能指标和约束条件的不同，将状态变量用不同阶次的多项式来表示，从而将连续时间状态微分方程组化成一组代数方程，转化为离散的非线性数学规划问题。包括

梯度法、共轭梯度法、变尺度法、直接配置方法等，近年来伪谱法在飞行器轨道优化中获得了广泛重视和关注。

2.3.3 其他方法

利用直接法变换后得到的最优控制问题往往是一类高维、非光滑优化问题，传统基于梯度的非线性规划方法往往会遇到病态梯度、初始点敏感和局部收敛问题。特别是对于一类终端时间可变和路径约束的最优控制问题，基于梯度的非线性规划算法会遇到更大的困难。因此，寻求新的基于非梯度的智能全局寻优方法，包括遗传算法、模拟退火算法、粒子群算法、蚁群算法、免疫算法等，是当前研究的热点。

3. 随机动态规划技术

确定性最优控制理论一般只适用于理想的无噪声情况，且要求所有状态都可以准确测量。由于实际中总是有噪声存在，并且给出的信息很少直接是我们需要的形式，因此需要基于状态估计进行随机系统的最优控制。这样的最优控制才是真实物理系统通常需要的。

与确定性系统的最优控制相比，随机系统的最优控制要复杂得多。根据采用的信息及结构的不同，随机最优策略有许多不同的形式，对应的理论方法差异也很多。本节仅讨论完全状态信息情况，即假定可以准确地获

取系统当前的状态。此时，如果控制策略取为当前状态的函数，一方面保证了物理系统因果属性，另外一方面还维持了系统的 Markov 特性。

动态规划很容易用于求解非线性随机系统的最优控制。下面将依次研究离散时间系统和连续时间系统的随机动态优化问题。

3.1 离散时间系统

考虑非线性随机系统

$$\mathbf{x}_{k+1} = \mathbf{f}_k(\mathbf{x}_k, \mathbf{u}_k) + \mathbf{G}_k \mathbf{w}_k \quad (57)$$

其中 \mathbf{w}_k 为白噪声序列, $\mathbf{f}_k(\cdot, \cdot)$, \mathbf{G}_k 的下标 k 表示依赖时间, 初始状态 \mathbf{x}_i 是随机量。针对系统在时间区间 $[i, N]$ 内的行为, 选取控制输入 \mathbf{u}_k , 最小化如下性能指标 (代价函数):

$$J_i(\mathbf{x}_i) = \phi(N, \mathbf{x}_N) + \sum_{k=i}^{N-1} L_k(\mathbf{x}_k, \mathbf{u}_k) \quad (58)$$

其中 ϕ 和 L_k 是确定性函数。

由于随机的过程噪声 w_k 和初始状态 x_i 导致 x_k 是随机量，故性能指标 $J_i(x_i)$ 也是随机量。要使所定义的控制问题不取决于特定的 x_i 或 w_k 序列，就不应最小化特定的 $J_i(x_i)$ ，而应最小化其平均值，即平均代价函数

$$j_i(x_i) = E[J_i(x_i)] \quad (59)$$

假设在 k 时刻 x_k 是准确可测的，下面进行问题的求解。利用贝叶斯定理，平均代价函数可写为

$$j_i = E_{x_i} \{E[J_i(x_i) | x_i]\} \quad (60)$$

其中 $E_{\mathbf{x}_i}$ 表示关于 \mathbf{x}_i 的期望值。当认为初始状态 \mathbf{x}_i 固定时, 可定义

$$\bar{J}_i(\mathbf{x}_i) = E[J_i(\mathbf{x}_i) | \mathbf{x}_i] \quad (61)$$

为固定初始状态下的平均代价函数, 或称条件期望代价。因此

$$j_i = E_{\mathbf{x}_i}[\bar{J}_i(\mathbf{x}_i)] \quad (62)$$

可以证明 (Åström, 1970)

$$j_i^* = \min_{\mathbf{u}_k}(j_i) = E_{\mathbf{x}_i} \left\{ \min_{\mathbf{u}_k} [\bar{J}_i(\mathbf{x}_i)] \right\} = E_{\mathbf{x}_i} [\bar{J}_i^*(\mathbf{x}_i)] \quad (63)$$

即“关于控制序列 \mathbf{u}_k 取极小”与“关于 \mathbf{x}_i 取期望”是可互换的。因此，最小化 j_i 可以等效为最小化在随机变量 \mathbf{x}_i 每个可能取值下的 $\bar{J}_i(\mathbf{x}_i)$ 。这也就是本节要介绍的方法。

下面建立随机系统的最优性原理。假设我们已经求出了在 \mathbf{x}_{k+1} 的每个容许值下的后续最优条件期望代价 $\bar{J}_{k+1}^*(\mathbf{x}_{k+1})$ 和最优控制序列 $\mathbf{u}_{k+1}^*, \mathbf{u}_{k+2}^*, \dots, \mathbf{u}_{N-1}^*$ ，也就是说最优控制问题在区间 $[k+1, N]$ 上已经解决。那么与确定性动态规划问题一样，要求解 $[k, N]$ 上的最优控制，我们首先将 $J_k(\mathbf{x}_k)$ 分解为两部分

$$J_k(\mathbf{x}_k) = L_k(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1}(\mathbf{x}_{k+1}) \quad (64)$$

始终假设在 k 时刻 \mathbf{x}_k 给定 (即可准确测得)。现在先假定 \mathbf{x}_{k+1} 也准确已知, 即给定, 在此条件下对上式取条件期望可得

$$E[J_k(\mathbf{x}_k) | \mathbf{x}_{k+1}] = E[L_k(\mathbf{x}_k, \mathbf{u}_k) | \mathbf{x}_{k+1}] + E[J_{k+1}(\mathbf{x}_{k+1}) | \mathbf{x}_{k+1}] \quad (65)$$

由于 \mathbf{x}_k 给定, 因而 $L_k(\mathbf{x}_k, \mathbf{u}_k)$ 是待求控制 \mathbf{u}_k 的确定性函数, 故有

$$E[J_k(\mathbf{x}_k) | \mathbf{x}_{k+1}] = L_k(\mathbf{x}_k, \mathbf{u}_k) + \bar{J}_{k+1}(\mathbf{x}_{k+1}) \quad (66)$$

根据最优性原理, 求 k 时刻的 \mathbf{u}_k^* 时, 仅需考察那些后续部分是最优的代价函数, 也就是

$$E[J_k(\mathbf{x}_k) | \mathbf{x}_{k+1}] = L_k(\mathbf{x}_k, \mathbf{u}_k) + \bar{J}_{k+1}^*(\mathbf{x}_{k+1}) \quad (67)$$

但实际上 \mathbf{x}_{k+1} 取决于随机过程噪声 \mathbf{w}_k , 所以并非固定。记关于 \mathbf{x}_{k+1} 的期望为 $E_{\mathbf{x}_{k+1}}$, 利用贝叶斯定理可得

$$\bar{J}_k(\mathbf{x}_k) = L_k(\mathbf{x}_k, \mathbf{u}_k) + E_{\mathbf{x}_{k+1}} [\bar{J}_{k+1}^*(\mathbf{x}_{k+1})] \quad (68)$$

注意 $L_k(\mathbf{x}_k, \mathbf{u}_k)$ 是确定性的。再次根据最优性原理, 可得在 k 时刻的最优条件代价为

$$\bar{J}_k^*(\mathbf{x}_k) = \min_{\mathbf{u}_k} \{ L_k(\mathbf{x}_k, \mathbf{u}_k) + E_{\mathbf{x}_{k+1}} [\bar{J}_{k+1}^*(\mathbf{x}_{k+1})] \} \quad (69)$$

该式称为完整状态信息下的离散时间非线性随机控制问题的 Bellman 方程, 由此得到的最优控制具有状态反馈的形式。由于 $L_k(\mathbf{x}_k, \mathbf{u}_k)$ 是确定性的,

上式也可写为

$$\bar{J}_k^*(\mathbf{x}_k) = \min_{\mathbf{u}_k} E_{\mathbf{x}_{k+1}} \left\{ L_k(\mathbf{x}_k, \mathbf{u}_k) + [\bar{J}_{k+1}^*(\mathbf{x}_{k+1})] \right\} \quad (70)$$

显然, 边界条件为

$$\bar{J}_N^*(\mathbf{x}_N) = E[\phi(N, \mathbf{x}_N) | \mathbf{x}_N] = \phi(N, \mathbf{x}_N) \quad (71)$$

这里假设了在 $k = N$ 时刻 \mathbf{x}_N 已知。根据 $\bar{J}_i^*(\mathbf{x}_i)$, 可得最优的平均代价函数

$$j_i^* = E_{\mathbf{x}_i} [\bar{J}_i^*(\mathbf{x}_i)] \quad (72)$$

Example 3.1 考虑标量随机系统

$$x_{k+1} = x_k + u_k + w_k$$

其中过程噪声 w_k 为白噪声，其概率密度函数如图 3 所示。性能指标取为

$$J_0(x_0) = (x_N - 1)^2 + \frac{1}{2} \sum_{k=0}^{N-1} u_k^2$$

简单起见令 $N = 2$ 。容许控制为

$$u_k = -0.5, 0, 0.5$$

容许状态为半整数

$$x_k = 0, \pm 0.5, \pm 1, \dots$$

求解控制输入 u_k 以最小化期望代价 $j_0 = E[J_0(x_0)]$ ，且考虑终端状态约束

$0 \leq x_N \leq 1.5$ ，即终端状态必须属于目标集

$$x_N \in \{0, 0.5, 1, 1.5\}$$

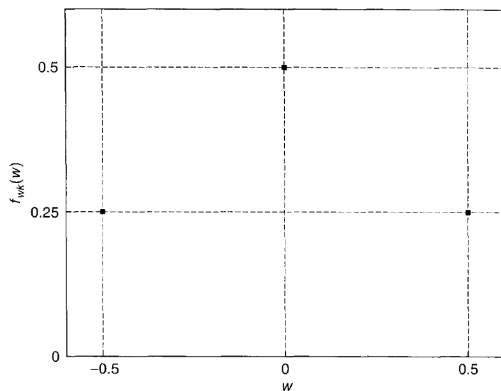


Figure 3: 过程噪声概率密度函数

[解] 应用贝尔曼方程 (69) 或 (70) 求解该最优控制问题, 可以构建表格形式的决策网格, 直观展示每一步决策过程。

(1) 首先考虑 $k = N = 2$, 计算终端代价 $\bar{J}_N(x_N)$ 。

由于在 $k = N$ 时不需要决策, 该代价也就等于 $\bar{J}_N^*(x_N)$ 。这一步只要简单对 x_2 的每一个容许值计算确定量

$$\bar{J}_2^*(x_2) = (x_2 - 1)^2$$

结果如表 1 所示。

(2) 再考虑 $k = N - 1 = 1$, 求 u_1^* 。

表 1: 终端代价

x_2	$\bar{J}_2(x_2) = (x_2 - 1)^2 = \bar{J}_2^*(x_2)$
1.5	0.25
1.0	0.0
0.5	0.25
0.0	1.0

对状态 x_1 的每个容许值, 施加每个容许的控制 u_1 , 并对得到的每个 x_2 计算代价 (注意, 准确说应该是 $J_1(x_1|x_2)$):

$$J_1(x_1) = \bar{J}_2^*(x_2) + \frac{u_1^2}{2}$$

此外还需计算每个 x_2 的概率 $P(x_2)$, 然后可计算容许期望代价

$$\bar{J}_1(x_1) = E_{x_2} [J_1(x_1)] = \sum_{x_2} J_1(x_1) P(x_2)$$

最后要针对每个 u_1 得到的 $\bar{J}_1(x_1)$ 进行比较, 确定最优的期望代价

$$\bar{J}_1^*(x_1) = \min_{u_1} [\bar{J}_1(x_1)]$$

并得到 u_1^* 。

上述过程如表2所示。先假设 $x_1 = 1.5$, 发现 $u_1 = 0.5$ 不可行, 因为它会导致

$$x_2 = x_1 + u_1 + w_1 = 2 + w_1$$

当 $w_1 = 0$ 或 $w_1 = 0.5$ 时将不满足容许状态。因此 $x_1 = 1.5$ 时, 控制 $u_1 = 0.5$ 不是容许控制。接下来分析控制 $u_1 = 0$ 。当 $x_1 = 1.5$ 且 $u_1 = 0$ 时, 有

$$x_2 = x_1 + u_1 + w_1 = 1.5 + w_1$$

由过程噪声概率密度图3可得

$$x_2 = \begin{cases} 2.0, & P = 0.25 \\ 1.5, & P = 0.5 \\ 1.0, & P = 0.25 \end{cases}$$

将该信息也记录在表2中。当选取 $u_1 = 0$ 时有 0.25 的概率发生 $x_2 = 2$, 这是不容许的, 所以在 $x_1 = 1.5$ 时该控制是不容许的。我们在 $J_1(x_1)$ 一列中用 \times 来表示, 意味着 $u_1 = 0$ 时已不需进一步计算, 因为 $x_1 = 1.5$ 时不

能采用该控制。现在对 $x_1 = 1.5$ 采用 $u_1 = -0.5$, 结果为

$$x_2 = x_1 + u_1 + w_1 = 1.0 + w_1$$

利用 w_1 的概率密度可算出 x_2 的可能取值及其概率 $P(x_2)$ 为

$$x_2 = \begin{cases} 1.5, & P = 0.25 \\ 1.0, & P = 0.5 \\ 0.5, & P = 0.25 \end{cases}$$

将该结果记录在表2中。因为此时所有 x_2 值都在目标集中, 故 $u_1 = -0.5$

是容许控制，可继续计算表2中其它信息。可算出对应代价函数为

$$J_1(x_1 = 1.5) = \frac{u_1^2}{2} + \bar{J}_2^*(x_2) = 0.125 + \bar{J}_2^*(x_2) = \begin{cases} 0.375, & x_2 = 1.5 \\ 0.125, & x_2 = 1.0 \\ 0.375, & x_2 = 0.5 \end{cases}$$

其中 $\bar{J}_2^*(x_2)$ 由表1查得。现在可求出取期望代价

$$\begin{aligned} \bar{J}_1(x_1 = 1.5) &= E_{x_2} [J_1(x_1)] \\ &= 0.375 \times 0.25 + 0.125 \times 0.5 + 0.375 \times 0.25 \\ &= 0.25 \end{aligned}$$

此时， $x_1 = 1.5$ 时所有 u_1 容许值（仅有 $u_1 = -0.5$ ）的 $\bar{J}_1(x_1)$ 已确定。

因此，有 $u_1^* = -0.5$ 且 $\bar{J}_1^*(x_1 = 1.5) = 0.25$ 。现在假设 $x_1 = 1.0$ ，并计算

$u_1 = 0, -0.5$ 时的 $x_2, P(x_2), J_1(x_1 = 1.0), \bar{J}_1(x_1 = 1.0)$, 可得

$$\bar{J}_1(x_1 = 1) = \begin{cases} 0.125, & u_1 = 0 \\ 0.5, & u_1 = -0.5 \end{cases}$$

如表2所示, 显然 $u_1 = 0.5$ 不容许, 因为当 $w_1 = 0.5$ 时将导致 $x_2 = x_1 + u_1 + w_1 = 2$ 。因此, 最小化 $\bar{J}_1(x_1)$ 应取

$$u_1^* = 0, \quad \bar{J}_1^*(x_1 = 1) = 0.125$$

类似地计算 $x_1 = 0.5$ 和 $x_1 = 0$ 的结果并填表, 仅考虑保证容许状态的控制值。注意表中还填有 $x_1 = 2$ 和 $x_1 = -0.5$, 但都为不容许状态, 因为此时没有控制能得到必然满足容许状态的终端状态。

表 2: 第一步优化

x_1	u_1	x_2	$P(x_2)$	$J_1(x_1) = \bar{J}_2^*(x_2) + \frac{u_1^2}{2}$	$\bar{J}_1(x_1)$	u_1^*	$\bar{J}_1^*(x_1)$
2.0	-0.5	2.0	0.25	×			
		1.5	0.5				
		1.0	0.25				
1.5	0.0	2.0	0.25	×			
		1.5	0.5				
		1.0	0.25				
	-0.5	1.5	0.25	0.375	0.25	-0.5	0.25
		1.0	0.5	0.125			

		0.5	0.25	0.375			
1.0	0.0	1.5	0.25	0.25	0.125	0.0	0.125
		1.0	0.5	0.0			
		0.5	0.25	0.25			
	-0.5	1.0	0.25	0.125	0.5		
		0.5	0.5	0.375			
		0.0	0.25	0.125			
0.5	0.5	1.5	0.25	0.375	0.25	0.5	0.25
		1.0	0.5	0.125			
		0.5	0.25	0.375			
	0.0	1.0	0.25	0.0	0.375		

		0.5	0.5		0.25			
		0.0	0.25		1.0			
0.0	0.5	1.0	0.25		0.125		0.5	0.5
		0.5	0.5		0.375			
		0.0	0.25		1.125			
-0.5	0.5	0.5	0.25					
		0.0	0.5					
		-0.5	0.25		×			

(3) 考虑 $k = 0$, 对每个可能的初始状态 x_0 求最优代价 $\bar{J}_0^*(x_0)$ 和 u_0^* 。

用上述方法完成表3, 需要的 $\bar{J}_1^*(x_1)$ 值由表2给出, 任何导致 x_1 超出范围的状态 x_0 和控制 u_0 都不允许。

表 3: 第 0 步优化

x_0	u_0	x_1	$P(x_1)$	$J_0(x_0) = \bar{J}_1^*(x_1) + \frac{u_0^2}{2}$	$\bar{J}_0(x_0)$	u_0^*	$\bar{J}_0^*(x_0)$
1.5	-0.5	1.5	0.25	0.375	0.3125	-0.5	0.3125
		1.0	0.5	0.25			
		0.5	0.25	0.375			
1.0	0.0	1.5	0.25	0.25	0.1875	0.0	0.1875
		1.0	0.5	0.125			
		0.5	0.25	0.25			

	-0.5	1.0	0.25	0.25	0.40625		
		0.5	0.5	0.375			
		0.0	0.25	0.625			
0.5	0.5	1.5	0.25	0.375	0.3125		
		1.0	0.5	0.25			
		0.5	0.25	0.375			
	0.0	1.0	0.25	0.125	0.28125	0.0	0.28125
		0.5	0.5	0.25			
		0.0	0.25	0.5			
0.0	0.5	1.0	0.25	0.25	0.40625	0.5	0.40625
		0.5	0.5	0.375			

0.0 0.25 0.625

(4) 所构建的表格已包含了最优反馈控制律。图4更清楚地展示了每个状态对应的最优控制和剩余代价。箭头表示所连接状态和控制的关系 $x_{k+1} = x_k + u_k$ ，该表示略去了噪声的影响。该图称为“随机极值场” (Bryson and Ho, 1975)，因为它具有向量场的形态。该图显示了每个 x_0 的最优条

件代价 $\bar{J}_0^*(x_0)$, 而为了计算最优期望代价 j_0^* , 还需要初始状态 x_0 的统计信息。假设已知

$$x_0 = \begin{cases} 0.0, & P = 0.5 \\ 0.5, & P = 0.25 \\ 1.0, & P = 0.125 \\ 1.5, & P = 0.125 \end{cases}$$

那么关于 x_0 的期望为

$$j_0^* = E_{x_0} [\bar{J}_0^*(x_0)] = \sum_{x_0} \bar{J}_0^*(x_0) P(x_0) = 0.3359$$

□

现在我們还需要讨论随机系统最优控制一个非常重要的特性。我們知道了怎么用 Bellman 方程求解最小化期望代价的最优控制, 并且通过例子

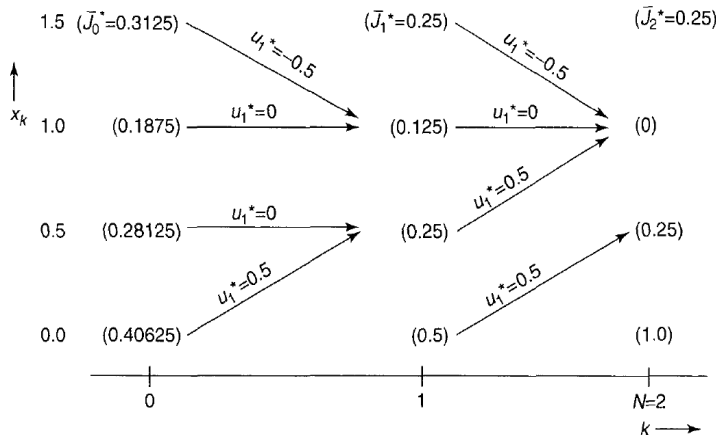


Figure 4: 随机极值场

展示了要保证终端状态 x_N 属于特定目标集是非常容易处理的。但是即使求出了最优控制，对给定初始状态 x_i 得到的终端状态值仍是未知的！也就是说并不能肯定对给定 x_i 应用 u_k^* 后的 x_N 值，因为 w_k 是未知的。不同于确定性控制问题，随机最优控制问题的最优控制序列和状态轨迹都不能由给定的 x_i 准确预测，这是因为极值场是随机的，它取决于过程噪声 w_k 的值。

根据求 u_k^* 时计算出的转移概率(即表2中的 $P(x_2)$ 和表3中的 $P(x_1)$), 我们可以使用 Chapman-Kolmogorov 方程来确定转移概率密度函数 $f(x_N|x_i)$ 。

为方便书写,令 $i = 0$ 。因为状态方程是马尔可夫过程,我们有 $f(x_2|x_1, x_0) = f(x_2|x_1)$ 等等, 因此

$$f(x_N|x_0) = \sum_{x_1, x_2, \dots, x_{N-1}} f(x_N|x_{N-1}) \cdots f(x_2|x_1) f(x_1|x_0) \quad (73)$$

Example 3.2 继续考虑例3.1 中的标量离散时间随机系统的动态规划问题, 设需要确定当 $x_0 = 0.5$ 且使用图4 的最优控制方案时, 终端状态 x_2 的可能值及其概率。

[解] 根据 $f(x_N|x_0)$ 的表达式 (73), 可得

$$f(x_2|x_0) = \sum_{x_1} f(x_2|x_1) f(x_1|x_0 = 0.5)$$

概率密度函数 $f(x_2|x_1)$ 和 $f(x_1|x_0)$ 可分别查表2和表3 (表中的 $P(x_2)$ 和 $P(x_1)$ 其实就是条件概率 $P(x_2|x_1)$ 和 $P(x_1|x_0)$), 称为转移概率。如果 $x_0 = 0.5$ 且应用 $u_0^* = 0$, 则根据表3, x_1 取值为

$$x_1 = \begin{cases} 1.0, & P = 0.25 \\ 0.5, & P = 0.5 \\ 0.0, & P = 0.25 \end{cases}$$

条件概率密度函数 $f(x_1|x_0 = 0.5)$ 如图3 所示。注意根据状态方程和

$u_0^* = 0$, 有

$$f(x_1|x_0 = 0.5) = f_{w_k}(w = 0.5)$$

$f(x_N|x_0)$ 现在可写为

$$f(x_2|x_0 = 0.5) = 0.25f(x_2|x_1 = 1) + 0.5f(x_2|x_1 = 0.5) + 0.25f(x_2|x_1 = 0)$$

接下来, 由表2按 u_1^* 确定转移概率密度函数 $f(x_2|x_1)$, 结果为

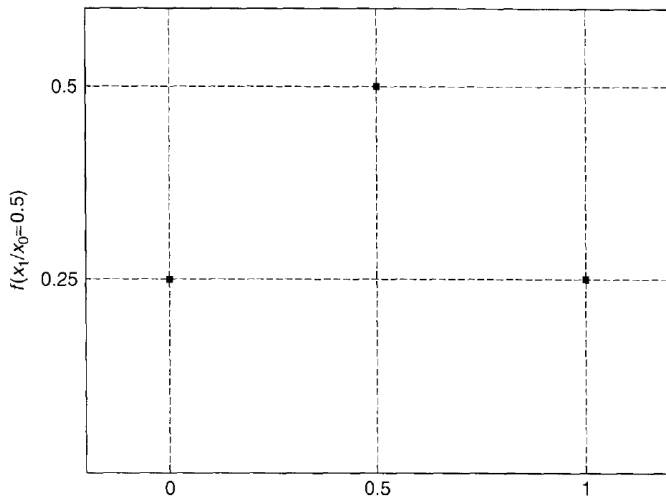
$$f(x_2|x_1 = 1) = \begin{cases} \frac{1}{4}, & x_2 = 1.5 \\ \frac{1}{2}, & x_2 = 1.0 \\ \frac{1}{4}, & x_2 = 0.5 \end{cases}$$

等等。于是，可得

$$f(x_2|x_0 = 0.5) = \begin{cases} \frac{3}{16}, & x_2 = 1.5 \\ \frac{7}{16}, & x_2 = 1.0 \\ \frac{5}{16}, & x_2 = 0.5 \\ \frac{1}{16}, & x_2 = 0.0 \end{cases}$$

现在我们讨论性能指标的另一种类型——希望最大化 x_N 在目标集 S (可能仅包含单个值 r_N , 即希望的或参考的终端状态) 内取值的概率。考虑如下性能指标:

$$J_i(x_i) = P(x_N \in S|x_i) \quad (74)$$

Figure 5: $x_0 = 0.5$ 下的转移概率密度函数

我们下面来讨论这种情况下最优性原理的形式。根据 Chapman-Kolmogorov 方程, 将 $J_k(x_k)$ 分解为两个部分

$$J_k(x_k) = \sum_{x_{k+1}} P(x_N \in \mathbb{S} | x_{k+1}) P(x_{k+1} | x_k, u_k) = \sum_{x_{k+1}} J_{k+1}(x_{k+1}) P(x_{k+1} | x_k, u_k) \quad (75)$$

其中 $P(x_{k+1} | x_k, u_k)$ 为施加控制 u_k 时从 x_k 到 x_{k+1} 的转移概率。根据最优性原理, 我们仅需考虑后续代价为最优的情况

$$J_k(x_k) = \sum_{x_{k+1}} P(x_{k+1} | x_k, u_k) J_{k+1}^*(x_{k+1}) \quad (76)$$

最大化容许代价即为

$$J_k^*(x_k) = \max_{u_k} \sum_{x_{k+1}} P(x_{k+1}|x_k, u_k) J_{k+1}^*(x_{k+1}) \quad (77)$$

这就是概率型性能指标问题的最优性原理的数学形式 (Elbert 1984)。

3.2 连续时间系统

对连续系统也可以应用离散随机最优性原理，只要将问题先离散化和数字化（包括过程噪声）。用这种方式求解随机系统最优控制非常适用于数字计算机编程，虽然维数灾难限制了可处理的系统规模和可达到的精度。

为了便于说明连续时间随机控制问题的一般求解原理，我们首先研究标量非线性随机系统：

$$\dot{x} = a(x, u, t) + w \quad (78)$$

其中， $w(t) \sim \mathcal{N}(0, \sigma^2)$ 为白噪声过程，初始状态 $x(t_0)$ 是随机的。性能指标为

$$J[x(t_0), t_0] = \phi[x(T), T] + \int_{t_0}^T L(x, u, t) dt \quad (79)$$

其中 ϕ 和 L 是给定的确定性函数。我们的问题是求解最优控制 $u^*(t)$ ，使性能指标的平均值最小。平均性能指标也称为期望代价函数，即

$$j(t_0) = E\{J[x(t_0), t_0]\} \quad (80)$$

此外，通常还要求终端状态满足如下约束条件：

$$\psi[x(T), T] = 0 \quad (81)$$

其中 ψ 是给定的确定性函数。这里我们假设最终时间 T 是固定的，并且在时刻 t 通过测量可以得到完整的状态信息 $x(t)$ 。

即使给定 $x(t_0)$ ，由于过程噪声的作用， $J[x(t_0), t_0]$ 也是随机的。我们定义给定 $x(t_0)$ 下的条件期望代价函数为

$$\bar{J}[x(t_0), t_0] = E[J[x(t_0), t_0] | x(t_0)] \quad (82)$$

根据贝叶斯法则，有

$$j(t_0) = E_{x(t_0)} \{ \bar{J}[x(t_0), t_0] \} \quad (83)$$

另外，类似离散时间问题，我们有

$$j^*(t_0) = E_{x(t_0)} \{ \bar{J}^*[x(t_0), t_0] \} \quad (84)$$

因此最优控制问题可以转化为最小化条件代价函数来解决。

首先，我们考察给定 x 和 t 条件下的状态增量

$$\Delta x = x(t + \Delta t) - x(t) \quad (85)$$

的条件概率密度函数。根据状态方程 (78)， Δt 为小量时有

$$\Delta x = a\Delta t + w\Delta t \quad (86)$$

其中, 过程噪声 $w\Delta t \sim \mathcal{N}(0, \sigma^2 \Delta t)$ 。因此, 可知 Δx 的概率密度函数为

$$f(\Delta x|x, t) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp\left[-\frac{(\Delta x - a\Delta t)^2}{2\sigma^2\Delta t}\right] \quad (87)$$

也就是说, 在 x 和 t 给定条件下, Δx 服从 $\mathcal{N}(a\Delta t, \sigma^2 \Delta t)$ 。注意, $x(t)$ 是 Markov 过程, Δx 只取决于当前状态。

在当前状态和时间 (x, t) 固定的情况下, 性能指标 $J[x(t), t]$ 可以分解为两部分

$$J(x, t) = J(x + \Delta x, t + \Delta t) + \int_t^{t+\Delta t} L(x, u, \tau) d\tau \quad (88)$$

对固定的 Δx 取条件期望, 可得

$$E[J(x, t)|\Delta x] = E\left[\int_t^{t+\Delta t} L(x, u, \tau)d\tau|\Delta x\right] + E[J(x + \Delta x, t + \Delta t)|\Delta x] \quad (89)$$

其中积分项是控制 $u(\tau)(t \leq \tau \leq t + \Delta t)$ 的确定性函数 (因为 $x(t)$ 给定), 且 x 和 Δx 固定表明 $x + \Delta x$ 固定, 因此

$$E[J(x, t)|\Delta x] = \int_t^{t+\Delta t} L(x, u, \tau)d\tau + \bar{J}(x + \Delta x, t + \Delta t) \quad (90)$$

根据最优性原理, 在 t 时刻确定最优策略仅需要考虑

$$E[J(x, t)|\Delta x] = \int_t^{t+\Delta t} L(x, u, \tau)d\tau + \bar{J}^*(x + \Delta x, t + \Delta t) \quad (91)$$

事实上 Δx 是依赖于过程噪声的随机变量, 因此应用贝叶斯法则, 并记关于 Δx 的期望为 $E_{\Delta x}$, 有

$$\begin{aligned}\bar{J}(x, t) &= E_{\Delta x} \left[\int_t^{t+\Delta t} L(x, u, \tau) d\tau + \bar{J}^*(x + \Delta x, t + \Delta t) \right] \\ &= \int_t^{t+\Delta t} L(x, u, \tau) d\tau + E_{\Delta x} [\bar{J}^*(x + \Delta x, t + \Delta t)]\end{aligned}$$

于是, 由最优性原理可得 (并将最后一项写为数学期望的定义形式)

$$\bar{J}^*(x, t) = \min_{\substack{u(\tau) \\ t \leq \tau \leq t+\Delta t}} \left[\int_t^{t+\Delta t} L(x, u, \tau) d\tau + \int \bar{J}^*(x + \Delta x, t + \Delta t) f(\Delta x | x, t) d(\Delta x) \right] \quad (92)$$

其中 $f(\Delta x|x, t)$ 即为上面推导的概率密度函数 (87), 即 $\Delta x \sim \mathcal{N}(a\Delta t, \sigma^2\Delta t)$ 。这就是完整状态信息下的连续时间随机系统最优性原理 的数学表示。

为了能具体求解最优的条件期望代价函数 $\bar{J}^*(x, t)$ 和最优控制 $u^*(t)$ 的解析表达式, 我们下面来推导 $\bar{J}^*(x, t)$ 满足的一组偏微分方程。针对 $\bar{J}^*(x, t)$ 的表达式 (92), 对 $L(x, u, \tau)$ 的积分取一阶近似, 并将 $\bar{J}^*(x + \Delta x, t + \Delta t)$ 在 (x, t) 处进行泰勒级数展开, 有

$$\begin{aligned}\bar{J}^*(x, t) = & \min_{\substack{u(\tau) \\ t \leq \tau \leq t + \Delta t}} \{L(x, u, \tau)\Delta t \\ & + \int [\bar{J}^*(x, t) + \bar{J}_x^* \Delta x + \bar{J}_t^* \Delta t \\ & + \frac{1}{2} \bar{J}_{xx}^* \Delta x^2 + \frac{1}{2} \bar{J}_{tt}^* \Delta t^2 + \bar{J}_{xt}^* \Delta x \Delta t + \cdots] f(\Delta x|x, t) d(\Delta x)\}\end{aligned}$$

可以看到 $\Delta t, \bar{J}^*$ 以及偏导数 $\bar{J}_x^*, \bar{J}_t^*, \bar{J}_{xx}^*, \bar{J}_{tt}^*, \bar{J}_{xt}^*$ 都可从积分中移出, 而剩下的数学期望要么是整个概率密度函数的区域下 (即等于 1), 要么是 $\Delta x \sim \mathcal{N}(a\Delta t, \sigma^2\Delta t)$ 的矩 (注意一阶近似下有 $E(\Delta x^2) = \sigma^2\Delta t$), 因此有

$$\begin{aligned}\bar{J}^*(x, t) = & \min_{\substack{u(\tau) \\ t \leq \tau \leq t + \Delta t}} [L(x, u, \tau)\Delta t + \bar{J}^*(x, t) \\ & + \bar{J}_x^* a \Delta t + \bar{J}_t^* \Delta t \\ & + \frac{1}{2} \bar{J}_{xx}^* \sigma^2 \Delta t + \frac{1}{2} \bar{J}_{tt}^* \Delta t^2 + \bar{J}_{xt}^* \Delta t a \Delta t + \cdots]\end{aligned}$$

把在 $t \leq \tau \leq t + \Delta t$ 上独立于 $u(\tau)$ 的项 $\bar{J}^*(x, t)$ 和 $\bar{J}_t^* \Delta t$ 从取最小中移出,

并忽略 Δt 的二阶以上项, 得到

$$0 = \bar{J}_t^* \Delta t + \min_{\substack{u(\tau) \\ t \leq \tau \leq t + \Delta t}} \left[L(x, u, \tau) \Delta t + \bar{J}_x^* a \Delta t + \frac{1}{2} \bar{J}_{xx}^* \sigma^2 \Delta t \right]$$

除以 Δt , 最后可得

$$-\frac{\partial \bar{J}^*}{\partial t} = \min_{\substack{u(\tau) \\ t \leq \tau \leq t + \Delta t}} \left[L(x, u, \tau) + \frac{\partial \bar{J}^*}{\partial x} a + \frac{1}{2} \frac{\partial^2 \bar{J}^*}{\partial x^2} \sigma^2 \right] \quad (93)$$

这就是我们需要的结果, 称为完整状态信息的随机 Hamilton-Jacobi-Bellman (HJB) 方程。其边界条件为

$$\bar{J}^*[x(T), T] = \phi[x(T), T] \quad (94)$$

且在超平面 $\psi[x(T), T] = 0$ 上。

对于向量随机系统

$$\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}, \mathbf{u}, t) + \mathbf{G}\mathbf{w} \quad (95)$$

其中 $\mathbf{x} \in R^n$ 且过程白噪声 $\mathbf{w}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}')$, 上述 HJB 方程 (93) 推广为

$$-\frac{\partial \bar{J}^*}{\partial t} = \min_{\mathbf{u}(t)} \left[L + \left(\frac{\partial \bar{J}^*}{\partial \mathbf{x}} \right)^T \mathbf{a} + \frac{1}{2} \text{tr} \left(\frac{\partial^2 \bar{J}^*}{\partial \mathbf{x}^2} \mathbf{G} \mathbf{Q}' \mathbf{G}^T \right) \right] \quad (96)$$

关于 HJB 方程, 完整严谨的数学论证需要引入伊藤随机微分相关理论。这里仅进行了直观解释和推导, 对于向量随机系统类似地有

$$\bar{J}(\mathbf{x}, t) = L(\mathbf{x}, \mathbf{u}, t) \Delta t + \bar{J}(\mathbf{x} + \Delta \mathbf{x}, t + \Delta t)$$

泰勒级数展开后可表达为

$$\begin{aligned}\bar{J}(\mathbf{x}, t) = & L(\mathbf{x}, \mathbf{u}, t) \Delta t + \bar{J}(\mathbf{x}, t) + \bar{J}_x(\mathbf{x}, t) \Delta \mathbf{x} + \bar{J}_t(\mathbf{x}, t) \Delta t \\ & + \frac{1}{2} \Delta \mathbf{x}^T \bar{J}_{xx}(\mathbf{x}, t) \Delta \mathbf{x} + \frac{1}{2} \bar{J}_{tt}(\mathbf{x}, t) (\Delta t)^2 + \bar{J}_{xt}(\mathbf{x}, t) \Delta \mathbf{x} \Delta t + \dots\end{aligned}$$

考虑 $\Delta t \rightarrow 0$, 忽略高价小后得

$$-\bar{J}_t(\mathbf{x}, t) = L(\mathbf{x}, \mathbf{u}, t) + \bar{J}_x(\mathbf{x}, t) \mathbf{a}(\mathbf{x}, \mathbf{u}, t) + \frac{1}{2} \text{tr}(\bar{J}_{xx}(\mathbf{x}, t) E[\Delta \mathbf{x} \Delta \mathbf{x}^T])$$

即

$$-\bar{J}_t(\mathbf{x}, t) = L(\mathbf{x}, \mathbf{u}, t) + \bar{J}_x(\mathbf{x}, t) \mathbf{a}(\mathbf{x}, \mathbf{u}, t) + \frac{1}{2} \text{tr}(\bar{J}_{xx}(\mathbf{x}, t) \mathbf{G} \mathbf{Q}' \mathbf{G}^T)$$

由此可以导出 (96)。

一旦求出并消去 $u(t)$, 随机 HJB 方程即变为 Kolmogorov 方程。原则上由 HJB 方程可以解出最优条件代价函数和最优控制的解析表达式, 得到状态反馈控制律。但实际上, 求解 HJB 方程是非常困难的, 一种可能的途径是采用现在流行的机器学习方法。当系统是线性系统、性能指标是二次型形式时, 上述 HJB 方程有解析解, 在实际中有广泛应用, 下面我们将专门讨论。

4. 连续时间线性二次型高斯问题

4.1 问题描述

考虑如下连续时间线性随机系统：

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{G}\mathbf{w} \quad (97)$$

其中，过程白噪声 $\mathbf{w}(t) \sim \mathcal{N}(0, \mathbf{Q}')$ ，初始状态 $\mathbf{x}(t_0) \sim \mathcal{N}(\bar{\mathbf{x}}_0, \mathbf{P}_0)$ 。相应性能指标为二次型

$$J[\mathbf{x}(t_0), t_0] = \frac{1}{2} \mathbf{x}^T(T) \mathbf{S}_T \mathbf{x}(T) + \frac{1}{2} \int_{t_0}^T (\mathbf{x}^T \mathbf{Q} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u}) dt \quad (98)$$

其中 $\mathbf{S}_T \geq 0, \mathbf{Q} \geq 0, \mathbf{R} > 0$ 为对称权重矩阵。一般情况下, 系统的模型矩阵和权重矩阵可以是时间的函数。希望求解 $[t_0, T]$ 上的控制 $\mathbf{u}^*(t)$ 以最小化平均代价函数

$$j(t_0) = E \{ J[\mathbf{x}(t_0), t_0] \} \quad (99)$$

此外, 假设 $\mathbf{x}(T)$ 自由, T 固定。

以上问题称为连续时间线性二次型高斯 (LQG) 控制问题。

显然, 对于 LQG 问题, 需要考虑两种情况: (1) 对于完整状态信息情况, 在 t 时刻状态向量 $\mathbf{x}(t)$ 准确已知, 故 $\mathbf{u}(t)$ 可以表示为 $\mathbf{x}(t)$ 的函数, 即状态反馈; (2) 对于非完全状态信息情况, 只能获得观测 (量测) 向量

$$\mathbf{z} = H\mathbf{x} + \mathbf{v} \quad (100)$$

其中量测噪声 $\mathbf{v}(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}')$ 。此时控制 $\mathbf{u}(t)$ 只能表示为状态估计 $\hat{\mathbf{x}}(t)$ 的反馈。

我们将要推导的结果是所有噪声为高斯分布时的最优控制。而当噪声不是高斯分布时, 它们是线性最小方差意义下的最优控制。

4.2 完全状态信息

4.2.1 基于 HJB 方程的最优解

根据 LQG 问题的性能指标 (98), 由随机系统的 HJB 方程 (96) 可知

$$-\bar{J}_t^* = \min_{\mathbf{u}(t)} \left[\frac{1}{2} \mathbf{x}^T \mathbf{Q} \mathbf{x} + \frac{1}{2} \mathbf{u}^T \mathbf{R} \mathbf{u} + \bar{J}_x^{*T} (\mathbf{A} \mathbf{x} + \mathbf{B} \mathbf{u}) + \frac{1}{2} \text{tr} (\bar{J}_{xx}^* \mathbf{G} \mathbf{Q}' \mathbf{G}^T) \right] \quad (101)$$

由于控制输入无约束, 最小化的驻点条件为

$$\mathbf{R} \mathbf{u} + \mathbf{B}^T \bar{J}_x^* = \mathbf{0} \quad (102)$$

由此可导出最优控制

$$\mathbf{u}^*(t) = -\mathbf{R}^{-1}\mathbf{B}^T\bar{J}_x^* \quad (103)$$

将其代入 HJB 方程 (101), 即得到 Kolmogorov 方程

$$-\bar{J}_t^* = \frac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x} - \frac{1}{2}\bar{J}_x^{*T}\mathbf{B}\mathbf{R}^{-T}\mathbf{B}^T\bar{J}_x^* + \bar{J}_x^{*T}\mathbf{A}\mathbf{x} + \frac{1}{2}\text{tr}(\bar{J}_{xx}^*\mathbf{G}\mathbf{Q}'\mathbf{G}^T) \quad (104)$$

类似于确定性 LQR 问题的求解, 我们采用试凑法, 假设

$$\bar{J}^*(\mathbf{x}, t) = \frac{1}{2}\left[\mathbf{x}^T\mathbf{S}\mathbf{x} + \text{tr}\int_t^T \mathbf{S}\mathbf{G}\mathbf{Q}'\mathbf{G}^T\text{d}\tau\right] \quad (105)$$

其中 $\mathbf{S}(t)$ 为待定的确定性对称矩阵函数。可见

$$\bar{J}_x^* = \mathbf{S}\mathbf{x}, \quad \bar{J}_t^* = \frac{1}{2}\mathbf{x}^T\dot{\mathbf{S}}\mathbf{x}, \quad \bar{J}_{xx}^* = \mathbf{S}$$

将此代入 Kolmogorov 方程 (104), 得到

$$0 = \frac{1}{2} \mathbf{x}^T \left(\dot{\mathbf{S}} + \mathbf{Q} + \mathbf{A}^T \mathbf{S} + \mathbf{S} \mathbf{A} - \mathbf{S} \mathbf{B} \mathbf{R}^{-T} \mathbf{B}^T \mathbf{S} \right) \mathbf{x} \quad (106)$$

现在将 $\mathbf{S}(t)$ 选择为下式的解

$$-\dot{\mathbf{S}} = \mathbf{A}^T \mathbf{S} + \mathbf{S} \mathbf{A} - \mathbf{S} \mathbf{B} \mathbf{R}^{-T} \mathbf{B}^T \mathbf{S} + \mathbf{Q} \quad (107)$$

该式称为连续时间 LQG 问题的 Riccati 方程, 边界条件为 $\mathbf{S}(T) = \mathbf{S}_T$ 。

最优控制可以表示为状态变量反馈的形式, 即

$$\mathbf{u}^*(t) = -\mathbf{R}^{-1} \mathbf{B}^T \mathbf{S} \mathbf{x}(t) \quad (108)$$

在 $[t, T]$ 上的最优条件代价函数 $\bar{J}^*(\mathbf{x}, t)$ 由 t 时刻的状态 $\mathbf{x}(t)$ 给出, 关于 $\mathbf{x}(t)$ 求期望可以得到最优平均代价函数

$$\begin{aligned} j^*(t) &= \frac{1}{2} E_{\mathbf{x}} \left[\mathbf{x}^T \mathbf{S} \mathbf{x} + \text{tr} \int_t^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T d\tau \right] \\ &= \frac{1}{2} \text{tr} \left[\mathbf{S}(t) \mathbf{X}(t) + \int_t^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T d\tau \right] \end{aligned} \quad (109)$$

其中

$$\mathbf{X}(t) = E [\mathbf{x}(t) \mathbf{x}^T(t)] \quad (110)$$

为 $\mathbf{x}(t)$ 的均方值, 即 $\mathbf{x}(t)$ 的自相关函数。可见, $[t_0, T]$ 上的最优平均代价函数为

$$j^*(t_0) = \frac{1}{2} \text{tr} \left[\mathbf{S}(t_0) \mathbf{X}(t_0) + \int_{t_0}^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T d\tau \right] \quad (111)$$

现在可以清楚过程噪声 $w(t)$ 的影响：随机 LQG 问题的解与确定性 LQR 问题几乎完全相同，唯一的差别是最优代价中多出了 $\frac{1}{2}\text{tr}(SGQ'G^T)$ 这一项，这代表了由协方差为 Q' 的过程噪声扰动所增加的代价。

4.2.2 另一求解方法

首先建立一些有用的表达式。设 $\phi[x(t)]$ 为 $x \in R^n$ 的任意二阶可微标量函数，随机变量 $x(t)$ 具有概率密度函数 $f_x(x, t)$ ，且 x 的时间更新过程为

$$\dot{x} = a(x, t) + G(x, t)w \quad (112)$$

其中 $w(t) \sim \mathcal{N}(0, Q')$ 。

定义

$$\hat{\phi}[\mathbf{x}(t)] = \int \phi(\mathbf{x}) f_x(\mathbf{x}, t) d\mathbf{x} \quad (113)$$

于是

$$\frac{d\hat{\phi}[\mathbf{x}(t)]}{dt} = \int \phi(\mathbf{x}) \frac{\partial f_x}{\partial t} d\mathbf{x}$$

根据 Fokker-Planck 方程

$$\frac{\partial}{\partial t} f_x(\mathbf{x}, t) = \sum_i \left\{ -\frac{\partial}{\partial x_i} [a_i f_x(\mathbf{x}, t)] + \frac{1}{2} \frac{\partial}{\partial x_i} \left[\mathbf{G} \mathbf{Q}' \mathbf{G}^T \frac{\partial f_x(\mathbf{x}, t)}{\partial \mathbf{x}} \right]_i \right\} \quad (114)$$

可得

$$\frac{d\hat{\phi}[\mathbf{x}(t)]}{dt} = \int \text{tr} \left[\phi(\mathbf{x}) \frac{\partial (f_x \mathbf{a})}{\partial \mathbf{x}} \right] d\mathbf{x} + \frac{1}{2} \int \phi(\mathbf{x}) \sum_{i,j=1}^n \frac{\partial^2 \left(f_x [\mathbf{G} \mathbf{Q}' \mathbf{G}^T]_{ij} \right)}{\partial x_i \partial x_j} d\mathbf{x}$$

进行分部积分，并利用一些关于求迹运算的矩阵恒等式，推导可得

$$\frac{d\hat{\phi}[\mathbf{x}(t)]}{dt} = \int \phi_x^T \mathbf{a} f_x dx + \frac{1}{2} \int \text{tr}(\mathbf{G}\mathbf{Q}'\mathbf{G}^T \phi_{xx}) f_x dx \quad (115)$$

由数学期望的定义，上式可表达为

$$\frac{d}{dt} E[\phi(\mathbf{x})] = E(\phi_x^T \mathbf{a}) + \frac{1}{2} \text{tr} E(\mathbf{G}\mathbf{Q}'\mathbf{G}^T \phi_{xx}) \quad (116)$$

下面将上述 $\phi[\mathbf{x}(t)]$ 拓展为 $\phi[\mathbf{x}(t), t]$ ，后者是 t 的显函数。显然

$$\begin{aligned} \frac{d}{dt} E[\phi(\mathbf{x}, t)] &= \frac{d}{dt} \int \phi(\mathbf{x}, t) f_x(\mathbf{x}) dx \\ &= \int \left[\frac{\partial \phi(\mathbf{x}, t)}{\partial t} f_x(\mathbf{x}) + \phi(\mathbf{x}, t) \frac{\partial f_x(\mathbf{x})}{\partial t} \right] dx \\ &= E \left[\frac{\partial \phi}{\partial t} \right] + \int \phi(\mathbf{x}, t) \frac{\partial f_x(\mathbf{x})}{\partial t} dx \end{aligned}$$

结合 (116) 式, 可知

$$\frac{d}{dt}E[\phi(\mathbf{x}, t)] = E(\phi_t) + E(\phi_x^T \mathbf{a}) + \frac{1}{2}\text{tr}E(\mathbf{G}\mathbf{Q}'\mathbf{G}^T \phi_{xx}) \quad (117)$$

取 $\phi(\mathbf{x}, t) = \mathbf{x}^T \mathbf{S} \mathbf{x}$, 注意到 $\mathbf{S} = \mathbf{S}(t)$, $\mathbf{a}(\mathbf{x}, t) = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$, 由 (117) 可导出

$$\begin{aligned} \frac{d}{dt}E[\mathbf{x}^T \mathbf{S} \mathbf{x}] &= E(\mathbf{x}^T \dot{\mathbf{S}} \mathbf{x}) + E[\mathbf{x}^T \mathbf{S} (\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}) + (\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u})^T \mathbf{S} \mathbf{x}] \\ &\quad + \text{tr}(\mathbf{S}\mathbf{G}\mathbf{Q}'\mathbf{G}^T) \end{aligned} \quad (118)$$

最后，注意到

$$\int_{t_0}^T \frac{d}{dt} E(\mathbf{x}^T \mathbf{S} \mathbf{x}) dt = E[\mathbf{x}^T(T) \mathbf{S}(T) \mathbf{x}(T)] - E[\mathbf{x}^T(t_0) \mathbf{S}(t_0) \mathbf{x}(t_0)] \quad (119)$$

取 $\mathbf{S}(T) = \mathbf{S}_T$ ，将以上表达式代入二次型性能指标 $J[\mathbf{x}(t_0), t_0]$ ，可得到平均代价函数的表达式

$$\begin{aligned} j(t_0) = & \frac{1}{2} E[\mathbf{x}^T(t_0) \mathbf{S}(t_0) \mathbf{x}(t_0)] \\ & + \frac{1}{2} E \int_{t_0}^T \left[\mathbf{x}^T (\dot{\mathbf{S}} + \mathbf{S} \mathbf{A} + \mathbf{A}^T \mathbf{S} + \mathbf{Q}) \mathbf{x} + \mathbf{x}^T \mathbf{S} \mathbf{B} \mathbf{u} + \mathbf{u}^T \mathbf{B}^T \mathbf{S} \mathbf{x} + \mathbf{u}^T \mathbf{R} \mathbf{u} \right] \\ & + \frac{1}{2} \text{tr} \int_{t_0}^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T dt \end{aligned}$$

若选取 $S(t)$ 使其在 $t \leq T$ 内满足 Riccati 方程

$$-\dot{S} = A^T S + S A - S B R^{-1} B^T S + Q, \quad S(T) = S_T \quad (120)$$

则 $j(t_0)$ 简化为

$$\begin{aligned} j(t_0) = & \frac{1}{2} E [x^T(t_0) S(t_0) x(t_0)] \\ & + \frac{1}{2} E \int_{t_0}^T [x^T S B R^{-1} B^T S x + x^T S B u + u^T B^T S x + u^T R u] dt \\ & + \frac{1}{2} \text{tr} \int_{t_0}^T S G Q' G^T dt \end{aligned}$$

式中前一个积分内是一个二次型，即

$$\begin{aligned} j(t_0) = & \frac{1}{2} E [\mathbf{x}^T(t_0) \mathbf{S}(t_0) \mathbf{x}(t_0)] \\ & + \frac{1}{2} E \int_{t_0}^T \|\mathbf{R}^{-1} \mathbf{B}^T \mathbf{S} \mathbf{x} + \mathbf{u}\|_{\mathbf{R}}^2 dt \\ & + \frac{1}{2} \text{tr} \int_{t_0}^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T dt \end{aligned}$$

注意 Riccati 方程就是为了给出积分内的二次型形式。由于只有这一项包含 $\mathbf{u}(t)$ ，所以 $j(t_0)$ 的最小值为

$$j^*(t_0) = \frac{1}{2} E [\mathbf{x}^T(t_0) \mathbf{S}(t_0) \mathbf{x}(t_0)] + \frac{1}{2} \text{tr} \int_{t_0}^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T dt \quad (121)$$

相应的最优控制为

$$\mathbf{u}^*(t) = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}(t)\mathbf{x}(t) \quad (122)$$

其中 $\mathbf{S}(t)$ 满足 Riccati 方程 (120)。此外 $j^*(T) = j(T)$ 。

可见，这种推导方法给出了基于 HJB 方程相同的结果。再次指出，除了因协方差为 \mathbf{Q}' 的过程噪声引起代价函数增大，导致 $j(t_0)$ 中增加了一项外，完整信息下随机 LQG 控制问题的解与确定性 LQR 问题相同。

引入 Kalman 控制增益

$$\mathbf{K}(t) = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}(t) \quad (123)$$

最优控制可写为

$$\mathbf{u}^*(t) = -\mathbf{K}(t)\mathbf{x}(t) \quad (124)$$

即线性时变状态反馈。

4.2.3 两点讨论

(1) 在确定性问题中，通过设计合适的 \mathbf{Q} 和 \mathbf{R} ，随着 $(T - t_0)$ 增大，LQR 的控制将使状态趋于 $\mathbf{0}$ 。但在随机情况下并非如此，由于过程噪声 $\mathbf{w}(t)$ 的存在，系统状态将持续地被噪声激励，稳态的下状态 $\mathbf{x}(t)$ 的行为由 $\mathbf{w}(t)$ 的扰动效果与最优控制之间的平衡来决定，状态均方值 $\mathbf{X}(t) = E[\mathbf{x}(t)\mathbf{x}^T(t)]$ 在稳态下不为 $\mathbf{0}$ 。

为求解均方状态 $\mathbf{X}(t)$, 考虑最优控制 (124) 作用下的闭环系统

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{BK}) \mathbf{x} + \mathbf{G} \mathbf{w} \quad (125)$$

由此可导出 Lyapunov 方程

$$\dot{\mathbf{X}} = (\mathbf{A} - \mathbf{BK}) \mathbf{X} + \mathbf{X} (\mathbf{A} - \mathbf{BK})^T + \mathbf{G} \mathbf{Q}' \mathbf{G}^T \quad (126)$$

其初始条件为 $\mathbf{X}(t_0) = E[\mathbf{x}(t_0) \mathbf{x}^T(t_0)] = \mathbf{P}_0 + \bar{\mathbf{x}}_0 \bar{\mathbf{x}}_0^T$, 注意 $\mathbf{x}_0 \sim \mathcal{N}(\bar{\mathbf{x}}_0, \mathbf{P}_0)$ 。

另外, 控制的均方值也能事先确定:

$$E(\mathbf{u} \mathbf{u}^T) = E(\mathbf{K} \mathbf{x} \mathbf{x}^T \mathbf{K}^T) = \mathbf{K} \mathbf{X}(t) \mathbf{K}^T \quad (127)$$

稳态时状态 $\mathbf{x}(t)$ 的均方值可通过如下 Lyapunov 代数方程得到:

$$\mathbf{0} = (\mathbf{A} - \mathbf{BK})\mathbf{X} + \mathbf{X}(\mathbf{A} - \mathbf{BK})^T + \mathbf{GQ}'\mathbf{G}^T \quad (128)$$

注意 Riccati 方程 (120) 是按时间倒推求解的, 且独立于 $\mathbf{x}(t)$, 因此可以离线求解得到 $\mathbf{S}(t)$, 之后得到 $\mathbf{K}(t)$; 然后可由给定初值 $\mathbf{X}(t_0)$ 按时间正向积分 (126) 计算出 $\mathbf{X}(t)$, 并由 (127) 计算 $E(\mathbf{u}\mathbf{u}^T)$; 而最优期望代价由 $j^*(t_0)$ 表达式 (121) 给出。

综上讨论, 在对系统应用最优控制之前, 我们就可以提前确定平均的代价函数、状态和控制的均方值, 使得可以在设计控制器时就评估控制器的性能和控制消耗。

(2) 假如状态反馈控制中的增益 $\mathbf{K}(t)$ 不是最优的而是其它值时, 那么可能得到什么样的平均代价函数 $j(t_0)$ 呢? 因为我们经常希望使用简单形式的次优反馈, 例如常值稳态反馈

$$\mathbf{K}_\infty = \mathbf{R}^{-1} \mathbf{B}^T \mathbf{S}_\infty \quad (129)$$

其中 \mathbf{S}_∞ 是 Riccati 方程 (107) 的稳态或极限解。此时, 我们希望确定非最优控制律下的 $j(t_0)$ 以判断其闭环行为是否可接受。

为此, 我们将非最优的反馈控制代入 $j(t_0)$ 表达式可得

$$\begin{aligned} j(t_0) = & \frac{1}{2} E [\mathbf{x}^T(t_0) \mathbf{S}(t_0) \mathbf{x}(t_0)] \\ & + \frac{1}{2} \text{tr} \int_{t_0}^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T dt \\ & + \frac{1}{2} E \int_{t_0}^T \mathbf{x}^T \left[\dot{\mathbf{S}} + \mathbf{S}(\mathbf{A} - \mathbf{B}\mathbf{K}) + (\mathbf{A} - \mathbf{B}\mathbf{K})^T \mathbf{S} + \mathbf{Q} + \mathbf{K}^T \mathbf{R} \mathbf{K} \right] \mathbf{x} dt \end{aligned}$$

现在选取 $\mathbf{S}(t)$ 使其在 $t \leq T$ 内满足 Lyapunov 方程

$$-\dot{\mathbf{S}} = \mathbf{S}(\mathbf{A} - \mathbf{B}\mathbf{K}) + (\mathbf{A} - \mathbf{B}\mathbf{K})^T \mathbf{S} + \mathbf{Q} + \mathbf{K}^T \mathbf{R} \mathbf{K} \quad (130)$$

它可由闭环系统方程

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{B}\mathbf{K}) \mathbf{x} + \mathbf{G} \mathbf{w} \quad (131)$$

得到。则任意具有 $u = -Kx$ 形式的反馈控制的平均代价函数等于

$$j(t_0) = \frac{1}{2} E [x^T(t_0) S(t_0) x(t_0)] + \frac{1}{2} \text{tr} \int_{t_0}^T S G Q' G^T dt \quad (132)$$

注意，这里的 Lyapunov 方程 (130) 边界条件仍然为 $S(T) = S_T$ 。要评价该控制在过程噪声下实现的状态调节效果如何，同样可以考察状态的均方值 $X(t)$ 。

Example 4.1 考虑标量随机控制系统

$$\dot{x} = ax + bu + gw$$

其中, 高斯白噪声 $w(t) \sim \mathcal{N}(0, q')$, 且 $x(0) \sim \mathcal{N}(\bar{x}_0, p_0)$ 。性能指标为

$$J = \frac{1}{2} s_T x^2(T) + \frac{1}{2} \int_0^T (qx^2 + ru^2) dt$$

假设状态 $x(t)$ 在每个时刻 t 都可准确测得。希望求解最优控制最小化平均代价函数

$$j(0) = E(J)$$

并求出对应的状态均方值。

[解] LQG 最优控制为

$$u^*(t) = -K(t)x(t)$$

其中 Kalman 控制增益由 (123) 可知为

$$K(t) = \frac{b}{r}s(t)$$

式中 $s(t)$ 满足 Riccati 方程 (120), 即

$$-\dot{s} = 2as - \frac{b^2}{r}s^2 + q, \quad s(T) = s_T$$

该方程容易通过分离变量求解。

在控制 $u^*(t)$ 作用下的均方状态 $X(t) = E[x^2(t)]$ 由 Lyapunov 方程 (126) 给出

$$\dot{X} = 2(a - bK)X + g^2q', \quad X(0) = p_0 + \bar{x}_0^2$$

由 (121) 式可知 $[0, T]$ 上的最优平均代价函数为

$$j^*(0) = \frac{1}{2} \left[s(0)X(0) + g^2 q' \int_0^T s(\tau) d\tau \right]$$

随机 LQG 的实际应用非常容易。首先将 Riccati 方程反向积分得到 $s(t)$ ，同时可以得到 $K(t)$ 并存储。实际上，反向积分可化为正向积分

$$\frac{ds}{d\tau} = 2as - \frac{b^2}{r}s^2 + q, \quad \tau \geq 0$$

其中 $s(0) = s_T$ ，然后将所得结果 $s(\tau)$ 按时间反转得到 $s(t)$ 。注意， $\tau = T - t$ 。 $X(t)$ 的 Lyapunov 方程需按时间正向积分，而闭环系统的仿真也按时间正向推进，因此可以同时进行。□

4.3 非完全状态信息与分离原理

4.3.1 基于状态估计的最优解

对于完全状态信息情况，我们时刻知道状态 $\boldsymbol{x}(t)$ 的准确值。而更一般的情况是非完全状态信息，此时我们仅有带高斯噪声的量测信息

$$\boldsymbol{z} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{v} \quad (133)$$

此时我们可以由 Kalman 滤波器给出系统状态的最优估计：

$$\dot{\hat{x}} = A\hat{x} + Bu + L(z - H\hat{x}), \quad \hat{x}(0) = \bar{x}_0 \quad (134)$$

$$L = PH^T R'^{-1} \quad (135)$$

$$\dot{P} = AP + PA^T + GQG^T - PH^T R'^{-1}HP, \quad P(t_0) = P_0 \quad (136)$$

其中 $P(t)$ 是估计误差协方差矩阵， $L(t)$ 是 Kalman 滤波增益。注意 $\hat{x}(t)$ 和 $P(t)$ 也是已知 t 时刻前控制输入和量测值下 $x(t)$ 的条件均值与协方差矩阵。

因此，现在当前时刻的控制只能是估计量 $\hat{x}(t)$ 的函数，而非未知状态 $x(t)$ 的函数。

为求解此时的最优控制，回顾上小节导出的 $j(t_0)$ 表达式：

$$j(t_0) = \frac{1}{2}E[\mathbf{x}^T(t_0)\mathbf{S}(t_0)\mathbf{x}(t_0)] + \frac{1}{2}E\int_{t_0}^T \|\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x} + \mathbf{u}\|_{\mathbf{R}}^2 dt \quad (137)$$

$$+ \frac{1}{2}\text{tr}\int_{t_0}^T \mathbf{S}\mathbf{G}\mathbf{Q}'\mathbf{G}^T dt$$

定义 t 时刻的已知观测数据集

$$\mathbf{Z}_t = \{\mathbf{z}(\tau) | t_0 \leq \tau \leq t\} \quad (138)$$

因此 $j(t_0)$ 中的二次型积分项可写为

$$\frac{1}{2}E\int_{t_0}^T \|\mathbf{R}^{-1}\mathbf{B}^T\mathbf{S}\mathbf{x} + \mathbf{u}\|_{\mathbf{R}}^2 dt = \frac{1}{2}E_{\mathbf{Z}_t}\int_{t_0}^T E\left[(\mathbf{u} + \mathbf{K}\mathbf{x})^T \mathbf{R}(\mathbf{u} + \mathbf{K}\mathbf{x}) | \mathbf{Z}_t\right] dt \quad (139)$$

其中 Kalman 控制增益 $K = R^{-1}B^T S$, E_{Z_t} 表示关于 Z_t 的期望。

注意到, 对于任意 $x \sim (\bar{x}, P_x)$, $R = R^T \geq 0$, 有

$$\begin{aligned} E(x^T R x) &= E\left[(x - \bar{x})^T R (x - \bar{x})\right] + \bar{x}^T R \bar{x} \\ &= \bar{x}^T R \bar{x} + \text{tr} E\left[R (x - \bar{x}) (x - \bar{x})^T\right] \\ &= \bar{x}^T R \bar{x} + \text{tr}(R P_x) \end{aligned}$$

该表达式在 $\bar{\mathbf{x}}$ 和 \mathbf{P}_x 分别为条件均值和协方差、 E 为条件期望时仍然成立。于是, (139) 可化为

$$\begin{aligned} \frac{1}{2} E \int_{t_0}^T \|\mathbf{R}^{-1} \mathbf{B}^T \mathbf{S} \mathbf{x} + \mathbf{u}\|_{\mathbf{R}}^2 dt &= \frac{1}{2} E_{Z_t} \int_{t_0}^T \left[(\mathbf{u} + \mathbf{K} \hat{\mathbf{x}})^T \mathbf{R} (\mathbf{u} + \mathbf{K} \hat{\mathbf{x}}) \right] dt \\ &\quad + \frac{1}{2} \text{tr} \int_{t_0}^T \mathbf{K}^T \mathbf{R} \mathbf{K} \mathbf{P} dt \end{aligned}$$

由此可知此时的最优控制应该为

$$\mathbf{u}^*(t) = -\mathbf{K}(t) \hat{\mathbf{x}}(t) \quad (140)$$

代入 $j(t_0)$ 可得最优的平均代价函数

$$\begin{aligned} j^*(t_0) = & \frac{1}{2} E [\mathbf{x}^T(t_0) \mathbf{S}(t_0) \mathbf{x}(t_0)] + \frac{1}{2} \text{tr} \int_{t_0}^T \mathbf{S} \mathbf{G} \mathbf{Q}' \mathbf{G}^T dt \\ & + \frac{1}{2} \text{tr} \int_{t_0}^T \mathbf{K}^T \mathbf{R} \mathbf{K} \mathbf{P} dt \end{aligned} \quad (141)$$

4.3.2 分离原理

上述结果非常重要，称为非完全信息问题的分离原理，也称为确定性等价原理。可以进一步表述如下：非完全状态信息下的随机 LQG 控制问题的最优解，具有确定性 LQR 问题最优控制一样的形式，只需将构成反馈的状态替换为 Kalman 滤波器给出的最优状态估计即可。相应的最优代价函

数 $j^*(t_0)$ 中，第一项来自于确定性 LQR 控制，第二项代表过程噪声引起的附加代价，最后一项是量测不准确带来的附加代价。

分离原理的重要性在于：LQG 控制设计过程可以由两个分离的步骤来完成，即分开设计 Kalman 滤波器和反馈控制。这意味着分别由确定性最优控制问题和最优估计问题得到的所有结果仍然有效。例如，如果过程噪声和测量噪声相关，只需要使用改进的滤波增益。如果噪声统计特性为非高斯，则分离原理得到的是依据最佳线性状态估计的最佳线性调节器。

LQG 调节器的结构框图如图6所示，其中包括了参考或指令输入 $r(t)$ ，而 $s(t)$ 是中间信号。调节器的主方程为

$$\dot{\hat{x}} = (A - LH)\hat{x} + Bu + Lz \quad (142)$$

$$u = -K\hat{x} + r \quad (143)$$

这表示动态输出反馈，即系统输出 $z(t)$ 和控制输入 $u(t)$ 进入动态系统 $\dot{\hat{x}} = (A - LH)\hat{x} + Bu + Lz$ ，其输出 $K\hat{x}$ 作为控制返回。附加动态的功能是重构出无法直接由输出测得的状态信息。其中，反馈动态系统输出的状态其实是原系统状态的最优估计，该输出反馈应区别于完整状态信息情况中的状态反馈。

对于随机系统，我们希望了解调节器能使状态保持多小，也就是确定

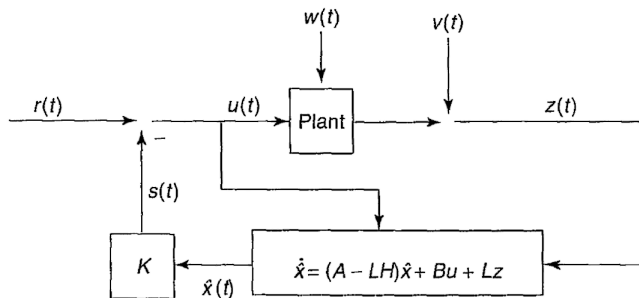


Figure 6: 非完全状态信息 LQG 调节器

所给 LQG 控制作用下的状态均方值 $\mathbf{X}(t) = E[\mathbf{x}(t)\mathbf{x}^T(t)]$ 。显然, 估计误差 $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t)$ 的动力学为

$$\dot{\tilde{\mathbf{x}}} = (\mathbf{A} - \mathbf{LH})\tilde{\mathbf{x}} + \mathbf{G}\mathbf{w} - \mathbf{L}\mathbf{v} \quad (144)$$

注意 $\tilde{\mathbf{x}}(t)$ 独立于控制, 这也是分离原理成立的原因之一。由上式, 可知估计误差的协方差满足

$$\dot{\mathbf{P}} = (\mathbf{A} - \mathbf{LH})^T \mathbf{P} + \mathbf{P}(\mathbf{A} - \mathbf{LH}) + \mathbf{GQ}'\mathbf{G}^T + \mathbf{L}\mathbf{R}'\mathbf{L}^T, \quad t \geq t_0 \quad (145)$$

现在引入残差

$$\tilde{\mathbf{z}}(t) = \mathbf{H}\tilde{\mathbf{x}}(t) + \mathbf{v}(t) \quad (146)$$

则状态估计的动力学方程可化为

$$\dot{\hat{\mathbf{x}}} = (\mathbf{A} - \mathbf{BK}) \hat{\mathbf{x}} + \mathbf{L}\tilde{\mathbf{z}} \quad (147)$$

由于 $\tilde{\mathbf{z}}(t)$ 是独立于 $\hat{\mathbf{x}}$ 的零均值、协方差为 \mathbf{R}' 的白噪声过程, 因此估计量的均方值 $\hat{\mathbf{X}}(t) = E[\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)]$ 依下式传播:

$$\dot{\hat{\mathbf{X}}} = (\mathbf{A} - \mathbf{BK})^T \hat{\mathbf{X}} + \hat{\mathbf{X}} (\mathbf{A} - \mathbf{BK}) + \mathbf{L}\mathbf{R}'\mathbf{L}^T \quad (148)$$

考虑到 $\mathbf{X}(t) = E[\mathbf{x}\mathbf{x}^T] = E[(\hat{\mathbf{x}} + \tilde{\mathbf{x}})(\hat{\mathbf{x}} + \tilde{\mathbf{x}})^T]$, $E(\tilde{\mathbf{x}}) = 0$, 因此

$$\mathbf{X}(t) = \hat{\mathbf{X}}(t) + \mathbf{P}(t) \quad (149)$$

由于状态的初始均值 \bar{x}_0 和协方差 P_0 已知, 由

$$X(t_0) = P_0 + \bar{x}_0 \bar{x}_0^T \quad (150)$$

可知

$$\hat{X}(t_0) = X(t_0) - P_0 = x_0 x_0^T \quad (151)$$

此即为 \hat{X} 传播方程 (148) 的初始条件。同时由 (141), 可将最优的平均代价函数写为

$$j(t_0) = \frac{1}{2} \text{tr} S(t_0) X(t_0) + \frac{1}{2} \text{tr} \int_{t_0}^T S G Q' G^T dt + \frac{1}{2} \text{tr} \int_{t_0}^T K^T R K P dt \quad (152)$$

此外，考察控制的均方值有助于预测需要付出的控制代价（能耗）。显然

$$E [\mathbf{u}(t)\mathbf{u}^T(t)] = \mathbf{K}(t)\hat{\mathbf{X}}(t)\mathbf{K}^T(t) \quad (153)$$

Example 4.2 (导弹滚动姿态调节器) 如图 (7) 所示，对于副翼操控的导弹滚动姿态运动，希望设计反馈控制器使滚动角 ϕ 趋近于 0，并使副翼偏转角 δ 及其变化率 $\dot{\delta}$ 满足物理限制。

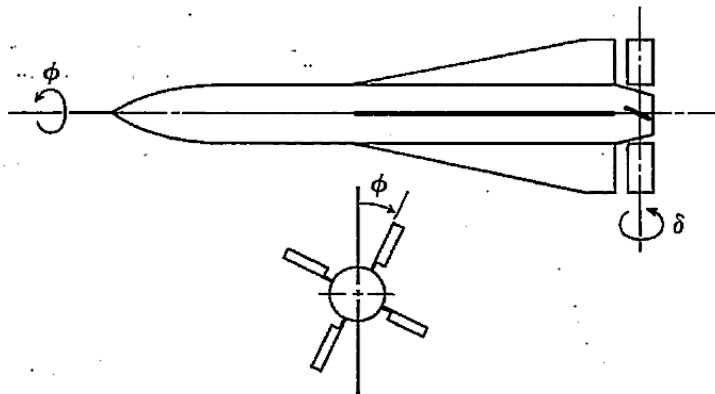


Figure 7: 导弹滚动姿态控制示意图

导弹滚动姿态运动方程可描述为

$$\begin{aligned}\dot{\delta} &= u \\ \dot{\omega} &= -\frac{1}{\tau}\omega + \frac{q}{\tau}\delta + n \\ \dot{\phi} &= \omega\end{aligned}$$

其中 τ 为滚动时间常数, q 为副翼效率, u 为控制指令, ω 为滚动角速度; n 为随机滚动干扰力矩, 假设为白噪声, 且 $E[n(t)] = 0$, $E[n(t)n(\tau)] = N\delta(t - \tau)$ 。

考虑拟最小化的二次型性能指标

$$J = \lim_{t_f - t_0 \rightarrow \infty} \frac{1}{2} \int_{t_0}^{t_f} \left(\frac{\phi^2}{\phi_0^2} + \frac{\delta^2}{\delta_0^2} + \frac{u^2}{u_0^2} \right) dt$$

其中, ϕ_0, δ_0, u_0 分别为 ϕ, δ, u 的最大可能值。

[解] 令 $\mathbf{x} = [\delta, \omega, \phi]^T$, 将模型写为标准线性二次型形式

$$\dot{\mathbf{x}} = \mathbf{F}(t)\mathbf{x} + \mathbf{G}(t)\mathbf{u} + \mathbf{w}(t)$$

$$J = \frac{1}{2} (\mathbf{x}^T \mathbf{S}_f \mathbf{x})_{t=t_f} + \frac{1}{2} \int_{t_0}^{t_f} (\mathbf{x}^T \mathbf{A} \mathbf{x} + \mathbf{u}^T \mathbf{B} \mathbf{u}) dt$$

其中

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 0 \\ \frac{g}{\tau} & -\frac{1}{\tau} & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \frac{1}{\delta_0^2} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{\phi_0^2} \end{bmatrix}, \quad \mathbf{B} = \frac{1}{u_0^2}$$

由 (120) 可知稳态 Riccati 方程为

$$0 = -SF - F^T S + SGB^{-1}G^T S - A$$

式中

$$S = \begin{bmatrix} S_{11} & S_{12} & S_{13} \\ S_{12} & S_{22} & S_{23} \\ S_{13} & S_{23} & S_{33} \end{bmatrix}$$

展开可得 6 个方程

$$-2\frac{q}{\tau}S_{12} + u_0^2 S_{11}^2 - \frac{1}{\delta_0^2} = 0$$

$$\frac{1}{\tau}S_{12} - S_{13} - \frac{q}{\tau}S_{22} + u_0^2 S_{11}S_{12} = 0$$

$$-\frac{q}{\tau}S_{23} + u_0^2 S_{11}S_{13} = 0$$

$$2\frac{1}{\tau}S_{22} - 2S_{23} + u_0^2 S_{12}^2 = 0$$

$$\frac{1}{\tau}S_{23} - S_{33} + u_0^2 S_{12}S_{13} = 0$$

$$u_0^2 S_{13}^2 - \frac{1}{\phi_0^2} = 0$$

令 $\sigma = u_0 S_{11}$, 则有

$$S_{11} = \frac{\sigma}{u_0}, \quad S_{12} = \frac{\tau}{2q} \left(\sigma^2 - \frac{1}{\delta_0^2} \right), \quad S_{13} = \frac{1}{u_0 \phi_0},$$

$$S_{23} = \frac{\tau \sigma}{q \phi_0}, \quad S_{22} = \frac{\tau^2 \sigma}{q \phi_0} - \frac{u_0^2 \tau^3}{8q^2} \left(\sigma^2 - \frac{1}{\delta_0^2} \right)^2, \quad S_{33} = \frac{\sigma}{q \phi_0} + \frac{u_0 \tau}{2q \phi_0} \left(\sigma^2 - \frac{1}{\delta_0^2} \right)$$

σ 需满足

$$\begin{aligned} \sigma^4 + \frac{4}{u_0 \tau} \sigma^3 + \left(\frac{4}{u_0^2 \tau^2} - \frac{2}{\delta_0^2} \right) \sigma^2 - \frac{4}{u_0 \tau} \left(\frac{1}{\delta_0^2} + \frac{2q}{u_0 \phi_0} \right) \sigma \\ + \frac{1}{\delta_0^4} - \frac{4}{u_0^2 \tau^2 \delta_0^2} - \frac{8q}{u_0^3 \tau^2 \phi_0} = 0 \end{aligned}$$

最优控制律为

$$\mathbf{u} = -\mathbf{C}\mathbf{x} = -u_0^2 \begin{bmatrix} S_{11} & S_{12} & S_{13} \end{bmatrix} \begin{bmatrix} \delta \\ w \\ \phi \end{bmatrix}$$

如果 $\tau = 1 \text{ s}$, $q = 10 \text{ s}^{-1}$, $\phi_0 = \frac{\pi}{180} \text{ rad}$, $\delta_0 = \frac{\pi}{12} \text{ rad}$, $u_0 = \pi \text{ rad/s}$, $Q = 0.35 \text{ rad}^2/\text{s}^3$, 则可解得 (注意需使 S 为实正定矩阵) $\sigma = 8.55$, $\mathbf{C} = \begin{bmatrix} 27 & 29 & 180 \end{bmatrix}^T$ 。进一步有

$$\mathbf{F} - \mathbf{G}\mathbf{C} = \begin{bmatrix} -27 & -29 & -180 \\ 10 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.35 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

由此可得

$$\mathbf{X} = \begin{bmatrix} 0.0333 & -0.0149 & -0.0026 \\ -0.0149 & 0.0259 & 0 \\ -0.0026 & 0 & 0.0003 \end{bmatrix}$$

$$E(u^2) = 7.4$$

进一步可求得状态量和控制量的均方根如下：

$$\sqrt{E(\delta^2)} = 0.17, \quad \sqrt{E(\phi^2)} = 0.017, \quad \sqrt{E(w^2)} = 0.16$$

$$\sqrt{E(u^2)} = 2.7$$

□

对于该问题, 如果仅有滚动角的量测 $z = \phi + v$, 其中 $E(v) = 0$, $E[v(t)v(\tau)] = R\delta(t - \tau)$, 大家可以分析该二次型最优滚动姿态调节器的平均性能。

Example 4.3 (保持方位角和滚动姿态的横向自动驾驶仪) 飞机的滚动、偏航、侧滑角运动(横向运动)相互耦合, 但可与俯仰、攻角(纵向运动)解

耦。如图 (8) 所示, 横向扰动运动方程 (五阶系统) 可描述为

$$\dot{\beta} + r = \frac{Y_{\beta}}{mV} (\beta + \beta_n) + \frac{g}{V} \phi$$

$$I_{zz}\dot{r} + I_{xz}\dot{p} = n_{\beta} (\beta + \beta_n) + n_r r + n_p p + n_{\delta_r} \delta_r$$

$$I_{xx}\dot{p} + I_{xz}\dot{r} = l_p p + l_{\beta} (\beta + \beta_n) + l_r r + l_{\delta_a} \delta_a$$

$$\dot{\phi} = p$$

$$\dot{\psi} = r$$

式中 β 为侧滑角, ψ 为偏航角, r 为偏航角速度, ϕ 为滚动角, p 为滚动角速度, δ_r 为方向舵偏角, δ_a 为副翼偏角; β_n 为阵风干扰引起的侧滑角扰动, 可视为白噪声 $E(\beta_n) = 0$, $E[\beta_n(t)\beta_n(\tau)] = N\delta(t - \tau)$ 。

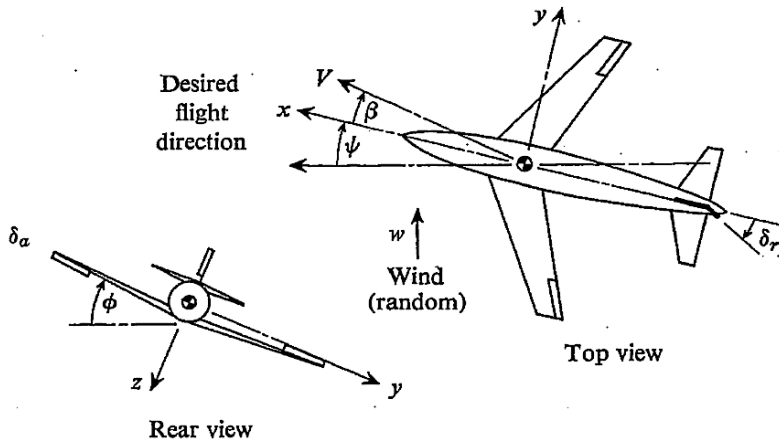


Figure 8: 飞机横向运动控制示意图

设性能指标为

$$J = \lim_{t_f - t_0 \rightarrow \infty} \frac{1}{2} \int_{t_0}^{t_f} \left(\frac{\delta_a^2}{\delta_{a0}^2} + \frac{\delta_r^2}{\delta_{r0}^2} + \frac{(\beta + \psi)^2}{\varepsilon_0^2} + \frac{\phi^2}{\phi_0^2} \right) dt$$

求解最优稳态调节器

$$\begin{bmatrix} \delta_r \\ \delta_a \end{bmatrix} = \begin{bmatrix} C_{r1} & C_{r2} & C_{r3} & C_{r4} & C_{r5} \\ C_{a1} & C_{a2} & C_{a3} & C_{a4} & C_{a5} \end{bmatrix} \begin{bmatrix} \beta \\ r \\ p \\ \phi \\ \psi \end{bmatrix}$$

并计算稳态均方根值 $\sqrt{E(\beta + \psi)^2}$, $\sqrt{E(\phi^2)}$, $\sqrt{E(\delta_a^2)}$, $\sqrt{E(\delta_r^2)}$ 。计算时

相关参数参考取值如下:

$$\begin{aligned}\frac{Y_\beta}{mV} &= -0.0297 \text{ s}^{-1}, \quad \frac{g}{V} = 0.0438 \text{ s}^{-1}, \quad \frac{I_{xz}}{I_{zz}} = -0.0423, \quad \frac{I_{xz}}{I_{xx}} = -0.106 \\ \frac{n_\beta}{I_{zz}} &= 0.379 \text{ s}^{-1}, \quad \frac{n_r}{I_{zz}} = -0.0096 \text{ s}^{-1}, \quad \frac{n_p}{I_{zz}} = -0.0125 \text{ s}^{-1}, \quad \frac{n_{\delta_r}}{I_{zz}} = -0.379 \text{ s}^{-2} \\ \frac{l_p}{I_{xx}} &= -0.790 \text{ s}^{-1}, \quad \frac{l_\beta}{I_{xx}} = -1.17 \text{ s}^{-2}, \quad \frac{l_r}{I_{xx}} = 0.129 \text{ s}^{-1}, \quad \frac{l_{\delta_a}}{I_{xx}} = 1.580 \text{ s}^{-1} \\ \delta_{a0} &= \delta_{r0} = \varepsilon_0 = \phi_0 = 1\end{aligned}$$

[解] 这里直接给出结果，供大家验证参考。

$$\begin{bmatrix} \delta_r \\ \delta_a \end{bmatrix} = \begin{bmatrix} -0.317 & 1.01 & 0.069 & 0.076 & 0.551 \\ 0.177 & 0.388 & 0.737 & 1.03 & 0.834 \end{bmatrix} \begin{bmatrix} \beta \\ r \\ p \\ \phi \\ \psi \end{bmatrix}$$

$$\sqrt{E(\beta + \psi)^2} = 0.60^\circ, \quad \sqrt{E(\phi^2)} = 2.87^\circ$$

$$\sqrt{E(\delta_a^2)} = 3.30^\circ, \quad \sqrt{E(\delta_r^2)} = 2.48^\circ$$

□

5. 离散时间线性二次型高斯问题

5.1 问题描述

类似于连续时间系统，离散时间线性二次型高斯控制问题经常简称离散时间 LQG 问题，或称为离散时间 LQG 调节器问题。研究的系统可描述为

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{G}\boldsymbol{w}_k, \quad k \geq i \quad (154)$$

其中, 过程白噪声 $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}')$, 初始状态 $\mathbf{x}_i \sim \mathcal{N}(\bar{\mathbf{x}}_i, \mathbf{P}_i)$ 。时间区间 $[i, N]$ 上的性能指标为

$$J_i(\mathbf{x}_i) = \frac{1}{2} \mathbf{x}_N^T \mathbf{S}_N \mathbf{x}_N + \frac{1}{2} \sum_{k=i}^{N-1} (\mathbf{x}_k^T \mathbf{Q} \mathbf{x}_k + \mathbf{u}_k^T \mathbf{R} \mathbf{u}_k) \quad (155)$$

其中, $\mathbf{S}_N \geq 0, \mathbf{Q} \geq 0, \mathbf{R} > 0$ 都是对称的权重矩阵。系统矩阵 $\mathbf{A}, \mathbf{B}, \mathbf{G}$ 和权重矩阵 \mathbf{Q}, \mathbf{R} 可以是时变的, 此时可加入下标 k 来描述, 例如 $\mathbf{A}_k, \mathbf{B}_k, \mathbf{Q}_k, \mathbf{R}_k$ 等。希望求解时间区间 $[i, N]$ 上的控制 \mathbf{u}_k^* , 最小化平均代价函数

$$j_i = E[J_i(\mathbf{x}_i)] \quad (156)$$

比较一般地, 认为 \mathbf{x}_N 自由、终端时间 N 固定。

下面依次给出完整和非完全状态信息下的结果, 它们是高斯噪声下的最优控制器, 也是任意噪声下的最佳线性控制律。

在此不进行推导，推导过程类似前一节，但会复杂一些，因为离散系统的曲率矩阵 $(\mathbf{B}^T \mathbf{S}_{k+1} \mathbf{B} + \mathbf{R})$ 和残差协方差 $(\mathbf{H} \mathbf{P}_k \mathbf{H}^T + \mathbf{R}')$ 形式复杂。

5.2 完全状态信息

当 k 时刻的状态 \mathbf{x}_k 精确可知时，LQG 调节器与确定性离散时间 LQR 的求解过程及结果基本相同。首先考虑 $k = N$ ，此时

$$j_N = \frac{1}{2} E(\mathbf{x}_N^T \mathbf{S}_N \mathbf{x}_N) = \frac{1}{2} E_{x_N} \{ E(\mathbf{x}_N^T \mathbf{S}_N \mathbf{x}_N) | \mathbf{x}_N \}$$
$$j_N^* = \frac{1}{2} \text{tr} [\mathbf{S}_N \mathbf{X}_N]$$

其中, $\mathbf{X}_N = E\mathbf{x}_N\mathbf{x}_N^T$ 。当 $k = N - 1$ 时, 有

$$\begin{aligned} j_{N-1} &= \frac{1}{2}E(\mathbf{x}_N^T \mathbf{S}_N \mathbf{x}_N + \mathbf{x}_{N-1}^T \mathbf{Q} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T \mathbf{R} \mathbf{u}_{N-1}) \\ &= \frac{1}{2}E[(\mathbf{A}\mathbf{x}_{N-1} + \mathbf{B}\mathbf{u}_{N-1})^T \mathbf{S}_N (\mathbf{A}\mathbf{x}_{N-1} + \mathbf{B}\mathbf{u}_{N-1}) + \mathbf{x}_{N-1}^T \mathbf{Q} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T \mathbf{R} \mathbf{u}_{N-1}] \\ &\quad + \frac{1}{2}E[\mathbf{w}_{N-1}^T \mathbf{G}^T \mathbf{S}_N \mathbf{G} \mathbf{w}_{N-1}] \end{aligned}$$

由此可导出

$$\mathbf{u}_{N-1}^* = -(\mathbf{B}^T \mathbf{S}_N \mathbf{B} + \mathbf{R})^{-1} \mathbf{B}^T \mathbf{S}_N \mathbf{x}_{N-1} = -\mathbf{K}_{N-1} \mathbf{x}_{N-1}$$

$$\mathbf{K}_{N-1} = (\mathbf{B}^T \mathbf{S}_N \mathbf{B} + \mathbf{R})^{-1} \mathbf{B}^T \mathbf{S}_N$$

$$j_{N-1}^* = \frac{1}{2} \text{tr} [\mathbf{S}_{N-1} \mathbf{X}_{N-1} + \mathbf{S}_N \mathbf{G} \mathbf{Q}' \mathbf{G}^T]$$

$$\mathbf{S}_{N-1} = \mathbf{A}^T \left[\mathbf{S}_N - \mathbf{S}_N \mathbf{B} (\mathbf{B}^T \mathbf{S}_N \mathbf{B} + \mathbf{R})^{-1} \mathbf{B}^T \mathbf{S}_N \right] \mathbf{A} + \mathbf{Q}$$

下面依次考虑 $i = N - 2, N - 3, \dots$, 根据最优性原理, 一般地可导出最优控制如下:

$$\mathbf{u}_k^* = -\mathbf{K}_k \mathbf{x}_k \quad (157)$$

其中

$$\mathbf{K}_k = (\mathbf{B}^T \mathbf{S}_{k+1} \mathbf{B} + \mathbf{R})^{-1} \mathbf{B}^T \mathbf{S}_{k+1} \mathbf{A} \quad (158)$$

称为 Kalman 控制增益。此外, \mathbf{S}_k 满足如下 Riccati 方程:

$$\mathbf{S}_k = \mathbf{A}^T \left[\mathbf{S}_{k+1} - \mathbf{S}_{k+1} \mathbf{B} (\mathbf{B}^T \mathbf{S}_{k+1} \mathbf{B} + \mathbf{R})^{-1} \mathbf{B}^T \mathbf{S}_{k+1} \right] \mathbf{A} + \mathbf{Q}, \quad k < N \quad (159)$$

边界条件为性能指标中终端权重系数矩阵 S_N 。最优平均代价函数为

$$j_i^* = \frac{1}{2} \text{tr} \left[S_i X_i + \sum_{k=i}^{N-1} S_{k+1} G Q' G^T \right] \quad (160)$$

式中, $X_k = E[x_k x_k^T]$, 即系统最优状态的均方值。上式右端第二项为过程噪声引起的附加代价函数。

为了考察 LQG 控制器调节状态的效果, 可以研究闭环系统状态的均方值。由最优控制 u_k^* 反馈作用下的闭环系统

$$x_{k+1} = (A - BK) x_k + G w_k \quad (161)$$

注意到 \mathbf{x}_k 与 \mathbf{w}_k 正交, 由此可得

$$\mathbf{X}_{k+1} = (\mathbf{A} - \mathbf{BK}) \mathbf{X}_k (\mathbf{A} - \mathbf{BK})^T + \mathbf{GQ}'\mathbf{G}^T, \quad k > i \quad (162)$$

初始值为

$$\mathbf{X}_i = \mathbf{P}_i + \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \quad (163)$$

有了最优状态的均方值 \mathbf{X}_k , 亦可导出最优控制的均方值

$$E [\mathbf{u}_k \mathbf{u}_k^T] = \mathbf{K}_k \mathbf{X}_k \mathbf{K}_k^T \quad (164)$$

和连续时间 LQG 问题类似, 不难发现, 离散时间 LQG 问题的最优状态均方值、最优控制均方值、最优平均代价函数等都可以首先离线求解。

5.3 非完全状态信息

当系统状态不完全准确知道时，考虑如下形式的量测方程：

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \quad (165)$$

其中，白噪声 $\mathbf{v}_k \sim \mathcal{N}(0, \mathbf{R}')$ 。此时，需要动态输出反馈。

类似连续时间 LGQ 问题的推导，我们可以证明分离原理对离散时间系统同样成立。即非完全状态信息下的 LQG 控制包含状态估计器和确定性 LQR 控制器两部分，但反馈的是最优状态估计 $\hat{\mathbf{x}}_k$ 而非准确的状态 \mathbf{x}_k 。

为便于应用和查询，我们将最优 LQG 调节器所有方程汇集如下。

动态输出反馈

$$\mathbf{u}_k = -\mathbf{K}_k \hat{\mathbf{x}}_k \quad (166)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{L}_k (\mathbf{z}_k - \mathbf{H} \hat{\mathbf{x}}_{k|k-1}) \quad (167)$$

$$\hat{\mathbf{x}}_{k+1|k} = \mathbf{A} \hat{\mathbf{x}}_k + \mathbf{B} \mathbf{u}_k, \quad \hat{\mathbf{x}}_i = \bar{\mathbf{x}}_i \quad (168)$$

控制与滤波增益

$$\mathbf{K}_k = (\mathbf{B}^\top \mathbf{S}_{k+1} \mathbf{B} + \mathbf{R})^{-1} \mathbf{B}^\top \mathbf{S}_{k+1} \mathbf{A} \quad (169)$$

$$\mathbf{L}_k = \mathbf{P}_{k|k-1} \mathbf{H}^\top (\mathbf{H} \mathbf{P}_{k|k-1} \mathbf{H}^\top + \mathbf{R}')^{-1} = \mathbf{P}_k \mathbf{H}^\top (\mathbf{R}')^{-1} \quad (170)$$

其他辅助方程

$$S_k = A^T S_{k+1} A - K_k^T (B^T S_{k+1} B + R) K_k + Q, \quad S_N \text{ 给定} \quad (171)$$

$$P_{k+1|k} = A P_k A^T + G Q' G^T, \quad P_i \text{ 给定} \quad (172)$$

$$P_k = P_{k|k-1} - L_k (H P_{k|k-1} H^T + R')^{-1} L_k^T \quad (173)$$

LQG 调节器的最优平均代价函数为

$$J_i^* = \frac{1}{2} \text{tr} \left[S_i X_i + \sum_{k=i}^{N-1} S_{k+1} G Q' G^T + \sum_{k=i}^{N-1} S_{k+1} B K_k P_k A^T \right] \quad (174)$$

右端第三项是测量不准确引起的附加代价。注意，如果系统模型、噪声协方差、性能指标权重矩阵是时变的，以上公式需要在相应的时变参数加下标 k 。

为了评估 LQG 控制器的调节效果以及控制能耗, A. E. Bryson 和 Y. C. Ho 提出了状态均方值和控制均方值的计算方法。首先定义估计误差

$$\tilde{\mathbf{x}}_{k|k-1} = \mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1}$$

$$\tilde{\mathbf{x}}_k = \mathbf{x}_k - \hat{\mathbf{x}}_k$$

考虑到 $\tilde{\mathbf{x}}_{k|k-1}$ 与 $\hat{\mathbf{x}}_{k|k-1}$ 相互正交, 可知状态的均方值为

$$\mathbf{X}_k = \mathbf{P}_{k|k-1} + \hat{\mathbf{X}}_{k|k-1} \quad (175)$$

由于一步预测 (先验估计) 的均方值 $\hat{\mathbf{X}}_{k|k-1} = E[\hat{\mathbf{x}}_{k|k-1}\hat{\mathbf{x}}_{k|k-1}^T]$ 还未知, 我

们首先考察一步预测的动态演化

$$\begin{aligned}\hat{\mathbf{x}}_{k+1|k} &= (\mathbf{A} - \mathbf{BK}_k) \hat{\mathbf{x}}_k \\ &= (\mathbf{A} - \mathbf{BK}_k) [\hat{\mathbf{x}}_{k|k-1} + \mathbf{L}_k (\mathbf{z}_k - \mathbf{H} \hat{\mathbf{x}}_{k|k-1})] \\ &= (\mathbf{A} - \mathbf{BK}_k) [\hat{\mathbf{x}}_{k|k-1} + \mathbf{L}_k \mathbf{H} \tilde{\mathbf{x}}_{k|k-1} + \mathbf{L}_k \mathbf{v}_k]\end{aligned}$$

注意到 $\tilde{\mathbf{x}}_{k|k-1}$, $\hat{\mathbf{x}}_{k|k-1}$, \mathbf{v}_k 两两正交, 于是可得

$$\hat{\mathbf{X}}_{k+1|k} = (\mathbf{A} - \mathbf{BK}_k) (\hat{\mathbf{X}}_{k|k-1} + \mathbf{L}_k \mathbf{H} \mathbf{P}_{k|k-1} \mathbf{H}^T \mathbf{L}_k^T + \mathbf{L}_k \mathbf{R}' \mathbf{L}_k^T) (\mathbf{A} - \mathbf{BK}_k)^T$$

代入 \mathbf{P}_k 的传播方程, 可得到一步预测均方值的递归关系

$$\hat{\mathbf{X}}_{k+1|k} = (\mathbf{A} - \mathbf{BK}_k) \left(\hat{\mathbf{X}}_{k|k-1} + \mathbf{P}_{k|k-1} - \mathbf{P}_k \right) (\mathbf{A} - \mathbf{BK}_k)^T \quad (176)$$

由于初始状态均值和协方差 $\bar{\mathbf{x}}_i, \mathbf{P}_i$ 已知, 形式上有

$$\hat{\mathbf{X}}_{i|i-1} = \mathbf{X}_i - \mathbf{P}_{i|i-1} = \mathbf{P}_i - \mathbf{P}_{i|i-1} + \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T$$

因此

$$\hat{\mathbf{X}}_{i+1|i} = (\mathbf{A} - \mathbf{BK}_i) \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T (\mathbf{A} - \mathbf{BK}_i)^T \quad (177)$$

用于启动 (176) 式 $\hat{\mathbf{X}}_{k+1|k}$ 的传播计算。从而由 (175) 便可确定 \mathbf{X}_k , 以此可以判断 LQG 控制器的调节行为是否满意。

对于控制输入的均方值, 因为

$$\mathbf{u}_k = -\mathbf{K}_k \hat{\mathbf{x}}_k = -\mathbf{K}_k (\hat{\mathbf{x}}_{k|k-1} + \mathbf{L}_k \mathbf{H} \tilde{\mathbf{x}}_{k|k-1} + \mathbf{L}_k \mathbf{v}_k)$$

所以

$$E[\mathbf{u}_k \mathbf{u}_k^T] = \mathbf{K}_k (\hat{\mathbf{X}}_{k|k-1} + \mathbf{P}_{k|k-1} - \mathbf{P}_k) \mathbf{K}_k^T \quad (178)$$

以此可以评估 LQG 控制器的能量消耗。

6. 非线性随机系统次优控制方法

动态规划技术在理论上可以求解非线性随机系统的最优控制问题，但实际应用非常困难。即使采用离散化状态和控制变量的方法，也需要十分高昂的计算代价，即“维数灾难”或“计算灾难”。为此需要根据具体问题发展近似处理方法。

6.1 摄动法

摄动方法是处理非线性随机系统控制问题最简单、实用的方法，其基本思想和实现步骤如下：

(1) 不计系统的不确定性和可能的干扰，按确定性最优控制理论求解最优的状态轨道 $\bar{x}(t)$ 及控制 $\bar{u}(t)$ ；

(2) 以 $\bar{x}(t)$ 及 $\bar{u}(t)$ 为基准对系统模型进行线性化，求解最优 (LQR) 的摄动控制 $\delta u^*(t) = \mathbf{K}_\delta(t)\delta x(t)$ ；

(3) 设计摄动系统最优滤波器，获得 $\delta \hat{x}(t)$ ；

(4) 综合次优控制： $u(t) = \bar{u}(t) + \mathbf{K}_\delta(t)\delta \hat{x}(t)$ 。

Example 6.1 对于如下非线性系统：

$$\dot{x}(t) = -x(t) + ax^3(t) + u(t) + w(t)$$

$$y(t) = x(t) + v(t)$$

设计控制策略使得输出 $y(t)$ 尽量接近设定值 y_d 。

[解] 考虑理想平衡状态

$$0 = -\bar{x} + a\bar{x}^3 + \bar{u}, \quad \bar{y} = \bar{x} = y_d,$$

可导出

$$\bar{u} = y_d + ay_d^3, \quad \bar{x} = y_d$$

以此为参考点，对原系统进行线性化，即

$$A = \left. \frac{\partial f}{\partial x} \right|_{\bar{x}, \bar{u}} = -1 + 3a\bar{x}^2 = -1 + 3ay_d^2 \quad B = \left. \frac{\partial f}{\partial u} \right|_{\bar{x}, \bar{u}} = 1$$

于是可得如下摄动系统：

$$\delta \dot{x}(t) = (-1 + 3ay_d^2)\delta x(t) + \delta u(t) + w(t)$$

$$\delta y(t) = \delta x(t) + v(t)$$

对于上述摄动系统，我们可以应用 LGQ 理论获得最优控制 $\delta u^*(t)$ 。最后施加到原系统的近似最优控制策略为 $u(t) = \bar{u} + \delta u^*(t)$ 。 \square

6.2 强迫分离法

除了上面讨论的摄动法，对于非线性随机控制系统，还可以直接应用 LGQ 理论中的分离原理。当然，对于一般的非线性随机系统，无法证明分离原理的正确性，因此最优性完全没有保证。但采用强迫分离的思想，可以快速建立非线性随机系统的可行控制策略。

强迫分离法的思想非常简单，主要步骤如下：

- (1) 忽略不确定性，设计确定性最优控制策略 $\hat{\mathbf{u}}(t) = \hat{\mathbf{u}}[\mathbf{x}(t)]$;
- (2) 设计非线性滤波器 (EKF、UKF、CKF 等)，获得系统状态估计 $\hat{\mathbf{x}}(t)$;
- (3) 实际施加控制策略： $\hat{\mathbf{u}}(t) = \hat{\mathbf{u}}[\hat{\mathbf{x}}(t)]$ 。

这里的 $\mathbf{u}(t)$, $\mathbf{x}(t)$ 分别表示的系统的控制与状态量。

7. 本章小结

最优控制理论在工程实践中具有广泛的应用价值，构成了现代控制理论的重要组成部分。本章针对连续时间系统和离散时间系统两种情况，在介绍确定性最优控制理论的基础上，较全面地讨论了随机系统的动态规划技术。在确定性最优控制理论部分，为了理论方法的完整性，对最优性原理、变分法和极大（小）值原理都进行了简单介绍，但动态规划理论中的最优性原理本章的重点。基于动态规划理论，重点介绍了线性二次型高斯问题，这是随机最优控制最成熟的部分，在理论研究和工程应用中都受到

了广泛关注。其中非常重要的一个内容是关于 LGQ 问题的分离原理，即确定等价原理。

我们遴选了多个飞行器制导与控制相关的例子，有助于大家把握最优控制理论的重要应用领域及相关概念。作为最优控制密切相关的微分对策理论，我们没有专门讨论，但通过一个空间飞行器制导问题讲解了微分对策的主要思想。

关于最优控制的数值计算，在过去几十年中得到了高速发展。本章没有讨论具体的数值算法，仅说明了一些基本概念。有需要的读者，建议通过实践来学习和把握相关的方法和算法。经验表明，实际的编程实践是学习最优控制数值计算最有效的途径。

对于非线性系统，不管是确定性非线性系统，还是随机非线性系统，对应的最优控制问题都比较困难，一直是有意义的研究主题。本章简要介绍了两种近似处理方法，可以帮忙我们快速获得非线性系统的近似最优解。近年来，随着人工智能、机器学习理论的高速发展，已经有若干基于智能理论的新尝试，获得了一些有价值的成果，参见我们及相关团队近年的学术论文。

最优控制理论，特别是随机系统的最优控制，理论成果和工程应用非常丰富，限于篇幅，这里仅介绍了若干基本概念和方法，许多主题不可能逐一论及，有兴趣的同学可以参阅相关参考文献。

Questions?