



# 随机最优控制理论

蔡远利 教授  
西安交通大学自动化学院

# 0. Outline

- 1 最优性原理 / 2
- 2 确定性最优控制基础 / 3
- 3 随机动态规划技术 / 46
- 4 连续时间线性二次型高斯问题 / 84
- 5 离散时间线性二次型高斯问题 / 124
- 6 非线性随机系统次优控制方法 / 137

# 1. 最优性原理

最优性原理 (principle of optimality) 是美国理查德·贝尔曼 (Richard Bellman) 教授 1957 年提出的: 对于一个最优策略, 在任何状态下, 无论过去的历史如何, 做出的选择都应该是从当前状态到最终状态的最优选择。



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

考虑非线性微分方程描述的系统

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

式中,  $n$  维向量时间函数  $\boldsymbol{x}(t)$  由  $m$  维向量时间函数  $\boldsymbol{u}(t)$  确定, 前者表示系统的状态向量, 后者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：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  $u(t) \in \mathcal{U}$  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, 忽略高阶小量可得

$$J^* [\mathbf{x}(t), t] = \min_{\mathbf{u}(t)} \{ L [\mathbf{x}(t), \mathbf{u}(t), t] \Delta t + J^* [\mathbf{x}(t), t] \\ + J_t^* [\mathbf{x}(t), t] \Delta t + J_x^* [\mathbf{x}(t), t] \mathbf{g} [\mathbf{x}(t), \mathbf{u}(t), t] \Delta t \}$$

由此可得

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

引入哈密顿 (Hamilton) 函数

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

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

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

$$\boldsymbol{\psi}[\mathbf{x}(t_f), t_f] = 0 \quad (10)$$

式中  $\boldsymbol{\psi}$  为  $q$  维向量函数。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。

引进拉格朗日乘子  $\boldsymbol{\xi}$  和  $\boldsymbol{\lambda}(t)$  (也称为协态变量), 将性能指标拓展为

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

定义哈密顿函数

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

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

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

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

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

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

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

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

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

该式即为最优控制的必要 (驻点) 条件。

上述方程可以汇总为

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

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

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

上述关系式在变分理论中称为欧拉-拉格朗日方程。

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

当  $\mathbf{u}(t)$  存在容许集  $\mathcal{U}$  (约束集), 在容许集边界  $\frac{\partial H}{\partial \mathbf{u}} = 0$  不再成立。庞特里亚金极小值原理可表述为

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

$$\dot{\mathbf{x}}(t) = \mathbf{g}[\mathbf{x}(t), \mathbf{u}(t), t], \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (21)$$

$$\dot{\boldsymbol{\lambda}}(t) = -\frac{\partial H}{\partial \mathbf{x}}, \quad \boldsymbol{\lambda}(t_f) = -\frac{\partial \varphi}{\partial \mathbf{x}(t_f)} - \boldsymbol{\xi}^T \frac{\partial \psi}{\partial \mathbf{x}(t_f)} \quad (22)$$

$$\frac{\partial \phi}{\partial t_f} = -H[\mathbf{x}(t_f), \mathbf{u}(t_f), \boldsymbol{\lambda}(t_f), t_f] \quad (23)$$

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

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

**Example 2.1 (Linear Quadratic Regulator, LQR)**

$$\dot{\mathbf{x}}(t) = A(t)\mathbf{x}(t) + B(t)\mathbf{u}(t), \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (24)$$

$$\min_{\mathbf{u}(t)} J = \frac{1}{2} \mathbf{x}^T(t_f) S_f \mathbf{x}(t_f) + \frac{1}{2} \int_{t_0}^{t_f} [\mathbf{x}^T Q \mathbf{x} + \mathbf{u}^T R \mathbf{u}] dt \quad (25)$$

其中,  $S_f \geq 0, Q \geq 0, R > 0$  为对称加权矩阵。

**Solution 2.1** 根据式 (7), 此时的哈密顿函数可表示为

$$H[\mathbf{x}(t), \mathbf{u}(t), J_x^*, t] = \frac{1}{2} [\mathbf{x}^T Q \mathbf{x} + \mathbf{u}^T R \mathbf{u}] + J_x^{*T} [A\mathbf{x} + B\mathbf{u}] \quad (26)$$

根据 HJB 方程 (8), 可导出

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

由此可解出

$$\mathbf{u}^* = -R^{-1}B^T J_x^* \quad (27)$$

将对应的  $H^*$  代入 HJB 方程 (8) 得到

$$-J_t^* = \frac{1}{2}\mathbf{x}^T Q \mathbf{x} - \frac{1}{2}J_x^{*T} B R^{-1} B^T J_x^* + J_x^{*T} A \mathbf{x} \quad (28)$$

边界条件为

$$J^* [\mathbf{x}(t_f), t_f] = \frac{1}{2}\mathbf{x}^T(t_f) S_f \mathbf{x}(t_f) \quad (29)$$

根据上述边界条件, 假设

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

其中  $P \geq 0$  为对称矩阵, 则有  $J_t^* = \frac{1}{2} \mathbf{x}^T(t) \dot{P}(t) \mathbf{x}(t)$ ,  $J_x^* = P(t) \mathbf{x}(t)$ , 代入 (28), 考虑到  $2\mathbf{x}^T P A \mathbf{x} = \mathbf{x}^T (P A + A^T P) \mathbf{x}$ , 可整理得到

$$\dot{P} + P A + A^T P - P B R^{-1} B^T P + Q = 0 \quad (31)$$

该式称为矩阵黎卡提 (Riccati) 微分方程, 其边界条件为  $P(t_f) = S_f$ 。

最后可得 LQR 问题的最优解为

$$\mathbf{u}^*(t) = -R^{-1} B^T P \mathbf{x}(t) \quad (32)$$

$$J^*[\mathbf{x}_0, t_0] = \frac{1}{2} \mathbf{x}_0^T P(t_0) \mathbf{x}_0 \quad (33)$$

以上是根据动态规划方法进行的求解，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

$$H[\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}, t] = \frac{1}{2} [\mathbf{x}^T Q \mathbf{x} + \mathbf{u}^T R \mathbf{u}] + \boldsymbol{\lambda}^T [A \mathbf{x} + B \mathbf{u}] \quad (34)$$

由此可知

$$\mathbf{u}^* = -R^{-1} B^T \boldsymbol{\lambda} \quad (35)$$

并得到如下两点边值问题

$$\dot{\mathbf{x}} = A \mathbf{x} - B R^{-1} B^T \boldsymbol{\lambda}, \quad \mathbf{x}(t_0) = \mathbf{x}_0 \quad (36)$$

$$\dot{\boldsymbol{\lambda}}(t) = -A^T \boldsymbol{\lambda} - Q \mathbf{x}, \quad \boldsymbol{\lambda}(t_f) = -S_f \mathbf{x}(t_f) \quad (37)$$

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

**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

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

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

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

**Solution 2.2** 将问题写为  $LQR$  标准形式, 有

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

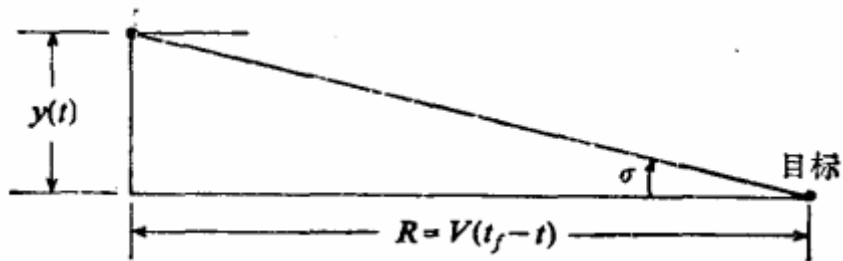


Figure 2: 拦截或交会问题示意图

则由二次型问题的结果可得

$$a(t) = -\Lambda_v(t)v(t) - \Lambda_y(t)y(t) \quad (38)$$

其中

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

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

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

在上述公式，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

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

对应的制导加速度指令为

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

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

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

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

$$a(t) = -3V\dot{\sigma} \quad (39)$$

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

$$\begin{aligned}\dot{\mathbf{r}}_p &= \mathbf{g}_p(t) + \mathbf{a}_p(t), & \dot{\mathbf{r}}_p &= \mathbf{v}_p \\ \dot{\mathbf{r}}_e &= \mathbf{g}_e(t) + \mathbf{a}_e(t), & \dot{\mathbf{r}}_e &= \mathbf{v}_e\end{aligned}$$

其中  $\mathbf{r}, \mathbf{v}, \mathbf{g}, \mathbf{a}$  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。

**Solution 2.3** 根据 LQR 理论, 可导出双方的最优控制策略分别为

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

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

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  $y(t)$ ，则运动方程为

$$\begin{aligned} \dot{v} &= a_p - a_e, & v(t_0) &= v_0 \\ \dot{y} &= v, & y(t_0) &= 0 \end{aligned}$$

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

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

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， $a_{pm} > a_{em}$ 。

### Solution 2.4 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

$$\begin{aligned}\dot{\lambda}_v &= -\lambda_y, & \lambda_v(t_f) &= 0 \\ \dot{\lambda}_y &= 0, & \lambda_y(t_f) &= y(t_f)\end{aligned}$$

由此可解算得到

$$\begin{aligned}\lambda_v(t) &= (t_f - t)y(t_f) \\ \lambda_y(t) &= y(t_f) = \text{Const.}\end{aligned}$$

根据极大（小）值原理，可知双方的最优控制为

$$\begin{aligned}a_p &= -a_{pm} \text{sgn}(\lambda_v) \\ a_e &= -a_{em} \text{sgn}(\lambda_v)\end{aligned}$$

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

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

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  $a_p = a_e + \frac{2v_0}{(t_f - t)}$ 。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。

## 2.2 离散时间系统

考虑如下非线性差分方程描述的系统

$$\mathbf{x}_{k+1} = \mathbf{g}(\mathbf{x}_k, \mathbf{u}_k), \quad k = 0, \dots, N-1 \quad (40)$$

给定初始状态  $\mathbf{x}_0$ , 其中  $n$  维向量  $\mathbf{x}_k$  由  $m$  维向量  $\mathbf{u}_k$  确定。

考虑性能指标

$$J_0(\mathbf{x}_0) = \varphi(\mathbf{x}_N, N) + \sum_{k=0}^{N-1} L_k(\mathbf{x}_k, \mathbf{u}_k) \quad (41)$$

此时, 最优控制问题可描述为求解控制序列  $\{u_k^*\} \in \mathcal{U}$  ( $\mathcal{U}$  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

$$J_k^*(\mathbf{x}_k) = \min_{\substack{\mathbf{u}_i \\ k \leq i < N}} \left[ \varphi(\mathbf{x}_N, N) + \sum_{i=k}^{N-1} L_i(\mathbf{x}_i, \mathbf{u}_i) \right] \quad (42)$$

根据最优性原理, 可知

$$J_k^*(\mathbf{x}_k) = \min_{\mathbf{u}_k} [L_k(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1}^*(\mathbf{x}_{k+1})] \quad (43)$$

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

$$H_k = L_k(\mathbf{x}_k, \mathbf{u}_k) + \boldsymbol{\lambda}_{k+1}^T \mathbf{g}(\mathbf{x}_k, \mathbf{u}_k) \quad (44)$$

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

$$\mathbf{u}_k^* = \arg \min_{\mathbf{u}_k \in \mathcal{U}} H_k(\mathbf{x}_k, \mathbf{u}_k, \boldsymbol{\lambda}_{k+1}) \quad (45)$$

$$\mathbf{x}_{k+1} = \mathbf{g}(\mathbf{x}_k, \mathbf{u}_k), \quad \mathbf{x}_0 \text{ 给定} \quad (46)$$

$$\boldsymbol{\lambda}_k = \frac{\partial H_k}{\partial \mathbf{x}_k}, \quad \boldsymbol{\lambda}_N = \frac{\partial \varphi}{\partial \mathbf{x}_N} \quad (47)$$

可见, 这是一个离散时间的两点边值问题。

**Example 2.5 (线性二次型调节器 (LQR) 问题)**

$$\mathbf{x}_{k+1} = A_k \mathbf{x}_k + B_k \mathbf{u}_k, \quad \mathbf{x}_0 \text{ 给定} \quad (48)$$

$$\min_{\mathbf{u}_k} J_0(\mathbf{x}_0) = \frac{1}{2} \mathbf{x}_N^T S_N \mathbf{x}_N + \frac{1}{2} \sum_{k=0}^{N-1} [\mathbf{x}_k^T Q_k \mathbf{x}_k + \mathbf{u}_k^T R_k \mathbf{u}_k] \quad (49)$$

其中,  $S_N \geq 0, Q \geq 0, R > 0$  为对称加权矩阵。

**Solution 2.5**  $k = N$  时, 由贝尔曼方程的边界条件可知

$$J_N^*(\mathbf{x}_N) = \frac{1}{2} \mathbf{x}_N^T S_N \mathbf{x}_N \triangleq \frac{1}{2} \mathbf{x}_N^T P_N \mathbf{x}_N \quad (50)$$

$k = N - 1$  时, 贝尔曼方程为

$$\begin{aligned} J_{N-1}^*(\mathbf{x}_{N-1}) &= \min_{u_{N-1}} \left[ \frac{1}{2} (\mathbf{x}_{N-1}^T Q_{N-1} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T R_{N-1} \mathbf{u}_{N-1}) + \frac{1}{2} \mathbf{x}_N^T P_N \mathbf{x}_N \right] \\ &= \min_{u_{N-1}} \left[ \frac{1}{2} (\mathbf{x}_{N-1}^T Q_{N-1} \mathbf{x}_{N-1} + \mathbf{u}_{N-1}^T R_{N-1} \mathbf{u}_{N-1}) \right. \\ &\quad \left. + \frac{1}{2} (A_{N-1} \mathbf{x}_{N-1} + B_{N-1} \mathbf{u}_{N-1})^T P_N (A_{N-1} \mathbf{x}_{N-1} + B_{N-1} \mathbf{u}_{N-1}) \right] \end{aligned}$$

由此可导出

$$\mathbf{u}_{N-1}^* = - (R_{N-1} + B_{N-1}^T P_N B_{N-1})^{-1} B_{N-1}^T P_N A_{N-1} \mathbf{x}_{N-1} \triangleq -K_{N-1} \mathbf{x}_{N-1} \quad (51)$$

将其代入  $J_{N-1}^*(\mathbf{x}_{N-1})$  可得

$$\begin{aligned} J_{N-1}^*(\mathbf{x}_{N-1}) &= \frac{1}{2} \mathbf{x}_{N-1}^T [Q_{N-1} + K_{N-1}^T R_{N-1} K_{N-1} \\ &\quad + (A_{N-1} + B_{N-1} K_{N-1})^T P_N (A_{N-1} + B_{N-1} K_{N-1})] \mathbf{x}_{N-1} \\ &\triangleq \frac{1}{2} \mathbf{x}_{N-1}^T P_{N-1} \mathbf{x}_{N-1} \end{aligned} \quad (52)$$

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

$$\mathbf{u}_k^* = -K_k \mathbf{x}_k, \quad J_k^*(\mathbf{x}_k) = \frac{1}{2} \mathbf{x}_k^T P_k \mathbf{x}_k \quad (53)$$

其中

$$K_k = (R_k + B_k^T P_{k+1} B_k)^{-1} B_k^T P_{k+1} A_k \quad (54)$$

$$P_k = Q_k + K_k^T R_k K_k + (A_k + B_k K_k)^T P_{k+1} (A_k + B_k K_k), \quad P_N = S_N \quad (55)$$

式 (55) 称为离散时间矩阵黎卡提 (Riccati) 方程。可见, LQR 问题归结为黎卡提方程的求解。

离散时间最优控制问题本质上是一个多级决策问题, 上述例题表明了基于贝尔曼动态规划技术求解离散时间最优控制问题的原理及过程, 核心思想是从最后一级最优决策开始, 逐步反向回扫。

**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

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

**Solution 2.6** (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

- (1) 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。
- (2) 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。
- (3) 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。

### 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式求解协态方程、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：

- (1) 间接法必须根据特定条件推导解析表达式，比较麻烦；
- (2) 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；
- (3) 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

- ♡ 直接方法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。
- ♡ 经典的直接法包括梯度法、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。
- ♡ 对于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。
- ♡ 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

- ♠ 这里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。
- ♠ 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马尔可夫特性。
- ♠ 动态规划很容易用于求解非线性随机系统的最优控制。

### 3.1 离散时间系统

考虑非线性随机系统

$$\mathbf{x}_{k+1} = \mathbf{g}_k(\mathbf{x}_k, \mathbf{u}_k) + G_k \mathbf{w}_k \quad (56)$$

其中  $\mathbf{w}_k$  为白噪声序列,  $\mathbf{g}_k(\cdot, \cdot)$ ,  $G_k$  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)

$$J_i(\mathbf{x}_i) = \phi(N, \mathbf{x}_N) + \sum_{k=i}^{N-1} L_k(\mathbf{x}_k, \mathbf{u}_k) \quad (57)$$

其中  $\phi$  和  $L_k$  是确定性函数。

## 3.1.1 离散时间系统贝尔曼方程

(1) 代价函数  $J_i(\mathbf{x}_i)$  也是随机量, 显然希望最小化其期望值, 即期望代价

$$j_i(\mathbf{x}_i) = E[J_i(\mathbf{x}_i)] \quad (58)$$

(2) 假设在  $k$  时刻  $\mathbf{x}_k$  是准确可测的, 利用贝叶斯定理, 期望代价可写为

$$j_i(\mathbf{x}_i) = E_{\mathbf{x}_i} \{E[J_i(\mathbf{x}_i) | \mathbf{x}_i]\} \quad (59)$$

其中  $E_{\mathbf{x}_i}$  表示关于  $\mathbf{x}_i$  的期望值。

(3) 定义固定初始状态下的期望代价 (也称条件期望代价)

$$\bar{J}_i(\mathbf{x}_i) = E[J_i(\mathbf{x}_i) | \mathbf{x}_i] \quad (60)$$

$$\implies j_i(\mathbf{x}_i) = E_{\mathbf{x}_i} [\bar{J}_i(\mathbf{x}_i)] \quad (61)$$

(4) 可以证明 (Åström, 1970)

$$j_i^*(\mathbf{x}_i) = \min_{\mathbf{u}_k} [j_i(\mathbf{x}_i)] = E_{\mathbf{x}_i} \left\{ \min_{\mathbf{u}_k} [\bar{J}_i(\mathbf{x}_i)] \right\} = E_{\mathbf{x}_i} [\bar{J}_i^*(\mathbf{x}_i)] \quad (62)$$

(5) 将  $J_k(\mathbf{x}_k)$  分解为两部分

$$J_k(\mathbf{x}_k) = L_k(\mathbf{x}_k, \mathbf{u}_k) + J_{k+1}(\mathbf{x}_{k+1}) \quad (63)$$

取条件期望可得

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

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

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

## (6) 根据最优性原理

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

利用贝叶斯法则

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

再次根据最优性原理，可得在  $k$  时刻的最优条件代价为

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

——完整状态信息下的离散时间非线性随机控制问题的贝尔曼方程。

(7) 考虑到  $L_k(\mathbf{x}_k, \mathbf{u}_k)$  是确定性的, 贝尔曼方程可写为

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

(8) 假设  $k = N$  时刻  $\mathbf{x}_N$  已知, 那么贝尔曼方程边界条件为

$$\bar{J}_N^*(\mathbf{x}_N) = E[\phi(N, \mathbf{x}_N) | \mathbf{x}_N] = \phi(N, \mathbf{x}_N) \quad (70)$$

(9) 根据  $\bar{J}_i^*(\mathbf{x}_i)$ , 可得最小期望代价

$$j_i^* = E_{\mathbf{x}_i} [\bar{J}_i^*(\mathbf{x}_i)] \quad (71)$$

**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

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

**Solution 3.1** 应用贝尔曼方程 (68) 或 (69) 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。

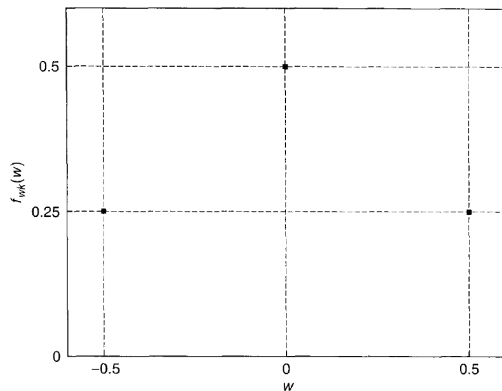


Figure 3: 过程噪声概率密度函数

Table 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

(2) 再考虑  $k = N - 1 = 1$ , 求  $u_1^*$ 。

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

$$x_2 = \begin{cases} 2.0, & P = 0.25 \\ 1.5, & P = 0.50 \\ 1.0, & P = 0.25 \end{cases}$$

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

$$x_2 = \begin{cases} 1.5, & P = 0.25 \\ 1.0, & P = 0.50 \\ 0.5, & P = 0.25 \end{cases}$$

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。

Table 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	×			

Table 2: 第一步优化 (续)

$x_1$	$u_1$	$x_2$	$P(x_2)$	$J_1(x_1) = \bar{J}_2^*(x_2) + \frac{u_1^2}{2}$	$\bar{J}_1(x_1)$	$u_1^*$	$\bar{J}_1^*(x_1)$
		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			

Table 2: 第一步优化 (续)

$x_1$	$u_1$	$x_2$	$P(x_2)$	$J_1(x_1) = \bar{J}_2^*(x_2) + \frac{u_1^2}{2}$	$\bar{J}_1(x_1)$	$u_1^*$	$\bar{J}_1^*(x_1)$
		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	0.5	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。

Table 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	1.0	0.25	0.25	0.40625		
			0.5	0.5	0.375			
		0.0	1.0	0.25	0.625			
			0.5	0.25	0.625			
0.5	0.5	1.5	0.25	0.375	0.3125			

Table 3: 第 0 步优化 (续)

$x_0$	$u_0$	$x_1$	$P(x_1)$	$J_0(x_0) = \bar{J}_1^*(x_1) + \frac{u_0^2}{2}$	$\bar{J}_0(x_0)$	$u_0^*$	$\bar{J}_0^*(x_0)$
		1.0	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  $x_{k+1} = x_k + u_k$ ，该图称为“随机极值场”。

随机极值场图显示了每个  $x_0$  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

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

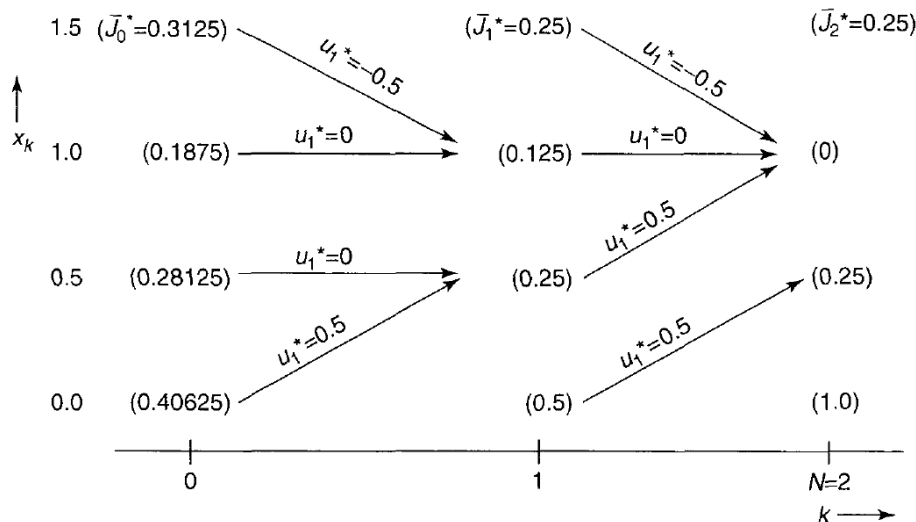


Figure 4: 随机极值场

**Remark 3.1** 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初始状态  $x_i$  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。

**Remark 3.2** 可以使用切普曼-柯尔莫哥洛夫方程来确定转移概率密度函数  $f(x_N|x_i)$ ，取  $i=0$ ，根据马尔可夫性，我们有

$$f(x_N|x_0) = \sum_{x_1, x_2, \dots, x_{N-1}} f(x_N|x_{N-1}) \cdots f(x_2|x_1) f(x_1|x_0) \quad (72)$$

**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。

**Solution 3.2** 根据  $f(x_N|x_0)$  的表达式 (72)，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

$$x_1 = \begin{cases} 1.0, & P = 0.25 \\ 0.5, & P = 0.50 \\ 0.0, & P = 0.25 \end{cases}$$

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

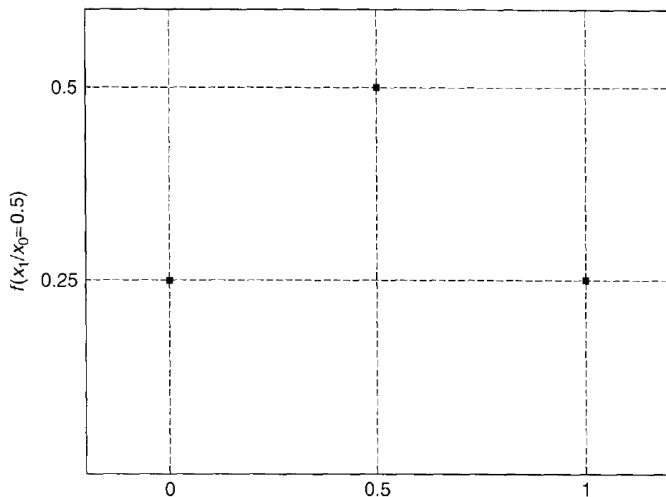
$$f(x_2|x_0 = 0.5) = 0.25f(x_2|x_1 = 1) + 0.5f(x_2|x_1 = 0.5) + 0.25f(x_2|x_1 = 0)$$

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

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

Figure 5:  $x_0 = 0.5$  下的转移概率密度函数

## 3.1.2 概率型性能指标

有时我们希望最大化  $x_N$  在目标集  $\mathbb{S}$  (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

$$J_i(x_i) = P(x_N \in \mathbb{S} | x_i) \quad (73)$$

根据切普曼-柯尔莫哥洛夫方程, 将  $J_k(x_k)$  分解为两个部分

$$\begin{aligned} J_k(x_k) &= \sum_{x_{k+1}} P(x_N \in \mathbb{S} | x_{k+1}) P(x_{k+1} | x_k, u_k) \\ &= \sum_{x_{k+1}} J_{k+1}(x_{k+1}) P(x_{k+1} | x_k, u_k) \end{aligned} \quad (74)$$

根据最优性原理，我们仅需考虑后续代价为最优的情况

$$J_k(x_k) = \sum_{x_{k+1}} P(x_{k+1}|x_k, u_k) J_{k+1}^*(x_{k+1}) \quad (75)$$

最大化容许代价即为

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

这就是**概率型性能指标问题的最优性原理**的数学形式。

## 3.2 连续时间系统

考虑标量非线性随机系统

$$\dot{x} = a(x, u, t) + w \quad (77)$$

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。

定义性能指标为

$$J[x(t_0), t_0] = \phi[x(T), T] + \int_{t_0}^T L(x, u, t) dt \quad (78)$$

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

$$j(t_0) = E\{J[x(t_0), t_0]\} \quad (79)$$

此外，通常还要求终端状态满足如下约束条件

$$\psi [x(T), T] = 0 \quad (80)$$

**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  $x(t)$ 。

(1) 定义给定  $x(t_0)$  下的**条件期望代价函数**为

$$\bar{J} [x(t_0), t_0] = E [J [x(t_0), t_0] | x(t_0)] \quad (81)$$

根据贝叶斯法则，有

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

(2) 类似离散时间问题, 我们有

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

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。

(3) 根据状态方程 (77),  $\Delta t$  为小量时有

$$\Delta x = a\Delta t + w\Delta t \quad (84)$$

其中, 过程噪声  $w\Delta t \sim \mathcal{N}(0, \sigma^2\Delta t)$ 。因此, 在  $x$  和  $t$  给定条件下,  $\Delta x$  服从  $\mathcal{N}(a\Delta t, \sigma^2\Delta t)$ 。即

$$f(\Delta x|x, t) = \frac{1}{\sqrt{2\pi\sigma^2\Delta t}} \exp \left[ -\frac{(\Delta x - a\Delta t)^2}{2\sigma^2\Delta t} \right] \quad (85)$$

注意,  $x(t)$  是马尔可夫过程,  $\Delta x$  只取决于当前状态。

## (4) 将性能指标分解为两部分

$$J(x, t) = J(x + \Delta x, t + \Delta t) + \int_t^{t+\Delta t} L(x, u, \tau) d\tau \quad (86)$$

⇒ (给定  $\Delta x$  求期望)

$$E[J(x, t) | \Delta x] = E \left[ \int_t^{t+\Delta t} L(x, u, \tau) d\tau | \Delta x \right] + E[J(x + \Delta x, t + \Delta t) | \Delta x] \quad (87)$$

⇒ (积分项是控制  $u(\tau)$  ( $t \leq \tau \leq t + \Delta t$ ) 的确定性函数)

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

⇒ (最优性原理)

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

$\implies (\Delta x$  是依赖于过程噪声的随机变量)

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

(5) 完整状态信息下的连续时间随机系统最优性原理

$$\begin{aligned}\bar{J}^*(x, t) &= \min_{\substack{u(\tau) \\ t \leq \tau \leq t + \Delta t}} \left[ \int_t^{t+\Delta t} L(x, u, \tau) d\tau \right. \\ &\quad \left. + \int \bar{J}^*(x + \Delta x, t + \Delta t) f(\Delta x | x, t) d(\Delta x) \right] \quad (90)\end{aligned}$$

其中  $f(\Delta x | x, t)$  即指  $\Delta x \sim \mathcal{N}(a\Delta t, \sigma^2\Delta t)$ 。

## (6) 一阶近似地

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

$\implies$  (根据  $\Delta x \sim \mathcal{N}(a\Delta t, \sigma^2 \Delta t)$ ,  $E\Delta x = a\Delta t$ ,  $E(\Delta x^2) = \sigma^2 \Delta t$ )

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

$\implies$  (忽略  $\Delta t$  二阶以上项)

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

(7) 完整状态信息条件下的随机哈密尔顿-雅可比-贝尔曼 (Hamilton-Jacobi-Bellman, HJB) 方程

$$-\frac{\partial \bar{J}^*}{\partial t} = \min_{\substack{u(\tau) \\ t \leq \tau \leq t + \Delta t}} \left[ L(x, u, \tau) + \frac{\partial \bar{J}^*}{\partial x} a + \frac{1}{2} \frac{\partial^2 \bar{J}^*}{\partial x^2} \sigma^2 \right] \quad (91)$$

边界条件为

$$\bar{J}^* [x(T), T] = \phi [x(T), T] \quad (92)$$

且在超平面  $\psi [x(T), T] = 0$  上。

**Remark 3.3** 对于向量随机系统

$$\dot{x} = A(x, u, t) + Gw \quad (93)$$

其中  $x \in R^n$  且过程白噪声  $w(t) \sim \mathcal{N}(0, Q')$ , 上述 HJB 方程 (91) 推广为

$$-\frac{\partial \bar{J}^*}{\partial t} = \min_{u(t)} \left[ L + \left( \frac{\partial \bar{J}^*}{\partial x} \right)^T A + \frac{1}{2} \text{tr} \left( \frac{\partial^2 \bar{J}^*}{\partial x^2} G Q' G^T \right) \right] \quad (94)$$

**Remark 3.4** 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。

**Remark 3.5 (直观解释和推导)** 对于向量随机系统类似地有

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

$\implies$  (泰勒级数展开)

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

$\implies$  (忽略高阶小)

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

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

**Remark 3.6** 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。

**Remark 3.7** 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。

# 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

$$\dot{\mathbf{x}} = A\mathbf{x} + B\mathbf{u} + G\mathbf{w} \quad (95)$$

其中, 过程白噪声  $\mathbf{w}(t) \sim \mathcal{N}(0, Q')$ , 初始状态  $\mathbf{x}(t_0) \sim \mathcal{N}(\bar{\mathbf{x}}_0, P_0)$ 。相应性能指标为如下二次型形式

$$J[\mathbf{x}(t_0), t_0] = \frac{1}{2}\mathbf{x}^T(T)S_T\mathbf{x}(T) + \frac{1}{2}\int_{t_0}^T (\mathbf{x}^T Q \mathbf{x} + \mathbf{u}^T R \mathbf{u}) dt \quad (96)$$

其中  $S_T \geq 0, Q \geq 0, R > 0$  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  $\mathbf{u}^*(t)$  以最小化期望代价函数

$$j(t_0) = E \{J[\mathbf{x}(t_0), t_0]\} \quad (97)$$

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（Linear Quadratic Gaussian, LQG）控制问题。

**Remark 4.1** 对于完整状态信息情况，在  $t$  时刻状态向量  $x(t)$  准确已知，故  $u(t)$  可以表示为  $x(t)$  的函数，即状态反馈。

**Remark 4.2** 对于非完全状态信息情况，只能获得观测（量测）向量

$$z = Hx + v \quad (98)$$

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

**Remark 4.3** 下面将研究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 (96), 由随机系统的 HJB 方程 (94) 可知

$$-\bar{J}_t^* = \min_{\mathbf{u}(t)} \left[ \frac{1}{2} \mathbf{x}^T Q \mathbf{x} + \frac{1}{2} \mathbf{u}^T R \mathbf{u} + \bar{J}_x^{*T} (A \mathbf{x} + B \mathbf{u}) + \frac{1}{2} \text{tr} (\bar{J}_{xx}^* G Q' G^T) \right] \quad (99)$$

由于控制输入无约束, 最小化的驻点条件为

$$R \mathbf{u} + B^T \bar{J}_x^* = 0 \quad (100)$$

由此可导出最优控制

$$\mathbf{u}^*(t) = -R^{-1} B^T \bar{J}_x^* \quad (101)$$

将其代入 HJB 方程 (99), 即得到柯尔莫哥洛夫方程

$$-\bar{J}_t^* = \frac{1}{2} \mathbf{x}^T Q \mathbf{x} - \frac{1}{2} \bar{J}_x^{*T} B R^{-1} B^T \bar{J}_x^* + \bar{J}_x^{*T} A \mathbf{x} + \frac{1}{2} \text{tr} (\bar{J}_{xx}^* G Q' G^T) \quad (102)$$

类似于确定性 LQR 问题的求解, 可以采用试凑法。

假设

$$\bar{J}^*(\mathbf{x}, t) = \frac{1}{2} \left[ \mathbf{x}^T S \mathbf{x} + \text{tr} \int_t^T S G Q' G^T d\tau \right] \quad (103)$$

其中  $S(t)$  为待定的确定性对称矩阵函数。

⇒ (偏导数)

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

⇒ (代入柯尔莫哥洛夫方程 (102))

$$0 = \frac{1}{2} \mathbf{x}^T \left( \dot{S} + Q + A^T S + SA - SBR^{-T} B^T S \right) \mathbf{x} \quad (104)$$

⇒ (黎卡提方程)

$$-\dot{S} = A^T S + SA - SBR^{-T} B^T S + Q \quad (105)$$

边界条件为  $S(T) = S_T$ 。

⇒ (状态变量反最优控制)

$$\mathbf{u}^*(t) = -R^{-1} B^T S \mathbf{x}(t) \quad (106)$$

⇒(最优平均代价函数)

$$\begin{aligned} j^*(t) &= \frac{1}{2} E_x \left[ \mathbf{x}^T S \mathbf{x} + \text{tr} \int_t^T S G Q' G^T d\tau \right] \\ &= \frac{1}{2} \text{tr} \left[ S(t) X(t) + \int_t^T S G Q' G^T d\tau \right] \end{aligned} \quad (107)$$

其中

$$X(t) = E [\mathbf{x}(t) \mathbf{x}^T(t)] \quad (108)$$

为  $\mathbf{x}(t)$  的均方值, 即  $\mathbf{x}(t)$  的自相关函数。

⇒( $[t_0, T]$  上的最优平均代价函数)

$$j^*(t_0) = \frac{1}{2} \text{tr} \left[ S(t_0) X(t_0) + \int_{t_0}^T S G Q' G^T d\tau \right] \quad (109)$$

**Remark 4.4** 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  $\frac{1}{2}tr(SGQ'G^T)$  这一项，这代表了由功率谱密度矩阵为  $Q'$  的过程噪声  $w(t)$  所增加的代价。

**Remark 4.5** 上述结果也可以根据 Fokker-Planck 方程导出。一般地

$$j(t_0) = \frac{1}{2}E[\mathbf{x}^T(t_0)S(t_0)\mathbf{x}(t_0)] + \frac{1}{2}E\int_{t_0}^T \|R^{-1}B^T S\mathbf{x} + \mathbf{u}\|_R^2 dt \\ + \frac{1}{2}tr\int_{t_0}^T SGQ'G^T dt$$

**Remark 4.6** 通常称

$$K(t) = R^{-1}B^T S(t) \quad (110)$$

为卡尔曼控制增益。最优控制可写为

$$\mathbf{u}^*(t) = -K(t)\mathbf{x}(t) \quad (111)$$

可见是线性时变状态反馈。

### 4.2.2 最优性能评估

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

(3) 在最优控制 (111) 作用下的闭环系统方程为

$$\dot{\boldsymbol{x}} = (A - BK) \boldsymbol{x} + G\boldsymbol{w} \quad (112)$$

由此可导出均方状态  $X(t)$  满足的李雅普诺夫方程

$$\dot{X} = (A - BK) X + X (A - BK)^T + GQ'G^T \quad (113)$$

初始条件为  $X(t_0) = E[\boldsymbol{x}(t_0) \boldsymbol{x}^T(t_0)] = P_0 + \bar{\boldsymbol{x}}_0 \bar{\boldsymbol{x}}_0^T$ 。

(4) 控制信号的均方值也能事先确定

$$E(\boldsymbol{u}\boldsymbol{u}^T) = E(K\boldsymbol{x}\boldsymbol{x}^T K^T) = KX(t)K^T \quad (114)$$

(5) 稳态时状态  $x(t)$  的均方值可通过如下李雅普诺夫代数方程得到

$$0 = (A - BK)X + X(A - BK)^T + GQ'G^T \quad (115)$$

(6) 黎卡提方程 (105) 是按时间倒推求解的, 且独立于  $x(t)$ , 因此可以离线求解得到  $S(t)$ , 之后得到  $K(t)$ ; 然后可由给定初值  $X(t_0)$  按时间正向积分 (113) 计算出  $X(t)$ , 并由 (114) 计算  $E(uu^T)$ ; 而最优期望代价由  $j^*(t_0)$  表达式 (109) 给出。

**Remark 4.7** 上述讨论表明, 在对系统应用随机最优控制之前, 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。

### 4.2.3 非最优控制策略

我们有时希望使用简单形式的次优反馈，例如常值稳态反馈：

$$K_{\infty} = R^{-1} B^T S_{\infty} \quad (116)$$

其中  $S_{\infty}$  是黎卡提方程 (105) 的稳态解。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。

(1) 任意具有  $\mathbf{u} = -K\mathbf{x}$  形式的反馈控制对应的闭环系统方程为

$$\dot{\mathbf{x}} = (A - BK)\mathbf{x} + G\mathbf{w} \quad (117)$$

(2) 可导出非最优的  $j(t_0)$  表达式为

$$j(t_0) = \frac{1}{2}E[\mathbf{x}^T(t_0)S(t_0)\mathbf{x}(t_0)] + \frac{1}{2}\text{tr} \int_{t_0}^T SGQ'G^T dt \\ + \frac{1}{2}E \int_{t_0}^T \mathbf{x}^T \left[ \dot{S} + S(A - BK) + (A - BK)^T S + Q + K^T RK \right] \mathbf{x} dt$$

(3) 选取  $S(t)$  使其在  $t \leq T$  内满足李雅普诺夫方程

$$-\dot{S} = S(A - BK) + (A - BK)^T S + Q + K^T RK, \quad S(T) = S_T \quad (118)$$

(4) 对应的非最优反馈控制平均代价函数等于

$$j(t_0) = \frac{1}{2}E[\mathbf{x}^T(t_0)S(t_0)\mathbf{x}(t_0)] + \frac{1}{2}\text{tr} \int_{t_0}^T SGQ'G^T dt \quad (119)$$

同样还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。

**Solution 4.1** LQG 最优控制为

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

其中卡尔曼控制增益由 (110) 可知为

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

式中  $s(t)$  满足黎卡提方程 (105), 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  $X(t) = E[x^2(t)]$  由李雅普诺夫方程 (113) 给出

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

由 (109) 式可知  $[0, T]$  上的最优平均代价函数为

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

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,  $\tau = T - t$ 。  
 $X(t)$  的李雅普诺夫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。

## 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

(1) 在非完全状态信息条件下，仅有带高斯噪声的量测信息

$$z = Hx + v \quad (120)$$

(2) 可以由卡尔曼滤波器给出系统状态的最优估计

$$\dot{\hat{x}} = A\hat{x} + Bu + L(z - H\hat{x}), \quad \hat{x}(0) = \bar{x}_0 \quad (121)$$

$$L = PH^T R'^{-1} \quad (122)$$

$$\dot{P} = AP + PA^T + GQ'G^T - PH^T R'^{-1}HP, \quad P(t_0) = P_0 \quad (123)$$

其中  $L(t)$  是卡尔曼滤波增益， $\hat{x}(t)$  和  $P(t)$  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。

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

(4) 回顾上小节导出的  $j(t_0)$  表达式

$$j(t_0) = \frac{1}{2} E [\mathbf{x}^T(t_0) S(t_0) \mathbf{x}(t_0)] + \frac{1}{2} E \int_{t_0}^T \|R^{-1} B^T S \mathbf{x} + \mathbf{u}\|_R^2 dt + \frac{1}{2} \text{tr} \int_{t_0}^T S G Q' G^T dt \quad (124)$$

(5) 定义  $t$  时刻的已知观测数据集

$$\mathbf{Z}_t = \{\mathbf{z}(\tau) | t_0 \leq \tau \leq t\} \quad (125)$$

(6)  $j(t_0)$  中的二次型积分项可写为

$$\frac{1}{2} E \int_{t_0}^T \|R^{-1} B^T S \mathbf{x} + \mathbf{u}\|_R^2 dt = \frac{1}{2} E_{\mathbf{Z}_t} \int_{t_0}^T E [(\mathbf{u} + K \mathbf{x})^T R (\mathbf{u} + K \mathbf{x}) | \mathbf{Z}_t] dt \quad (126)$$

其中卡尔曼控制增益  $K = R^{-1}B^T S$ ,  $E_{Z_t}$  表示关于  $Z_t$  的条件期望。

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

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

(8) 积分项 (126) 可化为

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

## (9) 最优控制

$$\mathbf{u}^*(t) = -K(t)\hat{\mathbf{x}}(t) \quad (127)$$

## (10) 最优平均代价函数

$$\begin{aligned} j^*(t_0) = & \frac{1}{2}E[\mathbf{x}^T(t_0)S(t_0)\mathbf{x}(t_0)] + \frac{1}{2}\text{tr}\int_{t_0}^T SGQ'G^T dt \\ & + \frac{1}{2}\text{tr}\int_{t_0}^T K^T R K P dt \end{aligned} \quad (128)$$

### 4.3.2 分离原理

- (1) 上述结果称为非完全信息问题的**分离原理**，也称为**确定性等价原理**。
- (2) **分离原理**表明：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卡尔曼滤波器给出的最优状态估计即可。
- (3) 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。
- (4) 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，即分开设计卡尔曼滤波器和反馈控制。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。

- (5) 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线性最优状态估计的线性最优调节器。

#### 4.3.3 LQG 调节器的结构及平均性能分析

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

$$\dot{\hat{\boldsymbol{x}}} = (A - LH)\hat{\boldsymbol{x}} + B\boldsymbol{u} + Lz \quad (129)$$

$$\boldsymbol{u} = -K\hat{\boldsymbol{x}} + \boldsymbol{r} \quad (130)$$

这表示**动态输出反馈**，而非直接状态反馈。

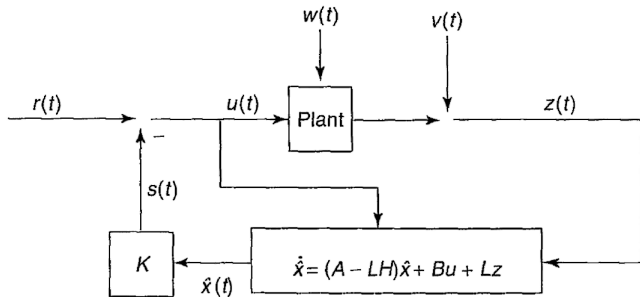


Figure 6: 非完全状态信息 LQG 调节器

估计误差  $\tilde{\mathbf{x}}(t) = \mathbf{x}(t) - \hat{\mathbf{x}}(t)$  的动态方程为

$$\dot{\tilde{\mathbf{x}}} = (A - LH)\tilde{\mathbf{x}} + G\mathbf{w} - L\mathbf{v} \quad (131)$$

可见  $\tilde{\mathbf{x}}(t)$  独立于控制（这也是分离原理成立的原因之一）。由此可导出估计误差协方差方程

$$\dot{P} = (A - LH)^T P + P(A - LH) + GQ'G^T + LR'L^T, \quad t \geq t_0 \quad (132)$$

引入残差（新息） $\tilde{\mathbf{z}}(t) = H\tilde{\mathbf{x}}(t) + \mathbf{v}(t)$ ，可将状态估计的动态方程化为

$$\dot{\hat{\mathbf{x}}} = (A - BK)\hat{\mathbf{x}} + L\tilde{\mathbf{z}} \quad (133)$$

由于  $\tilde{z}(t)$  独立于  $\hat{\boldsymbol{x}}$ , 可导出状态估计的均方值  $\hat{X}(t) = E[\hat{\boldsymbol{x}}(t)\hat{\boldsymbol{x}}^T(t)]$  的传播方程如下:

$$\dot{\hat{X}} = (A - BK)^T \hat{X} + \hat{X} (A - BK) + LR'L^T \quad (134)$$

初始条件为  $\hat{X}(t_0) = \boldsymbol{x}_0 \boldsymbol{x}_0^T$ 。

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

$$X(t) = \hat{X}(t) + P(t) \quad (135)$$

由此可知

$$X(t_0) = P_0 + \bar{\boldsymbol{x}}_0 \bar{\boldsymbol{x}}_0^T \quad (136)$$

于是由 (128) 可求出最优平均代价函数如下:

$$j^*(t_0) = \frac{1}{2} \text{tr}[S(t_0)X(t_0)] + \frac{1}{2} \text{tr} \int_{t_0}^T SGQ'G^T dt + \frac{1}{2} \text{tr} \int_{t_0}^T K^T R K P dt \quad (137)$$

此外, 控制信号的均方值表征了控制代价 (能耗)。显然

$$E [\mathbf{u}(t)\mathbf{u}^T(t)] = K(t)\hat{X}(t)K^T(t) \quad (138)$$

**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  $\phi$  趋近于 0, 并使副翼偏转角  $\delta$  及其变化率  $\dot{\delta}$  满足物理限制。

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

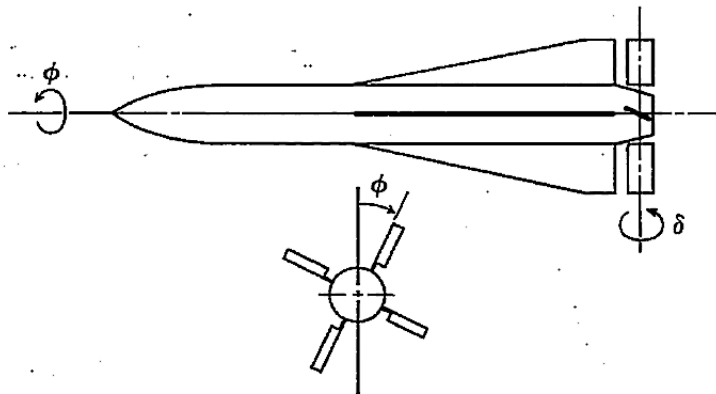


Figure 7: 导弹滚动姿态控制示意图

考虑拟最小化的二次型性能指标为

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

其中,  $\phi_0, \delta_0, u_0$  分别为  $\phi, \delta, u$  的最大可能值。

**Solution 4.2** 令  $\mathbf{x} = [\delta, \omega, \phi]^T$ , 将模型写为标准线性二次型形式

$$\dot{\mathbf{x}} = A(t)\mathbf{x} + B(t)\mathbf{u} + \mathbf{w}(t)$$

$$J = \frac{1}{2} (\mathbf{x}^T S_f \mathbf{x})_{t=t_f} + \frac{1}{2} \int_{t_0}^{t_f} (\mathbf{x}^T Q \mathbf{x} + \mathbf{u}^T R \mathbf{u}) dt$$

其中

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

由 (105) 可知稳态矩阵黎卡提方程为

$$0 = -SA - A^T S + SBR^{-1}B^T S - Q$$

式中

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

矩阵黎卡提方程展开可得 6 个方程

$$\begin{aligned} -2\frac{q}{\tau}s_{12} + u_0^2s_{11}^2 - \frac{1}{\delta_0^2} &= 0 \\ \frac{1}{\tau}s_{12} - s_{13} - \frac{q}{\tau}s_{22} + u_0^2s_{11}s_{12} &= 0 \\ -\frac{q}{\tau}s_{23} + u_0^2s_{11}s_{13} &= 0 \\ 2\frac{1}{\tau}s_{22} - 2s_{23} + u_0^2s_{12}^2 &= 0 \\ \frac{1}{\tau}s_{23} - s_{33} + u_0^2s_{12}s_{13} &= 0 \\ u_0^2s_{13}^2 - \frac{1}{\phi_0^2} &= 0 \end{aligned}$$

令  $\sigma = u_0 s_{11}$ , 则有

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

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

$$s_{33} = \frac{\sigma}{q \phi_0} + \frac{u_0 \tau}{2q \phi_0} \left( \sigma^2 - \frac{1}{\delta_0^2} \right)$$

$\sigma$  需满足

$$\sigma^4 + \frac{4}{u_0 \tau} \sigma^3 + \left( \frac{4}{u_0^2 \tau^2} - \frac{2}{\delta_0^2} \right) \sigma^2 - \frac{4}{u_0 \tau} \left( \frac{1}{\delta_0^2} + \frac{2q}{u_0 \phi_0} \right) \sigma$$

$$+ \frac{1}{\delta_0^4} - \frac{4}{u_0^2 \tau^2 \delta_0^2} - \frac{8q}{u_0^3 \tau^2 \phi_0} = 0$$

最优控制律为

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

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

$$A - BK = \begin{bmatrix} -27 & -29 & -180 \\ 10 & -1 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad Q' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.35 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

由此可得

$$X = \begin{bmatrix} 0.0333 & -0.0149 & -0.0026 \\ -0.0149 & 0.0259 & 0 \\ -0.0026 & 0 & 0.0003 \end{bmatrix}, \quad E(u^2) = 7.4$$

进一步可求得状态量和控制量的均方根如下

$$\sqrt{E(\delta^2)} = 0.17, \quad \sqrt{E(\phi^2)} = 0.017, \quad \sqrt{E(w^2)} = 0.16, \quad \sqrt{E(u^2)} = 2.7$$

对于该问题，如果仅有滚动角的量测  $z = \phi + v$ ，其中  $E(v) = 0, E[v(t)v(\tau)] = R'\delta(t - \tau)$ ，可以进一步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

$$\begin{aligned}\dot{\beta} + r &= \frac{Y_{\beta}}{mV} (\beta + \beta_n) + \frac{g}{V} \phi \\ I_{zz}\dot{r} + I_{xz}\dot{p} &= n_{\beta} (\beta + \beta_n) + n_r r + n_p p + n_{\delta_r} \delta_r \\ I_{xx}\dot{p} + I_{xz}\dot{r} &= l_p p + l_{\beta} (\beta + \beta_n) + l_r r + l_{\delta_a} \delta_a \\ \dot{\phi} &= p \\ \dot{\psi} &= r\end{aligned}$$

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

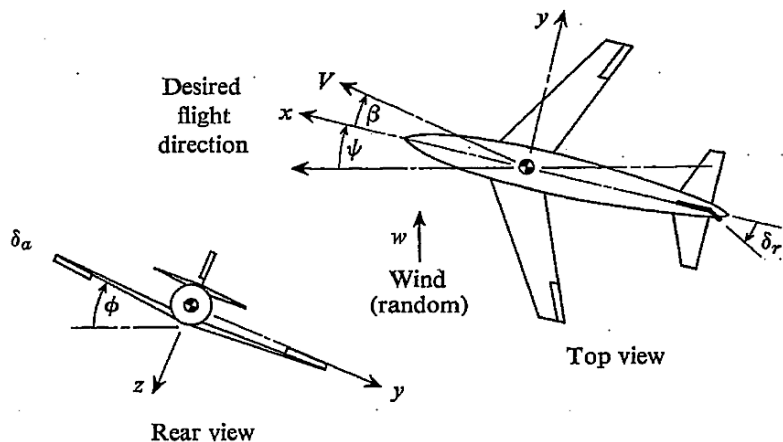


Figure 8: 飞机横向运动控制示意图

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

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

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:

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

**Solution 4.3** 这里直接给出结果，供大家验证参考。

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

$$\begin{aligned} \sqrt{E(\beta + \psi)^2} &= 0.60^\circ, & \sqrt{E(\phi^2)} &= 2.87^\circ \\ \sqrt{E(\delta_a^2)} &= 3.30^\circ, & \sqrt{E(\delta_r^2)} &= 2.48^\circ \end{aligned}$$

# 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

$$\mathbf{x}_{k+1} = F\mathbf{x}_k + B\mathbf{u}_k + G\mathbf{w}_k, \quad k \geq i \quad (139)$$

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

$$J_i(\mathbf{x}_i) = \frac{1}{2} \mathbf{x}_N^T S_N \mathbf{x}_N + \frac{1}{2} \sum_{k=i}^{N-1} (\mathbf{x}_k^T Q \mathbf{x}_k + \mathbf{u}_k^T R \mathbf{u}_k) \quad (140)$$

其中， $S_N \geq 0, Q \geq 0, R > 0$  都是对称的权重矩阵。

系统矩阵  $F, B, G$  和权重矩阵  $Q, R$  可以是时变的，此时可加入下标  $k$  来描述，例如  $F_k, B_k, Q_k, R_k$  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

$$j_i = E[J_i(\mathbf{x}_i)] \quad (141)$$

比较一般地，认为  $\mathbf{x}_N$  自由、终端时间  $N$  固定。

## 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

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

其中,  $X_N = E \mathbf{x}_N \mathbf{x}_N^T$ 。

当  $k = N - 1$  时, 有

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

由此可导出

$$\mathbf{u}_{N-1}^* = -(B^T S_N B + R)^{-1} B^T S_N F \mathbf{x}_{N-1} = -K_{N-1} \mathbf{x}_{N-1}$$

$$K_{N-1} = (B^T S_N B + R)^{-1} B^T S_N F$$

$$j_{N-1}^* = \frac{1}{2} \text{tr} [S_{N-1} X_{N-1} + S_N G Q' G^T]$$

$$S_{N-1} = F^T \left[ S_N - S_N B (B^T S_N B + R)^{-1} B^T S_N \right] F + Q$$

根据最优性原理，一般地可导出最优控制如下

$$\mathbf{u}_k^* = -K_k \mathbf{x}_k \quad (142)$$

其中

$$K_k = (B^T S_{k+1} B + R)^{-1} B^T S_{k+1} F \quad (143)$$

式 (143) 称为卡尔曼控制增益。此外， $S_k$  满足如下黎卡提方程

$$S_k = F^T \left[ S_{k+1} - S_{k+1} B (B^T S_{k+1} B + R)^{-1} B^T S_{k+1} \right] F + Q, \quad k < N \quad (144)$$

最优平均代价函数为

$$j_i^* = \frac{1}{2} \text{tr} \left[ S_i X_i + \sum_{k=i}^{N-1} S_{k+1} G Q' G^T \right] \quad (145)$$

式中,  $X_k = E[\mathbf{x}_k \mathbf{x}_k^T]$ , 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。

由最优控制  $\mathbf{u}_k^*$  反馈作用下的闭环系统

$$\mathbf{x}_{k+1} = (F - BK) \mathbf{x}_k + G\mathbf{w}_k \quad (146)$$

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

$$X_{k+1} = (F - BK) X_k (F - BK)^T + GQ'G^T, \quad k > i \quad (147)$$

初始值为

$$X_i = P_i + \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T \quad (148)$$

有了最优状态的均方值  $X_k$ , 可导出最优控制的均方值

$$E [\mathbf{u}_k \mathbf{u}_k^T] = K_k X_k K_k^T \quad (149)$$

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时先离线求解。

### 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

$$z_k = Hx_k + v_k \quad (150)$$

其中，白噪声  $v_k \sim \mathcal{N}(0, R')$ 。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  $\hat{x}_k$  而非准确的状态  $x_k$ 。

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

$$\mathbf{u}_k = -K_k \hat{\mathbf{x}}_k \quad (151)$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + L_k (\mathbf{z}_k - H \hat{\mathbf{x}}_{k|k-1}) \quad (152)$$

$$\hat{\mathbf{x}}_{k+1|k} = F \hat{\mathbf{x}}_k + B \mathbf{u}_k, \quad \hat{\mathbf{x}}_i = \bar{\mathbf{x}}_i \quad (153)$$

## 控制与滤波增益

$$K_k = (B^T S_{k+1} B + R)^{-1} B^T S_{k+1} F \quad (154)$$

$$L_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R')^{-1} = P_k H^T (R')^{-1} \quad (155)$$

## 其他辅助方程

$$S_k = F^T S_{k+1} F - K_k^T (B^T S_{k+1} B + R) K_k + Q, \quad S_N \text{ 给定} \quad (156)$$

$$P_{k+1|k} = F P_k F^T + G Q' G^T, \quad P_i \text{ 给定} \quad (157)$$

$$P_k = P_{k|k-1} - L_k (H P_{k|k-1} H^T + R')^{-1} L_k^T \quad (158)$$

## LQG 调节器的最优平均代价函数

$$j_i^* = \frac{1}{2} \text{tr} \left[ S_i X_i + \sum_{k=i}^{N-1} S_{k+1} G Q' G^T + \sum_{k=i}^{N-1} S_{k+1} B K_k P_k F^T \right] \quad (159)$$

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  $k$ 。

LQG 调节器的性能评估 根据估计误差的定义

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

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

考虑到  $\tilde{\mathbf{x}}_{k|k-1}$  与  $\hat{\mathbf{x}}_{k|k-1}$  相互正交, 可知**状态均方值**为

$$X_k = P_{k|k-1} + \hat{X}_{k|k-1} \quad (160)$$

由于一步预测 (先验估计) 的均方值  $\hat{X}_{k|k-1} = E[\hat{\mathbf{x}}_{k|k-1} \hat{\mathbf{x}}_{k|k-1}^T]$  还未知, 我们首先考察一步预测的动态演化

$$\begin{aligned} \hat{\mathbf{x}}_{k+1|k} &= (F - BK_k) \hat{\mathbf{x}}_k \\ &= (F - BK_k) [\hat{\mathbf{x}}_{k|k-1} + L_k (z_k - H \hat{\mathbf{x}}_{k|k-1})] \\ &= (F - BK_k) [\hat{\mathbf{x}}_{k|k-1} + L_k H \tilde{\mathbf{x}}_{k|k-1} + L_k \mathbf{v}_k] \end{aligned}$$

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

$$\hat{X}_{k+1|k} = (F - BK_k)(\hat{X}_{k|k-1} + L_k H P_{k|k-1} H^T L_k^T + L_k R' L_k^T)(F - BK_k)^T$$

代入  $P_k$  的传播方程, 可得到一步预测均方值的递归关系

$$\hat{X}_{k+1|k} = (F - BK_k) \left( \hat{X}_{k|k-1} + P_{k|k-1} - P_k \right) (F - BK_k)^T \quad (161)$$

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

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

因此

$$\hat{X}_{i+1|i} = (F - BK_i) \bar{\mathbf{x}}_i \bar{\mathbf{x}}_i^T (F - BK_i)^T \quad (162)$$

上式作为初值用于启动 (161) 式  $\hat{X}_{k+1|k}$  的传播计算。有了  $\hat{X}_{k+1|k}$ ，由 (160) 便可确定  $X_k$ ，以此可以判断 LQG 控制器的调节行为是否满意。

对于控制输入均方值，因为

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

所以

$$E [\mathbf{u}_k \mathbf{u}_k^T] = K_k (\hat{X}_{k|k-1} + P_{k|k-1} - P_k) K_k^T \quad (163)$$

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  $\bar{\mathbf{x}}(t)$  及控制  $\bar{\mathbf{u}}(t)$ ；

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

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

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

**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  $y_d$ 。

**Solution 6.1** 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

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

于是可得如下摄动系统

$$\begin{aligned}\delta \dot{x}(t) &= (-1 + 3ay_d^2)\delta x(t) + \delta u(t) + w(t) \\ \delta y(t) &= \delta x(t) + v(t)\end{aligned}$$

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  $u(t) = \bar{u} + \delta u^*(t)$ 。

## 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  $\mathbf{u}(t)$ ,  $\mathbf{x}(t)$  分别表示系统的控制与状态量。

## 6. 本章小结

- (1) 最优性原理是随机系统动态规划技术的核心;
- (2) LQG 理论是随机最优控制最重要和最成熟的内容;
- (3) 要求深入理解和掌握分离原理 (确定性等价原理);
- (4) 最优控制与微分对策理论在飞行器制导与控制等领域已经成为非常重要的工具;
- (5) 与人工智能、机器学习理论相结合, 是值得关注的方向发展。

Questions?