

2025–2026 学年秋冬学期
优化理论与算法研讨：动态优化专题

Lec 02: 最优控制

吴一航 yhwu_is@zju.edu.cn

浙江大学计算机科学与技术学院



CONTENT

目录

1. 最优控制问题

2. 极小值原理

3. 动态规划与 HJB 方程

停车问题

考虑一个汽车沿水平线移动的简单模型。设 $x \in \mathbb{R}$ 是汽车位置， u 是加速度，限定在有界区间 $U = [-1, 1]$ （负加速度对应制动）。汽车的动力学方程为 $\ddot{x} = u$ 。我们的目标是让汽车经过最短的时间在原点停车。

显然上述问题属于动态优化问题，因为我们需要为每个时刻选择一个加速度 $u(t)$ 以实现优化目标。不难总结出上述问题有两个关键变量，即**状态变量** $x(t)$ 和**控制变量** $u(t)$ ，因此这类问题被称为**最优控制问题**（optimal control problem）。要定义一个最优控制问题，我们需要三个关键要素，接下来逐一展开介绍：

1. 控制系统：描述状态变量如何随时间演化；
2. 成本泛函：衡量控制策略优劣的标准；
3. 目标集合：终端时间和状态需要满足的约束条件。

控制系统描述了状态变量 $x(t)$ 如何随时间演化，通常用微分方程表示：

$$\dot{x} = f(t, x, u), \quad x(t_0) = x_0.$$

例如在停车问题中，重记状态变量（汽车位置）为 x_1 ，并记汽车速度为 x_2 ，则控制系统为

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u \end{cases}$$

在一般问题中，状态变量 $x(t)$ 可以是多维向量（如停车问题），控制变量 $u(t)$ 也可以是多维向量，且控制变量通常受到约束，即 $u(t) \in U \subseteq \mathbb{R}^m$ 。

此外，还需要对 f 和 u 的性质做出一些假设，以保证控制系统解的存在唯一性：假设 f 关于 t 和 u 连续，且关于 x 满足 Lipschitz 条件；假设控制变量 u 关于 t 是可测且局部有界的函数。

最一般的成本泛函形式为

$$J(u) = \int_{t_0}^{t_f} L(t, x(t), u(t)) dt + K(t_f, x_f).$$

其中 t_f 和 $x_f = x(t_f)$ 分别为终端时间和终端状态, L 称为**运行成本** (running cost), K 称为**终端成本** (terminal cost)。在停车问题中, 我们的目标是最小化停车时间, 因此运行成本 $L \equiv 1$, 且无终端成本 $K \equiv 0$ 。

事实上成本泛函依赖于初始时间 t_0 、初始状态 x_0 以及终端时间 t_f , 但为了简化符号我们省略了这些依赖关系。此外, 证明后续极小值原理时需要假设 L 和 K 关于各自变量连续可微 (实际上不需要这么强, 但这里不展开详细的技术性说明)。

上述成本函数形式被称为 **Bolza 形式**, 在某些应用中也会使用其他形式, 例如 **Mayer 形式** (仅含终端成本) 和 **Lagrange 形式** (仅含运行成本)。事实上这些形式是等价的, 可以相互转换。

对于有终端成本 K 的问题，可以做如下改写转化为 Lagrange 形式：

$$\begin{aligned} K(t_f, x_f) &= K(t_0, x_0) + \int_{t_0}^{t_f} \frac{d}{dt} K(t, x(t)) dt \\ &= K(t_0, x_0) + \int_{t_0}^{t_f} [K_t(t, x(t)) + K_x(t, x(t)) \cdot f(t, x(t), u(t))] dt. \end{aligned}$$

另一方面，对于有运行成本 L 的问题，可以引入辅助状态变量 x^0 ，满足

$$\dot{x}^0 = L(t, x, u), \quad x^0(t_0) = 0,$$

则成本泛函可改写为 $J(u) = x^0(t_f) + K(t_f, x_f)$ 。

还有一个常见技巧是定义辅助状态变量 $x_{n+1} = t$ ，将时间 t 纳入状态变量

$$J(u) = \int_{t_0}^{t_f} L(x(t), u(t)) dt + K(x_f), \quad x \in \mathbb{R}^{n+1}.$$

最后需要定义目标集合，规定终端时间和状态需要满足的约束条件。一般地，目标集合可表示为

$$(t_f, x_f) \in S \subseteq [t_0, \infty) \times \mathbb{R}^n.$$

- 若 $S = [t_0, \infty) \times \mathbb{R}^n$, 则是自由时间自由端点问题;
- 若 $S = \{t_1\} \times \mathbb{R}^n$, 则是固定时间自由端点问题;
 - 若 $S = \{t_1\} \times \{x_1\}$, 则是固定时间固定端点问题;
- 若 $S = [t_0, \infty) \times \{x_1\}$, 则是自由时间固定端点问题;
 - 此类情况可以进一步推广为 $S = [t_0, \infty) \times S_1$, 其中 S_1 是 \mathbb{R}^n 的一个曲面 (流形);
- 通过 $x_{n+1} = t$ 的技巧, 可以将固定时间问题转化为自由时间问题, 同时端点需要增加一个维度及其约束。

最优控制可以解决的问题有很多，除去此前介绍的停车问题外，例如航天器的轨道转移问题如何做到燃料消耗最小，宏观经济学中的最优增长模型等，都是最优控制的经典应用。

读者可能会好奇，最优控制问题与此前介绍的变分问题有何关系？事实上，变分法中的最优曲线可以视为是通过某种控制系统生成的状态轨迹，因此变分问题可以很容易地转化为最优控制问题。事实上最优控制问题的适用范围更广，因为它允许更复杂的控制系统和约束条件，这些条件很难表达在变分问题中。

考虑一个质点在重力场中从点 $A(a, 0)$ 沿光滑曲线 $y(x)$ 滑向点 $B(b, y_1)$, 求使滑行时间最短的曲线。设质点位置为 $(x(t), y(t))$, 则能量守恒定律表明 $\dot{x}^2 + \dot{y}^2 = y$ (为简化讨论, 令 $g = 1/2$)。令 $u_1 = \dot{x}/\sqrt{y}$, $u_2 = \dot{y}/\sqrt{y}$, 则可以将最速降线问题表述为如下最优控制问题:

- 控制系统: 控制约束为 $u_1^2 + u_2^2 = 1$, 控制系统为

$$\begin{cases} \dot{x} = u_1 \sqrt{y} \\ \dot{y} = u_2 \sqrt{y} \end{cases}$$

- 成本泛函:

$$J(u) = \int_{t_0}^{t_1} 1 dt = t_1 - t_0.$$

- 目标集合: $(x(t_1), y(t_1)) \in S = [t_0, \infty) \times \{(b, y_1)\}$ 。

CONTENT

目录

1. 最优控制问题

2. 极小值原理

3. 动态规划与 HJB 方程

美苏冷战期间，军事、航天等领域对复杂动态系统的优化控制需求激增，变分法已无法面对此类复杂问题，促使最优控制理论迅速发展。

1952 年，美国数学家理查德·贝尔曼 (Richard Bellman) 加入兰德公司，提出了著名的动态规划方法（下一节介绍）。1955 年，苏联组建团队研究最优控制问题，以列夫·庞特里亚金 (Lev Pontryagin) 为首，带领他的两位年轻学生于 1956-1957 年间发展出极大值原理 (Pontryagin's maximum principle)，给出了最优控制问题取极大值的必要条件，适用于控制变量分段连续的问题，放宽了控制约束要求，因此是经典变分法的重要推广。

1960 年，首届国际自动控制联合会 (IFAC) 在莫斯科举行，极大值原理向全世界学术界展示，贝尔曼也参与了这次会议。1962 年，庞特里亚金师徒三人出版了《最优控制理论》(The Mathematical Theory of Optimal Processes) 一书，将问题改变为最小化问题，同时将该方法更名为**庞特里亚金极小值原理 (Pontryagin's minimum principle)**。

值得一提的是，庞特里亚金 13 岁就因为一次家里的汽油炉出了故障爆炸意外而失去了光明。母亲爱儿心切，便放下所有工作，白天送他上学、陪他上课，晚上帮他复习，把书上的内容念给他听。1925 年，17 岁的庞特里亚金由于成绩优异，被保送至他梦寐以求的莫斯科大学数学物理系，师从苏联著名拓扑学家帕维尔·亚历山德罗夫 (Pavel Aleksandrov) 院士。

正所谓名师出高徒，在他的引导下，庞特里亚金 19 岁便建立了著名的庞特里亚金对偶定理，揭示了局部紧交换群与其对偶群之间的同构关系，被认为是上世纪拓扑学最重要的成就之一。此后，庞特里亚金 27 岁博士毕业，并在同年成为莫斯科大学教授。他提出了配边理论的基本问题。这导致了 1940 年左右一个示性类理论的引入，现在被称为庞特里亚金类。

首先陈述 $f = f(x, u)$, $L = L(x, u)$, $K \equiv 0$ 且目标集 $S = [t_0, \infty] \times \{x_1\}$ 的基本固定端点控制问题的极小值原理。注意后续讨论均默认最优控制的存在性，尽管这并非总是成立的，但技术性较强故省略讨论。

注意此类问题中 f 和 L 不显示依赖于时间 t ，之前的讨论已经说明这可以通过对一般问题令 $x_{n+1} = t$ 来实现；此外，此前也讨论了如何将一般成本泛函转化为 $K \equiv 0$ 的情况。因此以上两点不构成实质性限制。唯一的限制在于目标集合为固定端点形式，因此之后也将推广到变端点情况。

极小值原理（基本固定端点控制问题）

令 $u^* : [t_0, t_f] \rightarrow U$ 为最优控制, $x^* : [t_0, t_f] \rightarrow \mathbb{R}^n$ 为对应的最优状态轨迹。则存在一个函数 $p^* : [t_0, t_f] \rightarrow \mathbb{R}^n$, 称为**伴随变量** (adjoint variable), 或**协态变量** (co-state variable), 以及一个常数 p_0^* , 满足对任意的 $t \in [t_0, t_f]$ 都有 $(p_0^*, p^*(t)) \neq 0$, 且满足以下条件:

1. 规范方程 (注意边值条件 $x^*(t_0) = x_0, \quad x^*(t_f) = x_1$):

$$\begin{cases} \dot{x}^* = H_p(x^*, u^*, p^*, p_0^*) \\ \dot{p}^* = -H_x(x^*, u^*, p^*, p_0^*) \end{cases}$$

其中哈密顿量 $H(x, u, p, p_0) = \langle p, f(x, u) \rangle + p_0 L(x, u)$;

2. 极值条件: 对任意固定的 t , 函数 $u \mapsto H(x^*(t), u, p^*(t), p_0^*)$ 在 $u = u^*(t)$ 处取得极小值, 即对任意的 $t \in [t_0, t_f]$ 都有

$$H(x^*(t), u^*(t), p^*(t), p_0^*) = \min_{u \in U} H(x^*(t), u, p^*(t), p_0^*);$$

3. 静态条件: $H(x^*, u^*, p^*, p_0^*) = 0, \forall t \in [t_0, t_f]$; 若终端时间 t_f 固定, 条件变为 $H(x^*, u^*, p^*, p_0^*) = \text{const}, \forall t \in [t_0, t_f]$ 。

考虑此前介绍的最速降线最优控制问题，应用极小值原理求解。显然最速降线问题属于基本固定端点控制问题。首先构造哈密顿量

$$H = \langle p, f(x, u) \rangle + p_0 L = p_1 u_1 \sqrt{y} + p_2 u_2 \sqrt{y} + p_0 \equiv 0.$$

由于 $u_1^2 + u_2^2 = 1$ ，根据极值条件， $(u_1^*(t), u_2^*(t))$ 应当与

$$(p_1^*(t), p_2^*(t))$$

共线但反向，结合约束有

$$(u_1^*(t), u_2^*(t)) = -\frac{1}{|p|} (p_1^*(t), p_2^*(t)).$$

其中 $|p| = \sqrt{p_1^2 + p_2^2}$ 。然后考察规范方程，第一条自动满足，第二条表明

$$\dot{p}_1 = 0,$$

$$\dot{p}_2 = -\frac{p_1 u_1 + p_2 u_2}{2\sqrt{y}} = \frac{|p|}{2\sqrt{y}}.$$

故 p_1 是一个常数，若 $p_1 = 0$ ，则 $u_1 \equiv 0$ ，则最速降线是垂直线；若 $p_1 \neq 0$ ，考虑最优曲线 $y(x)$ 的性质，不难得到

$$y'(x) = \frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{u_2}{u_1} = \frac{p_2}{p_1},$$

$$y''(x) = \frac{1}{p_1} \frac{dp_2}{dx} = -\frac{|p|^2}{2p_1^2 y}.$$

从而可以得到

$$1 + (y')^2 + 2yy'' = 0.$$

解这一微分方程即可得到摆线方程。

下面考虑变端点控制问题的极小值原理。相较于之前的基本固定端点控制问题，唯一的变化在于目标集合变为 $S = [t_0, \infty) \times S_1$ ，其中 S_1 是 \mathbb{R}^n 的一个 k 维曲面 ($k \leq n$)，这样的曲面可以通过如下等式约束定义：

$$S_1 = \{x \in \mathbb{R}^n : h_1(x) = \dots = h_{n-k}(x) = 0\}$$

极小值原理（基本变端点控制问题）

令 $u^* : [t_0, t_f] \rightarrow U$ 为最优控制， $x^* : [t_0, t_f] \rightarrow \mathbb{R}^n$ 为对应的最优状态轨迹。则存在一个函数 $p^* : [t_0, t_f] \rightarrow \mathbb{R}^n$ 和一个常数 p_0^* ，满足对任意的 $t \in [t_0, t_f]$ 都有 $(p_0^*, p^*(t)) \neq 0$ ，且满足以下条件：

1. 规范方程：与此前一致，边值条件 $x^*(t_0) = x_0, x^*(t_f) \in S$ ；
2. 极值条件：与此前一致；
3. 静态条件：与此前一致；
4. 横截条件： $p^*(t_f)$ 与 S_1 在 $x^*(t_f)$ 处的切空间正交，即对任意 $i = 1, \dots, n - k$ 都有 $\langle p^*(t_f), \nabla h_i(x^*(t_f)) \rangle = 0$ 。

极小值原理的证明略去，感兴趣的读者可以参考网站上给出的参考材料，其主要思想是考察扰动控制问题的特点。

回忆停车问题，记汽车位置为 x_1 ，并记汽车速度为 x_2 ，则控制系统为

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u \end{cases}$$

控制约束为 $u \in U = [-1, 1]$ ，运行成本 $L \equiv 1$ ，且无终端成本。目标集合为 $S = [t_0, \infty) \times \{0\}$ 。构造哈密顿量 $H = p_1 x_2 + p_2 u + p_0$ 。显然这一问题属于基本固定端点控制问题，根据规范方程有

$$\begin{aligned} \dot{p}_1^* &= 0, \\ \dot{p}_2^* &= -p_1^*. \end{aligned}$$

故第一行表明 p_1^* 是常数，记为 c_1 ，则第二行表明 $p_2^*(t) = -c_1 t + c_2$ ，其中 c_2 也是常数。根据极值条件，最优控制为

$$u^*(t) = \text{sgn}(p_2^*(t)) = \begin{cases} 1 & \text{if } p_2^*(t) < 0 \\ -1 & \text{if } p_2^*(t) > 0 \\ \text{arbitrary} & \text{if } p_2^*(t) = 0 \end{cases}$$

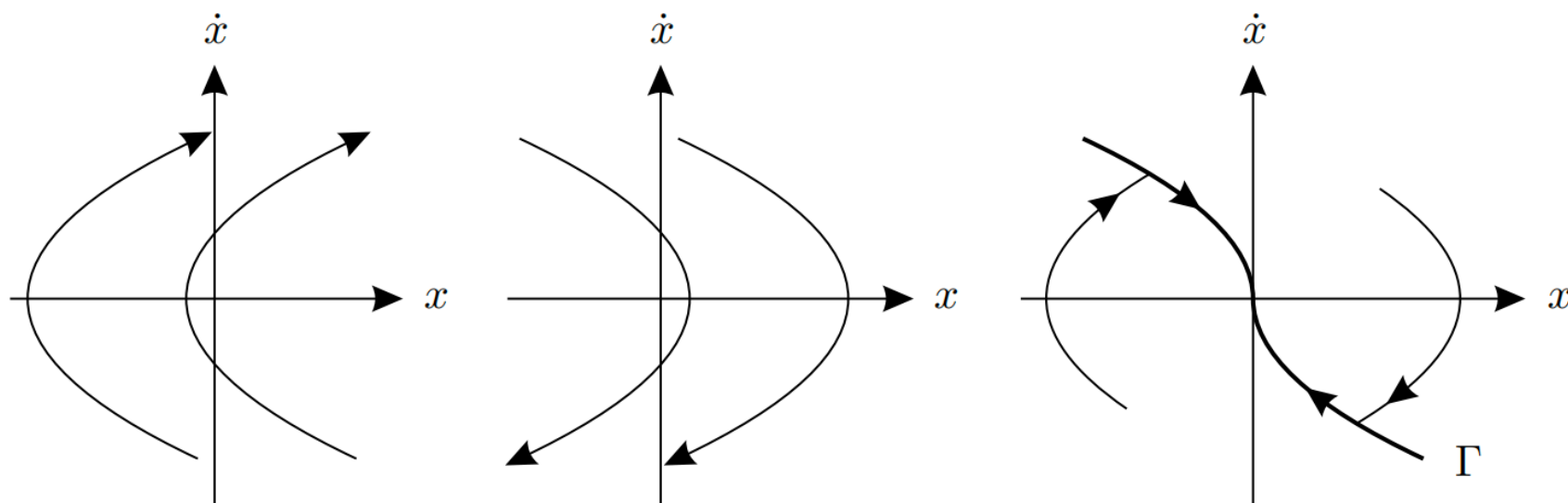
上式表明，当 $p_2^*(t) \neq 0$ 时，最优控制 $u^*(t)$ 确定，但当 $p_2^*(t) = 0$ 时，最优控制不唯一。我们自然希望最优控制是确定的，因此想知道 $p_2^*(t)$ 是否可能在某一区间上恒为零。注意到若 $p_2^*(t) \equiv 0$ ，则 $c_1 = c_2 = 0$ ，从而 $p^*(t) \equiv 0$ ，然而极小值原理中的静态条件表明这又会导致 $p_0^* = 0$ ，这与极小值原理中的 $(p_0^*, p^*(t)) \neq 0$ 矛盾。

因此 $p_2^*(t)$ 不可能在某一区间上恒为零，故 $c_1 \neq 0$ ，这表明 $p_2^*(t)$ 最多只能有一个零点。自然地，在唯一的零点上最优控制可以取 1 或 -1 以与前后的控制相连接。这表明在停车问题中，最优控制只能取 ± 1 ，并且最多切换一次。回归到其物理意义也就是，汽车只会选择使用最大或最小加速度。这种只在极值点取值的控制策略被称为 **Bang-Bang 控制**。

Bang-Bang 控制又称起停式控制，或开关控制，或砰砰控制。这一名字来源于其控制信号在最大值和最小值之间切换时电磁继电器发出的声音。

Bang-Bang 控制在实际应用中非常常见，一个典型的例子就是定频空调，温度高于设定温度则启动制冷降温，低于设定温度则停止工作。使用中可能会添加一个滞回区间减少频繁启停。

事实上根据起始状态（位置和速度）可以唯一确定最优控制策略。考虑在 (x, \dot{x}) 平面上绘制汽车的状态轨迹，当 $u \equiv 1$ 时，积分得到 $\dot{x}(t) = t + a$ ， $x(t) = 1/2t^2 + at + b$ （其中 a, b 为常数），从而可以得到 $x = 1/2\dot{x}^2 + c$ （其中 $c = b - a^2/2$ ），这给出了 (x, \dot{x}) 平面上的一族以 c 为参数的抛物线。同理，当 $u \equiv -1$ 时，可以得到另一族抛物线 $x = -1/2\dot{x}^2 + c$ 。



显然只有一组抛物线能够经过原点（即停在原点），即 $x = -1/2 |\dot{x}|\dot{x}$ ，这被称为**切换曲线(switching curve)**，即图中的曲线 Γ 。因此最优控制策略为从初始状态所在的抛物线出发，沿着该抛物线运动直到到达切换曲线，然后切换控制到相反的极值并沿着另一族抛物线运动直到到达原点。

考虑线性系统的控制问题，控制系统为

$$\dot{x} = Ax + Bu, \quad x(t_0) = x_0,$$

其中 $x \in \mathbb{R}^n, u \in \mathbb{R}^m$ ，控制约束为

$$U = \{u \in \mathbb{R}^m : u_i \in [-1, 1], i = 1, \dots, m\},$$

这里的上下界不影响之后的推导，此处归一化为 ± 1 仅为简化符号。成本泛函仍为总时间，即 $L \equiv 1, K \equiv 0$ ，目标集合为 $S = [t_0, \infty) \times \{x_1\}$ 。构造哈密顿量 $H = \langle p, Ax + Bu \rangle + p_0$ 。根据极小值条件，最优控制满足

$$\langle p^*(t), Bu^*(t) \rangle = \min_{u \in U} \langle p^*(t), Bu \rangle.$$

令 B 的第 i 列为 b_i ，则上式等价于对任意的 $i = 1, \dots, m$ 都有

$$\sum_{i=1}^n \langle p^*(t), b_i \rangle u_i^*(t) = \min_{u \in U} \sum_{i=1}^n \langle p^*(t), b_i \rangle u_i.$$

由于各个 u_i 的选取是相互独立的，故对任意的 $i = 1, \dots, m$ 都有

$$\langle p^*(t), b_i \rangle u_i^*(t) = \min_{u_i \in [-1, 1]} \langle p^*(t), b_i \rangle u_i.$$

从而可以得到

$$u_i^*(t) = \text{sgn}(\langle p^*(t), b_i \rangle) = \begin{cases} 1 & \text{if } \langle p^*(t), b_i \rangle < 0 \\ -1 & \text{if } \langle p^*(t), b_i \rangle > 0 \\ \text{arbitrary} & \text{if } \langle p^*(t), b_i \rangle = 0 \end{cases}$$

下面需要考虑 $\langle p^*(t), b_i \rangle$ 是否可能在某一区间上恒为零。根据规范方程，有 $\dot{p}^* = -A^T p^*$ ，因此 $p^*(t) = e^{A^T(t^*-t)} p^*(t^*)$ ，故

$$\langle p^*(t), b_i \rangle = \langle e^{A^T(t^*-t)} p^*(t^*), b_i \rangle = \langle p^*(t^*), e^{A(t^*-t)} b_i \rangle.$$

令 $\varphi(t) = \langle p^*(t), b_i \rangle$, 则 $\varphi(t)$ 是一个解析函数, 若 $\varphi(t)$ 在某一区间上恒为零, 则 $\varphi(t) \equiv 0$, 从而其在 t^* 处的任意阶导数 $\varphi^{(k)}(t^*) = \langle p^*(t^*), A^k b_i \rangle$ 也恒为零, 故

$$\langle p^*(t^*), b_i \rangle = \langle p^*(t^*), Ab_i \rangle = \dots = \langle p^*(t^*), A^{n-1}b_i \rangle = 0.$$

若 $b_i, Ab_i, \dots, A^{n-1}b_i$ 线性无关, 则上式表明 $p^*(t^*) = 0$, 再根据静态条件可知 $p_0^* = 0$, 这与极小值原理中的 $(p_0^*, p^*(t)) \neq 0$ 矛盾。因此若 $b_i, Ab_i, \dots, A^{n-1}b_i$ 线性无关, 则 $\langle p^*(t), b_i \rangle$ 不可能在某一区间上恒为零, 从而 $u_i^*(t)$ 在任意时间点要么取 1, 要么取 -1, 即线性系统的最优控制也是 Bang-Bang 控制。

CONTENT

目录

1. 最优控制问题

2. 极小值原理

3. 动态规划与 HJB 方程

考虑连续时间最优控制问题，为了简化讨论，首先考虑固定时间自由端点问题，即 $S = \{t_1\} \times \mathbb{R}^n$ 。写出成本泛函

$$J(t_0, x_0, u) = \int_{t_0}^{t_1} L(t, x(t), u(t)) dt + K(x(t_1)),$$

回忆动态规划思想，最重要的就是构造子问题然后应用最优性原理（递推式）。此时子问题显然就是 $[t_0, t_1]$ 的子区间，递推过程假定 $[t, t_1]$ 上的最优控制策略已知，故定义**值函数（value function）**

$$V(t, x) := \min_{u[t, t_1]} J(t, x, u).$$

其中 $u[t, t_1]$ 表示在区间 $[t, t_1]$ 上的可行控制。不难看出值函数 $V(t, x)$ 表示从时间 t 、状态 x 出发到终点的最小成本，因此有边值条件 $V(t_1, x) = K(x)$ 。此外，上式中的 \min 表示默认最小值存在。

基于离散动态规划的思想，可以写出连续时间的**最优性原理 (principle of optimality)**：

$$V(t, x) = \min_{u[t, t+\Delta t]} \left\{ \int_t^{t+\Delta t} L(s, x(s), u(s)) ds + V(t + \Delta t, x(t + \Delta t)) \right\}.$$

其中 $x(s) (s \in [t, t + \Delta t])$ 是由初始条件 $x(t) = x$ 和控制 $u(s)$ 决定的状态轨迹。不难看出，最优性原理给出了不同初始时间和状态下的值函数之间的关系。然而，这一关系在求解上并不直接，因此需要进一步推导出更为实用的形式，即通过无穷小分析得到一个偏微分方程。

首先下面的讨论均假设 $V \in C^1$ 。根据 x 与 u 的关系可知

$$x(t + \Delta t) = x(t) + f(t, x, u)\Delta t + o(\Delta t).$$

基于此可以展开 $V(t + \Delta t, x(t + \Delta t))$:

$$\begin{aligned} V(t + \Delta t, x(t + \Delta t)) &= V(t, x) + V_t(t, x)\Delta t \\ &\quad + \langle V_x(t, x), f(t, x, u)\Delta t \rangle + o(\Delta t). \end{aligned}$$

此外，还可以得到

$$\int_t^{t+\Delta t} L(s, x(s), u(s))ds = L(t, x, u)\Delta t + o(\Delta t).$$

将上述两式代入最优性原理，并整理得到

$$0 = \min_{u[t, t+\Delta t]} \{L(t, x, u)\Delta t + V_t(t, x)\Delta t + \langle V_x(t, x), f(t, x, u)\Delta t \rangle + o(\Delta t)\}.$$

两边除以 Δt 并令 $\Delta t \rightarrow 0^+$, 得到

$$0 = \min_{u \in U} \{L(t, x, u) + V_t(t, x) + \langle V_x(t, x), f(t, x, u) \rangle\}.$$

等价地, 可以写成

$$-V_t(t, x) = \min_{u \in U} \{L(t, x, u) + \langle V_x(t, x), f(t, x, u) \rangle\}.$$

这就是著名的**哈密顿-雅可比-贝尔曼 (Hamilton-Jacobi-Bellman, HJB) 方程**。结合边值条件 $V(t_1, x) = K(x)$, HJB 方程给出了值函数 $V(t, x)$ 的完整描述。

值得注意的是, **终端成本只在 HJB 方程的边值条件中出现, 而不在方程本身中体现**, 可以回顾证明过程进一步体会。此外, 当考虑更一般的目标集合 S 时, HJB 方程的形式保持不变, 但边值条件变为 $V(t, x) = K(t, x), \forall (t, x) \in S$ 。

停车问题

设 $x_1 \in \mathbb{R}$ 是汽车位置, x_2 是速度, $u \in U = [-1, 1]$ 是加速度。汽车动力学方程为 $\dot{x}_1 = x_2, \dot{x}_2 = u$, 目标是让汽车经过最短时间在原点停车。

不难写出 HJB 方程为

$$-V_t(t, x) = \min_{u \in [-1, 1]} \{V_{x_1}(t, x)x_2 + V_{x_2}(t, x)u + 1\}.$$

显然最优控制满足

$$u^*(t) = \operatorname{sgn}(V_{x_2}) = \begin{cases} 1 & \text{if } V_{x_2} < 0 \\ -1 & \text{if } V_{x_2} > 0 \\ \text{arbitrary} & \text{if } V_{x_2} = 0 \end{cases}$$

基于此简化 HJB 方程为 $-V_t(t, x) = V_{x_1}(t, x)x_2 - |V_{x_2}(t, x)| + 1$ 。

一般情况下, HJB 方程通常是非线性的偏微分方程, 求解较为困难。一些特殊情况可能可以简化为常微分方程, 此处不展开这些技术性的讨论。

此前的讨论都给出的是最优性的必要条件，然而事实上 HJB 方程是最优性的充分条件：假设 $\hat{V}(t, x)$ 满足 HJB 方程及边值条件， $\hat{u}(t)$ 是 HJB 取得最小值的控制策略，对应最优状态轨迹 $\hat{x}(t)$ （满足 $\hat{x}(t_0) = x_0$ ），则 $\hat{V}(t_0, x_0)$ 是最优成本，且 $\hat{u}(t)$ 是最优控制。

下面证明这一结论，首先根据 HJB 方程可知对任意的 $t \in [t_0, t_1]$ 都有

$$-\hat{V}_t(t, \hat{x}(t)) = L(t, \hat{x}(t), \hat{u}(t)) + \langle \hat{V}_x(t, \hat{x}(t)), f(t, \hat{x}(t), \hat{u}(t)) \rangle.$$

移项有

$$0 = L(t, \hat{x}(t), \hat{u}(t)) + \frac{d}{dt} \hat{V}(t, \hat{x}(t)).$$

等式两边对 t 积分，得到

$$0 = \int_{t_0}^{t_1} L(t, \hat{x}(t), \hat{u}(t)) dt + \hat{V}(t_1, \hat{x}(t_1)) - \hat{V}(t_0, \hat{x}(t_0)).$$

这等价于

$$\hat{V}(t_0, x_0) = \int_{t_0}^{t_1} L(t, \hat{x}(t), \hat{u}(t)) dt + K(\hat{x}(t_1)) = J(t_0, x_0, \hat{u}).$$

另一方面，若 u 是任意可行控制， x 是对应的初始状态与 \hat{x} 相同的状态轨迹，则根据 HJB 方程有

$$-\hat{V}_t(t, \hat{x}(t)) \leq L(t, x(t), u(t)) + \langle \hat{V}_x(t, x(t)), f(t, x(t), u(t)) \rangle.$$

类似的推导可以得到

$$\hat{V}(t_0, x_0) \leq \int_{t_0}^{t_1} L(t, x(t), u(t)) dt + K(x(t_1)) = J(t_0, x_0, u).$$

由此充分性得证。

HJB 方程可追溯至哈密顿的研究，后经雅可比在 1830 年代末改进。当时该方程作为最优性必要条件被广泛应用。其作为充分条件的应用由卡拉西奥多里在 1920 年代提出，其研究方法被称为变分法的“黄金法则”。

最优性原理可追溯至雅各布·伯努利 1697 年解决的最速降线问题。20 世纪 50 年代初，略早于贝尔曼时期，艾萨克斯 (Isaacs) 在微分博弈理论中将最优性原理形式化。动态规划这一术语由贝尔曼在 20 世纪 50 年代提出，他发表了多篇论文并出版了专著。贝尔曼的贡献在于认识到该方法能全局研究价值函数，并将其应用于求解变分法和最优控制领域的各类问题。

目前尚不清楚贝尔曼是否意识到其研究与变分法中的哈密顿-雅可比方程存在密切关联。这一联系在 20 世纪 60 年代初由卡尔曼明确指出，他是首位使用“HJB 方程”这一名称的学者。卡尔曼结合卡拉西奥多里和贝尔曼的思想，推导出最优控制的充分条件。

令人惊讶的是，当贝尔曼和卡尔曼在美国发表动态规划研究成果时，苏联也在同期独立发展极小值原理。因此探讨两者的关系显得十分自然。

为了得到 HJB 方程和极小值原理的联系，可以将 HJB 方程表达为另一种形式。首先，写下 HJB 方程

$$-V_t(t, x) = \min_{u \in U} \{ \langle V_x(t, x), f(t, x, u) \rangle + L(t, x, u) \}.$$

取 $p_0 = 1$ ，根据哈密顿量的定义 $H(t, x, u, p) = \langle p, f(t, x, u) \rangle + L(t, x, u)$ ，可以将上式写成

$$-V_t(t, x) = \min_{u \in U} H(t, x, u, V_x(t, x)).$$

即

$$H(t, x, u^*, V_x(t, x)) = \min_{u \in U} H(t, x, u, V_x(t, x)).$$

这与极小值原理中的极值条件一致，只要取 $p^*(t) = V_x(t, x^*(t))$ 即可。事实上也可以验证 $p^*(t)$ 满足规范方程，因此，HJB 方程蕴含了极小值原理。

事实真的如此吗？如果事实如此，则庞特里亚金所做的工作可以说是多此一举。然而事实并非如此：HJB 方程的推导依赖于值函数 $V(t, x) \in C^1$ ，而在实际问题中，值函数可能不具备光滑性，并且这类情况并非个例：当有终端成本和有界控制集合时，值函数往往不可微。这不仅使得 HJB 方程不能蕴含极小值原理，而且也使得 HJB 方程本身失去意义。为了解决这一问题，1980 年代，克鲁斯克尔（Crandall）和利奥尼（Lions）提出了粘性解（viscosity solution）的概念，从而推广了 HJB 方程的适用范围。

此外，还有一组要提及的概念是开环控制（open-loop control）和闭环控制（closed-loop control）。极小值原理事实上属于开环控制描述： $u^*(t) = \arg \max_{u \in U} H(t, x^*(t), u, p^*(t))$ ，而 $p^*(t)$ 由正则方程确定，故表达式中没有反馈机制体现；而 HJB 方程属于闭环控制描述： $u^*(t) = \arg \max_{u \in U} H(t, x^*(t), u, V_x(t, x^*(t)))$ ，因为 $V_x(t, x^*(t))$ 显然依赖于状态 $x^*(t)$ ，也间接依赖于控制 $u^*(t)$ 。

- HJB 方程是偏微分方程，求解较为困难，极小值原理通常更容易求解；
- HJB 方程给出了充分条件，而极小值原理仅给出必要条件；
- 因此综合二者求解最优控制问题的策略是：首先应用极小值原理得到候选最优控制策略，然后验证该策略对应的值函数是否满足 HJB 方程及边值条件，若满足则该策略为最优控制，否则需要重新寻找候选最优控制策略；
- 博弈论中的应用：微分博弈（differential game）是最优控制的经典应用，涉及多个决策者的动态系统控制问题。每个决策者都有自己的控制变量和目标函数，且各自的控制策略会影响系统的动态行为。