$$\begin{cases} \dot{x} = f(x(t), \alpha(t)) & t>0 \\ x(0) = x_0 \end{cases}$$

$\alpha(t): [0,+\infty) \to \mathbb{R}^m$

$$P[\alpha(\cdot)] = \int_0^T r(x(t),\alpha(t))dt + g(x(T))$$

PMP:

(ODE) $\dot{x}^*(t) = \nabla_p H(x^*(t), p^*(t), \alpha^*(t))$

(ADJ) $\dot{p}^*(t) = -\nabla_x H(x^*(t), p^*(t), \alpha^*(t))$

(PMP) $H(x^*(t), p^*(t), \alpha^*(t)) = \max_{\alpha \in A} H(x^*(t), p^*(t), \alpha)$

Chp 5 动态规划 (Dynamic programming)

核心思想：把一个控制问题嵌入到一族"从任意时刻、任意状态出发"的问题中，再用这一族的最优函数刻画最优控制。

Hamilton-Jacobi-Bellman 方程 (HJB 方程, Bellman PDE)

考虑 $\int_0^\infty \frac{\sin x}{x} dx$

定义参数化积分 $I(\alpha) = \int_0^\infty e^{-\alpha x} \frac{\sin x}{x} dx \qquad \alpha > 0$

对 α 求导 $I'(\alpha) = \frac{d}{d\alpha} \int_0^\infty e^{-\alpha x} \frac{\sin x}{x} dx$

$= \int_0^\infty \frac{d}{d\alpha} e^{-\alpha x} \frac{\sin x}{x} dx = \int_0^\infty (-x) e^{-\alpha x} \frac{\sin x}{x} dx$

$-J(\alpha) = -\int_0^\infty e^{-\alpha x} \sin x \, dx$

$\int u\,dv = uv - \int v\,du$

令 $u = e^{-\alpha x}$, $dv = \sin x\,dx$ 则 $du = -\alpha e^{-\alpha x} dx$, $v = -\cos x$

$$J(\alpha)=\int_0^\infty e^{-\alpha x}\sin x\,dx \qquad J(\alpha)=1-\alpha L(\alpha)$$

$$=\left[e^{-\alpha x}(-\cos x)\right]\Big|_0^\infty-\int_0^\infty(-\cos x)(-\alpha e^{-\alpha x})\,dx$$

$$=\left[-e^{-\alpha x}\cos x\right]\Big|_0^\infty-\alpha\int_0^\infty e^{-\alpha x}\cos x\,dx$$

$$=0-(-1)-\alpha\int_0^\infty e^{-\alpha x}\cos x\,dx$$

$$L(\alpha)=\int_0^\infty e^{-\alpha x}\cos x\,dx \qquad L(\alpha)=\alpha J(\alpha)$$

$$=\left[e^{-\alpha x}\sin x\right]\Big|_0^\infty+\alpha\int_0^\infty e^{-\alpha x}\sin x\,dx$$

$$=\alpha\int_0^\infty e^{-\alpha x}\sin x\,dx$$

联立可得 $J(\alpha)=1-\alpha(\alpha J(\alpha))=1-\alpha^2J(\alpha)$

$\therefore J(\alpha)=\frac{1}{1+\alpha^2}$

因此. $I'(\alpha)=-\frac{1}{1+\alpha^2}$

$$I(\alpha)=-\arctan\alpha+C \qquad I(\alpha)=\int_0^\infty e^{-\alpha x}\frac{\sin x}{x}\,dx \quad \alpha>0$$

当 $\alpha\to\infty$ 时，有 $I(\alpha)\to 0$，且 $\arctan\alpha\to\frac{\pi}{2}$

$$I(\infty)=-\frac{\pi}{2}+C=0 \Rightarrow C=\frac{\pi}{2}$$

因此 $\int_0^\infty \frac{\sin x}{x}\,dx=I(0)=\frac{\pi}{2}$

Remark: 通过引入参数 α，对 $\{I(\alpha)\}$ 做整体分析，更容易得到原来的值.

动态规划：不只看一个固定的初始状态和初始时间，而是考虑所有起点 (x,t).

考虑 $\begin{cases} \dot{x}(t) = f(x(t), \alpha(t)) & t \in [0, T] \\ x(0) = x_0 \end{cases}$

控制 $\alpha(\cdot) \in A \subset \mathbb{R}^m$　　A 是紧集

收益函数　$P[\alpha(\cdot)] = \int_0^T r(x(t), \alpha(t))\,dt + g(x(T))$

扩展到所有起始时刻 t 的子问题

$$\begin{cases} \dot{x}(s) = f(x(s), \alpha(s)) & s \in [t, T] \\ x(t) = x \in \mathbb{R}^n \end{cases}$$

收益函数　$P_{x,t}[\alpha(\cdot)] = \int_t^T r(x(s), \alpha(s))\,ds + g(x(T))$

Defn (价值函数, value function) 对每一个起点 $(x,t) \in \mathbb{R}^n \times [0,T]$,

价值函数为

$$v(x,t) = \sup_{\alpha(\cdot) \in A} P_{x,t}[\alpha(\cdot)] = \sup_{\alpha(\cdot) \in A} \left\{ \int_t^T r(x(s), \alpha(s))\,ds + g(x(T)) \right\}$$

当 $t = T$ 时,

$$v(x,T) = g(x).$$

Hamilton-Jacobi-Bellman 方程

假设: ① 控制集 A 为紧集

② 对每个 (x,t), $\sup_{\alpha(\cdot) \in A} P_{x,t}[\alpha(\cdot)]$ 可得 (存在最优控制)

③ $v \in C^1$ (v 一阶连续可微)

Thm (HJB 方程) 基于上述假设，价值函数 v 满足

$$v_t(x,t) + \max_{\alpha\in A}\left\{f(x,\alpha)\cdot\nabla_x v(x,t) + r(x,\alpha)\right\} = 0$$

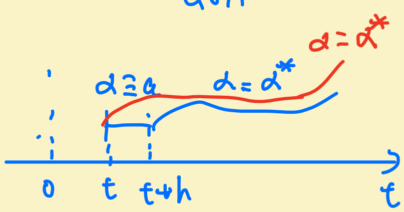
且 $\quad v(x,T) = g(x).$

Remark: 对应 Hamiltonian

$$H(x,p) := \max_{\alpha\in A}\left\{f(x,\alpha)\cdot p + r(x,\alpha)\right\}$$

$$\text{HJB} \Rightarrow \quad v_t(x,t) + H(x, \nabla_x v(x,t)) = 0$$

$$V(x,t) \geq \sup_{\alpha\in A}[P(\alpha)]$$



Proof: 给定任意 (x,t)，及常数 $h>0$，且 $h+t<T$. 在区间 $[t,t+h]$ 取一个常控制，即

$$\alpha(s) \equiv a \qquad s\in[t, t+h]$$

对应状态轨迹

$$\begin{cases} \dot{x}(s) = f(x(s), a) \\ x(t) = x \end{cases}$$

在 $s\in[t+h, T]$ 采用最优控制，最优收益为

$$v(x(t+h), t+h)$$

整体策略总收益为

$$\int_t^{t+h} r(x(s), a)\,ds + v(x(t+h), t+h)$$

有

$$v(x,t) \geq \int_t^{t+h} r(x(s), a)\,ds + v(x(t+h), t+h) \qquad ①$$

令 $h\to 0$，① 式可写为

$$\frac{1}{h}\int_t^{t+h} r(x(s), a)\,ds + \frac{v(x(t+h), t+h) - v(x,t)}{h} \leq 0$$

由于 $\quad \lim_{h\to 0} \dfrac{v(x(t+h), t+h) - v(x,t)}{h} = \nabla_x v(x,t)\cdot\dot{x}(t) + v_t(x,t)$

$$\lim_{h\to 0}\frac{1}{h}\int_t^{t+h} r(x(s), a)\,ds = r(x,a)$$

代回原式，可得 $\dot{x}(t)=f(x,a)$

$$r(x,a)+\nabla_x v(x,t)\cdot\dot{x}(t)+v_t(x,t)\le 0$$

$$\Rightarrow r(x,a)+\nabla_x v(x,t)\cdot f(x,a)+v_t(x,t)\le 0 \quad \forall a\in A$$

$$\Rightarrow \max_{a\in A}\{r(x,a)+\nabla_x v(x,t)\cdot f(x,a)+v_t(x,t)\}\le 0 \quad ②$$

令 α^* 是从 (x,t) 出发的最优控制，考虑 $s\in[t,t+h]$.

$$v(x,t)=\int_t^{t+h} r(x^*(s),\alpha^*(s))ds+v(x^*(t+h),t+h)$$

类似的，除以 h 并令 $h\to 0$，可得

$$v_t(x,t)+\nabla_x v(x,t)\cdot f(x,\alpha^*(t))+r(x^*,\alpha^*(t))=0$$

即存在 $\alpha^*=\alpha^*(t)\in A$ 使得.

$$v_t(x,t)+\nabla_x v(x,t)\cdot f(x,\alpha^*)+r(x^*,\alpha^*)=0$$

那么，可以找到一个最优值，使得.

$$\max_{a\in A}\{v_t(x,t)+\nabla_x v(x,t)\cdot f(x,a)+r(x,a)\}\ge 0 \quad ③$$

综合②、③，可得.

$$\max_{a\in A}\{v_t(x,t)+\nabla_x v(x,t)\cdot f(x,a)+r(x,a)\}=0$$

动态规划求最优控制的两步法

简单来说：先解HJB方程，再构造控制

Step 1：求解HJB方程，得到 v 的表达式：

$$\begin{cases} v_t(x,t)+\max_{a\in A}\{\nabla_x v(x,t)\cdot f(x,a)+r(x,a)\}=0 & 0\le t<T \\ v(x,T)=g(x) \end{cases}$$

Step 2: 由 V 构造最优反馈控制.

$$\alpha^* = \alpha(x,t) \in \arg\max_{a\in A} \{ f(x,a)\cdot \nabla_x V(x,t) + r(x,a) \}$$

然后考虑系统

$$\begin{cases} \dot{x}^*(s) = f(x^*(s), \alpha(x^*(s),s)) & s\in[t,T] \\ x^*(t) = x \end{cases}$$

$$\alpha^*(s) := \alpha(x^*(s),s)$$

Thm (最优性验证定理) 设 V 为 HJB 方程的解，则上述控制是从 (x,t) 出发的最优控制.

Proof: 以 (x,t) 为起点的子问题收益函数

$$P_{x,t}[\alpha^*(\cdot)] = \int_t^T r(x^*(s),\alpha^*(s))\,ds + g(x^*(T))$$

根据 HJB 方程

$$V_t(x^*(s),s) + \overbrace{f(x^*(s),\alpha^*(s))}^{\dot{x}^*(s)}\cdot \nabla_x V(x^*(s),s) + r(x^*(s),\alpha^*(s)) = 0$$

$$\Rightarrow V_t(x^*(s),s) + \nabla_x V(x^*(s),s)\cdot \dot{x}^*(s) = -r(x^*(s),\alpha^*(s)).$$

由于 $$\frac{d}{ds}V(x^*(s),s) = \nabla_x V(x^*(s),s)\cdot \dot{x}^*(s) + V_t(x^*(s),s)$$

$$\Rightarrow \frac{d}{ds}V(x^*(s),s) = -r(x^*(s),\alpha^*(s))$$

两边在 $[t,T]$ 上积分

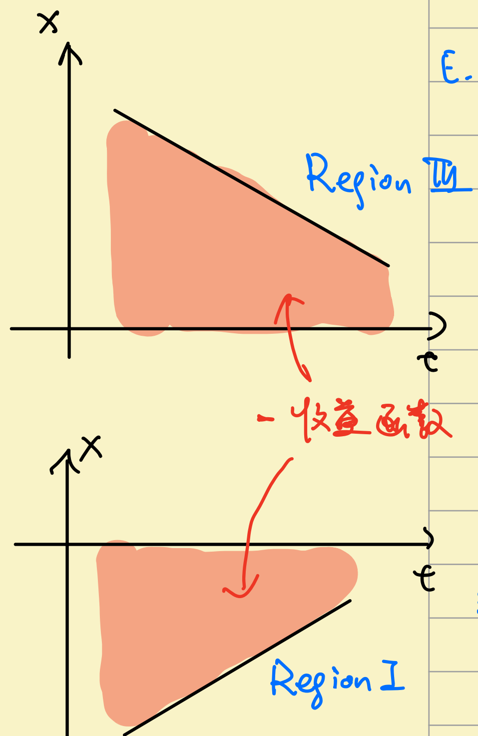
$$\begin{aligned} \int_t^T \frac{d}{ds}V(x^*(s),s) &= V(x^*(s),s)\Big|_t^T \\ &= V(x^*(T),T) - V(x^*(t),t) \\ &= -\int_t^T r(x^*(s),\alpha^*(s))\,ds \end{aligned}$$

根据 $V(x,T) = g(x)$，可得

$$g(x^*(T)) - v(x^*(t),t) = -\int_t^T r(x^*(s),\alpha^*(s))\,ds$$

化简可得

$$P_{x,t}[\alpha^*(\cdot)] = v(x,t)$$



E.g.: 三个速度的一维系统.

$$\begin{cases}\dot{x}(s)=\alpha(s) & 0\le t\le s\le 1\\ x(t)=x\end{cases}$$

控制集 $A=\{-1,0,1\}$

收益函数 $P_{x,t}(\alpha(\cdot)) = -\int_t^1 |x(s)|\,ds$.

价值函数 $v(x,t)=\sup_{\alpha\in A} P_{x,t}(\alpha(\cdot)) = -\inf_{\alpha\in A}\int_t^1 |x(s)|\,ds$

剩余时间为 $1-t$ ① 以速度 -1 运动，位置变化 $-(1-t)$

② 以速度 $+1$ 运动，位置变化 $(1-t)$

给出边界直线 $x=t-1$ 及 $x=1-t$，可将 $(x,t)\in\mathbb{R}\times[0,1]$ 分为三块

Region Ⅰ $=\{(x,t)\mid x<t-1,\ 0\le t\le 1\}$

Region Ⅱ $=\{(x,t)\mid |x|<1-t,\ 0\le t\le 1\}$

Region Ⅲ $=\{(x,t)\mid x>1-t,\ 0\le t\le 1\}$.

Region Ⅲ: 最优控制 $\alpha(s)\equiv -1 \quad s\in[t,1]$

$$x(s)=x-(s-t)>0$$

$$v(x,t) = -\int_t^1 x(s)\,ds = -\int_t^1 (x-(s-t))\,ds$$

$$= -\frac{1-t}{2}(2x+t-1)$$

Region Ⅰ: 最优控制 $\alpha(s)\equiv 1, \quad s\in[t,1]$

$$x(s)=x+(s-t)<0$$

$$v(x,t) = \int_t^1 x(s)\,ds = -\frac{1-t}{2}(-2x+t-1)$$

Region Ⅱ: 跑到原点的时间 $|x|$

$$\int_t^1 |x(s)|\,ds = \int_0^{|x|}(|x|-u)\,du = \frac{x^2}{2}$$

$$v(x,t) = -\frac{x^2}{2}$$

验证 HJB 方程.

HJB: $v_t(x,t) + \max_{a\in\{-1,0,1\}} \{a v_x(x,t) - |x|\} = 0$

$$\max_{a\in\{-1,0,1\}} a v_x \begin{cases} v_x & v_x > 0 \\ 0 & v_x = 0 \\ -v_x & v_x < 0 \end{cases} = |v_x|$$

因此

$$v_t(x,t) + \max_{a\in\{-1,0,1\}} \{a v_x(x,t) - |x|\} = 0$$

$$\Rightarrow v_t(x,t) + \max_{a\in\{-1,0,1\}} \{a v_x(x,t)\} - |x| = 0$$

$$\Rightarrow v_t(x,t) + |v_x(x,t)| - |x| = 0$$

当 Region I 时. $v(x,t) = -\frac{1-t}{2}(-2x+t-1)$

$$\left.\begin{aligned} v_t &= -(-x+t-1) \\ v_x &= 1-t \\ |v_x| &= 1-t \end{aligned}\right\}$$ 代 HJB 等式成立.

同理, Region II, 和 III 成立.

练习:
$$\begin{cases} \dot{x}(s) = \alpha(s) \\ x(t) = x \in \mathbb{R} \end{cases} \qquad 0 \le t \le s \le 2$$

控制集 $A = \{-1,0,1\}$

收益函数 $r(x,a) = -|x|$

① 仿照 5.2.1 求出 Region I, II, III.

② 求出三种情形下 $v(x,t)$ 表达式

③ 三种情形下, 验证 HJB 方程满足