Dynamic programming

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

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

价值函数: $v(x,t) = \sup_{\alpha(\cdot)\in A} P_{x,t}[\alpha(\cdot)]$.

HJB 方程: $\begin{cases} v_t(x,t) + \max_{\alpha\in A}\{f(x,\alpha)\cdot \nabla_x v(x,t) + r(x,\alpha)\} = 0 \\ \qquad\qquad \forall x \in \mathbb{R}^n,\ 0 \le t < T \\ v(x,T) = g(x), \end{cases}$

两步法: ① 求解 HJB 方程, 得到价值函数 $v(x,t)$

② 基 $v(x,t)$ 构造最优反馈控制 $\alpha^*(x,t)$

Chp 6 微分博弈问题 (Differential Games)

博弈论 (Game theory)

要素: 玩家 (player), 策略 (strategy), 收益 (Payoff).

关键: 自己的收益不仅取决于自己的行动, 还取决于其他人的行动.

研究: 多方决策, 相互影响的环境下:

- 各方如何决策
- 会出现什么结果

有限静态博弈 (normal-form game)

有限策略，一次同时行动，完全信息的静态博弈。

Defn: (有限静态博弈) 一个两人的有限静态博弈包括：

① 玩家的集合：$N=\{1,2\}$.

② 玩家 i 的策略集合 S_i，元素记为 $s_i \in S_i$

③ 收益函数 (payoff function)

$$u_1: S_1 \times S_2 \to \mathbb{R}$$

$$u_2: S_1 \times S_2 \to \mathbb{R}$$

两人有限策略的矩阵表示

		玩家B 策略1	玩家B 策略2
玩家A	策略1	(a_{11}, b_{11})	(a_{12}, b_{12})
	策略2	(a_{21}, b_{21})	(a_{22}, b_{22})

零和博弈 (zero-sum game) 存在函数 P，使得

$$u_1(s_1, s_2) = P(s_1, s_2) \qquad u_2(s_1, s_2) = -P(s_1, s_2)$$

一般博弈：u_1+u_2 不恒为常数

E.g.: **囚徒困境**. A 和 B 被分开审讯，每人有两种选择：

① 保持沉默

② 坦白并指证对方

		B 沉默	B 坦白
A	沉默	$(-1, -1)$	$(-10, 0)$
	坦白	$(0, -10)$	$(-5, -5)$

E.g. 协调博弈. 两个司机相向而行.

A \ B	往左	往右
往左	(1, 1)	(0, 0)
往右	(0, 0)	(1, 1)

Defn: (Best response, 最佳反应) 给定玩家B的策略 $s_2 \in S_2$, 我们称 $s_1^* \in S_1$ 是玩家A的对于 s_2 的最佳反应, 如果

$$u_1(s_1^*, s_2) \geq u_1(s_1, s_2) \quad \forall s_1 \in S_1$$

Defn: (Strict Dominant Strategy. 严格占优策略) 对于玩家A, 若存在策略 $s_1^* \in S_1$, 使得对任意 $s_1 \in S_1$, $s_2 \in S_2$, 都有

$$u_1(s_1^*, s_2) > u_2(s_1, s_2)$$

则称 s_1^* 为一个严格占优策略.

Nash 均衡 (Nash equilibrium)

Defn: (Nash equilibrium) 策略组合 $(s_1^*, s_2^*) \in S_1 \times S_2$ 称为一个纯策略 Nash 均衡, 如果对于每个玩家 $i=1,2$ 都有

$$u_i(s_i^*, s_{-i}^*) \geq u_i(s_i, s_{-i}^*) \quad \forall s_i \in S_i$$

$s_{-1}^* = s_2^*$, $s_{-2}^* = s_1^*$

Note: ① 任何玩家均无动力改变自己的策略.

② 即, Nash 均衡策略是对对手策略的最佳反应.

E.g.: 囚徒困境. (坦白, 坦白) 的策略是最佳反应, 所以是一个纯策略 Nash 均衡.

E.g.: 协调博弈. 对手选左, 我选左, 对手选右, 我选右. 所以 (左, 左), (右, 右) 都是纯策略 Nash 均衡.

Eg.: 正反面猜拳，出拳相同，玩家A胜；出拳不同，玩家B胜.

	正	反
正	(1, -1)	(-1, 1)
反	(-1, 1)	(1, -1)

设有任何纯策略立给对方知道的情况下是稳态.

P.S.: 仅存在混合策略稳态.

静态零和博弈.

设 $u_1(s_1, s_2) = P(s_1, s_2)$ $\quad u_2(s_1, s_2) = -P(s_1, s_2)$

- 玩家A：最大化 $P(s_1, s_2)$
- 玩家B：最小化 $P(s_1, s_2)$

上、下值函数

对玩家A：若先选策略 s_1，并认为对手会选择对自己最不利的策略 s_2，则他的收益为

$$\min_{s_2 \in S_2} P(s_1, s_2)$$ ← 玩家A考虑玩家B的选择.

进一步，玩家A会选 s_1 使上述最小值尽可能大

$$v := \max_{s_1 \in S_1} \min_{s_2 \in S_2} P(s_1, s_2)$$

称 v 为下值 (lower value)

对玩家B：若先选策略 s_2，并认为对手会选择对自己最不利的策略 s_1，

$$\max_{s_1 \in S_1} P(s_1, s_2)$$

玩家B会选 s_2 使上述最大值尽可能小，

$$u := \min_{s_2 \in S_2} \max_{s_1 \in S_1} P(s_1, s_2)$$

称 u 为上值 (upper value)

Thm: $v \le u$

Defn: (博弈有值, has a value) 如果 $v = u = V$, 则称该零和博弈有值, 并将 V 称为该博弈的值.

Thm (von Neumann 极大极小定理) 在允许混合策略的情形下, 对任意的有限两人零和博弈, 都有

$$\max_{\sigma_1} \min_{\sigma_2} E(P(\sigma_1, \sigma_2)) = \min_{\sigma_2} \max_{\sigma_1} E(P(\sigma_1, \sigma_2))$$

其中 σ_i 是玩家 i 的混合策略.

静态	→	动态
- 每个玩家在单一时刻做出一个选择		- 每个玩家有一个控制 $\alpha(\cdot), \beta(\cdot)$
- 收益函数		状态随时间演化为 $\dot{x} = f(x, \alpha, \beta)$
		- 收益函数 = 运行收益 + 终端收益

状态、控制、动力系统

时间: $0 \le t \le T$, 终端时间固定 T

控制: 玩家 A: $\alpha(\cdot) \in A$, 玩家 B: $\beta(\cdot) \in B$.

动力系统: $f: \mathbb{R}^n \times A \times B \to \mathbb{R}^n$

$$\begin{cases} \dot{x}(s) = f(x(s), \alpha(s), \beta(s)), & t \le s \le T \\ x(t) = x. \end{cases}$$

给定初始点 $x \in \mathbb{R}^n$.

收益函数: 给定一对控制 $(\alpha(\cdot), \beta(\cdot))$

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

两人零和微分博弈问题

玩家 A: 希望最大化 $P_{x,t}[\alpha(\cdot), \beta(\cdot)]$

玩家 B: 希望最小化 $P_{x,t}[\alpha(\cdot), \beta(\cdot)]$.

Ex.2. $\begin{cases} \dot{x}_1 = x_2(t) \\ \dot{x}_2 = 2\alpha \end{cases}$ $|\alpha| \le 1$ ⇒ 5.2.2

x_1：距离
x_2：速度
α：加速度

$(x_1, x_2) \to (0,0)$ 达到此状态所用时间为 τ.

收益函数：$P[\alpha(\cdot)] = -\tau = -\int_0^\tau 1\,dt$

运行收益：$r = -1$

价值函数：$V(x_1, x_2, t) = -\inf_{\alpha(\cdot)} \tau$

(a) HJB 方程：$V_t + \max_{\alpha}\{f \cdot V_x + r\} = 0$

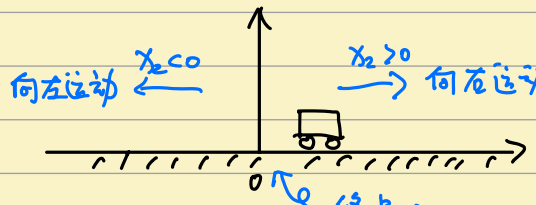
$\Rightarrow \max_{|\alpha|\le 1}\{f \cdot \nabla v + r\} = 0$ $\quad f(x_1, x_2, \alpha) = \begin{pmatrix} x_2 \\ 2\alpha \end{pmatrix}$

$\Rightarrow \max_{|\alpha|\le 1}\{x_2 \cdot V_{x_1} + 2\alpha \cdot V_{x_2} - 1\} = 0$

$\Rightarrow x_2 \cdot V_{x_1} + 2|V_{x_2}| - 1 = 0$ $\quad V(0,0) = 0$

(b) 给定 $\alpha(s) = -\mathrm{sgn}(x_2)$

Note：即施加最大负向加速度



Case 1：$x_2 > 0 \Rightarrow \alpha(s) = -1$

$\Rightarrow \begin{cases} \dot{x}_2 = 2\alpha = -2 \\ x_2(0) = x_2 > 0 \end{cases} \Rightarrow x_2(t) = x_2 - 2t$

令 $x_2(t) = 0 \Rightarrow 0 = x_2 - 2t \Rightarrow t_{stop} = \frac{x_2}{2}$

Case 2：$x_2 < 0 \Rightarrow \alpha(s) = 1$

$$\Rightarrow \begin{cases} \dot{x}_2 = 2\alpha = 2 \\ x_2(0) = x_2 < 0 \end{cases} \Rightarrow \underline{x_2(t) = x_2 + 2t}$$

令 $x_2(t) = 0 \Rightarrow 0 = x_2 + 2t \Rightarrow t_{stop} = -\frac{x_2}{2}$

综上 $t_{stop} = \frac{|x_2|}{2}$

接着算 $x_1(t_{stop})$ 已知 $\dot{x}_1 = x_2$, $x_1(0) = x_1$ $\Rightarrow x_1(t) = x_1 + \int_0^t x_2(s)\,ds$

Case 1: $x_2 > 0$, 则 $x_2(t) = x_2 - 2t$

$$x_1(t) = x_1 + \int_0^t (x_2 - 2s)\,ds$$

把 $t = t_{stop} = \frac{x_2}{2}$ 代入, 得

$$x_1(t_{stop}) = x_1 + \frac{x_2^2}{4}$$

Case 2: $x_2 < 0$, 同理, 得

$$x_1(t_{stop}) = x_1 - \frac{x_2^2}{4}$$

综上, 可得 $x_1(t_{stop}) = x_1 + \frac{1}{4}|x_2|x_2$

(c) 刚好在原点停下 $\therefore x_1(t_{stop}) = x_1 + \frac{1}{4}|x_2|x_2 = 0$