Dynamic Programming

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

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

value function: $v(x,t) = \sup_{\alpha(\cdot) \in A} P_{x,t}[\alpha(\cdot)]$

HJB 方程：

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

两步法求最优控制：① 解HJB方程，得到 $v(x,t)$
② 基于 $v(x,t)$ 构造 $\alpha^*(x,t)$

Game theory

① 零和博弈：存在函数 P，使得

玩家I：$u_1(s_1, s_2) = P(s_1, s_2)$

玩家II：$u_2(s_1, s_2) = -P(s_1, s_2)$

其中 $s_1 \in S_1$ 玩家I的策略，$s_2 \in S_2$ 玩家II的策略。

② Nash均衡：策略集合 $(s_1^*, s_2^*) \in S_1 \times S_2$ 称一个 Nash 均衡，如果对于每个玩家 $i=1,2$ 都有

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

③ 上下值函数　对于玩家 I　$v := \max_{s_1\in S_1} \min_{s_2\in S_2} P(s_1, s_2)$ → 下值函数

对于玩家 II　$u := \min_{s_2\in S_2} \max_{s_1\in S_1} P(s_1, s_2)$ → 上值函数

Thm: $v \le u$

Thm: (von Neuman minimax theorem)

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

微分博弈问题 (Differential Games)

时间: $0 \le t \le T$，终端时间 T

状态: $x(s) \in \mathbb{R}^n$

控制集合:　玩家 I: $A \subset \mathbb{R}^m$

　　　　　玩家 II: $B \subset \mathbb{R}^l$

状态变化函数: $f: \mathbb{R}^n \times A \times B \to \mathbb{R}^n$　　$f(x, \alpha, \beta) = \dot{x}$

Defn (控制) 给定起始时刻 t:

玩家 I 的控制: $\alpha(\cdot): [t, T] \to \mathbb{R}^m$

玩家 II 的控制: $\beta(\cdot): [t, T] \to \mathbb{R}^l$

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

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

玩家 I: 最大化 $P_{x,t}[\alpha, \beta]$

玩家 II: 最小化 $P_{x,t}[\alpha, \beta]$

典型双人零和微分博弈问题

控制集合与策略

Defn: (控制集合) 玩家I的所有控制组成的集合为

$$\mathcal{A}(t) := \{\alpha(\cdot) : [t,T] \to A \mid \alpha(\cdot) \text{ 可测}\}$$

玩家II的所有控制组成的集合为

$$\mathcal{B}(t) := \{\beta(\cdot) : [t,T] \to B \mid \beta(\cdot) \text{ 可测}\}.$$

Note: 非预见性 (non-anticipative) 策略：根据对手的控制轨迹，来给出自己的控制轨迹。但是，不能利用对手的未来信息。

Defn: (玩家I的策略) 映射 $\Phi : \mathcal{B}(t) \to \mathcal{A}(t)$ 称为玩家I的一个策略，如果对所有 $t \le s \le T$，只要

$$\beta(\tau) = \hat{\beta}(\tau), \quad t \le \tau \le s$$

对手的控制在 $[t,s]$ 完全一致

就有

① $$\Phi[\beta](\tau) \equiv \Phi[\hat{\beta}](\tau), \quad t \le \tau \le s.$$

玩家I的响应控制也完全相同

Defn: (玩家II的策略) 映射 $\Psi : \mathcal{A}(t) \to \mathcal{B}(t)$ 称为玩家II的一个策略，如果对所有 $t \le s \le T$，只要

$$\alpha(\tau) \equiv \hat{\alpha}(\tau) \quad t \le \tau \le s.$$

就有

② $$\Psi[\alpha](\tau) \equiv \Psi[\hat{\alpha}](\tau). \quad t \le \tau \le s.$$

Defn: (策略集合) 玩家I的所有非预见性策略组成的集合

$$\mathscr{A}(t) := \{\Phi : \mathcal{B}(t) \to \mathcal{A}(t) : \Phi \text{ 满足 } ①\}$$

玩家II的所有非预见性策略组成的集合

$$\mathscr{B}(t) := \{\Psi : \mathcal{A}(t) \to \mathcal{B}(t) : \Psi \text{ 满足 } ②\}.$$

术语:

控制：$\alpha(\cdot)$．$\beta(\cdot)$ 是"在每个时刻"选什么.(可视为一种记录)

控制的集合：$\mathcal{A}$, $\mathcal{B}$ 是在每个时刻可以选什么

策略：Φ, Ψ 是"面对对手的一整条控制轨迹，自己该如何应对"的规则

策略的集合：可以采取的策略有哪些

静态博弈

$$v = \max_{s_1} \min_{s_2} P(s_1, s_2)$$

上值函数和下值函数 (upper / lower value function)

Defn (下值函数) 下值函数的定义为 玩家II的策略

$$v(x,t) := \inf_{\Psi \in \mathcal{B}(t)} \sup_{\alpha(\cdot) \in \mathcal{A}} P_{x,t}[\alpha(\cdot), \Psi[\alpha](\cdot)]$$

理解:

1. 首先，玩家II先宣布一个策略Ψ
2. 玩家I观测到Ψ后，选择了自己的控制$\alpha(\cdot)$.
3. 执行时，玩家II的控制是基于玩家I的控制$\alpha(\cdot)$
4. 对于固定的策略，玩家I会最大化收益所选择控制 $\sup_\alpha$.
5. 然后玩家II会基于$\sup_\alpha$选择最有利的策略$\inf_\Psi$

Defn (上值函数) 上值函数的定义为

$$u(x,t) := \sup_{\Phi \in \mathcal{A}(t)} \inf_{\beta(\cdot) \in \mathcal{B}} P_{x,t}[\Phi[\beta](\cdot), \beta(\cdot)]$$

玩家I的策略.

理解:

1. 首先，玩家I先宣布一个策略Φ
2. 玩家II看到后选择$\beta(\cdot)$,
3. 执行时，玩家I的控制为$\Phi[\beta](\cdot)$.

4. 对于固定的 $\Phi\langle\beta\rangle(\cdot)$, 玩家 II 会选择 $\inf_\beta$

5. 然后玩家 I 会选择 $\sup_\Phi$

Thm: $v \le u$

Defn (the game has value, 博弈有值) 若 $v(x,t) = u(x,t)$ 对于所有的 (x,t) 都成立, 我们说这个博弈有值. 记

$$v(x,t) = u(x,t) = V(x,t)$$

动态规划与 Isaacs 方程

Thm (上值函数和下值函数的 PDE) 假设 u, v 在 (x,t) 上连续可微, 则有.

上值函数 u 满足

$$\begin{cases} u_t(x,t) + \min\limits_{b\in B}\max\limits_{a\in A}\{f(x,a,b)\cdot\nabla_x u(x,t) + r(x,a,b)\} = 0 \\ u(x,T) = g(x) \end{cases}$$

→ 玩家目标对立

下值函数 v 满足

$$\begin{cases} v_t(x,t) + \max\limits_{a\in A}\min\limits_{b\in B}\{f(x,a,b)\cdot\nabla_x v(x,t) + r(x,a,b)\} = 0 \\ v(x,T) = g(x) \end{cases}$$

Notation: PDE 意义下的 Hamiltonian 量:

$$H^+(x,p) = \min_{b\in B}\max_{a\in A}\{f(x,a,b)\cdot p + r(x,a,b)\}$$

$$H^-(x,p) = \max_{a\in A}\min_{b\in B}\{f(x,a,b)\cdot p + r(x,a,b)\}.$$

$H^+(x,p) = \min_{b\in B} \max_{a\in A}$

$H^-(x,p) = \max_{a\in A} \min_{b\in B}$

PDE简化

$$\begin{cases} u_t(x,t) + H^+(x,p) = 0 \\ u(x,T) = g(x) \end{cases}$$

$$\begin{cases} v_t(x,t) + H^-(x,p) = 0 \\ v(x,T) = g(x) \end{cases}$$

In general: $\max_a \min_b \{\cdots\} \le \min_b \max_a \{\cdots\}$,

因此: $H^-(x,p) \le H^+(x,p)$

Defn: (Isaacs条件 / minimax条件) 若对所有 (x,p) 有

$$H^+(x,p) = H^-(x,p)$$

则称该微分博弈问题满足Isaacs条件(或minimax条件).

博弈有值

$v = u = V(x,p)$

在该条件下,定义

$$H(x,p) := H^+(x,p) = H^-(x,p)$$

则上下值PDE统一

(PDE) $$\begin{cases} V_t(x,t) + H(x, D_x V(x,t)) = 0 \\ V(x,T) = g(x) \end{cases}$$

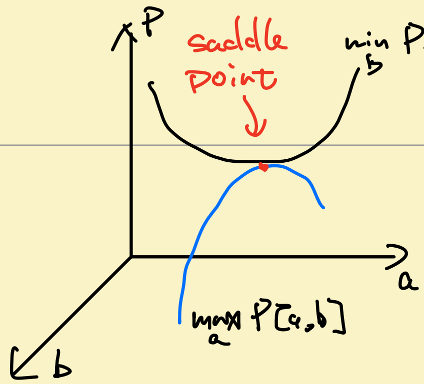
如何求解:

① 先解 (PDE), 得到 $V(x,p)$

② 构造最优控制: 在任意时间点及状态 (x,t)

- 玩家I选择控制 α^* 最大化 $f(x,a,b)\cdot p + r(x,a,b)$
- 玩家II选择控制 β^* 最小化 $f(x,a,b)\cdot p + r(x,a,b)$

对于任意 (x,t) 都能找到的 (α^*,β^*) 为最优的反馈策略.



Defn (鞍点, saddle point) 若一对控制 $(\alpha^*(\cdot), \beta^*(\cdot))$ 满足

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

$$\forall \alpha(\cdot) \in A(t),\ \beta(\cdot) \in B(t)$$

则称 (α^*, β^*) 为该博弈的一个鞍点.

Note：鞍点状态：微分博弈问题中的均衡局面.

E.x.

		Player II: L	Player II: R
Player I	L	(2, 2)	(0, 0)
	R	(0, 0)	(1, 1)

① 找出所有的 Nash 均衡
② 如果玩家间无法沟通，会出现什么情况？

① Player I: L → Player II: L
Player II: L → Player I: L　　(L, L)
Player I: R → Player II: R
Player II: R → Player I: R　　(R, R)

② 如果无法沟通，可能会出现社会总效用无法最大化的情况.