第十讲：微分博弈

最优控制的数学理论之六

张杰

人工智能学院
中国科学院大学

复杂系统管理与控制国家重点实验室
中国科学院自动化研究所

2017 年 10 月 19 日

Table of Contents

Table of Contents

最优控制问题

问题 1 (最优控制问题)

1. 被控对象的状态方程为
$$\dot{x}(t) = f(x(t), u(t), t), \quad x(t_0) = x_0.$$
2. 容许控制, $u \in \mathcal{U}, \quad x \in \mathcal{X}$.
3. 目标集, $x(t_f) \in \mathcal{S}$
$$\mathcal{S} = [t_0, \infty) \times \{x(t_f) \in \mathbb{R}^n : m(x(t_f), t_f) = 0\}$$
4. 求分段连续的 u，以最小化性能指标
$$J(u) = h(x(t_f), t_f) + \int_{t_0}^{t_f} g(x(t), u(t), t)\,\mathrm{d}t.$$

Pontryagin 极小值原理

定理 1 (庞特里亚金极小值原理)

上述问题得到最优控制 $u(t)$ 的必要条件为（TPBVP）

- *极值条件：对任意容许控制 $u'(t)$*

$$\mathcal{H}(x(t),u(t),p(t),t) \le \mathcal{H}(x(t),u'(t),p(t),t).$$

- *规范方程：*

$$\text{状态 (state) 方程:}\quad \dot{x}(t) = +\frac{\partial \mathcal{H}}{\partial p}(x(t),u(t),p(t),t),$$

$$\text{协态 (costate) 方程:}\quad \dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial x}(x(t),u(t),p(t),t).$$

- *边界条件 (用于处理目标集)：*

$$\left[\frac{\partial h}{\partial x}(x(t_f),t_f) - p(t_f)\right]\cdot \delta x_f + \left[\mathcal{H}(x(t_f),u(t_f),p(t_f),t_f) + \frac{\partial h}{\partial t}(x(t_f),t_f)\right]\delta t_f = 0.$$

Bellman 方程

定理 2 (Bellman 方程)

x_0 为初值 k_0 为初始时刻，最优控制下的性能指标记为“值函数”

$$V(x_0, k_0) = \min_{u \in U} J(u; x_0, k_0) \tag{1}$$

根据最优性原理，最优控制满足下列 Bellman 方程：

$$V(x(N), N) = h_D(x(N), N). \tag{2}$$

$$V(x(k), k) = \min_{u(k) \in U} \{g_D(x(k), u(k), k) + V(x(k+1), k+1)\}, \quad k = N-1, \ldots, 0. \tag{3}$$

Hamilton-Jacobi-Bellman 方程

定理 3 (Hamilton-Jacobi-Bellman 方程)

若最优控制问题有解，值函数是以 t_0 为初始时刻，x_0 为初始状态，在最优控制下的性能指标：

$$V(x_0, t_0) = \min_u J(u; x_0, t_0). \tag{4}$$

若值函数二阶连续可微，则如下 Hamilton-Jacobi-Bellman 方程（简称 HJB 方程）是最优控制的充分必要条件：

$$-\frac{\partial V}{\partial t}(x(t), t) = \min_{u(t)\in\mathbb{R}^m} \mathcal{H}\left(x(t), u(t), \frac{\partial V}{\partial x}(x(t), t), t\right), \tag{5}$$

$$V(x(t_f), t_f) = h(x(t_f), t_f)(\text{终端代价}) . \tag{6}$$

Table of Contents

例子：导弹攻击固定目标的最优控制

例 1 (导弹攻击固定目标的最优控制)

- 初始时刻导弹三维坐标 $\mathbf{x}_0$，速度为 $\mathbf{v}_0$，状态方程

$$\dot{\mathbf{x}}(t) = \mathbf{v}(t),\ \mathbf{x}(t_0) = \mathbf{x}_0. \tag{7}$$

$$\dot{\mathbf{v}}(t) = \mathbf{u}(t),\ \mathbf{v}(t_0) = \mathbf{v}_0. \tag{8}$$

- 终止条件：t_f 时刻导弹击中目标的坐标 $\mathbf{x}_f$，速度 $\mathbf{v}_f$ 自由
- 最小化性能指标，例如最小能量

$$J(\mathbf{u}) = \int_{t_0}^{t_f} \frac{1}{2}\|\mathbf{u}\|^2 dt \tag{9}$$

例子：导弹攻击移动目标的最优控制

例 2 (导弹攻击移动目标的最优控制)

- 导弹 (M) 状态方程和目标 (T) 状态方程分别为 ($\mathbf{u}_T$ 已知)

$$\dot{\mathbf{x}}_M(t) = \mathbf{v}_M(t),\ \dot{\mathbf{v}}_M(t) = \mathbf{u}_M(t). \tag{10}$$

$$\dot{\mathbf{x}}_T(t) = \mathbf{v}_T(t),\ \dot{\mathbf{v}}_T(t) = \mathbf{u}_T(t). \tag{11}$$

- 终止条件：t_f 时刻导弹击中目标，速度 $\mathbf{v}_f$ 自由

$$\mathbf{x}_M(t_f) = \mathbf{x}_T(t_f). \tag{12}$$

- 最小化性能指标，例如能量

$$J(\mathbf{u}_M) = \int_{t_0}^{t_f} \frac{1}{2}\|\mathbf{u}_M\|^2 dt \tag{13}$$

引入“相对位置”“相对速度”

令 $\mathbf{x} := \mathbf{x}_M - \mathbf{x}_T$, $\mathbf{v} := \mathbf{v}_M - \mathbf{v}_T$. 状态方程变为

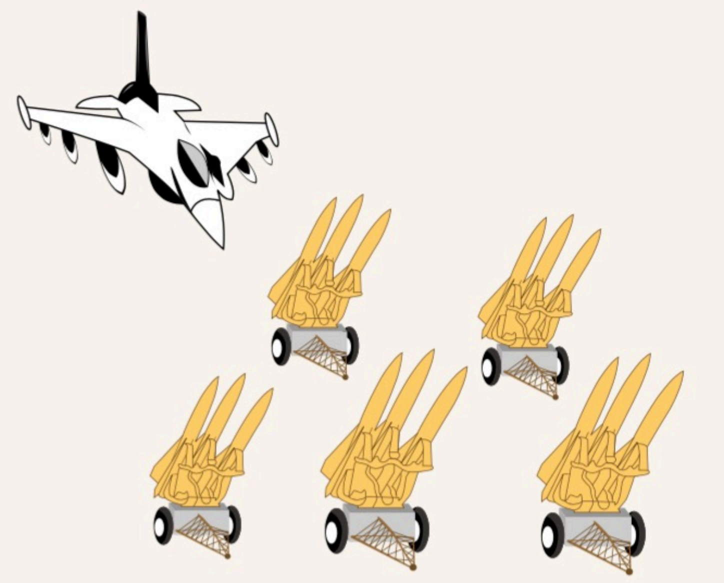
$$\dot{\mathbf{x}} = \mathbf{v}_M - \mathbf{v}_T = \mathbf{v}, \tag{14}$$

$$\dot{\mathbf{v}} = \mathbf{u}_M - \mathbf{u}_T. \tag{15}$$

终值条件 $\mathbf{x}(t_f) = 0$, $\mathbf{v}(t_f)$ free。性能指标不变

环境中有其他决策者，控制策略已知，则可化为最优控制问题

被攻击的目标也使用最优控制躲避?



微分博弈的发展

- 1928 年 (On the Theory of Games of Strategy)，1944 年 (Theory of games and economic behavior) 两篇著作中，John Von Neumann 和 Osker Morgenstern 创立博弈论
- 1951 年起，Rand 公司在美国空军资助下，Rufus Issacs 研究对抗双方都能自由决策行动的追逃问题，形成了微分博弈的最初研究成果
- 60-70 年代，微分博弈理论逐渐完善，得到微分博弈值函数存在性等基础结果；1965 年，Issacs 整理出版了第一部微分博弈同名专著。也称动态博弈
- Saridis 称之为“最坏情况设计”(1971 年)
- 2016 年，Google 公司的 AlphaGo 结合动态博弈和强化学习首次在围棋领域战胜人类世界冠军

从优化到博弈

定义 1 (函数极小值)

$\Omega \subset \mathbb{R}^N$ 是开集。称函数 $F \in C^1(\Omega)$ 在 x 达到局部极小值，若存在 $\epsilon > 0$ 使得:

$$F(x) \leq F(x'), \text{ if } \|x' - x\| < \epsilon, \forall x' \in \Omega.$$

定义 2 (纳什平衡 Nash Equilibrium, NE)

$F \in C^1(\Omega_1 \times \Omega_2)$，局中人 $i = 1, 2$ 的性能指标分别为 $F_i(x_1, x_2)$, $x_1 \in \Omega_1$,$x_2 \in \Omega_2$。x_1, x_2 是纳什平衡，若

$$F_1(x_1, x_2) \leq F_1(x_1', x_2), \forall x_1' \in \Omega_1, \tag{16}$$

$$F_2(x_1, x_2) \leq F_2(x_1, x_2'), \forall x_2' \in \Omega_2. \tag{17}$$

反应函数法求解博弈平衡

定义 3 (反应函数)

对于任意给定的 $x_2 \in \Omega$，映射 $R_1(x_2) = \text{argmin}_{x_1 \in \Omega} F_1(x_1, x_2)$ 称为局中人-1 的反应函数 (reaction function, or best response)

Remark 1 (反应函数法求解纳什平衡)

若 $x_1 = R_1(x_2), x_2 = R_2(x_1)$，可知 x_1, x_2 为纳什平衡。可通过联立博弈双方的反应函数求解博弈的纳什平衡

古诺博弈: 反应函数法求解纳什平衡

例 3 (古诺寡头竞争模型, Cournot Model)

两家公司 $i=1,2$ 生产同类产品，生产数量为 $q_i \geq 0$，生产成本为 $c(q_i)=cq_i$，市场上产品单价 $p(q)=a-q$ 与市场上的产品总量 $q=q_1+q_2$ 有关。

两家公司都希望最大化各自的净利润

$$V_1(q_1,q_2)=p(q_1+q_2)q_1-c(q_1), \tag{18}$$

$$V_2(q_1,q_2)=p(q_1+q_2)q_2-c(q_2). \tag{19}$$

1/2 求 Best-response

固定公司 2 产量 q_2，公司 1 产量 q_1 应满足一阶条件

$$0 = \frac{\partial V_1}{\partial q_1} = \dot{p}(q_1 + q_2)q_1 + p(q_1 + q_2) - c,$$

$$R_1(q_2) = \frac{a - q_2 - c}{2}.$$

类似的，固定公司 1 产量，可得公司 2 的反应函数

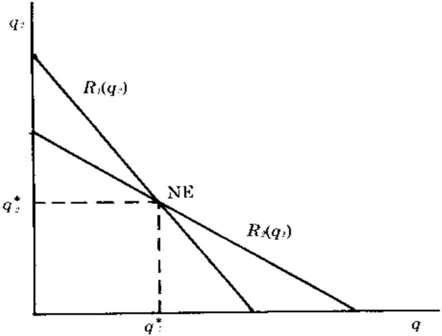
$$R_2(q_1) = \frac{a - q_1 - c}{2}.$$

$$q_1 = R_1(q_2) = \frac{a - q_2 - c}{2}, \quad q_2 = R_2(q_1) = \frac{a - q_1 - c}{2}.$$

2/2 根据 Best-response 求得 NE

联立两个公司的反应函数得到古诺模型的纳什平衡

$$q_1 = \frac{a-c}{3}, \; q_2 = \frac{a-c}{3}. \tag{20}$$



斯坦伯格模型

问题 2 (斯坦伯格模型, Stackelberg Model)

“领导者”率先实施并公布策略,“跟随者”随后实施策略，则称为斯坦伯格模型。同样考虑古诺模型中双寡头竞争的例子，但局中人 1 先发，局中人 2 后发

Remark 2 (反应函数法求解斯坦伯格平衡)

跟随者采用策略 $x_2 = R_2(x_1)$ 时，领导者性能指标（或效用函数）中已经不再包含其他人的策略，只需求解以自己策略为自变量的最优化问题即可

计算斯坦伯格平衡

固定领导者的产量 q_1，跟随者的反应函数为

$$R_2(q_1) = \frac{a - q_1 - c}{2}.$$

尽管领导者无法观测对手具体实施的策略，然而他可以据此得到对手的反应函数，于是

$$V_1(q_1, R_2(q_1)) = p(q_1 + R_2(q_1))q_1 - c(q_1) = p(\frac{a + q_1 - c}{2})q_1 - cq_1,$$

$$0 = \frac{\partial V_1}{\partial q_1} = \dot{p}(\frac{a + q_1 - c}{2})\frac{q_1}{2} + p(\frac{a + q_1 - c}{2}) - c$$

其最优策略应满足一阶条件：

$$q_1 = \frac{a - c}{2}, \ q_2 = R_2(q_1) = \frac{a - c}{4}. \tag{21}$$

Table of Contents

最优控制问题

问题 (典型最优控制问题)

1. 被控对象的状态方程为

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

2. 容许控制, $u \in U$
3. 目标集, $x(t_f) \in S$
4. 最小化性能指标

$$J(u) = h(x(t_f), t_f) + \int_{t_0}^{t_f} g(x(t), u(t), t)dt.$$

从最优控制到微分博弈

问题 3 (微分博弈问题)

1. 博弈双方的状态方程为

$$\dot{x}(t) = f(x(t), u_1(t), u_2(t), t), \ x(t_0) = x_0. \tag{22}$$

2. 容许控制 $u_1 \in U_1$, $u_2 \in U_2$
3. 二者均最小化各自的性能指标

$$J_1(u_1, u_2) = h_1(x(t_f), t_f) + \int_{t_0}^{t_f} g_1(x(t), u_1(t), u_2(t), t)dt \tag{23}$$

$$J_2(u_1, u_2) = h_2(x(t_f), t_f) + \int_{t_0}^{t_f} g_2(x(t), u_1(t), u_2(t), t)dt. \tag{24}$$

一些假定

对于任意容许的 t, x，$\forall q_1, q_2 \in \mathbb{R}^n$，方程组有唯一解

$$\xi = \operatorname*{argmin}_{u_1 \in U_1}\{q_1 \cdot f(x, u_1, \eta, t) + g_1(x, u_1, \eta, t)\} \tag{25}$$

$$\eta = \operatorname*{argmin}_{u_2 \in U_2}\{q_2 \cdot f(x, \xi, u_2, t) + g_2(x, \xi, u_2, t)\} \tag{26}$$

假定微分博弈的参与人具有如下共同知识：

- 共同的初始状态 x_0
- 共同的状态方程 f，以及 x, u_1, u_2 的容许集合
- J_1, J_2 的形式，即 g_1, g_2, h_1, h_2

微分博弈平衡的形式

定义 4 (开环形式的微分博弈)

除共同知识外，局中人均无法观测系统状态或他人策略，则形成开环形式的微分博弈

定义 5 (马尔可夫形式的微分博弈)

局中人可观测系统状态，形成马尔可夫形式微分博弈

定义 6 (斯坦伯格形式的微分博弈)

领导者不观测系统状态，跟随者可观测系统状态，则形成斯坦伯格形式的微分博弈

两人零和微分博弈的开环形式平衡

定理 4 (两人零和微分博弈的开环形式平衡)

1. 博弈双方的状态方程为
$$\dot{x}(t) = f(x(t), u_1(t), u_2(t), t), \; x(t_0) = x_0. \tag{27}$$
2. 容许控制 $u_1(t) \in U_1$, $u_2(t) \in U_2$
3. 局中人 1 最小化性能指标，局中人 2 最大化性能指标
$$J(u_1, u_2) = h(x(t_f), t_f) + \int_{t_0}^{t_f} g(x(t), u_1(t), u_2(t), t)dt \tag{28}$$

定义 Hamiltonian

$$\begin{aligned} \mathcal{H}(x(t), u_1(t), u_2(t), p(t), t) := g(x(t), u_1(t), u_2(t), t) \\ + p^T(t) f(x(t), u_1(t), u_2(t), t), \end{aligned} \tag{29}$$

两人零和微分博弈的开环形式平衡

定理 4 (两人零和微分博弈的开环形式平衡)

该微分博弈的开环形式平衡 $u_1(t) \in U_1, u_2(t) \in U_2$ *满足极值条件:*

$$\begin{aligned}\mathcal{H}(x(t), u_1(t), u_2(t), p(t), t) &= \min_{u_1} \max_{u_2} \mathcal{H}(x(t), u_1(t), u_2(t), p(t), t) \\ &= \max_{u_2} \min_{u_1} \mathcal{H}(x(t), u_1(t), u_2(t), p(t), t)\end{aligned}$$

状态 (state) 方程: $\dot{x}(t) = +\frac{\partial \mathcal{H}}{\partial p}(x(t), u_1(t), u_2(t), p(t), t),$

协态 (costate) 方程: $\dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial x}(x(t), u_1(t), u_2(t), p(t), t).$

边界条件: $[\frac{\partial h}{\partial x}(x(t_f), t_f) - p(t_f)] \cdot \delta x_f$

$$+ [\mathcal{H}(x(t_f), u_1(t_f), u_2(t_f), p(t_f), t_f) + \frac{\partial h}{\partial t}(x(t_f), t_f)]\delta t_f = 0.$$

Table of Contents

追逃微分博弈

例 (追逃微分博弈)

- 导弹 (1) 状态方程和目标 (2) 状态方程分别为

$$\dot{x}_1(t) = v_1(t),\ \dot{v}_1(t) = u_1(t),\ x_1(0) = -2,\ v_2(0) = 1. \tag{30}$$

$$\dot{x}_2(t) = v_2(t),\ \dot{v}_2(t) = u_2(t),\ x_2(0) = 0,\ v_2(0) = 2. \tag{31}$$

- 导弹要在终止时刻 $t_f = 2$ 命中目标，$1/E_1$ 表示能量的权重

$$J_1(u_1, u_2) = \frac{b}{2}|x_1(t_f) - x_2(t_f)|^2 + \int_{t_0}^{t_f} \frac{1}{2E_1} u_1(t)^2 dt. \tag{32}$$

- 目标要在终止时刻 $t_f = 2$ 远离导弹，$E_2 < E_1$

$$J_2(u_1, u_2) = -\frac{b}{2}|x_1(t_f) - x_2(t_f)|^2 + \int_{t_0}^{t_f} \frac{1}{2E_2} u_2(t)^2 dt. \tag{33}$$

零和追逃博弈的平衡 1/9

解：（转化为零和博弈形式）

引入状态将原问题转化为两人零和追逃博弈

$$x = x_1 - x_2,\ v = v_1 - v_2.$$
$$\dot{x}(t) = \dot{x}_1(t) - \dot{x}_2(t) = v_1 - v_2 = v.\ x(0) = -2 - 0 = -2.$$
$$\dot{v}(t) = \dot{v}_1(t) - \dot{v}_2(t) = u_1 - u_2.\ v(0) = 1 - 2 = -1.$$

以及统一的性能指标

$$J(u_1, u_2) = \frac{b}{2}x(t_f)^2 + \int_{t_0}^{t_f} \{\frac{1}{2E_1}u_1^2(t) - \frac{1}{2E_2}u_2^2(t)\}dt \tag{34}$$

追逐者希望最小化 $J(u_1, u_2)$，逃跑者则希望将其最大化

零和追逃博弈的平衡, 2/9

解：（计算 Hamiltonian, 考察极值条件）

Hamiltonian 与两者的控制均有关

$$\begin{aligned}&\mathcal{H}(x(t), v(t), u_1(t), u_2(t), p_1(t), p_2(t), t)\\ =&\frac{1}{2E_1}u_1(t)^2 - \frac{1}{2E_2}u_2(t)^2 + p_1(t)v(t) + p_2(t)(u_1(t) - u_2(t)).\end{aligned}$$

极值条件为

$$\frac{\partial \mathcal{H}}{\partial u_1} = 0 \Rightarrow u_1(t) = -p_2(t)E_1, \tag{35}$$

$$\frac{\partial \mathcal{H}}{\partial u_2} = 0 \Rightarrow u_2(t) = -p_2(t)E_2. \tag{36}$$

零和追逃博弈的平衡, 3/9

解：（将极值条件代入规范方程）

$$\dot{x}(t) = v(t) \tag{37}$$

$$\dot{v}(t) = -p_2(t)(E_1 - E_2) \tag{38}$$

$$\dot{p}_1(t) = -\frac{\partial \mathcal{H}}{\partial x} = 0, \Rightarrow p_1(t) = c_1 \tag{39}$$

$$\dot{p}_2(t) = -\frac{\partial \mathcal{H}}{\partial v} = -p_1(t), \Rightarrow p_2(t) = -c_1 t + c_2 \tag{40}$$

零和追逃博弈的平衡, 4/9

解：（处理边界条件）

t_f*fixed*, x_f,*free*。边界必要条件

$$0=\frac{\partial h}{\partial x}(x(t_f),v(t_f),t_f)-p_1(t_f)=bx(t_f)-p_1(t_f), \tag{41}$$

$$0=\frac{\partial h}{\partial v}(x(t_f),v(t_f),t_f)-p_2(t_f)=-p_2(t_f). \tag{42}$$

以及初值 $x(t_0)=x_0, v(t_0)=v_0$.

$$\begin{aligned} p_2(t) &= -c_1 t + c_2,\ p_2(t_f)=0, \Rightarrow c_2 = c_1 t_f \\ p_2(t) &= -c_1(t-t_f) \end{aligned} \tag{43}$$

下面把 p_2 代入状态方程

零和追逃博弈的平衡, 5/9

解：(把协态变量代回状态方程)

把协态变量 $p_2(t) = -c_1(t - t_f)$ 代入状态方程得到

$$\dot{v}(t) = -p_2(t)(E_1 - E_2) = (E_1 - E_2)c_1(t - t_f)$$

$$v(t) = (E_1 - E_2)c_1 \frac{(t - t_f)^2}{2} - (E_1 - E_2)c_1 \frac{(t_0 - t_f)^2}{2} + v_0 \tag{44}$$

再把 v 代入 $\dot{x} = v$

$$\begin{aligned} x(t) =& (E_1 - E_2)c_1 \frac{(t - t_f)^3}{6} - (E_1 - E_2)c_1 \frac{(t_0 - t_f)^3}{6} \\ &+ [-(E_1 - E_2)c_1 \frac{(t_0 - t_f)^2}{2} + v_0](t - t_0) + x_0 \end{aligned} \tag{45}$$

零和追逃博弈的平衡 6/9

解：（代入边界条件）

代入边界条件 $bx(t_f) - p_1(t_f) = 0$ 得到

$$c_1 = \frac{x_0 + v_0(t_f - t_0)}{1/b + (E_1 - E_2)(t_f - t_0)^3/3}, x(t_f) = c_1/b$$

得到微分博弈平衡

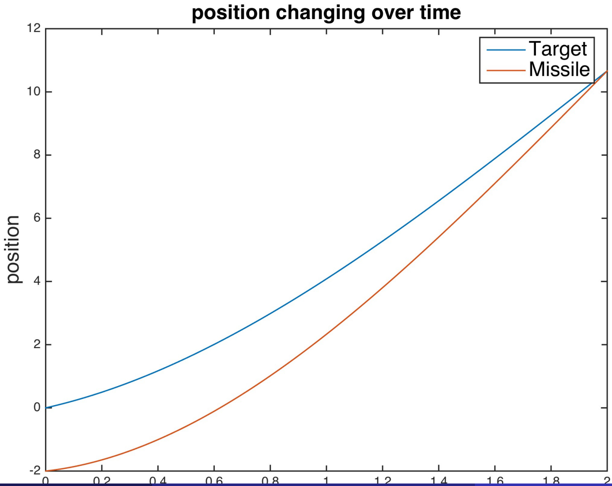
$$u_1(t) = -\frac{E_1(t_f - t)(x_0 + v_0(t_f - t_0))}{1/b + (E_1 - E_2)(t_f - t_0)^3/3}, \tag{46}$$

$$u_2(t) = \frac{E_2}{E_1} u_1(t). \tag{47}$$

$E_1 > E_2$ 时，令 $b \to \infty$，$x(t_f) \to 0$，命中

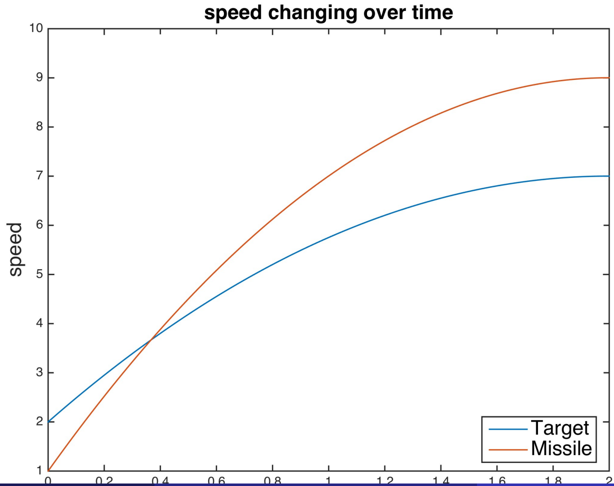
零和追逃博弈的平衡, 7/9

Figure: 零和追逃博弈的平衡：位置-时间



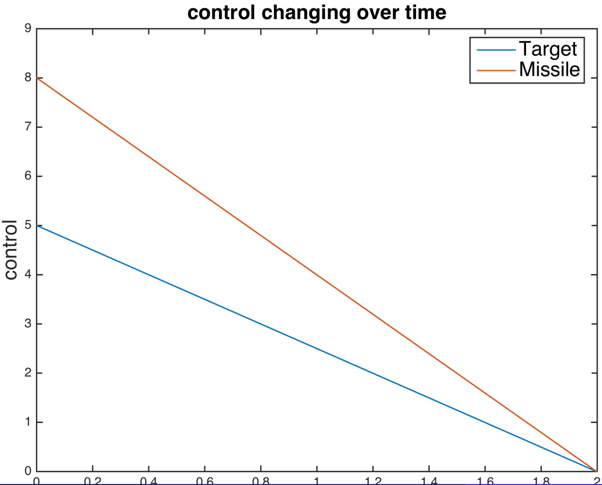
零和追逃博弈的平衡, 8/9

Figure: 零和追逃博弈的平衡：速度-时间



零和追逃博弈的平衡, 9/9

Figure: 零和追逃博弈的平衡：控制-时间 $J_1 = 26.706, J_2 = 26.651$



微分博弈平衡求解的过程和缺陷

Remark 3 (博弈求解的过程)

- 固定对方策略，求解己方最优策略
- 固定己方策略，求解对方最优策略
- 联立方程组求解博弈平衡

- 微分博弈旨给出的是一种“按照最坏情况打算”的控制律，仅在“想象”对方符合理性人假定属实的情况下达到最优
- 下节开始：对未知信息的其他“想象”