最优控制的数学理论与智能方法

最优控制方法总结

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 11 月 9 日

Table of Contents

Table of Contents

最优控制问题

问题 (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

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

离散时间最优控制问题

问题 1 (离散时间最优控制问题)

状态变量为 $x(k): \mathbb{N} \to \mathbb{R}^n$，控制变量为 $u(k): \mathbb{N} \to \mathbb{R}^m$

(1) 被控对象的状态方程

$$x(k+1) = f_D(x(k), u(k), k), \quad k = 0, \ldots, N-1. \tag{1}$$

(2) 容许控制:

$$u(k) \in U, \quad x(k) \in X. \tag{2}$$

(3) 目标集:

$$x(N) \in \mathcal{S}. \tag{3}$$

(4) 性能指标:

$$J(u; x(k), k) = h_D(x(N), N) + \sum_{i=k}^{N-1} g_D(x(i), u(i), i). \tag{4}$$

状态方程的处理

Remark 1 (连续状态方程 v.s. 离散状态方程)

- 连续状态方程可通过离散化转化为离散状态方程

Remark 2 (已知精确 v.s. 未知不精确)

- 已知精确状态方程的最优控制问题可通过数学方法求解
- 未知或不精确状态方程的系统可通过智能方法近似求解

约束条件

定义 1 (终端时刻和终端状态)

- t_f 时刻系统终止,“固定终端时刻”; 否则称“自由终端时刻”
- 状态变量在 t_f 的取值必须为 $x(t_f) = x_f$ 称为“固定终端状态, fixed”; 对 $x(t_f)$ 无约束称为“自由终端状态, free”

定义 2 (状态变量和控制变量的约束条件)

- “等式约束, equality constraints”
- “不等式约束, inequality constraints”

等式约束
转化成等

Remark 3

处理状态、控制约束是最优控制及其扩展方法(预测控制)的重要优势

性能指标

定义 3 (最优控制的性能指标)

最优控制最小化或最大化某一性能指标，

$$J(u) = h(x(t_f), t_f) + \int_{t_0}^{t_f} g(x(t), u(t), t)dt, \tag{5}$$

在讨论多人问题时，每人可有不同的性能指标

$$J_1(u_1, u_2) = h_1(x(t_f), t_f) + \int_{t_0}^{t_f} g_1(x(t), u_1(t), u_2(t), t)dt \tag{6}$$

$$J_2(u_1, u_2) = h_2(x(t_f), t_f) + \int_{t_0}^{t_f} g_2(x(t), u_1(t), u_2(t), t)dt. \tag{7}$$

二次型性能指标

$$J = \frac{1}{2}x^T(t_f)Hx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}[x^T(t)Q(t)x(t) + u^T(t)R(t)u(t)]dt.$$

- $\frac{1}{2}x^T(t_f)Hx(t_f)$
  H 半正定，取值越大则终止状态越接近原点
- $x^T(t)Q(t)x(t)$
  $Q(t)$ 半正定，取值越大则状态尽早接近原点
- $u^T(t)R(t)u(t)$
  $R(t)$ 正定，取值越大则能量损耗越小

能量最优，时间最优

例 1 (能量最优)

$$J = \int_{t_0}^{t_f} u^2(t)dt. \tag{8}$$

例 2 (时间最优)

$$J = \int_{t_0}^{t_f} dt = t_f - t_0. \tag{9}$$

控制策略的形式

定义 4 (闭环控制)

若控制变量形如:

$$u(t) = \phi(x(t), t), \tag{10}$$

则称其为闭环控制

定义 5 (开环控制)

若控制变量形如:

$$u(t) = \phi(x(t_0), t), \tag{11}$$

则称其为开环控制

Table of Contents

求泛函变分

定义 6 (泛函的变分)

若泛函增量可写为变分的线性泛函和变分的高阶无穷小两个部分:

$$\Delta J(x, \delta x) = \delta J(x, \delta x) + g(x, \delta x) \cdot \|\delta x\|, \tag{12}$$

- 前项 $\delta J(x, \delta x)$ 是 δx 的线性泛函
- 后项是 δx 的高阶无穷小

则称 δJ 是 J 对于 x 的变分, 称 J 对 x 可微

引理 1 ($J(x)$ 可微时通过求导计算变分)

泛函 $J(x(t))$ 的变分满足

$$\delta J(x, \delta x) = \frac{d}{d\alpha} J(x + \alpha \delta x)\Big|_{\alpha=0}. \tag{13}$$

定理：泛函极值一阶条件

定理 1 (泛函极值一阶条件)

$x \in M$， M 是一类函数的开集合，泛函 J 对 x 可微。若 x 使 J 取极值，则对任意容许的 δx 有

$$\delta J(x, \delta x) = 0. \tag{14}$$

[其中容许的 δx 指，若 $x \in \Omega$ 则 $x + \delta x \in \Omega$]

Remark 4

泛函变分可对比函数导数，上述泛函极值的一阶条件则可对比函数极值的一阶条件：$\forall \Delta x$，需满足 $\dot{F}(x)\Delta x = 0$

欧拉-拉格朗日方程

例 3 (最简变分问题)

求 $x(t):[t_0,t_f]\to\mathbb{R}^n$，满足 $x(t_0)=x_0$, $x(t_f)=x_f$，最小化：

$$J(x)=\int_{t_0}^{t_f} g(x(t),\dot{x}(t),t)\,\mathrm{d}t, \tag{15}$$

其中函数 g 取值于 $\mathbb{R}$，二阶连续可微，J 是从 $[t_0,t_f]\to\mathbb{R}^n$ 的连续可微函数全体到 $\mathbb{R}$ 的映射，是一个泛函。

$$\frac{\partial g}{\partial x}(x(t),\dot{x}(t),t)-\frac{d}{dt}[\frac{\partial g}{\partial \dot{x}}(x(t),\dot{x}(t),t)]=0$$

欧拉-拉格朗日方程的特殊情况求解

$$\frac{\partial g}{\partial x}(x(t),\dot{x}(t),t)-\frac{\mathrm{d}}{\mathrm{d}t}\Big[\frac{\partial g}{\partial \dot{x}}(x(t),\dot{x}(t),t)\Big]=0.$$

$$\frac{\partial g}{\partial x}-\frac{\partial^2 g}{\partial \dot{x}\partial x}\dot{x}-\frac{\partial^2 g}{\partial \dot{x}\partial \dot{x}}\ddot{x}-\frac{\partial^2 g}{\partial \dot{x}\partial t}=0$$

二阶方程化简为一阶?

- Case 1. No $\dot{x}$, i.e. $g=g(x,t)$.
- Case 2. No x, i.e. $g=g(\dot{x},t)$.
- Case 3. No t, i.e. $g=g(x,\dot{x})$.

欧拉-拉格朗日方程与哈密尔顿函数

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

以状态变量的“方向”为“控制变量” $u(t)$

$$\begin{cases} \dfrac{\partial g}{\partial x}(x(t), u(t), t) - \dfrac{\mathrm{d}}{\mathrm{d}t}\left[\dfrac{\partial g}{\partial u}(x(t), u(t), t)\right] = 0, \\ u(t) = \dot{x}(t). \end{cases} \tag{16}$$

引入哈密尔顿函数

定义哈密尔顿函数:

$$\mathcal{H}(x,u,p,t) := g(x,u,t) + p \cdot f(x,u,t). \tag{17}$$

通过对哈密尔顿函数简单计算，立即可得:

$$\frac{\partial \mathcal{H}}{\partial p} = u$$
$$\frac{\partial \mathcal{H}}{\partial x} = \frac{\partial g}{\partial x}$$
$$\frac{\partial \mathcal{H}}{\partial u} = p + \frac{\partial g}{\partial u}.$$

取 $x(t), u(t)$ 为上述一阶常微分方程组-(16) 的解，再定义该最优控制问题的协态（costate）为:

$$p(t) := -\frac{\partial g}{\partial u}(x(t), u(t), t).$$

计算规范方程

可得状态和协态的导数分别为:

$$\dot{x}(t) = u(t) = \frac{\partial \mathcal{H}}{\partial p}(x(t), \dot{x}(t), p(t), t),$$
$$\dot{p}(t) = \frac{\mathrm{d}}{\mathrm{d}t}[-\frac{\partial g}{\partial u}(x(t), u(t), t)].$$

由欧拉-拉格朗日方程-(16),协态变量的导数可进一步化简为使用哈密尔顿函数表示的:

$$\begin{aligned}\dot{p}(t) &= -\frac{\partial g}{\partial x}(x(t), u(t), t)\\ &= -\frac{\partial \mathcal{H}}{\partial x}(x(t), u(t), t).\end{aligned}$$

以及:

$$\frac{\partial \mathcal{H}}{\partial u}(x(t), u(t), t) = p(t) + \frac{\partial g}{\partial u}(x(t), u(t), t) = 0.$$

变分问题的“极小值原理”——初识

$$0 = \frac{\partial \mathcal{H}}{\partial u} \tag{18}$$

$$\dot{x} = +\frac{\partial \mathcal{H}}{\partial p} \tag{19}$$

$$\dot{p} = -\frac{\partial \mathcal{H}}{\partial x}. \tag{20}$$

初识“庞特里亚金极小值原理”

- 控制变量为状态变化率
- 无约束
- 连续可微

有等式约束函数极值——拉格朗日乘子

对于有等式约束 $f(x_1, x_2) = 0$ 情况下求 $F(x_1, x_2)$ 极值问题，引入拉格朗日乘子 λ

$$\bar{F}(x_1, x_2, \lambda) = F(x_1, x_2) + \lambda f(x_1, x_2) \tag{21}$$

$F(x_1, x_2)$ 取极值的必要条件是

$$\frac{\partial \bar{F}}{\partial \lambda} = f(x_1, x_2) = 0 \tag{22}$$

$$\frac{\partial \bar{F}}{\partial x_1} = 0 \tag{23}$$

$$\frac{\partial \bar{F}}{\partial x_2} = 0 \tag{24}$$

拉格朗日乘子法处理约束

Remark 5

- 微分方程约束【拉格朗日乘子法】

$$f(x(t), \dot{x}(t), t) = 0 \tag{25}$$

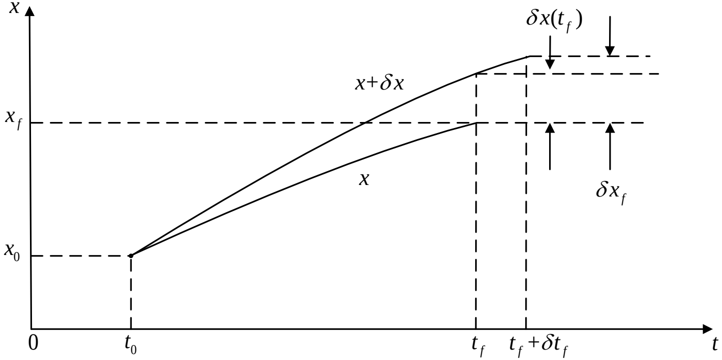
- 积分约束【构造积分引入新变量】

$$\int_{t_0}^{t_f} b(x(t), \dot{x}(t), t) dt = B \tag{26}$$

消除变分依赖

$$\delta\dot{x}(t) = \frac{d}{dt}\delta x(t) \tag{27}$$

$$\delta x_f \approx \delta x(t_f) + \dot{x}(t_f)\delta t_f \tag{28}$$



三种形式的性能指标

例 4 (Lagrange 形式)

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t) dt$$

例 5 (Mayer 形式)

$$J(x) = h(x(t_f), t_f)$$

例 6 (Bolza 形式)

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

变分法求解最优控制问题 – 2nd

$$\text{极值条件}:0 = \frac{\partial \mathcal{H}}{\partial u}(x(t), u(t), t). \tag{29}$$

$$\text{状态方程}:\dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial x}(x(t), u(t), t). \tag{30}$$

$$\text{协态方程}:\dot{x}(t) = +\frac{\partial \mathcal{H}}{\partial p}(x(t), u(t), t). \tag{31}$$

以及边界条件:

$$0 = \left[\frac{\partial h}{\partial x}(x(t_f), t_f) - p(t_f)\right] \cdot \delta x_f. \tag{32}$$

$$0 = \left[\frac{\partial h}{\partial t}(x(t_f), t_f) + \mathcal{H}(x(t_f), u(t_f), t_f)\right] \delta t_f. \tag{33}$$

四种特殊情况；拉格朗日乘子法处理一般目标集

稳态 Hamiltonian

定理 2 (终端时刻固定，稳态 Hamiltonian)

t_f fixed, 稳态系统最优控制的 Hamiltonian 满足

$$\mathcal{H}(x(t), u(t), p(t)) = c_1, \ \forall t \in [t_0, t_f]. \tag{34}$$

其中 c_1 为常数

定理 3 (终端时刻自由，稳态 Hamiltonian)

t_f free, 稳态系统最优控制的 Hamiltonian 满足

$$\mathcal{H}(x(t), u(t), p(t)) = 0, \ \forall t \in [t_0, t_f]. \tag{35}$$

经典变分法求解最优控制的缺陷

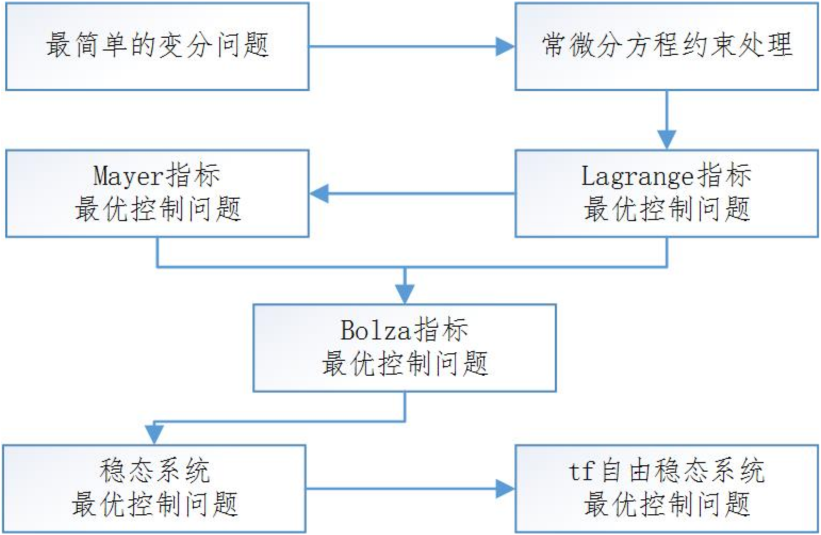
Remark 6 (经典变分法求解最优控制的缺陷)

- 状态变量或控制变量有不等式约束难以处理
- 控制变量不连续无法处理

扩大解空间，状态变量分段连续可微，控制变量分段连续

Table of Contents

回忆：变分法求解最优控制的思路



极小值原理的思路

$$\Delta J = 0 \Rightarrow\Rightarrow\Rightarrow \Delta J \geq 0$$

- 稳态系统 Mayer 指标的最优控制问题
- 稳态系统 Bolza 指标的最优控制问题
- 时变系统的最优控制问题

Pontryagin 极小值原理. 3rd

定理 4 (庞特里亚金极小值原理)

放大解空间，考察 $\Delta J \geq 0$，最优控制 $u(t)$ 的必要条件为（TPBVP）

- 极值条件：对任意容许控制 $u'(t)$
$$\mathcal{H}(x(t), u(t), p(t), t) \leq \mathcal{H}(x(t), u'(t), p(t), t).$$
- 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)：
$$\left[\frac{\partial h}{\partial x}(x(t_f), t_f) - p(t_f)\right] \cdot \delta x_f$$
$$+ \left[\mathcal{H}(x(t_f), u(t_f), p(t_f), t_f) + \frac{\partial h}{\partial t}(x(t_f), t_f)\right]\delta t_f = 0.$$

极值原理求解最优控制的过程

Remark 7 (极值原理求解最优控制的过程)

- 构造 Hamiltonian
- 求容许控制极值条件，以协态状态表示最优控制
- 最优控制代入规范方程，得到关于最优状态、协态的微分方程组
- 根据边界条件和初值获得微分方程组的边界条件
- 直接求解 或 使用数值方法求解 两点边值问题

打靶法，Single/Sequential Shooting Method

- 初始化：$s = x(t_0)$
- 打靶：求解常微分方程初值问题（IVP），得 $x(t_f; s)$
- 得终端时刻的误差：$c(s) := x_f - x(t_f; s)$
- 使用非线性规划方法重复上述过程求解 $c(s) = 0$

- IVP 求解：微分化为差分

时间最短控制

问题 2 (时间最短控制)

状态变量 $x(t):[t_0,t_f]\to\mathbb{R}^n$ 分段连续可微，控制变量 $u(t):[t_0,t_f]\to\mathbb{R}^m$ 分段连续。状态初值 $x(t_0)=x_0$。状态方程为

$$\dot{x}(t)=A(x(t),t)+B(x(t),t)u(t). \tag{36}$$

容许控制为对任意的 $t\in[t_0,t_f]$，

$$|u_i(t)|\le 1,\quad i=1,2,\ldots,m. \tag{37}$$

具有自由终端时刻的目标集

$$\mathcal{S}=[t_0,\infty)\times\{x(t_f):m(x(t_f),t_f)=0\}. \tag{38}$$

要最小化的性能指标为达到目标集所用时间 $J(u)=t_f-t_0$

Bang-Bang 控制原理

定义 7 (正常 (normal) 时间最短控制问题)
$p(t)b_i(x(t),t)=0$ 仅在可数时刻成立，则称时间最短控制问题是正常的

定理 5 (Bang-Bang 控制原理)

若时间最短控制问题是正常的，则最优控制的每个分量 $u_i(t)\in\mathbb{R}$, $i=1,2,\ldots,n$，在最大值 $+1$ 和最小值 -1 之间切换

$$u_i(t)=-\operatorname{sign}\{p^{\mathrm{T}}(t)b_i(x(t),t)\}.$$

其中 $\operatorname{sign}(y)$ 为符号函数 $\mathbb{R}\to\mathbb{R}$，取值为 y 的正负号，即

$$\operatorname{sign}(y)\stackrel{def}{=}\begin{cases}+1, & y\geq 0,\\ -1, & y<0.\end{cases}$$

例子：线性定常系统的时间最短控制

例 7 (线性定常系统的时间最短控制)

时间最短控制，初值 t_0, x_0，自由的 t_f 位于原点。状态方程

$$\dot{x}_1(t) = x_2(t) \tag{39}$$

$$\dot{x}_2(t) = u(t). \tag{40}$$

控制约束为

$$|u(t)| \leq 1. \tag{41}$$

最小化性能指标

$$J(u) = \int_{t_0}^{t_f} dt = t_f - t_0 \tag{42}$$

计算 Hamiltonian, 考察极值条件

Hamiltonian

$$\mathcal{H}(x(t), u(t), p(t)) = 1 + p_1(t)x_2(t) + p_2(t)u(t). \tag{43}$$

极值条件

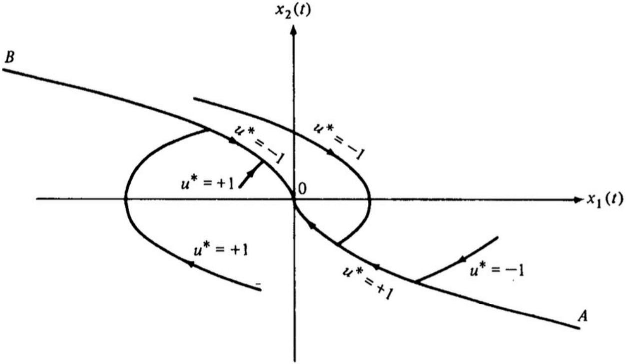
$$p_2(t)u(t) \leq p_2(t)u'(t), \forall u \in U, t \in [t_0, t_f], \tag{44}$$

于是

$$u(t) = -\operatorname{sign}[p_2(t)] \tag{45}$$

时间最短控制的切换曲线

最优轨迹应先依最大/最小控制至切换曲线 $s(x(t)) := x_1(t) + \frac{1}{2}x_2(t)|x_2(t)| = 0$ 切换为最小/最大控制至结束



闭环形式的时间最短控制

$$u(t) = \begin{cases} -\operatorname{sign}[s(x(t))], & s(x(t)) \neq 0, \\ -\operatorname{sign}[x_2(t)], & s(x(t)) = 0. \end{cases} \tag{46}$$

线性二次型最优控制

问题 3 (线性二次型最优控制)

状态方程

$$\dot{x}(t) = A(t)x(t) + B(t)u(t). \tag{47}$$

最小化性能指标

$$J = \frac{1}{2}x^T(t_f)Hx(t_f) + \frac{1}{2}\int_{t_0}^{t_f}[x^T(t)Q(t)x(t) + u^T(t)R(t)u(t)]dt. \tag{48}$$

t_f *fixed*, $x(t_f)$ *free*
其中 H 和 $Q(t)$ 是实对称半正定矩阵，$R(t)$ 是实对称正定矩阵

使用极值原理可求得线性二次型的 闭环形式最优控制

极值原理求解最优控制

- 极值原理得开环控制，一般问题无法保证闭环解
- 求解有不等式约束最优控制时处理较为复杂

极值原理是处理最优控制的有效方法，可处理各种非线性问题及约束条件。对于特定问题可解得 闭环形式最优控制 。然而，对于长时间运行的非线性、有约束系统，BVP 问题求解复杂度较高；对一般问题也无法保证求得闭环形式最优控制。预测控制可用于改善上述问题

Table of Contents

预测控制（线性/非线性、有约束/无约束）4nd

模型预测控制

Remark 8

模型预测控制是一系列控制方法的统称。都具有如下特征:

1. 预测模型
   利用预测模型预测系统在一定控制作用下未来的动态行为。应确保能快速求解
2. 滚动优化
   对预测模型求解一段时间内的开环最优控制，并实施当前时刻的控制变量；下一采样时刻，重新获取状态作为新的初值，滚动时间窗口重复上述最优控制求解
3. 反馈矫正
   在求解滚动优化前，系统首先利用反馈信息矫正预测模型

线性预测控制

问题 4 (线性预测控制)

线性状态方程

$$x(k;k) = x(k) \tag{49}$$

$$x(i+1;k) = Ax(i;k) + Bu(i;k),\ i = k, k+1, \dots, N-1 \tag{50}$$

预测时段 N，设计二次性能指标

$$\begin{aligned} J(u;k) = & \frac{1}{2} x^T(k+N;k) H x(k+N;k) \\ & + \frac{1}{2} \sum_{i=k}^{k+N-1} [x^T(i;k) Q x(i;k) + u^T(i;k) R u(i;k)]. \end{aligned} \tag{51}$$

线性预测控制求解

对于无约束的预测控制问题，在每个时刻仅需求解最优化问题

$$J(u;k) = \frac{1}{2}U^T(k)\widetilde{Q}U(k) - U^T(k)\widetilde{\mathcal{B}}x(k) + \frac{1}{2}x^T(k)\widehat{Q}x(k)$$

于是

$$U(k) = \widetilde{Q}^{-1}\widetilde{\mathcal{B}}x(k) \tag{52}$$

考虑预测控制的滚动时域

$$u(k;k) = [I, 0, \ldots, 0]\widetilde{Q}^{-1}\widetilde{\mathcal{B}}x(k) \tag{53}$$

模型预测控制的补偿策略性能分析

- 终端零约束
  有限时域优化问题中，加入条件 $x(k+N;k)=0$，若 $x(k+i;k)=0$，$i=N,N+1,\ldots$ 则 MPC 可保证稳定
- 终端集约束
  若在终端时刻将状态控制到 0 的小邻域，假定其后实施稳态的反馈最优控制，则 MPC 可保证稳定
- 终端代价函数–ADP
  若可精确得知最优控制问题的状态值函数，将其作为终端代价，则 MPC 可精确最优
- 其他调整性能指标的方法

Table of Contents

动态规划方法

离散：Bellman 方程，

$$V(x(k),k) = \min_{u(k)\in U}\{g_D(x(k),u(k),k) + V(x(k+1),k+1)\}$$

$$k = k_0,\ldots,N-1 \tag{54}$$

$$V(x(N),N) = h_D(x(N),N). \tag{55}$$

连续：HJB 方程，

$$-\frac{\partial V}{\partial t}(x(t),t) = \min_{u(t)\in\mathbb{R}^m} \mathcal{H}(x(t),u(t),\frac{\partial V}{\partial x}(x(t),t),t), \tag{56}$$

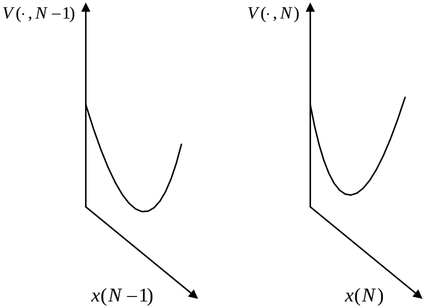
$$V(x(t_f),t_f) = h(x(t_f),t_f). \tag{57}$$

直接倒推求解 1/2

$$V(x(N),N) = h_D(x(N),N)$$

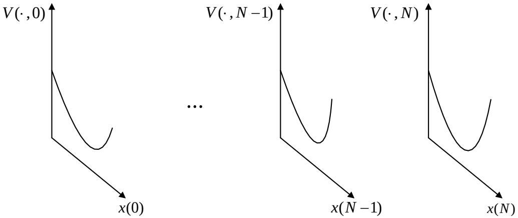
$$V(x(N-1),N-1) = \min_{u(N-1)\in U}\left\{g_D(x(N-1),u(N-1),N-1) + V(x(N),N)\right\}$$

$$= \min_{u(N-1)\in U}\left\{g_D(x(N-1),u(N-1),N-1) + V(f_D(x(N-1),u(N-1)),N)\right\}.$$



直接倒推求解 2/2

$$
\begin{aligned}
V(x(k),k) &= \min_{u(k)\in U}\Big\{g_D(x(k),u(k),k)+V(x(k+1),k+1)\Big\} \\
&= \min_{u(k)\in U}\Big\{g_D(x(k),u(k),k)+V(f_D(x(k),u(k)),k+1)\Big\}.
\end{aligned}
$$



1/3 HJB⇒PMP. 5nd

上述证明过程中出现了极值条件，即对于最优控制 $x(t), u(t)$

$$\mathcal{H}(x(t), u(t), \frac{\partial V}{\partial x}(x(t), t), t) \leq \mathcal{H}(x(t), u'(t), \frac{\partial V}{\partial x}(x(t), t), t).$$

若我们能由此推得 $p = \partial V / \partial x$ 满足协态方程和边界条件，则推得极值原理！

$$\dot{p}(t) = -\frac{\partial \mathcal{H}}{\partial x},$$

即，

$$\frac{d}{dt}[\frac{\partial V}{\partial x}] = -\frac{\partial \mathcal{H}}{\partial x}.$$

假定 V 二次连续可微，考察 $n = m = 1$，终端状态时间 free。有

$$\frac{\partial^2 V}{\partial x \partial t} = \frac{\partial^2 V}{\partial t \partial x}, \quad -\frac{\partial V}{\partial t} = g + \frac{\partial V}{\partial x} f$$

2/3 协态方程

$$
\begin{aligned}
\frac{d}{dt}[\frac{\partial V}{\partial x}(x(t),t)] &= \frac{\partial^2 V}{\partial x^2}(x(t),t)\frac{dx}{dt} + \frac{\partial}{\partial t}[\frac{\partial V}{\partial x}(x(t),t)] \\
&= \frac{\partial^2 V}{\partial x^2} f + \frac{\partial}{\partial x}[\frac{\partial V}{\partial t}] \\
&= \frac{\partial^2 V}{\partial x^2} f - \frac{\partial}{\partial x}[g + \frac{\partial V}{\partial x} f] \\
&= \frac{\partial^2 V}{\partial x^2} f - [\frac{\partial g}{\partial x} + \frac{\partial^2 V}{\partial x^2} f + \frac{\partial V}{\partial x}\frac{\partial f}{\partial x}] \\
&= -\frac{\partial g}{\partial x} - \frac{\partial V}{\partial x}\frac{\partial f}{\partial x} \\
&= -\frac{\partial \mathcal{H}}{\partial x}(x(t), u(t), \frac{\partial V}{\partial x}, t)
\end{aligned}
$$

3/3 边界条件

边界条件

$$V(x(t_f), t_f) = h(x(t_f), t_f).$$

于是

$$\frac{\partial V}{\partial x}(x(t_f), t_f) = \frac{\partial h}{\partial x}(x(t_f), t_f),$$

终端状态自由的边界条件。直接令 HJB 方程在 t_f 取值，

$$\begin{aligned} 0 &= \frac{\partial V}{\partial t}(x(t_f), t_f) + \min_{\xi} \mathcal{H}(x(t_f), \xi, \frac{\partial V}{\partial x}(x(t_f), t_f), t_f) \\ &= \frac{\partial V}{\partial t}(x(t_f), t_f) + \mathcal{H}(x(t_f), u(t_f), \frac{\partial V}{\partial x}(x(t_f), t_f), t_f) \end{aligned}$$

4/4 倒推求解最优控制

$$V(x(N),N)=\frac{1}{2}x^{\mathrm{T}}(N)K(N)x(N). \tag{58}$$

对 $k=N-1,\ldots,0$,

$$u(k)=F(k)x(k), \tag{59}$$

$$V(x(k),k)=\frac{1}{2}x^{\mathrm{T}}(k)K(k)x(k). \tag{60}$$

其中

$$K(N)=H, \tag{61}$$

$$F(k)=-[R(k)+B^{\mathrm{T}}(k)K(k+1)B(k)]^{-1}B^{\mathrm{T}}(k)K(k+1)A(k), \tag{62}$$

$$\begin{aligned}K(k)=&\ Q(k)+F^{\mathrm{T}}(k)R(k)F(k)\\&+[A(k)+B(k)F(k)]^{\mathrm{T}}K(k+1)[A(k)+B(k)F(k)].\end{aligned} \tag{63}$$

连续动态规划求解线性二次型 5/5

闭环形式的最优控制满足 Riccati 微分方程

$$
\begin{aligned}
0 &= \dot{K}(t) + Q(t) - K(t)B(t)R^{-1}(t)B^T(t)K(t) \\
&\quad + K(t)A(t) + A^T(t)K(t) \qquad (64) \\
K(t_f) &= H \qquad (65) \\
u(t) &= -R^{-1}(t)B^T(t)K(t)x(t). \qquad (66)
\end{aligned}
$$

动态规划求解最优控制

- 离散化模型面临维数灾难
- HJB 方程一般难以求解
- HJB 方程对值函数有可微的要求

对于相对简单的系统，尤其是状态空间离散的系统，动态规划方法可获得闭环形式的最优控制。对于无约束的线性二次型问题，Bellman 方程和 HJB 方程都容易求解。然而，对于较为复杂的状态空间或较长时间的问题，动态规划方法将面临维数灾难。自适应动态规划可处理这些问题

Table of Contents

智能体的基本要素

在强化学习中，智能体可能具有如下要素

- 策略 Policy：建模智能体的行为
- 值函数 Value function：建模智能体对状态和/或控制的估值
- 模型 Model：智能体对环境的表示 representation

Policy (控制策略，控制律)

控制策略policy从状态到控制的映射。本课考察稳态策略

- 确定策略

$$a = \pi(s)$$

- 随机策略

$$\pi(a|s) = P(A_t = a|S_t = s)$$

若求得行动值函数 $q_*(s,a)$，有确定性最优策略

$$\pi_*(a|s) = \begin{cases} 1 & \text{if } a = \text{argmax}_{a\in A}\, q_*(s,a), \\ 0 & \text{otherwise.} \end{cases} \tag{67}$$

Value Function （值函数）

一个控制策略的值函数value function定义为期望累积收益

$$v_\pi(s) = E_\pi(G_t | S_t = s)$$
$$q_\pi(s, a) = E_\pi[G_t | S_t = s, A_t = a]$$

其中，

$$G_t := R_{t+1} + \gamma R_{t+2} + \ldots = \sum_{i=0}^{\infty} \gamma^i R_{t+i+1}$$

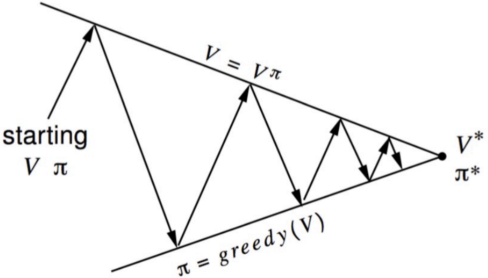
- 一个控制策略的值函数用于估计这个策略下特定状态的优劣
- 同时也是对这个策略的评价
- 最优值函数，或简称值函数，是任意策略的值函数的极大值

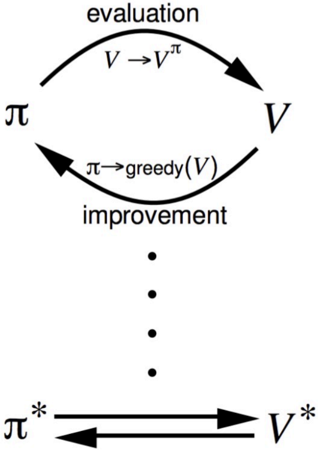
Model（模型）

智能体用模型model估计下一时刻的环境状态和奖励，对于未知的随机系统常用状态转移矩阵和期望收益表示

$$\mathcal{P}(s,a,s') = P(S_{t+1} = s' | S_t = s, A_t = a)$$
$$\mathcal{R}(s,a) = E(R_{t+1} | S_t = s, A_t = a)$$

Policy Iteration，策略迭代





Policy evaluation Estimate v_π

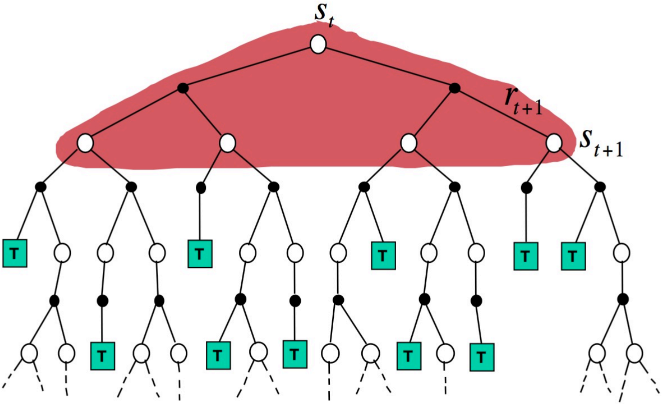
Iterative policy evaluation

Policy improvement Generate $\pi' \geq \pi$

Greedy policy improvement

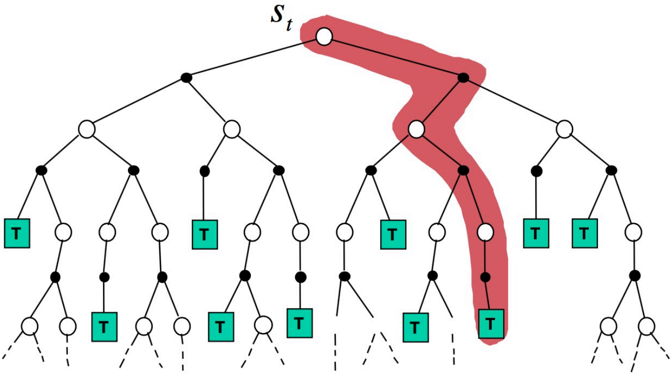
Dynamic Programming

$$V(S_t) \leftarrow E_\pi[R_{t+1} + \gamma V(S_{t+1})]$$



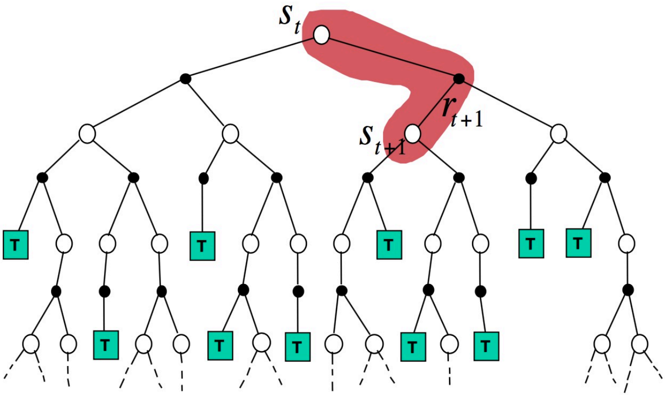
Monte-Carlo

$$V(S_t) \leftarrow V(S_t) + \alpha(G_t - V(S_t))$$



Temporal-Difference

$$V(S_t) \leftarrow V(S_t) + \alpha(R_{t+1} + \gamma V(S_{t+1}) - V(S_t))$$



贪婪和 $\epsilon-$ 贪婪

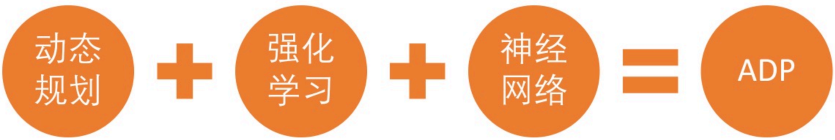
贪婪策略

$$\pi(s) = \operatorname*{argmax}_{a \in A} Q(s, a)$$

$\epsilon-$ 贪婪策略，ϵ 概率随机行动，$1-\epsilon$ 贪婪行动

$$\pi(a|s) = \begin{cases} \epsilon/m + 1 - \epsilon & \text{if } a = \text{argmax}_{a \in A} Q(s, a) \\ \epsilon/m & \text{otherwise} \end{cases}$$

自适应动态规划



HDP92: Heuristic Dynamic Programming

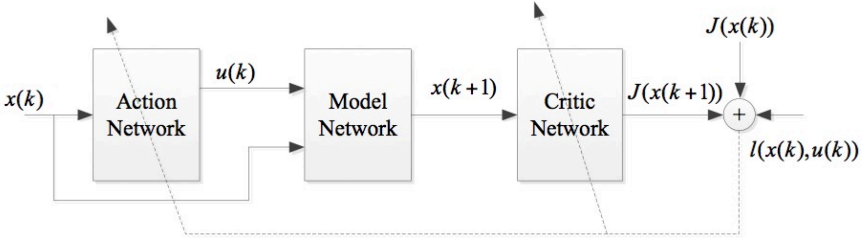


Figure: Werbos, 1992

值迭代自适应动态规划

对于值迭代自适应动态规划方法，初始的近似值函数可定义为

$$V_0(\cdot) = 0 \tag{68}$$

对于 $i = 0, 1, \ldots$，值迭代自适应动态规划方法如下迭代

$$u_i(x_k) = \underset{u_k}{\operatorname{argmin}}\{g_D(x_k, u_k) + V_i(x_{k+1})\} \tag{69}$$

$$V_{i+1}(x_k) = g_D(x_k, u_i(x_k)) + V_i(f_D(x_k, u_i(x_k))) \tag{70}$$

策略迭代自适应动态规划

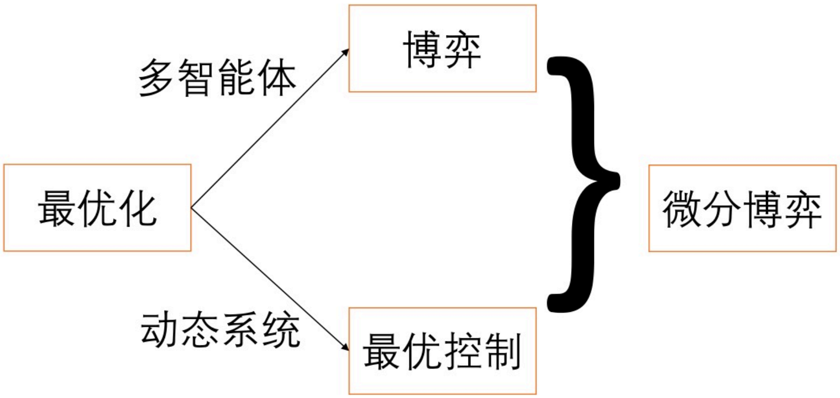
对于策略迭代自适应动态规划方法，初始于一个给定的容许控制律 (admissible control law) $u_0(x_k)$. 对于 $i = 1, 2, \ldots,$ 策略迭代自适应动态规划方法如下迭代

$$V_i(x_k) = g_D(x_k, u_i(x_k)) + V_i(f_D(x_k, u_i(x_k))) \tag{71}$$

$$u_{i+1}(x_k) = \underset{u_k}{\operatorname{argmin}}\{g_D(x_k, u_k) + V_i(x_{k+1})\} \tag{72}$$

Table of Contents

微分博弈



反应函数法求解博弈平衡

定义 8 (反应函数)

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

Remark 9 (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

Remark 10 (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

两人零和微分博弈的开环形式平衡

定理 6 (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
$$\dot{x}(t) = f(x(t), u_1(t), u_2(t), t),\ x(t_0) = x_0. \tag{73}$$
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
$$J(u_1, u_2) = h(x(t_f), t_f) + \int_{t_0}^{t_f} g(x(t), u_1(t), u_2(t), t)dt \tag{74}$$

定义 Hamiltonian
$$\begin{aligned}\mathcal{H}(x(t), u_1(t), u_2(t), p(t), t) := {} & g(x(t), u_1(t), u_2(t), t) \\ & + p^T(t) f(x(t), u_1(t), u_2(t), t),\end{aligned} \tag{75}$$

两人零和微分博弈的开环形式平衡

定理 6 (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:*

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

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

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

$$\begin{aligned}\text{边界条件:}\ &[\frac{\partial h}{\partial x}(x(t_f), t_f) - p(t_f)] \cdot \delta x_f \\ &+ [\mathcal{H}(x(t_f), u_1(t_f), u_2(t_f), p(t_f), t_f) + \frac{\partial h}{\partial t}(x(t_f), t_f)]\delta t_f = 0.\end{aligned}$$

博弈和倒推

这个理论涉及的与其说是一种从博弈的开始来看为最好的对策，不如说是一种从博弈的结局看来为最好的对策。在博弈的最后一着中，如果有可能，一个博弈参与者总是力求走能获胜的一着，其次要走能得平局的一着。他的对手，在走他这一着的前面一着时，总是力求要取一种着法，使得他不能走这获胜或得平局的一着……依次倒着推下去，都是如此。

—— 维纳

《控制论：或关于在动物和机器中控制和通讯的科学》第二版