数理经济学
(动态规划原理)

王鸣晖

西南财经大学

数学学院

wangmh@swufe.edu.cn

2022 年 05 月 23 日

动机

数学思想

一个小数学问题 $\Rightarrow$ 放进更一般的问题中 $\Rightarrow$ 求解更一般的问题 $\Rightarrow$ 求解了小问题.

一个例子

分析 PDE 解的存在唯一性 (不光滑) $\Rightarrow$ Sobolev 空间弱导数 $\Rightarrow$ Sobolev 弱解 $\Rightarrow$ Sobolev 嵌入定理 $\Rightarrow$ 原始问题的解.

微积分的例子

求解如下积分:

$$\int_0^{+\infty} \frac{\sin x}{x} dx.$$

微积分的例子

定义如下函数:

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

则有:

$$I'(\alpha) = \int_0^{+\infty} (-x) e^{-\alpha x} \frac{\sin x}{x} dx = \int_0^{+\infty} -e^{-\alpha x} \sin x dx = -\frac{1}{\alpha^2 + 1}.$$

这就是说：

$$I(\alpha) = -\arctan \alpha + C.$$

又由于有:

$$\lim_{\alpha \to +\infty} I(\alpha) = 0 \Rightarrow \lim_{\alpha \to +\infty} -\arctan \alpha + C = 0 \Rightarrow C = \frac{\pi}{2}.$$

故有:

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

对于固定的最优终止时间 $T>0$, 其对应的受控方程为:

$$\begin{cases} \frac{dX(s)}{ds} = f(X(s),\alpha(s)), & (0<s<T) \\ X(0) = x_0. \end{cases} \tag{ODE}$$

此外, 由该方程对应的损失泛函 (payoff functional) 为:

$$P[\alpha(\cdot)] = \int_0^T r(X(s),\alpha(s))ds + g(X(T)). \tag{P}$$

此外, 定义允许可测集 (admissible controls set) 如下:

$$\mathcal{A} = \{\alpha : [0,+\infty) \to A | \alpha(\cdot)\text{可测}\}.$$

这里 $A \in \mathscr{R}^m$.

更一般的最优控制问题

现在，我们将上述受控方程(ODE)和损失泛函(P)放入更一般的问题中，即：对任意的开始时间 $t < T$ 及其初始值 $X(t) = x$，有：

$$\begin{cases} \frac{dX(s)}{ds} = f(X(s), \alpha(s)), \quad (t < s < T) \\ X(t) = x. \end{cases} \tag{E-15-1}$$

及

$$P_{x,t}[\alpha(\cdot)] = \int_t^T r(X(s), \alpha(s))ds + g(X(T)). \tag{E-15-2}$$

定义

对 $x \in \mathscr{R}^n$ 和 $0 \le t \le T$，定义值函数 (value function) 如下：

$$v(x,t) = \sup_{\alpha(\cdot) \in \mathcal{A}} P_{x,t}[\alpha(\cdot)] \quad (x \in \mathscr{R}^n, 0 \le t \le T). \tag{E-15-3}$$

且有：$v(x,T) = g(x)$, $x \in \mathscr{R}^n$.

定理—1(HJB 方程)

假设 A 是一个紧集，且值函数 v 是一个关于 (x,t) 的 C^1 函数, 则 v 是下面非线性方程的解:

$$v_t(x,t)+\max_{a\in A}\{f(x,a)\cdot\nabla_x v(x,t)+r(x,a)\}=0,\quad (x\in\mathscr{R}^n, 0\leq t\leq T) \tag{HJB}$$

上述方程对应的终值条件为:

$$v(x,T)=g(x),\quad x\in\mathscr{R}^n.$$

注记—1

方程(HJB)可以被写为如下形式:

$$v_t(x,t)+H(x,\nabla_x v)=0. \quad (x\in\mathscr{R}^n, 0\leq t\leq T)$$

这里, Hamiltonian 函数为:

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

且 $x,p\in\mathscr{R}^n$.

定理—1 的证明

对应定理—1 的证明，主要分为如下三步:

第一步: 给定常数 $h>0$, 取 A 中任意的常数 $a \in A$ 且在时间区间 $t \le s \le t+h$ 中考虑如下的常数控制:

$$\alpha(\cdot)=a.$$

则此时动态系统的起点来到了 $x(t+h)$ 且 $t+h<T$, 且假设在 $t+h \le s \le T$ 中, 控制函数均为最优控制函数.
此时，对于 $t \le s \le t+h$, 我们可以得到:

$$\begin{cases} \frac{dx(s)}{ds}=f(x(s),a), \\ x(t)=x. \end{cases}$$

且此时对应的损失函数为:

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

此外，从时刻 $t+h$ 到 T 的值函数为

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

则根据损失函数 v 的定义，可以得到总的损失为:

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

另一方面，从初值 (x,t) 出发的最大损失函数为 $v(x,t)$，因此成立:

$$v(x,t) \geq \int_t^{t+h} r(x(s),a)ds + v(x(t+h),t+h). \tag{E-15-4}$$

定理—1 的证明 (接上)

第二步：将不等式(E-15-4)写成如下微分形式：

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

取 $h \to 0$，则有：

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

但由于 $x(\cdot)$ 是如下方程的解：

$$\begin{cases} \frac{dx(s)}{ds}=f(x(s),a), & (t\le s\le t+h) \\ x(t)=x. \end{cases}$$

故可得：

$$v_t(x,t)+f(x,a)\cdot\nabla_x v(x(t),t)+r(x,a)\le 0. \tag{E-15-5}$$

此时, 对 $\forall a \in A$, 由不等式(E-15-5)可以得到:

$$\max_{a\in A}\{v_t(x,t)+f(x,a)\cdot\nabla_x v(x(t),t)+r(x,a)\}\leq 0. \tag{E-15-6}$$

第三步: 说明在不等式(E-15-6)取等号时, 得到最大值函数. 假设 $\alpha^*(\cdot)$ 和 $x^*(\cdot)$ 是控制问题的最优解. 则在 $t\leq s\leq t+h$ 时, 损失为:

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

则在剩余时间的损失函数为 $v(x^*(t+h),t+h)$. 由此, 总的损失为:

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

则可以得到:

$$\frac{v(x^*(t+h),t+h)-v(x,t)}{h}+\frac{1}{h}\int_t^{t+h} r(x^*(s),\alpha^*(s))ds=0.$$

取 $h\to 0$ 且假设 $\alpha^*(t)=a^*\in A$, 则有:

$$v_t(x,t)+\nabla_x v(x,t)\cdot \underbrace{\dot{x}^*(t)}_{f(x,a^*)}+r(x,a^*)=0;$$

故有:

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

对 $a^*\in A$ 成立. 由此, 证明了 HJB 方程.

动态规划原理的使用步骤如下:

(1) 求解对应的 HJB 方程, 即求解值函数 v;

(2) 运用值函数和 HJB 方程设计最优回馈控制 $\alpha^*(\cdot)$. 具体而言, 对任意点 $x \in \mathscr{R}^n$ 和对任意时刻 $0 \le t \le T$, 和

$$\alpha(x,t) = a \in A$$

使得 HJB 方程最大化. 也就是说, 可以选择 $\alpha(x,t)$ 使得:

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

(3) 假设 $\alpha(\cdot,t)$ 是一个充分光滑的控制函数, 则求解如下 ODE:

$$\begin{cases} \frac{dx^*(s)}{ds} = f(x^*(s),\alpha(x^*(s),s)), & (t \le s \le t+h) \\ X(t) = x. \end{cases}$$

(4) 定义回馈控制为: $\alpha^*(s) = \alpha(x^*(s),s)$.

现在, 我们来证明上述步骤 (1) -(4) 所得到的结果是最优控制函数.

验证性定理 (Verification of optimalily)

上述步骤 (4) 中定义的最优控制函数 $\alpha^*(\cdot)$ 是最优的.

证明:

首先, 可以得到:

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

动态规划原理

证明（接上）

则根据上述步骤 (4) 中关于 $\alpha(\cdot)$ 的定义，成立:

$$
\begin{aligned}
P_{x,t}[\alpha^*(\cdot)] &= \int_t^T (-v_t(x^*(s),s) - f(x^*(s),\alpha^*(s)) \cdot \nabla_x v(x^*(s),s))ds + g(x^*(T)) \\
&= -\int_t^T v_t(x^*(s),s) + \nabla_x v(x^*(s),s) \cdot \dot{x}^*(s)ds + g(x^*(T)) \\
&= -\int_t^T \frac{d}{ds} v(x^*(s),s)ds + g(x^*(T)) \\
&= -v(x^*(T),T) + v(x^*(t),t) + g(x^*(T)) \\
&= -g(x^*(T)) + v(x^*(t),t) + g(x^*(T)) \\
&= v(x,t) = \sup_{\alpha(\cdot)\in\mathcal{A}} P_{x,t}[\alpha^*(\cdot)].
\end{aligned}
$$

此时，成立:

$$
P_{x,t}[\alpha^*(\cdot)] = \sup_{\alpha(\cdot)\in\mathcal{A}} P_{x,t}[\alpha^*(\cdot)].
$$

例子 (线性二次调节器)

对给定的矩阵 $M, B, D \in \mathcal{M}^{n\times n}$, $N \in \mathcal{M}^{n\times m}$ 和 $C \in \mathcal{M}^{m\times m}$, 假设 B, C, D 是对称矩阵且有 B, D 半正定, C 正定. 特别地, 还要求矩阵 C 可逆. 现在考虑如下线性系统:

$$\begin{cases} \frac{dX(s)}{ds} = Mx(s) + N\alpha(s), \quad (t \le s \le T) \\ x(t) = x. \end{cases} \tag{ODE}$$

此外, 上述系统对应的最大化目标泛函为:

$$P_{x,t}[\alpha(\cdot)] = -\int_t^T x(s)^T Bx(s) + \alpha(s)^T C\alpha(s) ds - x(T)^T Dx(T). \tag{P}$$

这里, 控制集的取值为 $A = \mathscr{R}^m$.

例子 (线性二次调节器)

由动态规划原理, 可以得到如下的 HJB 方程:

$$\begin{cases} \frac{\partial v}{\partial t} + \max_{a \in \mathscr{R}^m}\{f \cdot \nabla_x v + r\} = 0 & \text{in} \quad \mathscr{R}^n \times [0, T]; \\ v = g & \text{on} \quad \mathscr{R}^n \times (t = T). \end{cases}$$

这里有:

$$f = Mx + N\alpha, \quad r = -x^T Bx - a^T Ca, \quad g = -x^T Dx.$$

则此时可将上述 HJB 方程从新写为:

$$\frac{\partial v}{\partial t} + \max_{a \in \mathscr{R}^m}\{(\nabla v)^T Na - a^T Ca\} + (\nabla v)^T Mx - x^T Bx = 0, \tag{HJB}$$

和终值条件:

$$v(x, T) = -x^T Dx.$$

例子 (线性二次调节器)

为了求得最大化上述函数，定义:

$$Q(a) = (\nabla v)^T Na - a^T Ca,$$

且有:

$$\frac{\partial Q}{\partial a_j} = \sum_{i=1}^{n} v_{x_i} n_{ij} - 2a_i c_{ij} = 0. \quad (j = 1, \cdots, n)$$

则可以发现 $(\nabla_x v)^T N = 2a^T C \Rightarrow 2Ca = N^T \nabla_x v$. 又由于 C 是对称可逆矩阵，故有:

$$a = \frac{1}{2} C^{-1} N^T \nabla_x v.$$

现将上述结果代回(HJB)方程，可得:

$$\frac{\partial v}{\partial t} + \frac{1}{4} (\nabla v)^T N C^{-1} N^T \nabla v + (\nabla v)^T Mx - x^T Bx = 0. \tag{PDE}$$

即求得最大化的 HJB 方程.

假设方程(PDE)的解为如下形式:

$$v(x,t)=x^TK(t)x \tag{E-15-7}$$

这里 $K(t)$ 是一个取值为 $n\times n$ 的对称矩阵的矩阵映射. 且根据终值条件有:

$$K(T)=-D. \tag{E-15-8}$$

现在计算:

$$\frac{\partial v}{\partial t}=x^T\dot{K}(t)x,\qquad \nabla_x v=2K(t)x.$$

将上述式子和 $v=x^TK(T)x$ 代回方程(PDE), 可得:

$$x^T\{\dot{K}(t)+K(t)NC^{-1}N^TK(t)+2K(t)M-B\}x=0. \tag{E-15-9}$$

例子 (线性二次调节器)

由于:

$$\begin{aligned} 2x^T KMx &= x^T KMx + [x^T KMx]^T \\ &= x^T KMx + x^T M^T Kx. \end{aligned}$$

故可将(E-15-9)变为:

$$x^T\{\dot{K}(t) + K(t)NC^{-1}N^T K(t) + K(t)M + M^T K(t) - B\}x = 0.$$

由此可知, 对于 $\forall t \in [0, T]$, $K(t)$ 满足如下的矩阵值 Riccati 方程:

$$\dot{K}(t) + K(t)NC^{-1}N^T K(t) + K(t)M + M^T K(t) - B = 0. \tag{R}$$

由此可知, 如果求得方程(R)和终值条件(E-15-8), 就可求得 $K(t)$, 进而得到值函数 v. 此外, 最优的反馈控制为

$$\alpha_0(t) = C^{-1}N^T K(t)x_0(t).$$