清華大学
Tsinghua University

运筹学

第八讲 动态规划

魏韡

2025年5月28日

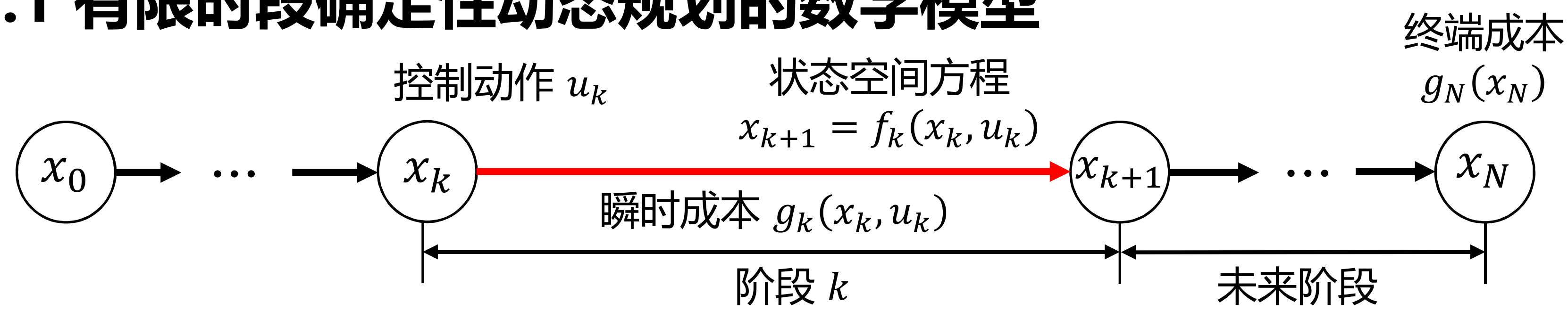
主要内容

1. 确定性动态规划

2. 应用举例

3. 随机动态规划

1.1 有限时段确定性动态规划的数学模型



有限时段确定性动态规划的基本要素

- 状态空间方程

$$x_{k+1}=f_k(x_k,u_k)$$

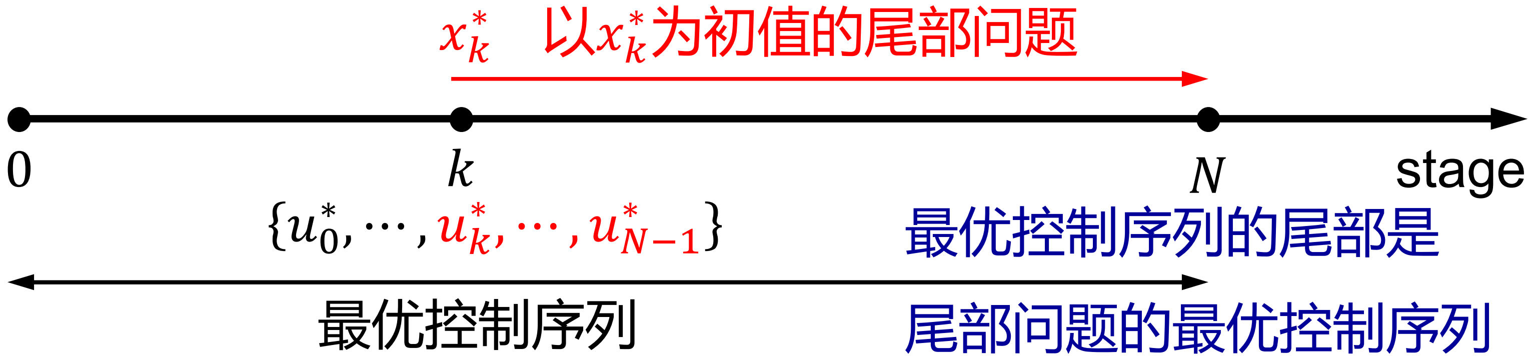
  其中x_k是状态, u_k是控制动作，$U_k(x_k)$是动作的可行集
- 目标函数:

$$J(x_0;u_0,\cdots,\ u_{N-1})=g_N(x_N)+\sum_{k=0}^{N-1}g_k(x_k,u_k)$$

- 给定x_0, 求使目标函数最小的控制序列 $u=\{u_0,\cdots,u_{N-1}\}$
- 确定性优化模型

$$J^*(x_0)=\min_{u}\ \{J(x_0;u)\ |u_k\in U_k(x_k),x_{k+1}=f_k(x_k,u_k),\forall k\}$$

1.2 最优性原理



最优性原理

- 设最优控制序列为 $\{u_0^*, \cdots, u_{N-1}^*\}$，相应状态序列为$\{x_1^*, \cdots, x_N^*\}$. 以 x_k^* 为初值，以阶段 k 到 N 的剩余成本

$$J\left(x_k^*; u_k, \cdots, u_{N-1}\right) = g_k(x_k^*, u_k) + \sum_{m=k+1}^{N-1} g_m(x_m, u_m) + g_N(x_N)$$

为目标函数的尾部最优控制问题，其最优控制序列为原问题最优控制序列的尾部 $\{u_k^*, \cdots, u_{N-1}^*\}$

1.3 逆推法构建最优值函数

基本思想

- 系统当前状态是 x_k，在$U_k(x_k)$中所有可行的动作u_k下系统的下一个状态是 $x_{k+1} = f_k(x_k, u_k)$.
- 假设以x_{k+1}所有可能值为初值的尾部问题最优值(即$J(x_{k+1}^*)$)已知
- 阶段 k 的最优控制 u_k 应使 $g_k(x_k, u_k) + J_{k+1}^*\big(f_k(x_k, u_k)\big)$ 最小

动态规划算法: 构造以x_k为初值的尾部问题的最优值函数 $J_k^*(x_k)$

- 从最后一个时段开始，令 $J_N^*(x_N) = g_N(x_N), \forall x_N$;
- 反向递归 $k = N - 1:0$，令

$$J_k^*(x_k) = \min_{u_k \in U_k(x_k)} \left[g_k(x_k, u_k) + J_{k+1}^*\big(f_k(x_k, u_k)\big)\right], \forall x_k$$

- 最终得到以 x_0 为初值的原问题最优值为

$$J^*(x_0) = J_0^*(x_0)$$

1.4 前推获取最优解

前推获取最优控制序列 [已知 $J_1^*(x_1),\cdots,J_N^*(x_N)$]

- 从初始状态x_0出发 $u_0^* \in \arg\min_{u_0\in U_0(x_0)}\left[g_0(x_0,u_0)+J_1^*\big(f_0(x_0,u_0)\big)\right]$
- 更新系统状态 $x_1^* = f_0(x_0,u_0^*)$，求取最优控制动作

$$u_1^* \in \arg\min_{u_1\in U_1(x_1^*)}\left[g_1(x_1^*,u_1)+J_2^*\big(f_1(x_1^*,u_1)\big)\right]$$

- 更新系统状态$x_2^* = f_2(x_2^*,u_2^*)$，以此类推，直至阶段N

值空间近似: 用参数化 $J_k^\star(x_k,\theta)$ 近似 $J_k^*(x_k)$ (近似动态规划/强化学习)

- 观测系统当前状态 $x_k^\star$，求取控制动作

离线学习

$$u_k^\star \in \arg\min_{u_k\in U_k(x_k^*)}\left[g_k(x_k^\star,u_k)+J_{k+1}^\star\big(f_k(x_k^\star,u_k)\big)\right],$$

在线决策

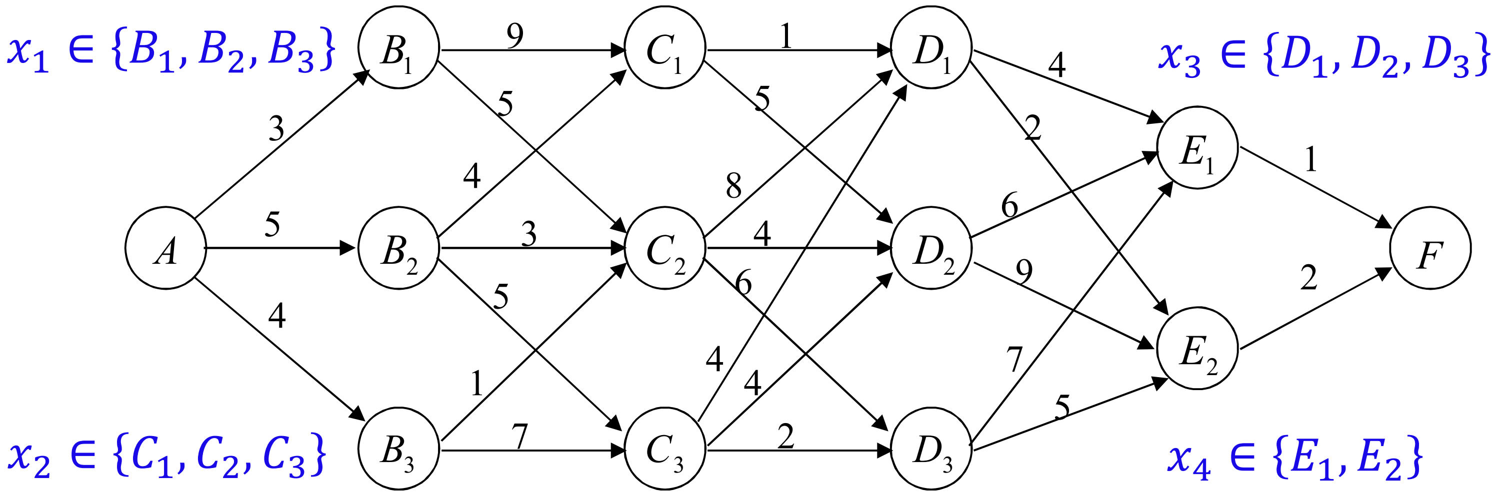
- 系统转移到新的状态$x_{k+1}^\star = f_k(x_k^\star,u_k^\star)$，重复以上步骤

主要内容

1. 确定性动态规划
2. 应用举例
3. 随机动态规划

2.1 最短路径问题

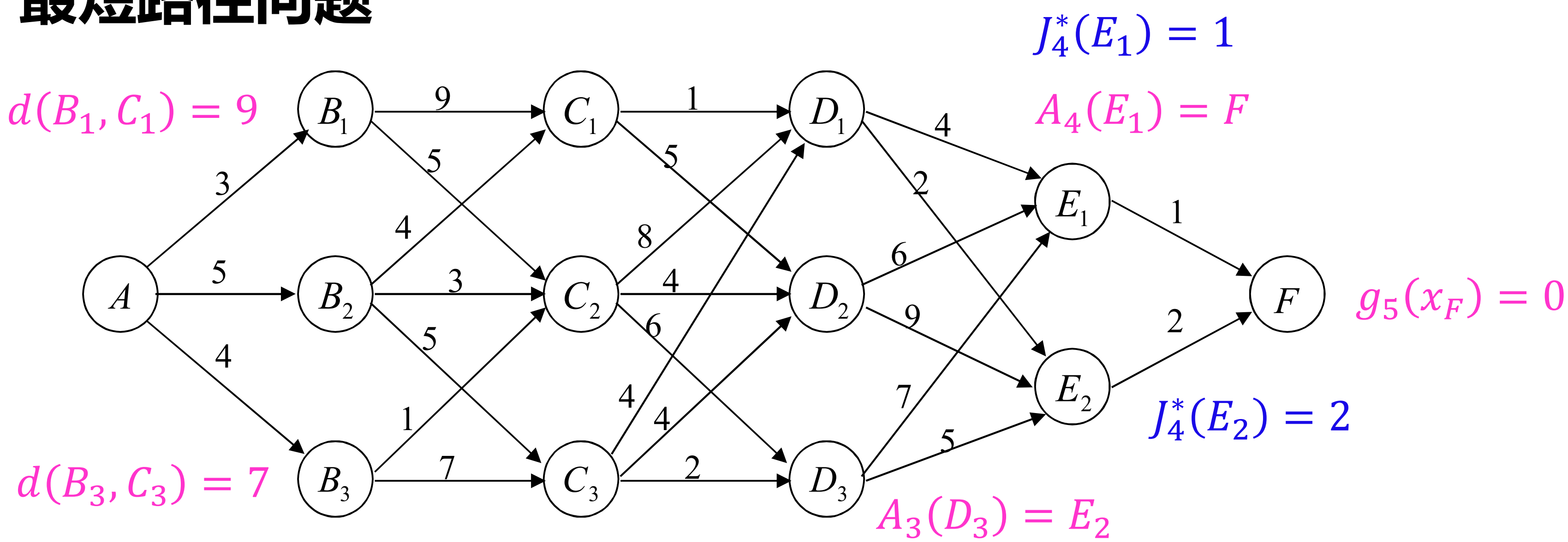
每个路段的长度如图所示. 求从A到F的最短路径。



动态规划建模

- 节点: 状态 x_k
- 边: 状态x_k下的控制动作 u_k
- 拓扑: $x_{k+1} = f_k(x_k, u_k)$
- 边长: 瞬时成本 $g_k(x_k, u_k)$
- 末端成本: $g_N(x_F) = 0$
- 初始状态: x_A
- 目标函数: A至F的总距离

2.1 最短路径问题

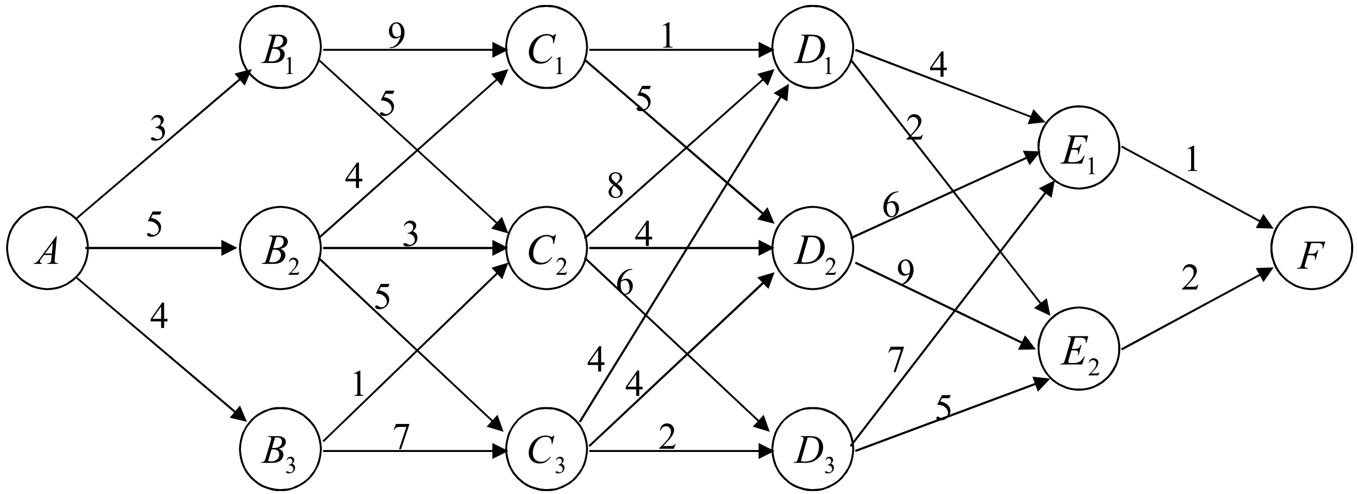


简洁的动态规划表达

- $d(x, y)$: 相邻两点 x 和 y 的距离（瞬时成本）
- $J_k^*(x_k)$: 阶段 k 从点 x 到 F 的最短距离（剩余最优成本）
- 根据最优性原理，从点 x 出发应到达的下一个点是：

$$y^*(x) \in \arg\min_{y \in Y(x)} [d(x, y) + J_{k+1}^*(y)]$$

2.1 最短路径问题



反向递归

$$J_3^*(D_1) = \min\begin{Bmatrix} d(D_1,E_1)+J_4^*(E_1) \\ d(D_1,E_2)+J_4^*(E_2) \end{Bmatrix} = \min\begin{Bmatrix} 4+1 \\ 2+2 \end{Bmatrix} = 4, u_3^*(D_1) = E_2$$

$$J_3^*(D_2) = \min\begin{Bmatrix} d(D_2,E_1)+J_4^*(E_1) \\ d(D_2,E_2)+J_4^*(E_2) \end{Bmatrix} = \min\begin{Bmatrix} 6+1 \\ 9+2 \end{Bmatrix} = 7, u_3^*(D_2) = E_1$$

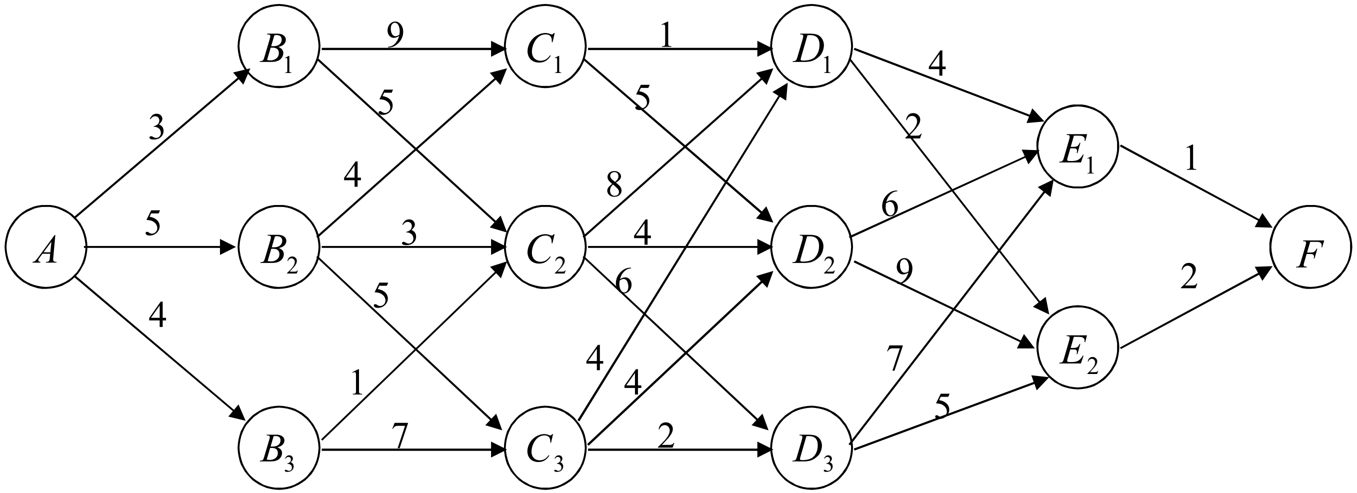
$$J_3^*(D_3) = \min\begin{Bmatrix} d(D_3,E_1)+J_4^*(E_1) \\ d(D_3,E_2)+J_4^*(E_2) \end{Bmatrix} = \min\begin{Bmatrix} 7+1 \\ 5+2 \end{Bmatrix} = 7, u_3^*(D_3) = E_2$$

第 4 阶段

$$J_4^*(E_1) = 1$$

$$J_4^*(E_2) = 2$$

2.1 最短路径问题



反向递归

$$J_2^*(C_1) = \min\begin{Bmatrix} d(C_1,D_1)+J_3^*(D_1) \\ d(C_1,D_2)+J_3^*(D_2) \end{Bmatrix} = \min\begin{Bmatrix} 1+4 \\ 5+7 \end{Bmatrix} = 5, u_2^*(C_1) = D_1$$

$$J_2^*(C_2) = \min\begin{Bmatrix} d(C_2,D_1)+J_3^*(D_1) \\ d(C_2,D_2)+J_3^*(D_2) \\ d(C_2,D_3)+J_3^*(D_3) \end{Bmatrix} = \min\begin{Bmatrix} 8+4 \\ 4+7 \\ 6+7 \end{Bmatrix} = 11, u_2^*(C_2) = D_2$$

$$J_2^*(C_3) = \min\begin{Bmatrix} d(C_3,D_1)+J_3^*(D_1) \\ d(C_3,D_2)+J_3^*(D_2) \\ d(C_3,D_3)+J_3^*(D_3) \end{Bmatrix} = \min\begin{Bmatrix} 4+4 \\ 4+7 \\ 2+7 \end{Bmatrix} = 8, u_2^*(C_3) = D_1$$

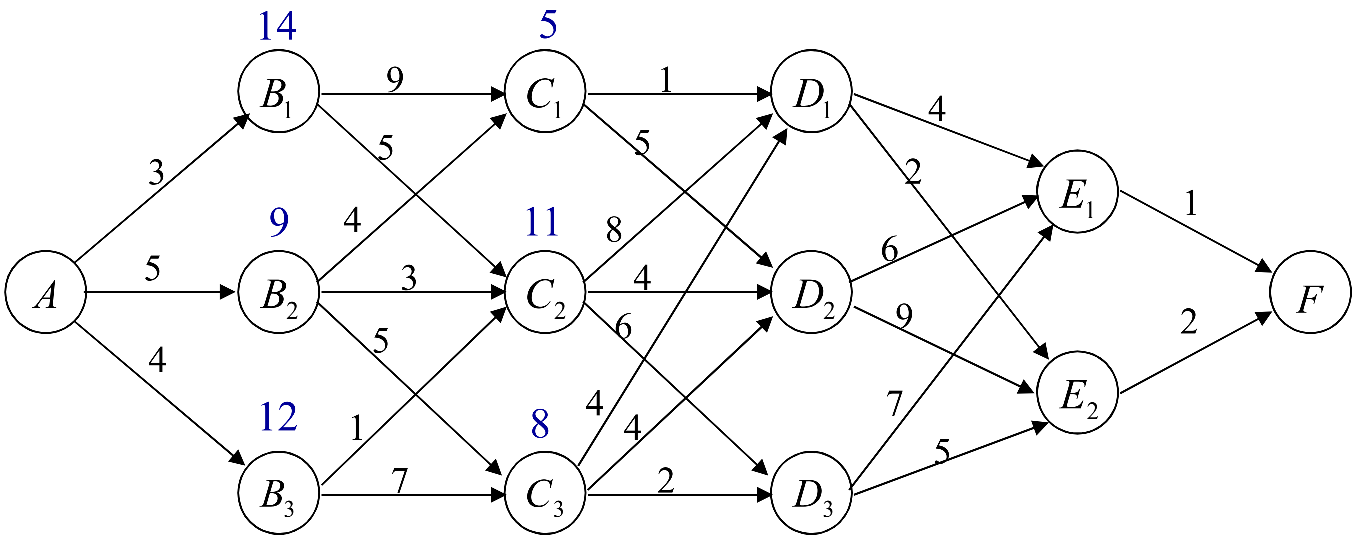
第 3 阶段

$J_3^*(D_1) = 4$

$J_3^*(D_2) = 7$

$J_3^*(D_3) = 7$

2.1 最短路径问题



反向递归

$J_1^*(B_1) = 14, u_1^*(B_1) = C_1$； $J_1^*(B_2) = 9, u_1^*(B_2) = C_1$； $J_1^*(B_3) = 12, u_1^*(B_3) = C_2$

$J_0^*(A) = \min\{3 + 14, 5 + 9, 4 + 12\} = 14, u_0^*(A) = B_2$

前推获取最短路径

- 最短路径: $A \to B_2 \to C_1 \to D_1 \to E_2 \to F \ (l = 14)$
- 最近邻法: $A \to B_1 \to C_2 \to D_2 \to E_1 \to F \ (l = 19)$

2.2 最优控制问题

离散线性系统的初值为x_0，状态方程为

$$x_{k+1} = x_k + u_k, \qquad k = 0,1,\cdots,N-1$$

控制 u 无约束。控制指标为

$$J = \min\left\{\frac{1}{2}cx_N^2 + \frac{1}{2}\sum_{k=0}^{N-1} u_k^2\right\}$$

求 $N = 2$ 时的最优控制序列$[u_0, u_1]$.

解

在最优控制序列 $[u_0, u_1]$ 作用下，系统状态序列 $[x_0, x_1, x_2]$.

当 $N = 2$ 时，系统状态从 x_1 在 u_1 作用下转移到 x_2

$$J(x_1; u_1) = \frac{1}{2}cx_2^2 + \frac{1}{2}u_1^2 = \frac{1}{2}c(x_1 + u_1)^2 + \frac{1}{2}u_1^2 \quad J_1^*(x_1) = \min_{u_1} J(x_1; u_1)$$

$$\frac{\partial J_1}{\partial u_1} = c(x_1 + u_1) + u_1 = 0 \Rightarrow u_1^* = \frac{-cx_1}{1+c}, x_2 = \frac{x_1}{1+c}, J_1^* = \frac{cx_1^2}{2(1+c)}$$

2.2 最优控制问题

离散线性系统的初值为x_0，状态方程为

$$x_{k+1} = x_k + u_k, \qquad k = 0,1,\cdots,N-1$$

控制 u 无约束。控制指标为

$$J = \min\left\{\frac{1}{2}cx_N^2 + \frac{1}{2}\sum_{k=0}^{N-1} u_k^2\right\}$$

求 $N = 2$ 时的最优控制序列$[u_0, u_1]$.

解

当 $N = 1$ 时，系统状态从 x_0 在 u_0 作用下转移到 x_1

$$J_0^*(x_0) = \min_{u_0}\left\{\frac{1}{2}u_0^2 + J_1^*\right\} = \min_{u_0}\left\{\frac{1}{2}u_0^2 + \frac{c}{2(1+c)}(x_0+u_0)^2\right\}$$

$$\frac{\partial J_0}{\partial u_0} = \frac{c(x_0+u_0)}{(1+c)} + u_0 = 0 \Rightarrow u_0^* = \frac{-cx_0}{1+2c}, x_1 = \frac{1+c}{1+2c}x_0$$

最优控制与最优值: $u_0^* = \dfrac{-cx_0}{1+2c}, u_1^* = \dfrac{-cx_0}{1+2c}, J_0^* = \dfrac{cx_0^2}{2(1+2c)}$

2.3 线性规划

用动态规划法求解以下线性规划

$$\begin{aligned} \max \quad & -2x_1 + 3x_2 \\ \text{s.t.} \quad & -3x_1 - 2x_2 \le 2 \\ & -2x_1 + 6x_2 \le 3 \\ & x_1, x_2 \in [0,1] \end{aligned}$$

解

将该线性规划视为两阶段决策过程

$$\max_{x_1 \in [0,1]} \left\{ -2x_1 + \max_{x_2 \in [0,1] \cap D(x_1)} 3x_2 \right\}$$

$$D\left(x_1\right) = \left\{ x_2 \,\middle|\, 1 - \frac{3x_1}{2} \le x_2 \le \frac{1}{2} + \frac{x_1}{3} \right\}$$

$D(x_1) \neq \emptyset$ 要求 $x_1 \geq 3/11$

2.3 线性规划

解

第二阶段问题以第一阶段变量 x_1 为参数

$$\max \quad 3x_2$$
$$\text{s.t. } 1-\frac{3x_1}{2} \le x_2 \le \frac{1}{2}+\frac{x_1}{3},\ \ 0 \le x_2 \le 1$$

由于 $x_1 \le 1$, x_2 的最大值即区间上界 $1/2+x_1/3$, 代入第一阶段问题

$$\max_{x_1\in[0,1]}\left\{-2x_1+\max_{x_1\in[3/11,1]}\left\{\frac{3}{2}+x_1\right\}\right\}=\max_{x_1\in[3/11,1]}\left\{\frac{3}{2}-x_1\right\}=\frac{27}{22}\quad (x_1,x_2)=\left(\frac{3}{11},\frac{13}{22}\right)$$

- 本课程学过哪几种求解线性规划的方法?
- 哪种计算效率最高?为什么?
- 静态优化和动态优化的本质区别是什么?

2.4 抢险救灾问题

A县220kV变电站在地震中损坏，电网公司组织了3支抢修队从附近B县出发分3路前往A县修复该变电站。任何一支抢修队若能顺利到达均可完成抢修。由于通往A县的3条道路均有不同程度损毁，3支抢修队不能到达变电站的概率分别为0.4，0.6和0.8，都不能到达变电站的概率为0.192。为了提高成功率，电网公司将B县供电局2台道路清障车配备给抢修队使用。3支抢修队配备不同数量的道路清障车后，不能到达变电站的概率如表所示

求清障车分配方案，使得抢修成功概率最大。记x_i是分配到第i支抢修队的道路清障车数量，$p_i(x_i)$是对应的不能到达变电站的概率，最优分配方案是以下问题的最优解

$$\begin{aligned} \min \quad & p_1(x_1)p_2(x_2)p_3(x_3) \\ \text{s.t.} \quad & x_1 + x_2 + x_3 = 2 \\ & x_1, x_2, x_3 \in \{0,1,2\} \end{aligned}$$

用动态规划求解该问题。

未达概率		1队	2队	3队
道路清障车数量	0	0.40	0.60	0.80
	1	0.20	0.40	0.50
	2	0.15	0.20	0.30

阶段：$k = 1, 2, 3$

状态s_k：可供分配的清障车数

动作x_k：分给k队的清障车数

2.4 抢险救灾问题

$v_k^*(s_k)$: 第 k 队有s_k台清障车可供分配时后k队都不能到达的最小概率

$$v_k^*\left(s_k\right)=\min_{x_k\in\{0,\cdots,s_k\}} p_k\left(x_k\right)\cdot v_{k+1}^*\left(s_k-x_k\right)$$

$$v_3^*\left(s_3\right)=\min_{x_3\in\{0,\cdots,s_3\}} p_3\left(x_3\right)$$

未达概率		1队	2队	3队
道路清障车数量	0	0.40	0.60	0.80
	1	0.20	0.40	0.50
	2	0.15	0.20	0.30

$Q_k(s_k,x_k)$: 从s_k台车中分配给 k 队x_k台，后k队都不能到达的最小概率

s_3	$v_3^*(s_3)$	x_3^*
0	0.8	0
1	0.5	1
2	0.3	2

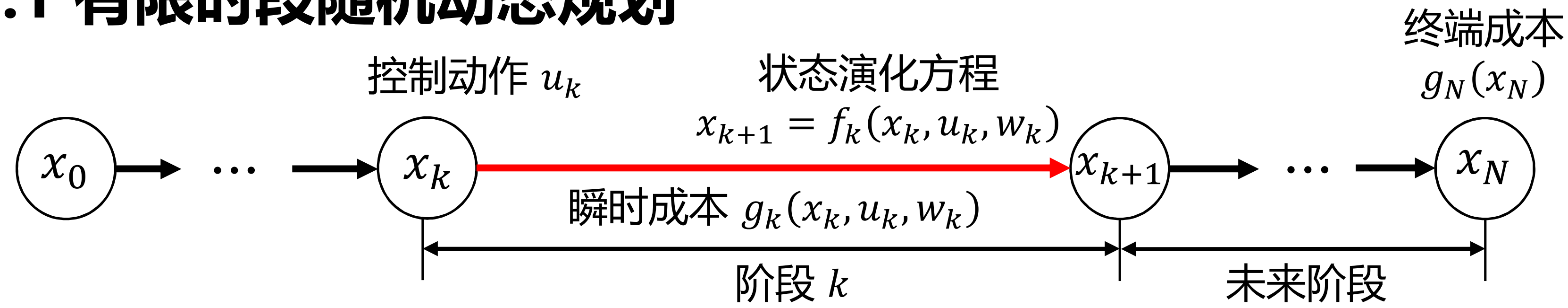
s_2	$Q_2(s_2,x_2)=p_2(x_2)\cdot v_3^*(s_2-x_2)$			$v_2^*(s_2)$	x_2^*
	$x_2=0$	$x_2=1$	$x_2=2$		
0	0.6·0.8	×	×	0.48	0
1	0.6·0.5	0.4 · 0.8	×	0.30	0
2	0.6·0.3	0.4 · 0.5	0.2 · 0.8	0.16	2

s_1	$Q_1(s_1,x_1)=p_1(x_1)\cdot v_2^*(s_1-x_1)$			$v_1^*(s_1)$	x_1^*
	$x_1=0$	$x_1=1$	$x_1=2$		
2	0.4 · 0.16	0.2 · 0.3	0.15 · 0.48	0.060	1

最优分配方案：
1队：1台
2队：0台
3队：1台

主要内容

3.1 有限时段随机动态规划



有限时段随机动态规划的基本要素

- 状态演化方程: $x_{k+1} = f_k(x_k, u_k, w_k), w_k \sim P_k(\cdot|x_k, u_k)$
  其中 x_k 是系统状态, u_k 是动作，$U_k(x_k)$ 是动作的可行集
- 闭环反馈控制策略:
  $$\pi = \{\mu_0, \cdots, \mu_{N-1}\}, u_k = \mu_k(x_k) \in U_k(x_k), k = 0{:}\, N-1$$
- 以x_0为初值在控制策略 π 下的期望成本
  $$J_\pi(x_0) = \mathbb{E}\left\{g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, \mu_k(x_k), w_k)\right\}$$
- 最优策略与最优值
  $$J_{\pi^*}(x_0) = \min_{\pi\in\Pi} J_\pi(x_0) = J^*(x_0)$$

3.1 有限时段随机动态规划

与确定性问题有何不同?

确定性问题

- 数值优化问题（开环）

$$\min_{x,u} \; J(x_0,u)$$

$$\text{s.t. } x_{k+1} = f_k(x_k,u_k), u_k \in U_k(x_k)$$

随机问题-函数优化

- 策略是控制规律（函数）

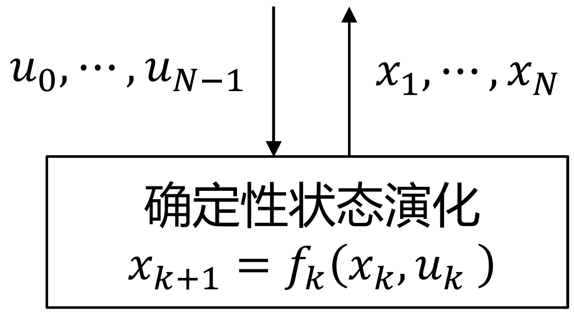
$$\pi = \{\mu_0,\cdots,\mu_{N-1}\}$$

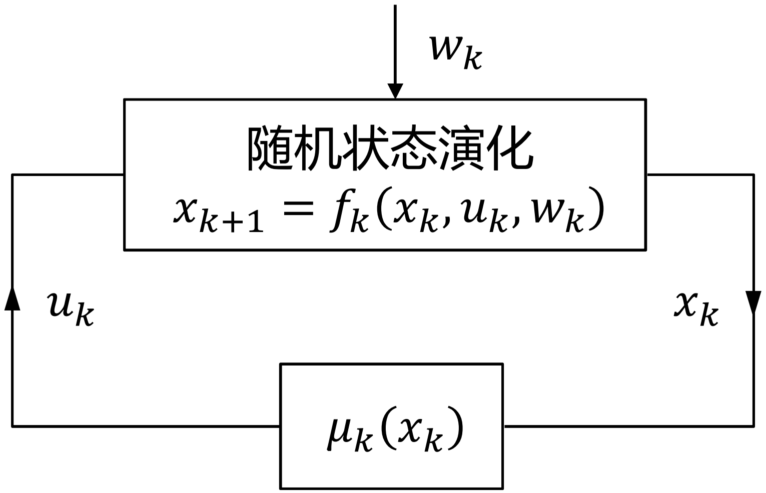
- 可行策略集

$$\pi \in \Pi \Rightarrow \mu_k(x_k) \in U_k(x_k)$$

- 泛函优化（闭环）

$$J_{\pi^*}(x_0) = \min_{\pi\in\Pi} J_\pi(x_0)$$





3.2 随机动态规划算法

动态规划算法: 构造以x_k为初值的尾部问题的最优值函数 $J_k^*(x_k)$

- 从最后一个时段开始，令 $J_N^*(x_N) = g_N(x_N), \forall x_N;$
- 反向递归 For $k = N-1:0$, 令

$$J_k^*(x_k) = \min_{u_k \in U_k(x_k)} \mathbb{E}_{w_k}\left[g_k(x_k, u_k, w_k) + J_{k+1}^*\left(f_k(x_k, u_k, w_k)\right)\right], \forall x_k$$

- 最终得到以 x_0 为初值的原问题最优值为 $J^*(x_0) = J_0^*(x_0)$

前推获取最优控制序列 [已知 $J_1^*(x_1), \cdots, J_N^*(x_N)$]

- 从 x_0 出发，令 $k = 0, \cdots, N-1$, 观测状态 x_k,
- 选择使期望值最小的动作

$$\mu_k(x_k) \in \arg\min_{u_k \in U_k(x_k)} \mathbb{E}_{w_k}\left\{g_k(x_k, u_k, w_k) + J_{k+1}^*\left(f_k(x_k, u_k, w_k)\right)\right\}$$

- 近似方法: 用参数化的 $\tilde{J}_k$ 近似 J_k^*, 对期望$\mathbb{E}\{\cdot\}$做近似

3.3 Q-因子

定义

- 随机动态规划的 Q-因子

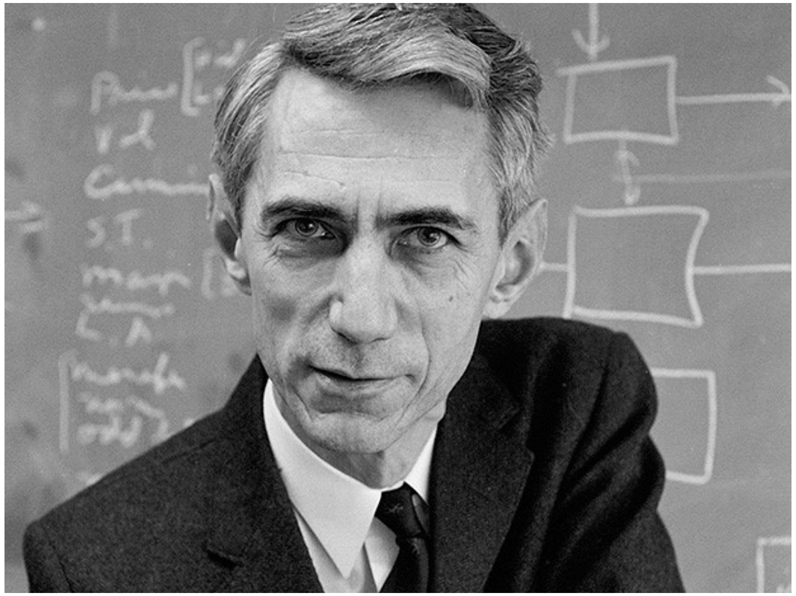
$$Q_k^*(x_k,u_k)=\mathop{\mathbb{E}}_{w_k}\left\{g_k(x_k,u_k,w_k)+J_{k+1}^*\big(f_k(x_k,u_k,w_k)\big)\right\}$$

- Q-因子和值函数的关系

$$J_k^*(x_k)=\min_{u_k\in U_k(x_k)}Q_k^*(x_k,u_k)$$

Q-因子的递归关系

$$Q_k^*(x_k,u_k)=\mathop{\mathbb{E}}_{w_k}\left\{\begin{matrix} g_k(x_k,u_k,w_k)+ \\ \min\limits_{u_{k+1}\in U_{k+1}(f_k(x_k,u_k,w_k))}Q_{k+1}^*\big(f_k(x_k,u_k,w_k),u_{k+1}\big)\end{matrix}\right\}$$



We know the past but cannot control it,

We control the future but cannot know it.

Claude Shannon