CSE 691 习题

德梅萃 · P. 博赛卡斯（Dimitri P. Bertsekas）著
李宇超（Yuchao Li）译

习题 1 [Ber17, 习题 2.1] 考虑由节点（node）$1, \ldots, 6$ 以及连接它们的边（edge）构成的图（graph）如图 1所示。请采用动态规划算法计算节点 $1, \ldots, 5$ 到节点 6 的最短路径。采用编程或者手算方式均可。*提示*：在此问题中，阶段数目 N 应当设为多少？每阶段中应当包含哪些状态呢？

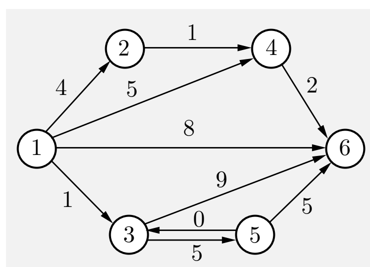

图 1: 习题 1 中涉及的图。标注于边旁的数值表示边长。

习题 2 考虑习题 1 中的最短路径问题。请采用策略前展算法（rollout）给出该问题的近似解。*提示*：可以采用贪心策略作为策略前展中的启发式方法。例如，当处于节点 3 时，可选的下一个节点包括了节点 5 和节点 6。贪心策略比较前往这两个节点的边的长度（即 5 和 9），并选择前往边长较短的后续节点（即对应于边长 5 的节点 5）。

习题 3 [Ber17, 例 3.5.1] 某智力竞赛共有 N 道题目，记作题目 $1, 2, \ldots, N$。参赛者可以自由选择其答题次序，当答对题目 i 时，参赛者可以得 R_i 的奖励，并继续回答后续问题。一旦某题目回答错误，参赛者便不可以回答后续问题。小明答对题目 i 的概率为 p_i，那么他应当如何安排答题顺序从而使他期望的收益最大化呢？请采用动态规划给出该问题的解析解。*提示*：在此问题中，什么是状态、控制和系统呢？

习题 4 [Ber20, 习题 1.5] 本习题的目的是通过一维的线性二次型问题

来体现策略迭代与牛顿法的等效性。在此问题中，系统为 $f(x,u)=x+bu$，阶段费用为 $g(x,u)=x^2+ru^2$，其中 $b\neq 0$ 且 $r>0$。

(a) 请验证贝尔曼方程

$$Kx^2=\min_u\left[x^2+ru^2+K(x+bu)^2\right]$$

可以写作等效形式 $H(K)=0$，其中

$$H(K)=K-\frac{rK}{r+b^2K}-1.$$

(b) 现考虑策略迭代算法

$$K_k=\frac{1+rL_k^2}{1-(1+bL_k)^2},$$

其中

$$L_{k+1}=-\frac{bK_k}{r+b^2K_k},$$

且 $\mu(x)=L_0x$ 为起始策略。证明该算法等效于通过牛顿法

$$K_{k+1}=K_k-\left(\frac{\partial H(K_k)}{\partial K}\right)^{-1}H(K_k)$$

求解贝尔曼方程 $H(K)=0$。

(c) 请将上述结论推广到系统为 $f(x,u)=ax+bu$，阶段费用为 $g(x,u)=qx^2+ru^2$，其中 $a,b\neq 0$ 且 $q,r>0$ 的情况。

参考文献

[Ber17] Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific, 4 edition, 2017.

[Ber20] Dimitri P. Bertsekas. *Rollout, Policy Iteration, and Distributed Reinforcement Learning*. Athena Scientific Belmont, MA, 2020.