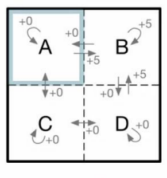
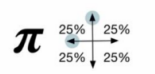
作业3

例题：一个简单有限马尔可夫决策过程

$$v_\pi(s)=\sum_a \pi(a|s)\sum_{s'}\sum_r p(s',r|s,a)[r+\gamma v_\pi(s')] \text{ for all } s\in \mathcal{S}$$



- 状态空间：$\mathcal{S}=\{s_A,s_B,s_C,s_D\}$
- r=0.7，奖励如图所示
- 问题：求解状态价值函数（即计算每个状态对应的价值）
- 解法：解决线性方程组即可(n个未知量，n个线性方程组)



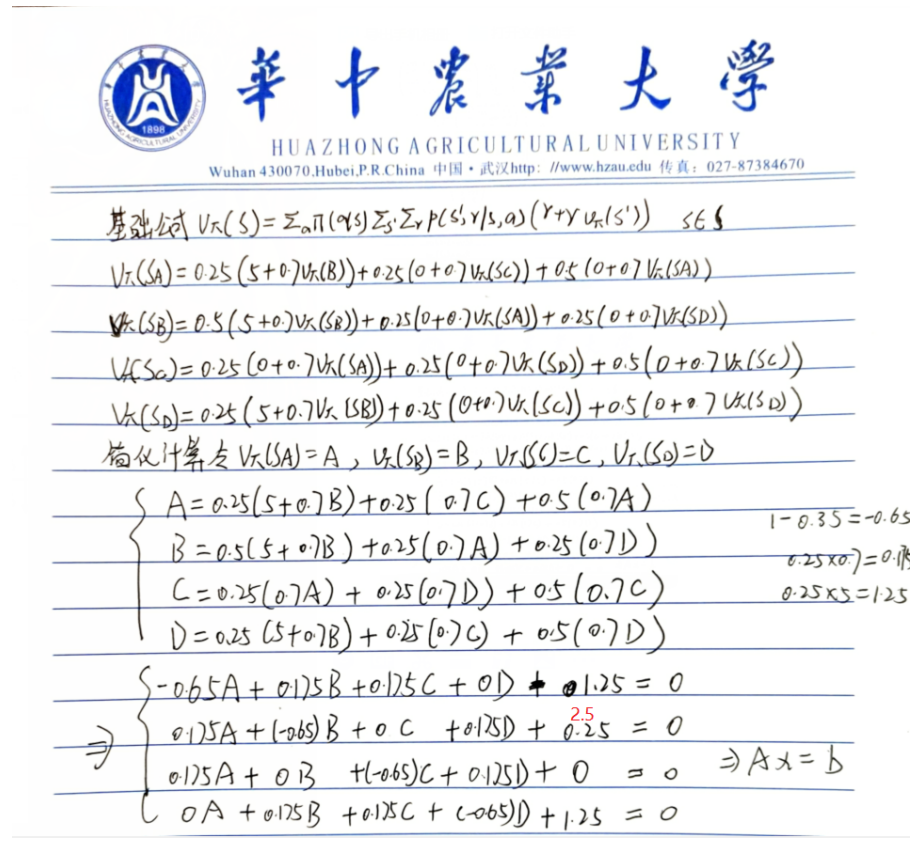
$$\begin{cases} v_\pi(s_A)=0.25[5+0.7v_\pi(s_B)]+0.25[0+0.7v_\pi(s_C)]+0.5[0+0.7v_\pi(s_A)] \\ v_\pi(s_B)=0.5[5+0.7v_\pi(s_B)]+0.25[0+0.7v_\pi(s_A)]+0.25[0+0.7v_\pi(s_D)] \\ v_\pi(s_C)=0.25[0+0.7v_\pi(s_A)]+0.25[0+0.7v_\pi(s_D)]+0.5[0+0.7v_\pi(s_C)] \\ v_\pi(s_D)=0.25[0+0.7v_\pi(s_B)]+0.25[0+0.7v_\pi(s_C)]+0.5[0+0.7v_\pi(s_D)] \end{cases}$$

$$v_\pi(s)=\sum_a \pi(a|s)\sum_{s'}\sum_r p(s',r|s,a)\left[r+\gamma v_\pi(s')\right]$$

$$\pi,p,\gamma \longrightarrow \boxed{\text{Linear System Solver}} \longrightarrow v_\pi$$

$$\begin{cases} v_\pi(s_C)=2.2 \\ v_\pi(s_D)=4.2 \end{cases}$$

卢东黎 2021317220603



- 题目：智能体与环境进行一个回合，获得如下奖励序列（$H=5$），

$$R_1=-1,R_2=2,R_3=6,R_4=3,R_5=2$$

- 问题：$\gamma=0.5$，计算G_0，G_1，$G_2,\ldots,G_5$ (提示：从后往前计算)