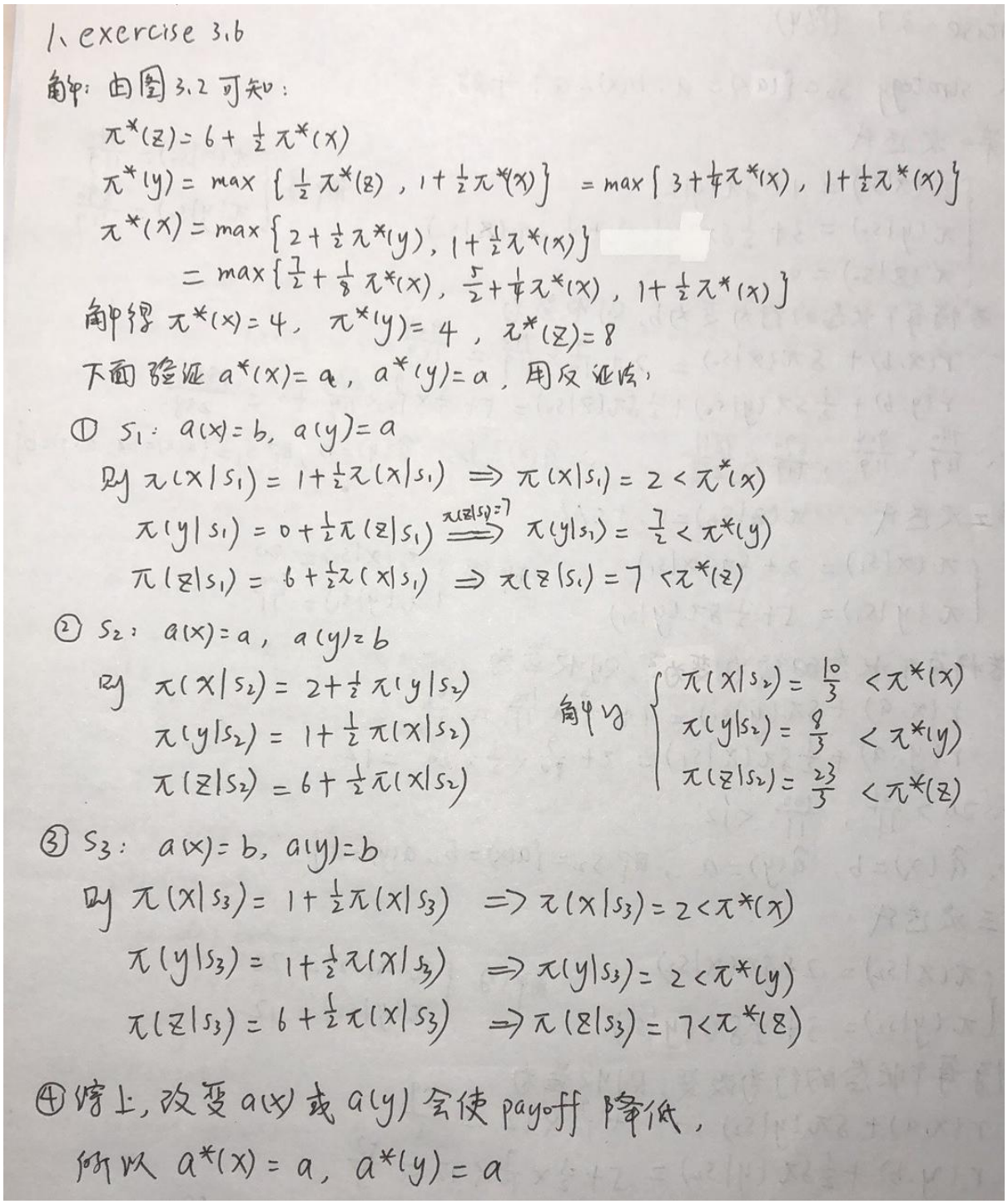
高级算法作业 2

TRY 计算机科学与技术

1. Exercise 3.6



1、exercise 3.6

解：由图 3.2 可知：

$\pi^*(z) = 6 + \frac{1}{2}\pi^*(x)$

$\pi^*(y) = \max\{\frac{1}{2}\pi^*(z), 1+\frac{1}{2}\pi^*(x)\} = \max\{3+\frac{1}{4}\pi^*(x), 1+\frac{1}{2}\pi^*(x)\}$

$\pi^*(x) = \max\{2+\frac{1}{2}\pi^*(y), 1+\frac{1}{2}\pi^*(x)\}$

$= \max\{\frac{7}{2}+\frac{1}{8}\pi^*(x), \frac{5}{2}+\frac{1}{4}\pi^*(x), 1+\frac{1}{2}\pi^*(x)\}$

解得 $\pi^*(x)=4$，$\pi^*(y)=4$，$\pi^*(z)=8$

下面验证 $a^*(x)=a$，$a^*(y)=a$，用反证法：

① s_1：$a(x)=b$，$a(y)=a$

则 $\pi(x|s_1) = 1+\frac{1}{2}\pi(x|s_1) \Rightarrow \pi(x|s_1) = 2 < \pi^*(x)$

$\pi(y|s_1) = 0+\frac{1}{2}\pi(z|s_1) \xrightarrow{\pi(z|s_1)=7} \pi(y|s_1) = \frac{7}{2} < \pi^*(y)$

$\pi(z|s_1) = 6+\frac{1}{2}\pi(x|s_1) \Rightarrow \pi(z|s_1) = 7 < \pi^*(z)$

② s_2：$a(x)=a$，$a(y)=b$

则 $\pi(x|s_2) = 2+\frac{1}{2}\pi(y|s_2)$

$\pi(y|s_2) = 1+\frac{1}{2}\pi(x|s_2)$

$\pi(z|s_2) = 6+\frac{1}{2}\pi(x|s_2)$

解得 $\begin{cases}\pi(x|s_2) = \frac{10}{3} < \pi^*(x)\\ \pi(y|s_2) = \frac{8}{3} < \pi^*(y)\\ \pi(z|s_2) = \frac{23}{3} < \pi^*(z)\end{cases}$

③ s_3：$a(x)=b$，$a(y)=b$

则 $\pi(x|s_3) = 1+\frac{1}{2}\pi(x|s_3) \Rightarrow \pi(x|s_3) = 2 < \pi^*(x)$

$\pi(y|s_3) = 1+\frac{1}{2}\pi(x|s_3) \Rightarrow \pi(y|s_3) = 2 < \pi^*(y)$

$\pi(z|s_3) = 6+\frac{1}{2}\pi(x|s_3) \Rightarrow \pi(z|s_3) = 7 < \pi^*(z)$

④ 停止，改变 $a(x)$ 或 $a(y)$ 会使 payoff 降低，

所以 $a^*(x)=a$，$a^*(y)=a$

2. Exercise 3.7 (P64)

2、exercise 3.7 (用 Pb1 Figure 3.2 完成)

解：(1) 从 strategy $s_0 = \{a(x)=a, a(y)=a\}$ 开始。$\delta = \frac{1}{2}$

①第一次迭代：

$$\begin{cases} \pi(x|s_0) = 2 + \delta\pi(y|s_0) \\ \pi(y|s_0) = 0 + \delta\pi(z|s_0) \\ \pi(z|s_0) = 6 + \delta\pi(x|s_0) \end{cases}$$ 解得 $$\begin{cases} \pi(x|s_0) = 4 \\ \pi(y|s_0) = 4 \\ \pi(z|s_0) = 8 \end{cases}$$

若将每个状态的行为改变，则收益为

$r(x,b) + \delta\pi(x|s_0) = 1 + \frac{1}{2}\times 4 = 3$

$r(y,b) + \delta\pi(x|s_0) = 1 + \frac{1}{2}\times 4 = 3$

∵ $3<4$, $3<4$

∴ 改变 action 不会导致更好的 payoff ∴ $\hat{a}(x)=a$, $\hat{a}(y)=a$

∴ $s^* = \{a(x)=a, a(y)=a\}$ 是 optimal strategy

(2) 从 strategy $s_0 = \{a(x)=a, a(y)=b\}$ 开始。$\delta = \frac{1}{2}$

①第一次迭代：

$$\begin{cases} \pi(x|s_0) = 2 + \delta\pi(y|s_0) \\ \pi(y|s_0) = 1 + \delta\pi(x|s_0) \\ \pi(z|s_0) = 6 + \delta\pi(x|s_0) \end{cases}$$ 解得 $$\begin{cases} \pi(x|s_0) = \frac{10}{3} \\ \pi(y|s_0) = \frac{8}{3} \\ \pi(z|s_0) = \frac{23}{3} \end{cases}$$

若将每个状态的行为改变，则收益为

$r(x,b) + \delta\pi(x|s_0) = 1 + \frac{1}{2}\times\frac{10}{3} = \frac{8}{3}$

$r(y,a) + \delta\pi(z|s_0) = 0 + \frac{1}{2}\times\frac{23}{3} = \frac{23}{6}$

∵ $\frac{8}{3} < \frac{10}{3}$, $\frac{23}{6} > \frac{8}{3}$ ∴ $\hat{a}(x)=a$, $\hat{a}(y)=a$, 即 $s_1 = \{a(x)=a, a(y)=a\}$

②第二次迭代：

由(1)可知，$\hat{a}(x)=a$, $\hat{a}(y)=a$ 是 optimal strategy

∴ $s^* = \{a(x)=a, a(y)=a\}$ 是 optimal strategy