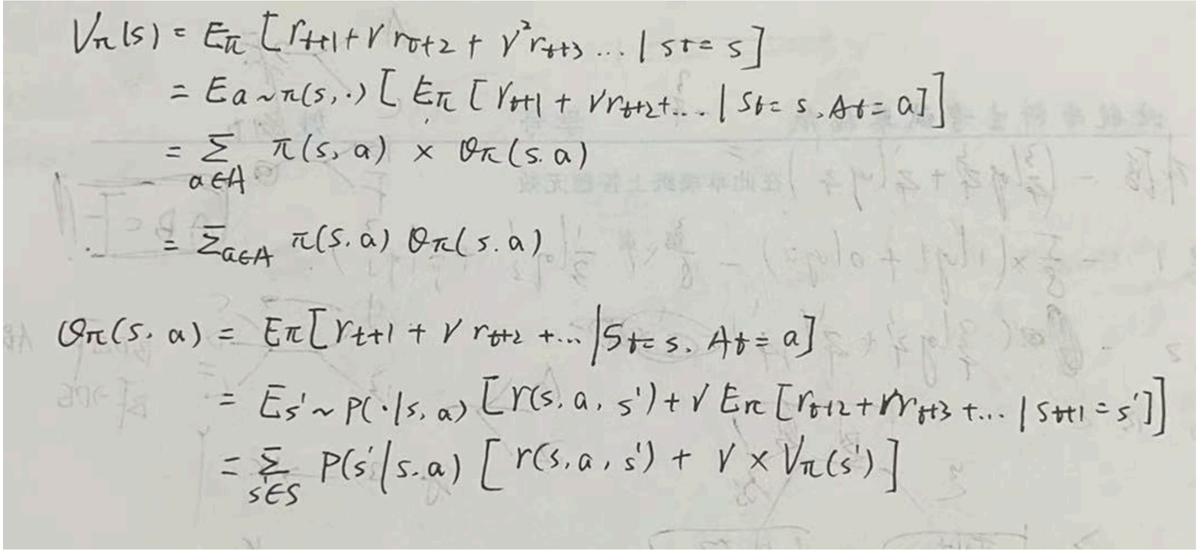
A4-QA

Q1

- G0 = 4
- G1 = 6
- G2 = 8
- G3 = 4
- G4 = 2
- G5 = 0

Q2

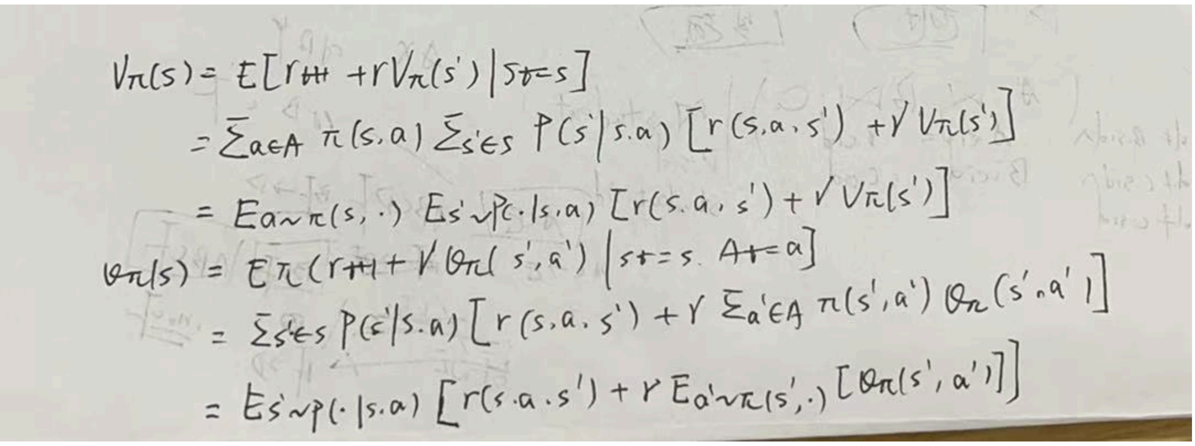
1



$$\begin{aligned}
V_\pi(s) &= E_\pi[r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} \dots | s_t = s] \\
&= E_{a\sim\pi(s,\cdot)}[E_\pi[r_{t+1} + \gamma r_{t+2} + \dots | S_t = s, A_t = a]] \\
&= \sum_{a\in A} \pi(s,a) \times Q_\pi(s.a) \\
&= \Sigma_{a\in A} \pi(s.a)\, Q_\pi(s.a)
\end{aligned}$$

$$\begin{aligned}
Q_\pi(s.a) &= E_\pi[r_{t+1} + \gamma r_{t+2} + \dots | S_t = s. A_t = a] \\
&= E_{s'\sim P(\cdot|s,a)}[r(s.a,s') + \gamma E_\pi[r_{t+2} + \gamma r_{t+3} + \dots | S_{t+1} = s']] \\
&= \sum_{s'\in S} P(s'|s.a)[r(s,a,s') + \gamma \times V_\pi(s')]
\end{aligned}$$

2

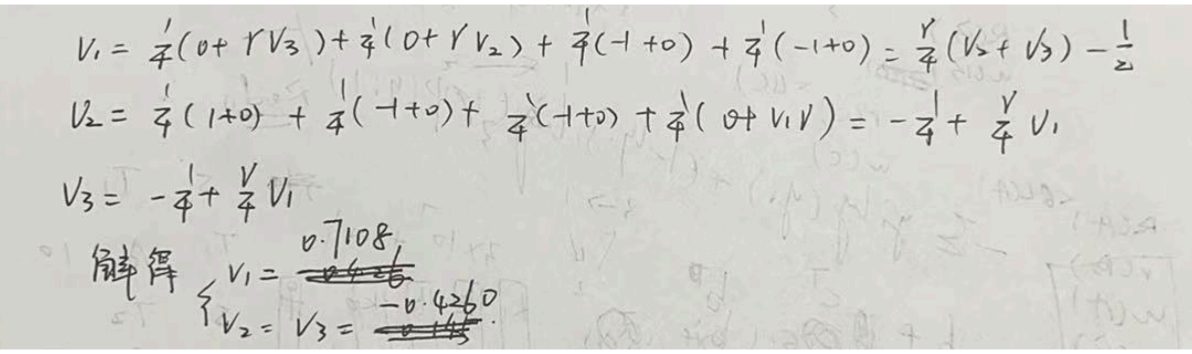


$$\begin{aligned}
V_\pi(s) &= E[r_{t+1} + \gamma V_\pi(s') | S_t = s] \\
&= \Sigma_{a\in A} \pi(s,a) \Sigma_{s'\in S} P(s'|s.a)[r(s,a,s') + \gamma V_\pi(s')] \\
&= E_{a\sim\pi(s,\cdot)} E_{s'\sim P(\cdot|s,a)}[r(s.a,s') + \gamma V_\pi(s')]
\end{aligned}$$

$$\begin{aligned}
Q_\pi(s) &= E_\pi(r_{t+1} + \gamma Q_\pi(s',a')) | s_t = s. A_t = a] \\
&= \Sigma_{s'\in S} P(s'|s.a)[r(s,a.s') + \gamma \Sigma_{a'\in A} \pi(s',a') Q_\pi(s',a')] \\
&= E_{s'\sim P(\cdot|s.a)}[r(s.a.s') + \gamma E_{a'\sim\pi(s',\cdot)}[Q_\pi(s',a')]]
\end{aligned}$$

Q3

1



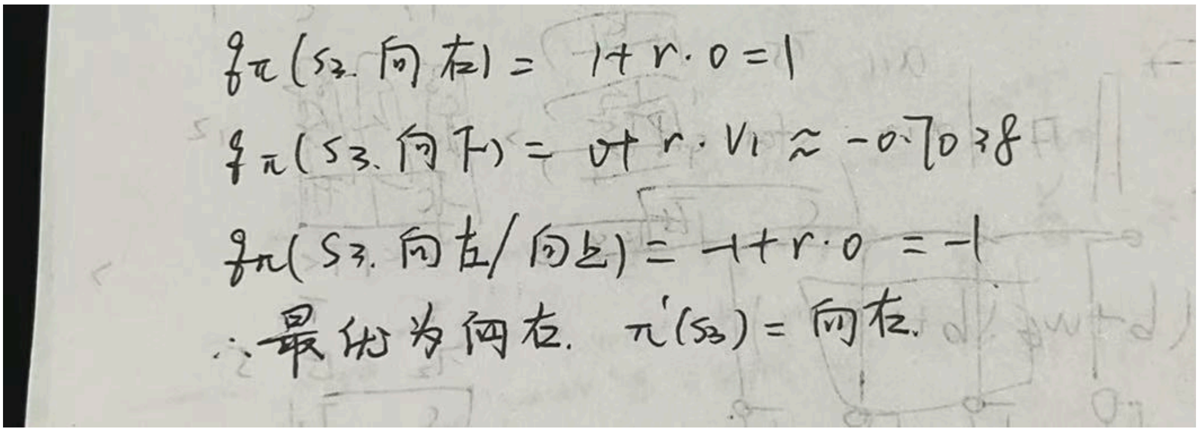
$V_1 = \frac{1}{4}(0+\gamma V_3)+\frac{1}{4}(0+\gamma V_2)+\frac{1}{4}(-1+0)+\frac{1}{4}(-1+0)=\frac{\gamma}{4}(V_2+V_3)-\frac{1}{2}$

$V_2 = \frac{1}{4}(1+0)+\frac{1}{4}(-1+0)+\frac{1}{4}(-1+0)+\frac{1}{4}(0+V_1\gamma)=-\frac{1}{4}+\frac{\gamma}{4}V_1$

$V_3 = -\frac{1}{4}+\frac{\gamma}{4}V_1$

解得 $\begin{cases} V_1 = \overset{0.7108}{\cancel{0.426}} \\ V_2 = V_3 = \overset{-0.4260}{\cancel{0.145}}. \end{cases}$

2



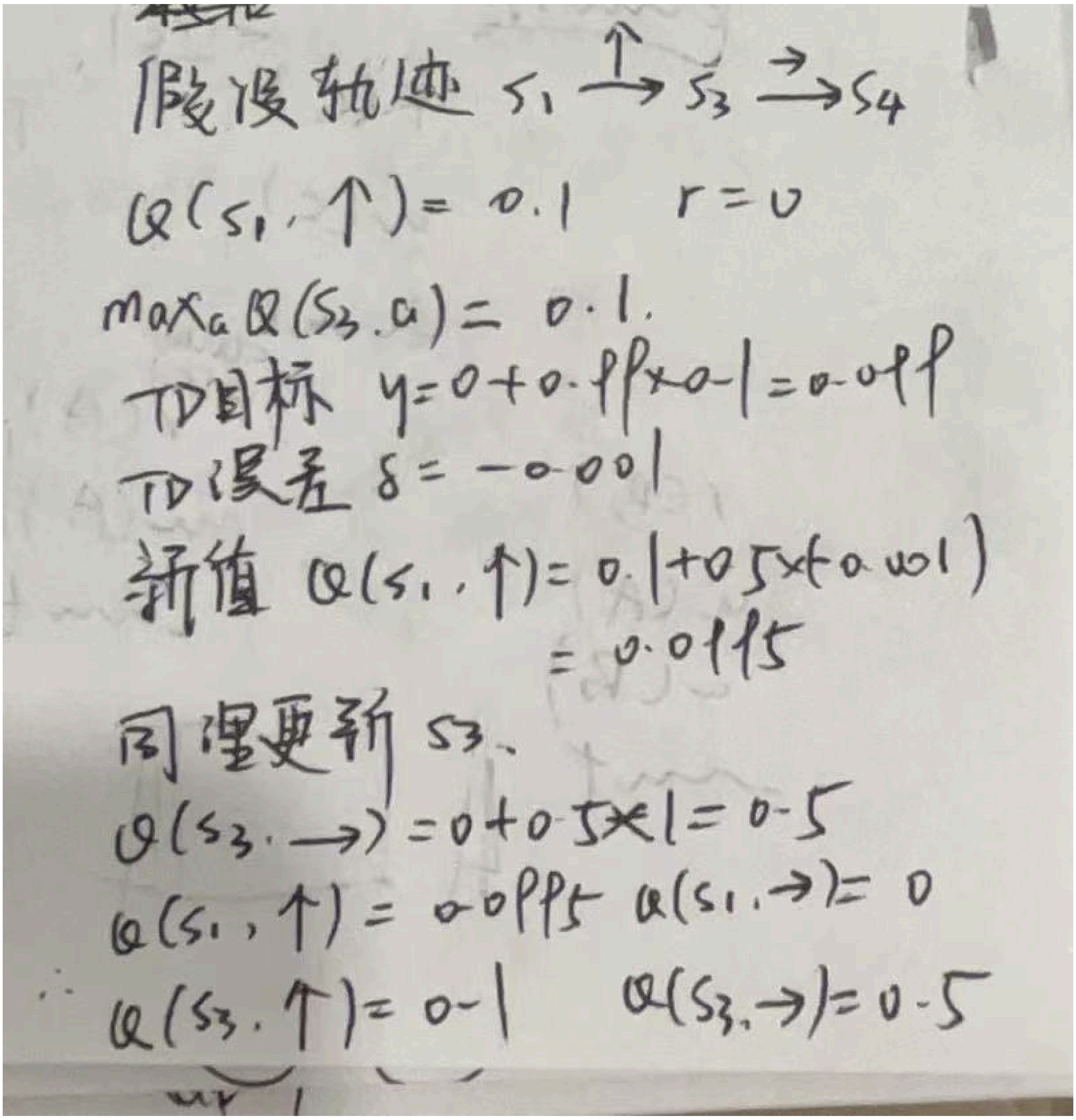
q_π(S3. 向右) = $1+r\cdot 0=1$

q_π(S3. 向下) = $0+r\cdot V_1 \approx -0.7038$

q_π(S3. 向左/向上) = $-1+r\cdot 0 = -1$

∴最优为向右. $\pi'(S_3)$ = 向右.

3



假设轨迹 $s_1 \xrightarrow{\uparrow} s_3 \xrightarrow{\rightarrow} s_4$

$Q(s_1, \uparrow) = 0.1$ $r = 0$

$\max_a Q(s_3, a) = 0.1.$

TD目标 $y = 0 + 0.99 \times 0.1 = 0.099$

TD误差 $\delta = -0.001$

新值 $Q(s_1, \uparrow) = 0.1 + 0.5 \times (-0.001)$

$= 0.0995$

同理更新 s3。

$Q(s_3, \rightarrow) = 0 + 0.5 \times 1 = 0.5$

$Q(s_1, \uparrow) = 0.0995$ $Q(s_1, \rightarrow) = 0$

$\therefore Q(s_3, \uparrow) = 0.1$ $Q(s_3, \rightarrow) = 0.5$

最后策略为 s_1 时向上，s_3 时向右