Chapter 4

- Bellman optimality equations

$$V_*(s) = \max_{q} E \left[R_{++} + \gamma V_* \right]$$

$$(4.1) V_{*}(s) = \max_{q} E \left[R_{t+1} + \gamma V_{*}(s_{t+1}) \middle| S_{t} = s, A_{t} = a \right]$$

$$= \max_{\alpha} \sum_{s',r} \rho(s',r|s,\alpha) \left[r + \gamma V_*(s')\right]$$

$$(4.2) \quad g_{*}(s,a) = E \left[R_{++} + y \max_{a'} g_{*}(s_{++}, a') \middle| S_{+} = s, A_{+} = a \right]$$

$$= \sum_{s',r} p(s',r|s,a) \left[r + \gamma \stackrel{\text{max}}{a'} q_*(s',a')\right]$$

$$(4.4) \quad V_{\pi}(s) = \sum_{\alpha} \pi(\alpha|s) \sum_{s,r} \rho(s',r|s,a) \left[r + \gamma V_{\pi}(s')\right]$$

(4.5)
$$V_{k+1}(s) = \sum_{q} \pi(a|s) \sum_{s',r} p(s',r|s,q) [r + \gamma V_{k}(s')].$$

Iterative policy evaluation - the sequence $\{V_k\}$ converges to V_T as $k \to \infty$

Chapter 4

Chapter 4

Policy improvement

(46)
$$q_{T}(s,a) = \sum_{s,r} \rho(s,r|s,a) \Big[r + \gamma V_{T}(s') \Big]$$

Policy π' is as good as or better than π if

(4.7) $q_{T}(s,\pi'(s)) \geq V_{T}(s)$

(4.8) $V_{T'}(s) \geq V_{T}(s)$

Greedy policy

$$= arg_{max} \sum_{s,r} \rho(s,r|s,a) \Big[r + \gamma V_{T}(s') \Big]$$

Suppose π' is as good as but not better than π
 $V_{T}(s) = a \sum_{s,r} \rho(s,r|s,a) \Big[r + \gamma V_{T}(s') \Big]$

i. $V_{T'}$ must be V_{X} , both π and π' are optimal $v_{T}(s,\pi'(s)) = v_{T}(s,\pi'(s)) = v_{T}(a|s) q_{T}(s,a)$

C for stochastic policies

Policy iteration

Once a policy, T, has been improved using VII to yield a better policy, TI', we can then compute VI' and improve if again to yield an even better TI'!

This method can be used to find an optimal policy.

- Value iteration

4.10)

 $V_{k+1}(s) = \max_{a} \sum_{s',r} \rho(s',r|s,a) \left[r + \gamma V_{k}(s')\right]$

Gombines policy evaluation and policy improvement

Generalized policy evaluation (GPI) - the general idea of letting policy evaluation and policy improvement processes interact independenty of the details of the two processes.

$$q_{T}(s,a) = \sum_{s,r} p(s',r|s,a) \left[r + \gamma V_{T}(s')\right]$$

4.4

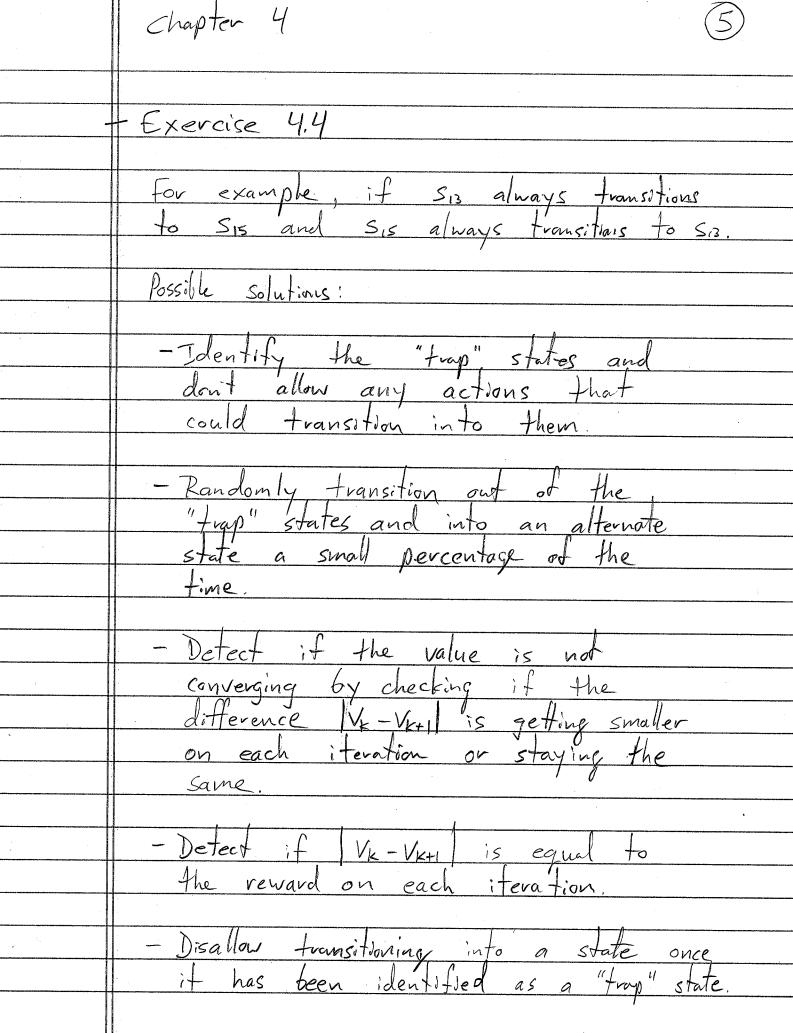
(4.3)
$$V_{tr}(s) = E_{tr} \left[P_{t+1} + \gamma V_{tr}(S_{t+1}) \middle| S_{t} = S \right]$$

$$V_{\pi}(s) = \sum_{a} \pi(a|s) \sum_{s,r} \rho(s',r|s,a) \left[r + \gamma V_{\pi}(s')\right]$$

$$g_{tt}(s,a) = \sum_{s,r} p(s',r|s,a) \left[r + \gamma \sum_{ar} \pi(a'|s) g_{tt}(s',a') \right]$$

(4.5)
$$V_{k+1}(s) = \sum_{\alpha} \pi(\alpha|s) \sum_{s,r} p(s',r|s,\alpha) \left[r + \gamma V_{k}(s')\right]$$

$$g_{K+1}(s,a) = \sum_{s,r} p(s',r|s,a) \left[r + \gamma \sum_{a'} \pi(a'|s') g_{K}(s',a')\right]$$



$$\pi_{*}(s) = \underset{s,r}{\operatorname{argmax}} \sum_{s,r} p(s,r|s,q) \left[r + \gamma V_{*}(s')\right]$$

$$T_*(s) = \underset{\alpha}{avgmax} q_*(s, a)$$

$$V_{*}(s) = \sum_{s,r} p(s,r|s,\pi_{*}(s)) \left[r + \gamma V_{*}(s)\right]$$

$$q_*(s,a) = \sum_{s',r} p(s',r|s,a) \left[r + \gamma q_*(s',\pi_*(s'))\right]$$

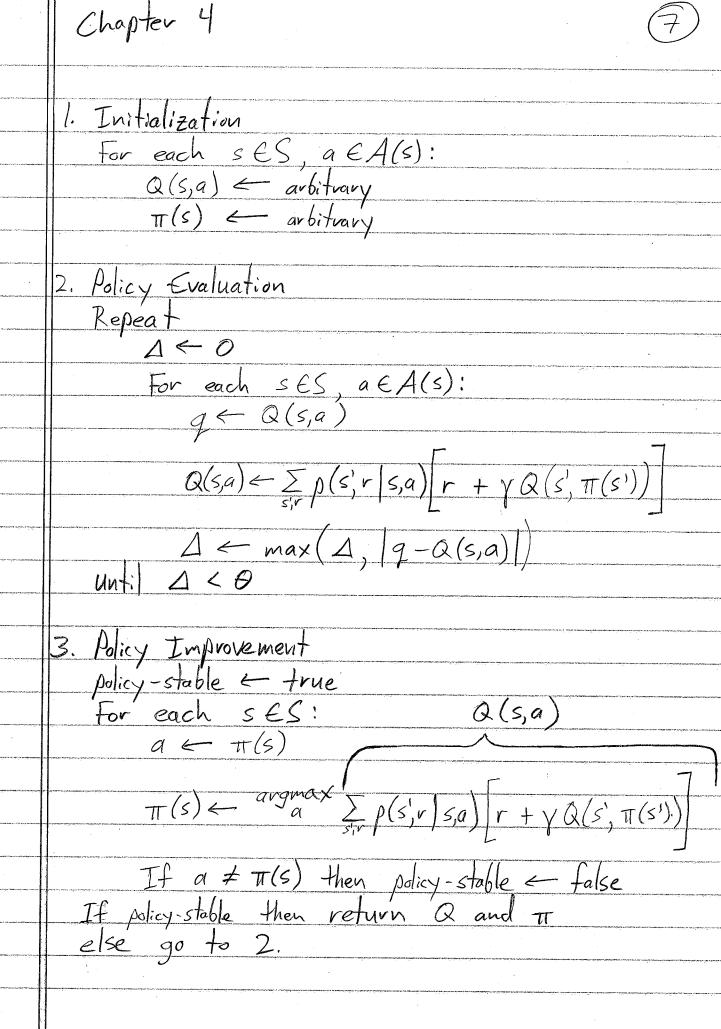
$$\pi'(s) = \underset{s,r}{\operatorname{argmax}} \sum_{s,r} p(s,r|s,a) \left[r + \gamma V_{\pi}(s') \right]$$

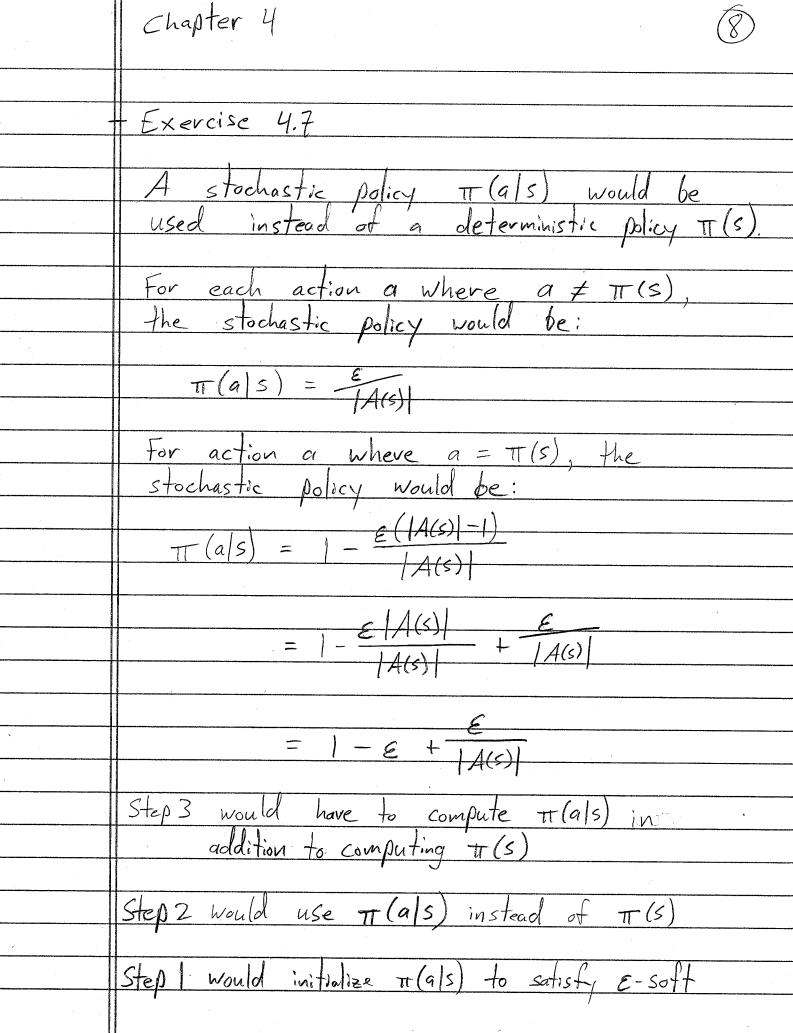
$$\pi'(s) = \underset{a}{\operatorname{argmax}} q_{\pi}(s, a)$$

$$V_{k+1}(s) = \sum_{s',r} \rho(s',r|s,\pi(s)) \left[r + \gamma V_k(s')\right]$$

$$q_{KH}(s,a) = \sum_{s,r} p(s',r|s,a) \left[r + \gamma q_{K}(s',\pi(s'))\right]$$

(cont.)





$$V_{k+1}(s) = \max_{\alpha} E[R_{t+1} + \gamma V_k(S_{t+1}) | S_{t} = s, A_{t} = a]$$

$$= \sum_{\alpha} \sum_{s,r} p(s,r|s,a) \left[r + \gamma V_{k}(s') \right]$$

$$g_{k+1}(s,a) = E\left[R_{++} + \gamma \stackrel{max}{a'} g_{k}(S_{++1},a')\right] S_{t}=s, A_{+}=a$$

$$=\sum_{s,r}p(s,r|s,a)\left[r+\gamma\max_{a'}q_{\kappa}(s,a')\right]$$