1.

(a)

(1) Want to show: $V_*(s) = \max_a Q_*(s,a)$

By definition,

$$V^\pi(s) = \mathbb{E}_\pi[G_t \mid S_t = s]$$

$$Q^\pi(s,a) = \mathbb{E}_\pi[G_t \mid S_t = s, a_t = a]$$

$$\Rightarrow V^\pi(s) = \sum_{a \in A} \pi(a|s) Q^\pi(s,a)$$

$$= \mathbb{E}[Q^\pi(s,a)] \leq \max_a Q^\pi(s,a) \quad \forall \pi$$

$\therefore V_*(s) \leq \max_a Q_*(s,a)$

If $V_*(s) < \max_a Q_*(s,a)$, then

$\exists \hat{\pi} = \arg\max_a Q_*(s,a)$

s.t. $V^{\hat{\pi}}(s) > V^*(s)$, which is impossible since π^* is optimal.

$\therefore \underline{V_*(s) = \max_a Q_*(s,a)}$ #

(2) Want to show: $Q_*(s,a) = R_s^a + \gamma \sum_{s'} P_{ss'}^a V_*(s')$.

By definition,

$$Q^\pi(s,a) = R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V^\pi(s')$$

$$\Rightarrow Q_*(s,a) = \max_\pi Q^\pi(s,a) = \max_\pi \left\{R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V^\pi(s')\right\}$$

$$= R_s^a + \max_\pi \left\{\gamma \sum_{s' \in S} P_{ss'}^a V^\pi(s')\right\}$$

$$= R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a \max_\pi V^\pi(s')$$

$$= \underline{R_s^a + \gamma \sum_{s' \in S} P_{ss'}^a V_*(s')}$$ #

(b)

Want to show : $\| T^*Q - T^*Q' \|_\infty \le \gamma \| Q - Q' \|_\infty$.

$$\| T^*Q - T^*Q' \|_\infty = \max_{s,a} \left| [T^*Q](s,a) - [T^*Q'](s,a) \right|$$

$$= \max_{s,a} \left| \left[R_s^a + \gamma \sum_{s'} P_{ss'}^a \max_{a'} Q(s',a') \right] - \left[R_s^a + \gamma \sum_{s'} P_{ss'}^a \max_{a'} Q'(s',a') \right] \right|$$

$$= \gamma \cdot \max_{s,a} \left| \sum_{s'} P_{ss'}^a \left(\max_{a'} Q(s',a') - \max_{a'} Q'(s',a') \right) \right|$$

$$\le \gamma \cdot \max_{s,a} \left| \sum_{s'} P_{ss'}^a \max_{a'} \left(Q(s',a') - Q'(s',a') \right) \right|$$

$$\le \gamma \cdot \max_{s,a} \left| \sum_{s'} P_{ss'}^a \max_{s'',a'} \left(Q(s'',a') - Q'(s'',a') \right) \right|$$

$$= \gamma \cdot \max_{s,a} \left| \max_{s'',a'} \left(Q(s'',a') - Q'(s'',a') \right) \right|$$

$$= \gamma \cdot \max_{s'',a'} \left| Q(s'',a') - Q'(s'',a') \right|$$

$$= \gamma \| Q - Q' \|_\infty$$

Therefore, T^* is a γ-contraction operator. #

2.

(a)

Assume U, V, A are continuous random variables with PDFs f_U, f_V, f_A

$\because A$ is independent of U and V,

$\Rightarrow f_{U+A}(z) = \int_{-\infty}^{\infty} f_U(z-a) f_A(a)\, da$

$f_{V+A}(z) = \int_{-\infty}^{\infty} f_V(z-a) f_A(a)\, da$, for $z \in R$

Let $f_{U,V}(u,v)$ be a joint PDF of U, V, that satisfies

$f_U(u) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\, dv$

and $f_V(v) = \int_{-\infty}^{\infty} f_{U,V}(u,v)\, du$

Let $x = \cancel{U+A}\ u+a$
$y = \cancel{V+A}\ v+a$

$$f_{U+A}(x) \stackrel{?}{=} \int_{-\infty}^{\infty} f_{U+A,V+A}(x,y)\, dy = \int_{-\infty}^{\infty} f_U(x-a) f_A(a)\, da$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{U,V}(u+a-a, v+a-a) f_A(a)\, dv\, da$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{U,V}(u,v) f_A(a)\, dv\, da.$$

$$f_{V+A}(y) = \int_{-\infty}^{\infty} f_{U+A,V+A}(x,y)\, dx = \int_{-\infty}^{\infty} f_V(x-a) f_A(a)\, da.$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{U,V}(u+a-a, v+a-a) f_A(a)\, du\, da.$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{U,V}(u,v) f_A(a)\, du\, da.$$

$$\|U-V\|_P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |u-v| f_{U,V}(u,v)\, du\, dv$$

$$\|(A+U)-(A+V)\|_P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x-y|\, f_{U+A,V+A}(x,y)\, dx\, dy$$

$=$ Let $\int du = \int f_{U+A,V+A}(x,y)\, dx \rightarrow u = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{U,V}(u,v) f_A(a)\, dv\, da.$

$v = |x-y| \longrightarrow dv =$

(b)

Want to show: $B^\pi : \mathcal{Z} \to \mathcal{Z}$ is a γ-contraction operator in $\bar{d}_p$.

Consider $Z_1, Z_2 \in \mathcal{Z}$. By definition,

$$\bar{d}_p(B^\pi Z_1, B^\pi Z_2) = \sup_{x,a} d_p(B^\pi Z_1(x,a), B^\pi Z_2(x,a)) \quad \cdots\cdots ①$$

By the properties of d_p,

$$\begin{aligned} & d_p(B^\pi Z_1(x,a), B^\pi Z_2(x,a)) \\ &= d_p(R(x,a) + \gamma P^\pi Z_1(x,a), R(x,a) + \gamma P^\pi Z_2(x,a)) \\ &\le \gamma\, d_p(P^\pi Z_1(x,a), P^\pi Z_2(x,a)) \\ &\le \gamma \sup_{x',a'} d_p(Z_1(x',a'), Z_2(x',a')) \end{aligned}$$

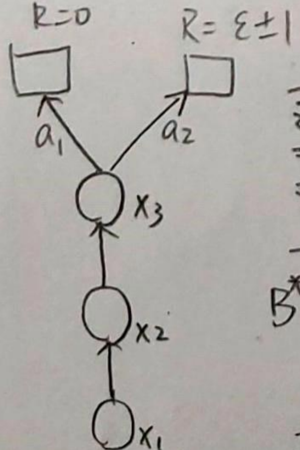
$\cdots\cdots$ by the definition given in the original paper: $P^\pi Z(x,a) := Z(X', A')$, $X' \sim P(\cdot \mid x,a)$, $A' \sim \pi(\cdot \mid X')$.

Combining with ①,

$$\begin{aligned} \bar{d}_p(B^\pi Z_1, B^\pi Z_2) &= \sup_{x,a} d_p(B^\pi Z_1(x,a), B^\pi Z_2(x,a)) \\ &\le \gamma \sup_{x',a'} d_p(Z_1(x',a'), Z_2(x',a')) \\ &= \gamma \bar{d}_p(Z_1, Z_2) \end{aligned}$$

Therefore, $B^\pi : \mathcal{Z} \to \mathcal{Z}$ is a γ-contraction operator in $\bar{d}_p$ #.

(c)



	x_1	x_2	x_3, a_1	x_3, a_2
Z^*	$\varepsilon \pm 1$	$\varepsilon \pm 1$	0	$\varepsilon \pm 1$
Z	$\varepsilon \pm 1$	$\varepsilon \pm 1$	0	$-\varepsilon \pm 1$
$B^* Z$	0	0	0	$\varepsilon \pm 1$

$$\bar{d}_1(Z, Z^*) = d_1(Z(x_3, a_2), Z^*(x_3, a_2)) = 2\varepsilon.$$

$d_1(B^* Z, B^* Z^*) = \frac{1}{2}|1-\varepsilon| + \frac{1}{2}|1+\varepsilon| > 2\varepsilon$ for a sufficiently small ε, which shows that it is not a contraction operator. #

4.

(a)

Want to show: $\mathbb{E}_{\tau \sim p_\mu^{\pi_\theta}}\left[\sum_{t=0}^{\infty} \gamma^t f(s_t, a_t)\right] = \frac{1}{1-\gamma} \mathbb{E}_{s \sim d_\mu^{\pi_\theta}, a \sim \pi_\theta(\cdot|s)}[f(s,a)]$

$$RHS = \frac{1}{1-\gamma} \sum_s d_\mu^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) f(s,a)$$

$$= \frac{1}{1-\gamma} \sum_s \sum_{s_0} \mu(s_0) d_{s_0}^{\pi}(s) \sum_a \pi_\theta(a|s) f(s,a)$$

$$= \frac{1}{\cancel{1-\gamma}} \sum_s \sum_{s_0} \mu(s_0) (\cancel{1-\gamma}) \sum_{t=0}^{\infty} \gamma^t P(s_t = s | s_0, \pi_\theta) \sum_a \pi_\theta(a|s) f(s,a)$$

$$= \sum_s \sum_{t=0}^{\infty} \gamma^t P(s_t = s | \mu, \pi_\theta) \sum_a \pi_\theta(a|s) f(s,a)$$

$$= \sum_s \sum_a \sum_{t=0}^{\infty} \gamma^t P(s_t = s, a_t = a | \mu, \pi_\theta) f(s,a)$$

$$= \mathbb{E}_{\tau \sim p_\mu^{\pi_\theta}}\left[\sum_{t=0}^{\infty} \gamma^t f(s,a)\right]$$

$= LHS$

#

(b)

Want to show : $\nabla_\theta V^{\pi_\theta}(\mu) = \mathbb{E}_{\tau \sim p_\mu^{\pi_\theta}}\left[\sum_{t=0}^{T-1} \gamma^t A^{\pi_\theta}(s_t, a_t) \nabla_\theta \log \pi_\theta(a_t|s_t)\right]$

By definition, $A^{\pi_\theta}(s,a) = Q^{\pi_\theta}(s,a) - V^{\pi_\theta}(s)$

$\Rightarrow \nabla_\theta V^{\pi_\theta}(\mu) = \mathbb{E}_{\tau \sim p_\mu^{\pi_\theta}}\left[\sum_{t=0}^{T-1} \gamma^t (Q^{\pi_\theta}(s_t, a_t) - V^{\pi_\theta}(s_t)) \nabla_\theta \log \pi_\theta(a_t|s_t)\right]$

By Eq.(7),

$$\mathbb{E}_{\tau \sim p_\mu^{\pi_\theta}}\left[\sum_{t=0}^{T-1} \gamma^t V^{\pi_\theta}(s_t) \cdot \nabla_\theta \log \pi_\theta(a_t|s_t)\right]$$

$$= \frac{1}{1-\gamma} \mathbb{E}_{s \sim d^{\pi_\theta}} \mathbb{E}_{a \sim \pi_\theta(\cdot|s)}\left[V^{\pi_\theta}(s) \cdot \nabla_\theta \log \pi_\theta(a|s)\right]$$

$$= \frac{1}{1-\gamma} \sum_s d^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) \cdot V^{\pi_\theta}(s) \cdot \nabla_\theta \log \pi_\theta(a|s)$$

$$= \frac{1}{1-\gamma} \cdot \sum_s d^{\pi_\theta}(s) V^{\pi_\theta}(s) \sum_a \pi_\theta(a|s) \nabla_\theta \log \pi_\theta(a|s)$$

$$= \frac{1}{1-\gamma} \cdot \sum_s d^{\pi_\theta}(s) V^{\pi_\theta}(s) \nabla_\theta \underbrace{\sum_a \pi_\theta(a|s)}_{\to 0} = 0$$

$\therefore$ The introduction of $V^{\pi_\theta}(s)$ does not change the expectation #