(a) Show that $\sum_{t}^{T} E_{\tau \sim p_\theta}\left[\nabla_\theta \log \pi_\theta(a_t|s_t)\, b(s_t)\right] = 0$

$$p_\theta(\tau) = p_\theta(s_t, a_t)\, p_\theta\left(\frac{\tau}{s_t, a_t} \,\Big|\, s_t, a_t\right)$$

$$E_{\tau \sim p_\theta}\left[\nabla_\theta \log \pi_\theta(a_t|s_t)\, b(s_t)\right]$$

$$= \int_\tau \nabla_\theta \log \pi_\theta(a_t|s_t)\, b(s_t)\, p(\tau)\, d\tau$$

$$= \iint \nabla_\theta \log \pi_\theta(a_t|s_t)\, b(s_t)\, p_\theta(s_t, a_t)\, p_\theta\left(\frac{\tau}{s_t, a_t} \,\Big|\, s_t, a_t\right) d(s_t, a_t)\, d\left(\frac{\tau}{s_t, a_t}\right)$$

$$= \int \nabla_\theta \log \pi_\theta(a_t|s_t)\, b(s_t)\, p_\theta(s_t, a_t)\, d(s_t, a_t) \underbrace{\int p_\theta\left(\frac{\tau}{s_t, a_t}\right) d\left(\frac{\tau}{s_t, a_t}\right)}_{1}$$

$\nabla p(x) = p(x) \nabla \log p(x)$

$p(x,y) = p(x)\, p(y|x)$

$p_\theta(a_t|s_t) = \pi_\theta(a_t|s_t)$

$$= \iint \frac{\nabla_\theta \pi_\theta(a_t|s_t)}{\pi_\theta(a_t|s_t)}\, b(s_t)\, \underbrace{p_\theta(s_t, a_t)}_{p_\theta(a_t|s_t)\, p(s_t) = \pi_\theta(a_t|s_t)\, p(s_t)}\, ds_t\, da_t$$

$$= \iint \nabla_\theta \pi_\theta(a_t|s_t)\, b(s_t)\, p(s_t)\, ds_t\, da_t$$

$$= \int p(s_t)\, b(s_t)\, ds_t\ \nabla_\theta \int \pi_\theta(a_t|s_t)\, da_t$$

$$= \int p(s_t)\, b(s_t)\, ds_t\ \nabla_\theta 1$$

$$= 0$$

(b) $$p_\theta(\tau) = p_\theta(s_{0:t}, a_{0:t-1})\, p_\theta(s_{t+1:T}, a_{t:T} \mid s_{0:t}, a_{0:t-1})$$

show $$\sum_{t=0}^{T} E_{\tau \sim p_\theta(\tau)} [\nabla_\theta \log \underbrace{p_\theta(a_t \mid s_t)}_{= \pi_\theta(a_t \mid s_t)}\, b(s_t)] = 0$$

$$P(s_t, a_t \mid \pi) = \sum_{s_0} \sum_{a_0} \sum_{s_1} \cdots \sum_{s_{t+1}} \sum_{a_{t+1}} P(s_0, a_0, s_1, a_1 \ldots s_{t+1}, a_{t+1}, \ldots \mid \pi)$$

$$\sum_t \sum_\tau p_\theta(\tau) \nabla_\theta \log p_\theta(a_t \mid s_t)\, b(s_t)$$

$$= \sum_t \sum_{s_{0:t}} \sum_{a_{0:t-1}} \sum_{s_{t+1:T}} \sum_{a_{t:T}} P(s_{0:t}, a_{0:t-1})\, P(s_{t+1:T}, a_{t:T} \mid s_{0:t}, a_{0:t-1}) \cdot \nabla_\theta \log p_\theta(a_t \mid s_t)\, b(s_t)$$

$$= \sum_t E_{s_{0:t}, a_{0:t-1}} \left[b(s_t)\, E_{s_{t+1:T}, a_{t:T}} [\nabla_\theta \log p_\theta(a_t \mid s_t) \mid \underbrace{s_{0:t}, a_{0:t-1}}] \right]$$

Markov property means $s_{t+1:T}, a_{t:T}$ only depends on s_t

$$E_{s_{t+1:T}, a_{t:T}} [\nabla_\theta \log p_\theta(a_t \mid s_t) \mid s_t]$$

$$= \sum_{s_{t+1}} \sum_{s_{t+2}} \cdots \sum_{s_T} \cdots \sum_{a_t} \cdots \sum_{a_T} p_\theta(s_{t+1} \mid s_t, a_t)\, p_\theta(s_{t+2} \mid s_{t+1}, a_{t+1}) \cdots p_\theta(a_t \mid s_t)\, p_\theta(a_{t+1} \mid s_{t+1})$$
$$\cdots p_\theta(a_T \mid s_T)\, p_\theta(s_T \mid s_{T-1}, a_{T-1}) \cdot \nabla_\theta \log p_\theta(a_t \mid s_t)$$

$$= \sum_{a_t} p_\theta(a_t \mid s_t) \nabla_\theta \log p_\theta(a_t \mid s_t) \underbrace{\sum_{s_{t+1}} p(s_{t+1} \mid s_t, a_t)}_{\to 1} \sum_{s_{t+2}} \cdots\cdots$$

all sum to 1 since main term doesn't use any variables besides s_t, a_t

$$= \sum_{s_t} \sum_{a_t} \cancel{p_\theta(a_t|s_t)} \underbrace{\nabla_\theta \log p_\theta(a_t|s_t)}_{\frac{\nabla p_\theta(a_t|s_t)}{\cancel{p_\theta(a_t|s_t)}}} p_\theta(s_t) b(s_t)$$

$$\nabla p(x) = p(x) \nabla \log p(x)$$

$$= \sum_{s_t} \sum_{a_t} \nabla_\theta p_\theta(a_t|s_t) \, p_\theta(s_t) \, b(s_t)$$

$$= \sum_{s_t} b(s_t) p_\theta(s_t) \nabla_\theta \cancelto{1}{\sum_{a_t} p_\theta(a_t|s_t)}$$

$$= \sum_{s_t} b(s_t) p_\theta(s_t) \cancelto{0}{\nabla_\theta 1}$$

$$= 0$$