hl
$\mathcal{E}(\pi_l)\pi_l\pi_h$

$$\eta(\pi_h) = V_h(s_0^h) = E_{s_0^h, a_0^h, \ldots}\left[\sum_{t=0,k,2k,\ldots} \gamma_h^{t/k} r_h(s_t^h, a_t^h)\right], where s_0^h \sim \rho_0^h(s_0^h), a_t^h \sim \pi_h(s_t^h), s_{t+1} \sim P(s_{t+1}^h | s_t^h, a_t^h, \mathcal{E}(\pi_l)) \tag{1}$$

$\pi_h\pi_l$
$\eta(\pi_h)RR$
$\frac{\gamma_h V_h(s_{t+k}^h) - V_h(s_t^h)}{k}\pi_l$
$?V_\pi(s_0)?JV_\pi(s_0)RE_{s,a\sim\pi}[R(s,a)]$
R
$\bigtriangledown J(\theta) =$
$\bigtriangledown E_{s\sim\pi}[\sum_a \pi(s,a)R(s,a)] =$
$\bigtriangledown E_{s,a\sim\pi}[R(s,a)].$
$_l(s_{t+i}^l, a_{t+i}^l)|_{i=0,1,\ldots,k-1} =$
$\frac{V_h(s_{t+k}^h) - V_h(s_t^h)}{k}.$
$E_{s,a\sim\pi}[R(s,a)]E_{s,a\sim\pi_l,\pi_h}[R_l(s_l,a_l)]\pi_h\pi_h$
$_l owE_{s^l,a^l\sim\pi_l,\pi_h}[R_l(s^l,a^l)] =$
$\frac{1}{k}E_{s^h,a^h\sim\pi_l,\pi_h}[\gamma_h V_h(s_{t+k}^h) -$
$V_h(s_t^h)].$
$A(s,a)$
$_t, A_t) =$
$Q(S_t, A_t) -$
$V(S_t) =$
$R(S_t, A_t) +$
$\gamma V(S_{t+1}) -$
$V(s).$
$_t, A_t) =$
$0, \forall t \neq$
$t_{end}.$
$_t, A_t) =$
$\gamma V(S_{t+1}) -$
$V(s).$
$_e xpectationeq : sparse_a dvantageTRPOE_{s^l,a^l\sim\pi_l,\pi_h}[R_l(s^l,a^l)] =$
$\frac{1}{k}E_{s^h,a^h\sim\pi_l,\pi_h}[A^h(s_t^h, a_t^h)].$
$\tilde{\eta}(\pi_h)\pi_l\tilde{\pi}_l\pi_h$**?Lemma**
1
$\tilde{\eta}(\pi_h) =$
$\eta(\pi_h) +$
$E_{s_0^h, a_0^h, \ldots\sim\pi_h, \mathcal{E}(\tilde{\pi_l})}\left[\sum_{t=0,k,2k,\ldots} \gamma_h^{t/k} A_{\pi_h}(s_t^h, a_t^h)\right].$
$_e nvironment_a dvantage$
ρ
$\rho_{\pi_h}(s^h) =$
$\sum_{t=0,k,2k,\ldots} \gamma_h^{t/k} P(s_t^h =$
$s|\mathcal{E}(\pi_l))$
$\rho_{\pi_h}\tilde{\rho}_{\pi_h}$
$_e nvironment_a dvantageTRPO22 \sum_{s^h} \tilde{\rho}_{\pi_h}(s^h) \sum_{a^h} \pi_h(a^h|s^h) A_{\pi_h}(s^h, a^h).$
$_0^h =$
$s^h) =$
$P(s_1^h =$
$s^h) =$
$\ldots =$
$P(s_i^h =$
$s^h)$
$P(s_i^h =$
$s^h)$
$_o bjective\tilde{\rho}_{\pi_h}(s^h) =$
$\frac{1}{1-\gamma_h} P(s_i^h =$
$s^h)$
$_e nvironment_a dvantage\tilde{\eta}(\pi_h) =$
$\eta(\pi_h) +$
$\frac{1}{1-\gamma_h} E_{s^h,a^h\sim\pi_h,\mathcal{E}(\tilde{\pi_l})}\left[A_{\pi_h}(s^h, a^h)\right].$
$_e xpectation_i s_a dvantageeq : accurate_o bjective_e xpectation_f ormTRPO\eta(\pi_h)$
$\pi_h\pi_l\pi_l\pi_h\eta(\pi_h)$
$\pi_h\pi_l\eta(\pi_h)\pi_l\pi_h$
1
$\tilde{\eta}(\pi_h) =$
$\eta(\pi_h) +$
$\left[$