AHAC Theory

Thursday, March 28, 2024 10:15

Core assumptions

1. $\pi(\cdot|s) : \mathbb{R}^n, \mathbb{R}^d \mapsto [0,1]^m$ is continuously differentiable and Lipshitz smooth w.r.t. L2 norm
   $\Rightarrow \| \nabla_\theta \pi(a|s) \|_2 \leq B_\pi$
2. $f(s,a) : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}^n$ is cont. differentiable and Lipschitz smooth in both arguments w.r.t. L2 norm
   $\Rightarrow \| \nabla_s f(s,a) \|_2 \leq B_f \qquad \| \nabla_a f(s,a) \|_2 \leq B_f$
3. $r(s,a) : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}$ is cont. differentiable and Lipshitz smooth in both arguments w.r.t. L2 norm
   $\Rightarrow \| \nabla_s r(s,a) \|_2 \leq B_r \qquad \| \nabla_a r(s,a) \|_2 \leq B_r$
4. Zeroth-order gradients use the baseline
   $\hat{\nabla}^{[0]}_\theta J(\theta) = \sum_{t=1}^{H} \left(r(s_t, a_t) - b \right) \nabla_\theta \log \pi_\theta(a_t|s_t)$

In the paper we define and bound

$$B = \| \mathbb{E}\left[\nabla^{[1]}_\theta J(\theta) \right] - \mathbb{E}\left[\nabla^{[0]}_\theta J(\theta) \right] \|$$

as pointed out by Reviewer 6LLu, $B \neq 0$. Instead we want to bound the sample error

$$B = \| \hat{\nabla}^{[1]}_\theta J(\theta) - \hat{\nabla}^{[0]}_\theta J(\theta) \|_2$$

$$= \| \frac{1}{N} \sum_{i=1}^{N} \hat{\nabla}^{[1]} J_i(\theta) - \frac{1}{N} \sum_{i=1}^{N} \hat{\nabla}^{[0]} J_i(\theta) \|_2$$

$$= \frac{1}{N} \| \sum_{i=1}^{N} \left(\hat{\nabla}^{[1]} J_i(\theta) - \hat{\nabla}^{[0]} J_i(\theta) \right) \|_2$$

$$\leq \| \hat{\nabla}^{(1)} J(\theta) - \hat{\nabla}^{[0]} J(\theta) \|_2 \qquad (1)$$

1. drop sample subscript for simplicity
2. assume $a \sim \pi_\theta(\cdot | \text{sg}(s))$ for easier derivation

where $\text{sg}(\cdot)$ is stop gradient operation (Xu et. al. 2024)

Aside: Why do this?

Example with $r(s_1, a_1)$. The assumption cancels out the real terms!

$$\begin{aligned}
\nabla_\theta r(s_1, a_1) = {} & \nabla_{a_1} r(s_1, a_1) \nabla_\theta \pi(a_1|s_1) \\
& + \nabla_{a_2} r(s_2, a_2) \nabla_{s_2} f(s_1, a_1) \nabla_{a_1} f(s_1, a_1) \nabla_\theta \pi(a_1|s_1) \\
& + \nabla_{a_2} r(s_2, a_2) \nabla_{s_2} f(s_1, a_1) \nabla_{a_1} f(s_1, a_1) \nabla_{s_1} f(s_1, a_1) \nabla_\theta \pi(a_1|s_1) \\
& + \nabla_{a_3} r(s_3, a_3) \nabla_{s_3} f(s_2, a_2) \nabla_{s_2} f(s_1, a_1) \nabla_{a_1} f(s_1, a_1) \nabla_\theta \pi(a_1|s_1) \\
& + \nabla_{s_2} r(s_2, a_2) \nabla_{a_2} f(s_1, a_1) \nabla_\theta \pi(a_1|s_1) \\
& + \nabla_{s_3} r(s_3, a_3) \nabla_{s_3} f(s_2, a_2) \nabla_{s_2} f(s_1, a_1) \nabla_{a_1} f(s_1, a_1) \nabla_\theta \pi(a_1|s_1)
\end{aligned}$$

Issues wrt Reviewer 4 (E23g)

1. Expanding $\nabla_\theta J(\theta)$ is incorrect
2. We implicitly assumed $a_t \sim \pi_\theta(\cdot|s_t)$ but never stated it

$$\hat{\nabla}^{(1)}_\theta J(\theta) - \hat{\nabla}^{[0]}_\theta J(\theta)$$

$$= \nabla_\theta \sum_{t=1}^{H} r(s_t, a_t) - \sum_{t=1}^{H} \left(r(s_t, a_t) - b \right) \nabla_\theta \log \pi(a_t|s_t)$$

$$= \sum_{t=1}^{H} \nabla_{a_t} r(s_t, a_t) \nabla_\theta \pi(a_t|s_t) + \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t)$$

$$- \sum_{t=1}^{H} \left(r(s_t, a_t) - b \right) \nabla_\theta \log \pi(a_t|s_t)$$

$$= \sum_{t=1}^{H} \nabla_\theta \pi_\theta(a_t|s_t)^\top \left(\nabla_{a_t} r(s_t, a_t) - \left(r(s_t, a_t) - b \right) \pi_\theta(a_t|s_t)^{-1} \right) + \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t)$$

where X^{-1} is Moore-Penrose inverse

If we assume $b = r(s_t, 0)$ we can exploit $r(\cdot,\cdot)$ Lipshitz i.e. $\| \nabla r(s,a) \|_2 \leq B_r$

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2} \| y - x \|_2^2$$

$$f(0) \leq f(x) - \nabla f(x)^\top x + \frac{L}{2} \| x \|_2^2$$

$$f(x) - f(0) \geq \nabla f(x)^\top x - \frac{L}{2} \| x \|_2^2$$

$$f(x) - f(0) \geq \left(\nabla f(x)^\top - \frac{L}{2} x^\top \right) x \qquad (2)$$

$$= \sum_{t=1}^{H} \nabla_\theta \pi_\theta(a_t|s_t)^\top \left(\nabla_{a_t} r(s_t, a_t) - \left(r(s_t, a_t) - r(s_t, 0) \right) \pi_\theta(a_t|s_t)^{-1} \right) + \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t) \qquad \text{apply } b = \sum_{t=1}^{H} r(s_t, 0)$$

$$\leq \sum_{t=1}^{H} \nabla_\theta \pi_\theta(a_t|s_t)^\top \left(\nabla_{a_t} r(s_t, a_t) - \left(\nabla r(s_t, a_t) - \frac{B_r}{2} \pi_\theta(a_t|s_t) \right) \right) + \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t) \qquad \text{apply } (2)$$

$$= \frac{B_r}{2} \sum_{t=1}^{H} \nabla_\theta \pi_\theta(a_t|s_t)^\top \pi_\theta(a_t|s_t) + \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t) \qquad (3)$$

Plug (3) into (1)

$$B \leq \| \hat{\nabla}^{[1]}_\theta J(\theta) - \hat{\nabla}^{[0]}_\theta J(\theta) \|$$

$$= \| \frac{B_r}{2} \sum_{t=1}^{H} \nabla_\theta \pi_\theta(a_t|s_t)^\top \pi_\theta(a_t|s_t) + \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t) \|_2$$

$$\leq \| \frac{B_r}{2} \sum_{t=1}^{H} \nabla_\theta \pi_\theta(a_t|s_t)^\top \pi_\theta(a_t|s_t) \|_2 + \| \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t) \|_2$$

since $\| \nabla_\theta \pi_\theta(a_t|s_t) \|_2 \leq B_\pi$ and $\| \pi_\theta(a_t|s_t) \| \leq 1$

$$\leq \frac{1}{2} H B_r B_\pi + \| \left(\sum_{t'=1}^{t} \left(\prod_{k=t'+1}^{t} \nabla_{s_k} f(s_k, a_k) \right) \nabla_{a_{t'}} f(s_{t'}, a_{t'}) \right)^\top \nabla_\theta \pi(a_{t'}|s_{t'}) \Big)^\top \nabla_{s_t} r(s_t, a_t) \|_2$$

$$\leq \frac{1}{2} H B_r B_\pi + (H-1) B_r B_\pi B_f^{H-1}$$

$$\leq \frac{1}{2} H B_r B_\pi + H B_r B_\pi B_f^{H-1}$$

$$= H B_r B_\pi \left(1/2 + B_f^{H-1} \right)$$

Critically the most important term above is B_f^{H-1} which implies that non-smooth dynamics contribute the most to this upper bound

Summary of changes:

1. Need to work in L2 norms for Lipshitz property to work
2. Drop $|r(s,a)| \leq 1$ assumption
3. Assume action-less baseline $b = \sum_{t=1}^{H} r(s_t, 0)$
4. Assume $a \sim \pi_\theta(\cdot | \text{sg}(s))$
5. More explicitly note that $r(s,a)$ and $f(s,a)$ are Lipshitz in both arguments