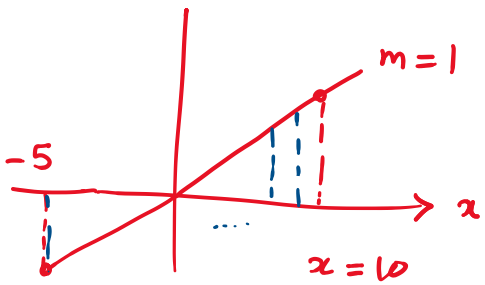


$$\min_x f \quad x \geq -5$$

$$\nabla f = \begin{cases} 1 & 0.5 \\ -1 & 0.5 \end{cases}$$

$$\nabla_x f = 1 + X$$

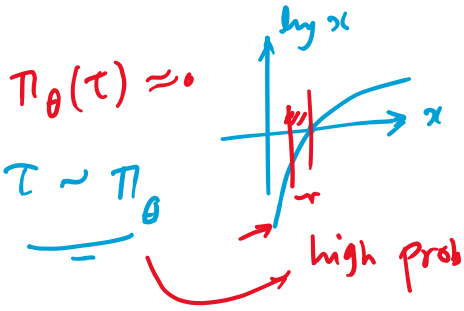
$$X = \begin{cases} 2 & 0.5 \\ -2 & 0.5 \end{cases}$$

$\eta = 1$

$x \leftarrow x - 1$ true grad.

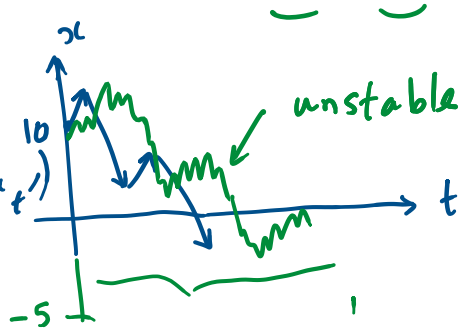
$x \leftarrow x - 1 \pm 2$ stoch. grad.

$-3 \quad +1$



$\pi_\theta(\tau) \approx 0$

$\tau \sim \pi_\theta$



$$X = \nabla \log \pi(a_t|s_t) \sum_{t'=t}^{T} r(s_{t'}, a_{t'}) \quad \overset{?}{=}$$

$\tau = s_1, a_1, s_2, a_2 \quad \cdots \quad s_t, a_t$

γ

$$\mathbb{E}_{\tau}\left\{ \nabla \log \pi(a_t|s_t) \sum_{t'=1}^{t-1} r(s_{t'}, a_{t'}) \right\} = 0$$

$\mathbb{E}(g(x,\gamma)) = \mathbb{E}_x \mathbb{E}_{\gamma|x}\, g(x,\gamma)$

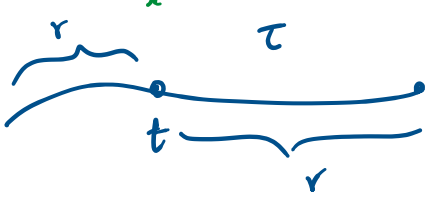
$$= \mathbb{E}_{s_1, a_1, \cdots, s_{t-1}, a_{t-1}} \mathbb{E}_{s_t, a_t | s_1, a_1 \cdots s_{t-1} a_{t-1}} \left\{ \nabla \log \pi(a_t|s_t) \sum_{t'=1}^{t-1} r(s_{t'}, a_{t'}) \right\}$$

$$= \mathbb{E}_{s_t, a_t} \left\{ \sum_{t'=1}^{t-1} r(s_{t'}, a_{t'}) \; \mathbb{E}_{s_t, a_t | s_{t-1}, a_{t-1}} \nabla_\theta \log \pi_\theta(a_t|s_t) \right\}$$

g

$$= \mathbb{E}_{s_t | s_{t-1}, a_{t-1}} \mathbb{E}_{a_t|s_t} \nabla_\theta \log \pi_\theta(a_t|s_t)$$

$$\int \nabla \pi / \pi \cdot \pi \, da_t \quad \leftarrow \quad \int g(a_t) \cdot \pi(a_t|s_t) \, da_t \quad \nabla \pi_\theta / \pi_\theta$$



$\int \nabla \pi_\theta \, da_t$

$Y \overset{?}{\perp\!\!\!\perp} X$

$\mathbb{E}[Y] = 0$

$\nabla_\theta \int \pi_\theta \, da_t = 0$

$Var[X + Y] \geq Var[X]$

$Var[X+Y] = Var[X] + Var[Y] + 2\,Cov(X, Y)$

$Cov(X, Y) = 0$

On-policy

bad

$\tau_1 \cdots \tau_N \leftarrow \pi_\theta$

bad

$\theta' \leftarrow \theta +$ noisy grad.

θ''